Fine phonetic detail in the production and perception of reduced pronoun and auxiliary combinations in British English

Emanuela Buizza

PhD

The University of York

Language and Linguistic Science

September 2021

Abstract

This thesis investigates the production and perception of reduced function words, and the role of fine phonetic detail in spoken word recognition. In particular, it focusses on the phonetic features that convey linguistic information and maintain the contrast between pronoun and auxiliary combinations in British English.

The thesis is divided into two parts. The first part reports on a qualitative and quantitative analysis of pronoun and cliticised auxiliary combinations, such as *I'd*, *she's*, *you'll*, in reduced speech. Data of high acoustic quality and high degree of reduction were elicited in a controlled phonological and prosodic environment. The auditory and acoustic analyses of the data collected revealed that in reduced speech, function words retain essential phonetic features that constitute the identity of the target words.

The comparison between the phonetic features of contrasting paradigms that convey linguistic information, such as *you'll* versus *you'd*, and *she's* versus *she was*, revealed that the contrast is maintained by a combination of acoustic parameters, of which resonances and duration are the most prominent.

The second part investigates the role of fine phonetic detail in the intelligibility of reduced function words, and the perceptual salience of selected acoustic parameters. A perception experiment indicated that listeners are sensitive to the fine phonetic detail that maintains the contrast in reduced speech, and that they can correctly identify highly reduced words even when they are presented in a minimal semantic context. A further investigation into the perceptual salience of duration and resonances in spoken word recognition indicated that the resonances play a primary role in the correct identification of reduced speech.

Besides contributing to our knowledge of reduced speech, and confirming the role of fine phonetic detail in speech intelligibility, this thesis highlights the importance of carrying out a qualitative analysis and using a non-segmental approach in the analysis of reduced speech.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Richard Ogden for his endless patience and kindness, and for sharing his knowledge and cats with me.

Thanks are due to the ESRC for the financial support, and to the participants of the experiments for their contribution, without which this research would not have been possible.

I am grateful to Huw Llewelyn-Jones for his help in the Recording Studio, and to Catherine Laing for her help with statistics.

I would like to thank my friends Melinda, Catherine, Lorena, Sara, and Lupo for their support and encouragement, and for bearing with me despite all.

No words can express my gratitude towards my parents for their love and support. It has been a long and painful journey for them too.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author.

This work has not previously been presented for an award at this, or any other,

University. All sources are acknowledged as References.

Material from Chapter 5 was previously published in:

- Buizza, E. (2019). Maintaining contrast in reduced speech in Standard Southern British English. *Proceedings of the 19th International Congress of Phonetic Sciences, ICPhS.*

Emanuela Buizza

September 2021

“It is not possible to have too much phonetic detail.”

Kelly and Local (1989: 26)

1. Introduction and literature review

The main aim of this research is to investigate the production and perception of reduced speech. In particular, the focus of this research is on the fine phonetic detail that remains in the signal in highly reduced speech, and how it helps maintain the intelligibility of reduced speech.

Reduction is a widespread phenomenon in everyday speech to the extent that it is considered "the normal way to communicate" (Warner and Tucker 2011: 1615). Several studies have shown that reduced speech is mostly intelligible and that reduction rarely hinders communication (Ernestus and Warner, 2011). The two main factors that play a role in maintaining the intelligibility of reduced speech are the context, and the acoustic cues remaining in the signal. Bearing in mind the importance of the contextual factors in speech understanding, this thesis focusses on the role of phonetic detail in reduced speech. In particular, starting from the assumption that fine phonetic detail conveys crucial linguistic information which is available to perception even in highly reduced speech, this research investigates the production and perception of reduced pronoun and auxiliary combinations in English.

Several reasons led to the choice of pronouns and auxiliary verbs as object of this investigation. English auxiliary verbs are known to have a wide range of variation. Several lexical, morpho-syntactic, and phonological factors affect their phonetic realisation. Some of these factors are: their lexical category, their lexical frequency and occurrence in predictable collocations, their distribution, the limited set of contrasting sounds in their system, and their restricted paradigmatic system of contrast.

This chapter is divided into three parts. The first part introduces the concept of reduction and the phonetic properties that characterise reduced speech (Section 1.1). It reports on the research on the intelligibility of reduced speech (Section 1.1.1), and the main linguistic factors that influence the type and degree of reduction in everyday speech (Section 1.1.2).

The second part of this chapter examines the main characteristics of English auxiliary verbs (Section 1.2), focussing on the aspects that are most relevant for this research: the weak forms of auxiliaries (Section 1.2.1), and the phonetic features and phonological accounts of English auxiliaries (Section 1.2.3).

The third part of this chapter explains the theoretical approach of the thesis (Section 1.3). It briefly describes the main aspects of Firthian Prosodic Analysis that are relevant to the present work (Section 1.3.1), and the importance of fine phonetic detail in any meaningful investigation of speech reduction (Section 1.3.3). The chapter concludes with the research questions that will be addressed by this thesis (Section 1.4).

1.1. Phonetic reduction

The broad term *reduction* refers to patterns of variation of speech sounds in connected speech, along several phonetic parameters. It is commonly used to indicate processes that are characterised by articulatory undershoot (Lindblom, 1963; Bauer, 2008) and associated with a decrease in time and articulatory effort (Barry and Andreeva, 2001). Some of the phonetic characteristics of reduced sounds and stretches of speech are: temporal reduction, vowel centralisation and monophthongisation, more open consonant articulation, deletion of phonetic features or segments, and increase in coarticulatory features (Kohler, 1991; Gahl, Yao and Johnson, 2012). Figure 1 shows an example of a reduced stretch of speech. The utterance *but do you know the vegetation* is uttered by a speaker of RP. The first part, *but do you know* is highly reduced.

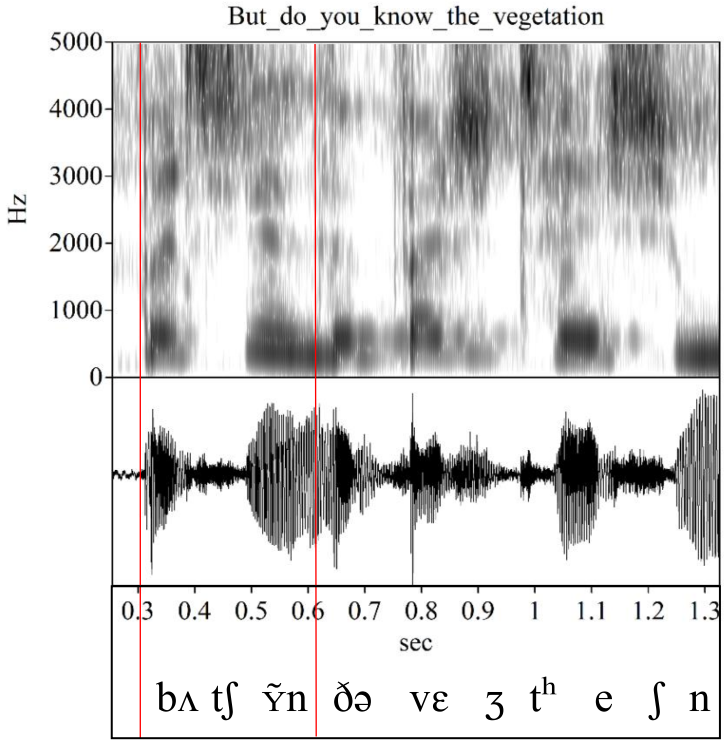


Figure 1. Spectrogram and waveform of the phrase *but do you know the vegetation* uttered by a speaker of RP English. The first three words *but do you know* are highly reduced compared to the rest of the phrase.

Figure 1 shows an example of reduction at the beginning of the phrase *but do you know the vegetation*. The red lines in the figure delimit the stretch of speech *but do you know* from the burst of /b/ in *but* to the burst of the phonological fricative /ð/ (which is realised with a burst followed by a short portion of friction). The duration of *but do you know* is 309 ms. The duration of *the vegetation* is 730 ms. The palatal approximant in *you* merges with the preceding plosive of *do*, while the vowel in *you* is realised as a short nasalised near-close near-front rounded vowel. The word *know* is realised as a dental nasal, as the nasal consonant assimilates to the place of articulation of the sound that follows. Instances of reduction such as the one described here are common in everyday speech. In fact, reduction is a widespread phenomenon cross-linguistically (Barry and Andreeva, 2001). While much of the literature has focussed on categorical processes of reduction such as deletion (see e.g. Johnson, 2004; Ernestus, 2014), reduction is a gradient rather than a categorical phenomenon (Kohler, 1999; Ernestus and Warner, 2011) – between the least and the most reduced forms of a word there is a wide range of realisations on a continuum between the two extremes (Lindblom, 1990; Bybee, 2001; Lavoie, 2002; Kohler, 1999, 2011).

This variation can be explained with reference to the principle of language economy: the tendency to reduce the articulatory effort by the speaker as long as it does not impair speech understanding by the listener (Lindblom, 1990). In Lindblom's H&H theory (where H&H refers to 'hyper'- and 'hypo'-articulation), the range of phonetic variation found in speech is explained as a "continual tug-of-war between the demands on the output on the one hand and system-based constraints on the other" (Lindblom, 1990: 420). According to the H&H theory, speakers adapt their production to the perceptual needs of the listeners (the output constraint) and in relation to the contextual information available, thus varying their production (the system-based constraint) on a continuum between hyper- and hypo-speech (Lindblom, 1990).

The same theory has been implemented by Aylett and Turk (2004) in their Smooth Signal Redundancy Hypothesis. According to their theory, communication is a balance between two types of redundancy: language redundancy – which can be summarised as contextual information – and acoustic redundancy – the acoustic information contained in the acoustic signal. In communication these two redundancies are in inverse relationship: the more contextual information is available (e.g. high predictability of a word given the context), the less acoustic information is needed (reduction can be high) and vice-versa (Aylett and Turk, 2004).

A related phenomenon is the spreading of non-segmental correlates of sounds to neighbouring sounds. This phenomenon is widespread cross-linguistically (Barry and Andreeva, 2001) and can be explained with reference to coarticulatory features: "the pervasive, systematic, reciprocal influences among contiguous and often non-contiguous speech segments" (Farnetani and Recasens, 1999: 31). Coarticulatory features can spread across several segments or even syllables, as observed by Heid and Hawkins (2000) in their study on long-domain /r/ and /l/ coarticulation. Common coarticulatory features are the characteristic colouring or resonances of sonorant consonants – nasals and liquids (see e.g. Tunley, 1999; West, 1999a, 1999b; Carter and Local, 2007); and secondary articulations such as palatalisation, velarisation, labialisation, nasalisation (Kohler, 1998, 1999; Wesener, 2001). Moreover, coarticulatory features can affect adjacent sounds even when the sound that triggers them is reduced or apparently deleted. Kohler (2011) and Niebuhr and Kohler (2011) suggest that this phenomenon is due to sounds being more weakly articulated in reduced speech. In reduced speech, the movements in the vocal tract are more loosely coordinated, so that the overlap of adjacent sounds increases, resulting in segmental features becoming *prosodic* (Kohler, 1999, 2011).

In Articulatory Phonology (Browman and Goldstein, 1989, 1990, 1992), a great deal of variation in word forms is explained in terms of two interacting factors: the "reduction in the magnitude of individual gestures" and the "increase in overlap among gestures" (Browman and Goldstein, 1989: 214). This view implies that articulatory gestures are not absent altogether, or *deleted*, but, in Browman and Goldstein's (1989) words, they are either *blended* (in the case of apparent assimilation) or *hidden* (in the case of apparent deletion). The increase in gesture overlap over time accounts for the increase in coarticulatory features typically found in patterns of reduction – the articulatory gestures of any given sound start in anticipation of the sound itself and continue after it, influencing the neighbouring sounds. The influence on adjacent sounds has important implications for speech perception, as it means that "information about that segment is available to perception longer than would be the case if all cues were confined inside its boundaries" (Kühnert and Nolan, 1999: 9). This raises the question of whether information is lost during reduction or, on the contrary, enhanced by its spreading over longer time domains. For example, in the English word *can't* the nasal consonant often undergoes apparent deletion (it is not identifiable as a time-limited segment). Some of its features, however, such as the opening of the velopharyngeal port, are still articulated and can be perceived in the nasalisation of the preceding vowel (Barry and Andreeva, 2001). This feature, which might be described as 'reduction', aids word recognition and speech understanding rather than hindering them, thus challenging the view of reduction as loss of information.

1.1.1. Intelligibility of reduced speech

Research has shown that phonetic reduction is a pervasive phenomenon in everyday speech (e.g. Johnson, 2004; Ernestus and Warner, 2011; Warner and Tucker, 2011). Although the degree of reduction is related to several aspects, such as the speaking style and rate (van Son and Pols, 1990, 1999), the speech situation (Ernestus, Hanique and Verboom, 2015), and the communicative function (Local, Kelly and Wells, 1986), extreme (or "massive") reduction, including deletion of entire syllables, is common in conversational speech (Johnson, 2004). Despite this, speech remains mainly intelligible (Ernestus and Warner, 2011). It has been shown that listeners rely on multiple types of cues for the correct interpretation of reduced speech, including the semantic and syntactic contexts (Janse and

Ernestus, 2011; van de Ven, Ernestus and Schreuder, 2012), the situational context (Aguilar and Machuca, 1995; Aguilar, 1999), and the acoustic cues remaining in the signal (Ernestus, Baayen and Schreuder, 2002; Janse, Nooteboom and Quené, 2007; Warner, Fountain and Tucker, 2009; Kohler and Niebuhr, 2011; Niebuhr and Kohler, 2011; Mitterer, 2011; Pitt, Dilley and Tat, 2011).

Work by Ernestus and colleagues focussed on the role of the context in spoken word recognition and its relation to the degree of reduction (e.g. Ernestus et al., 2002; Mitterer and Ernestus, 2006; Janse and Ernestus, 2011; van de Ven et al., 2012). Ernestus et al. (2002) showed that in spoken word recognition there is an inverse correlation between the degree of reduction and the amount of context given. They tested the recognition of low, medium and high reduction in three different degrees of contextual information: in isolation, in their phonological context (neighbouring vowels and intervening consonants), and in their full prosodic phrase or sentence. They found that while tokens with low and medium reduction could be identified in all contexts, highly reduced tokens needed their full sentential context to be correctly identified. Janse and Ernestus (2011) investigated which elements of the context facilitate word comprehension by testing the role of the acoustic cues and the semantic/syntactic information contained in the context. Although they found that both acoustic cues and the semantic/syntactic context had an effect on word recognition, the results of their experiments suggested that the role of the contextual acoustic cues is crucial for word comprehension while the semantic/syntactic context by itself does not facilitate word recognition.

The importance of the acoustic residues of reduced sounds for speech comprehension has been demonstrated by Kohler and colleagues (Kohler, 1999; Kohler and Niebuhr, 2011; Niebuhr and Kohler, 2011). Kohler and Niebuhr focussed on the role of long-domain resonances – which Kohler (1999) terms articulatory prosodies – in speech understanding. They claimed that articulatory features such as palatality and nasality, which can spread over long stretches of speech, are crucial for the correct interpretation of reduced speech. In one of their experiments, Kohler and Niebuhr (2011) manipulated the degree of palatality and the duration of nasality in a sentence such as "ich *kann Ihnen das* ja mal sagen" ('I can mention this to you'). The sentence is grammatical with or without the function word *Ihnen* (which means *to you*). By manipulating the duration of the stretch of speech *kann (Ihnen) das* and its palatality, they tested the identification of the words in the sentence without the influence of the semantic/syntactic context. That is, whether listeners interpret the sentence as containing the word *Ihnen* or not, is solely due to the acoustic cues in the signal. They found that the articulatory prosody of palatality has a strong effect on whether the utterance is interpreted as containing the word *Ihnen* or not. However, the effect of the duration of the nasal consonant was less strong. They report that "[w]hen palatality is strong […], nasal duration has very little influence on *Ihnen* judgements; when palatality is weak […] or absent […], duration can only weakly compensate for it" (Kohler and Niebuhr, 2011: 25). Niebuhr and Kohler (2011) refer to the articulatory prosodies as the 'phonetic essence' of function words. They state that while segments constitute the phonetic essence of less reduced word forms, the phonetic essence of highly reduced words is retained in their articulatory prosodies (Niebuhr and Kohler, 2011). The phonetic essence of a word constitutes its identity, which is maintained in speech regardless of the degree of reduction. The phonetic essence of a word must always be available to perception for successful lexical access. In reduced speech, the phonetic essence of a word is manifested in the articulatory prosodies, which are crucial for speech understanding (Kohler and Niebuhr, 2011).

Several linguistic and extra-linguistic factors influence the type and extent of reduction (Lavoie, 2002). The following sections describe the main linguistic factors that are known to affect the degree of reduction.

1.1.2. Linguistic factors influencing reduction

Several factors play a role in the type and degree of reduction found in casual speech. This section describes the main lexical, morpho-syntactic, phonological and prosodic factors that influence the phonetic realisation of speech sounds.

1.1.2.1. Word category

Words can be divided into two classes: function words and content words (Selkirk, 1996). While function words (also called grammatical words) have a grammatical role in language, content words (also called lexical words) carry a semantic meaning (Bell, Brenier, Gregory, Girand and Jurafsky, 2009: 92). The group of function words is a closed class and includes items such as prepositions, auxiliaries, modals, conjunctions, and determiners. The group of content words is an open class, and includes nouns, verbs and adjectives (Selkirk, 1996: 187).

Function words and content words differ in many aspects. Some of the differences that are most relevant for the present research are:

- function words do not usually receive phrasal stress, unless uttered in isolation or in contrastive position (Selkirk, 1996);
- function words have several forms that can be classified into two types – strong and weak forms – while lexical words have only one type of form (Selkirk, 1996);

- frequency and predictability of function words are higher than those of content words (Bell et al., 2009);
- the two classes undergo different connected speech processes (Local, 2003);
- function words have a wider range of realisations and can be more reduced than content words (Lavoie, 2002).

One of the most important aspects for this research is the wide range of variability found in function words, but not in content words. It has been shown that there is a wider range of phonetic variation in function words (see e.g. Manuel, 1995; Ogden 1998, 1999; Hawkins and Smith, 2001; Local, 2003) than in comparable content words (see e.g. Lavoie, 2002; Local, 2003; Baker, 2008; Clayards, Gaskell and Hawkins, 2021). An example of this is word-final /m/. As stated by Local, word-final /m/ in words such as *time* and *lime* does not assimilate to the place of articulation of a following consonant, but maintains its labiality in any phonological context (2003: 333). However, when found in a function word such as the finite form of TO BE, as in *I'm*, the nasal consonant does assimilate to a following consonant, and can thus have as many realisations as the contexts it occurs in. This issue will be discussed in more detail in Section 4.1.4. At word level, Lavoie examined the phonetic differences between phonologically similar function and content words by comparing the surface form of *him* and *hymn* in the sentences *give him a hymn* and *give a hymn to him* (Lavoie, 2002: 177). Although the underlying forms of *him* and *hymn* are both /hɪm/, their realisations in the above-mentioned sentences, as transcribed by Lavoie in broad phonemic transcription, are /gɪvm̩əhɪm/ and /gɪvəhɪmtum̩/, where the function word can be reduced to a syllabic consonant, while the content word retains all its segments (Lavoie, 2002: 177).[1]

A thorough analysis of the realisations of pairs of function and content words such as *for* and *four* in both laboratory and spontaneous speech led Lavoie to conclude that there is more variation in the phonetic realisation of function words, and that *for* "exhibits extreme reduction possibilities not seen in the number" (Lavoie, 2002: 175). Moreover, despite the traditional description of function words as having two forms – weak and strong forms – Lavoie claimed that there is not a categorical dichotomy of realisations between the two extreme forms of *for* but a continuum of surface realisations "resulting from gestural interaction" from the most reduced to the least reduced form (Lavoie, 2002: 176).

1.1.2.2. **Frequency**

Research has shown that the frequency with which an item (sound, word or phrase) recurs in speech affects its phonetic realisation (Bybee, 2001). The more frequently a word occurs, the more likely it is to undergo processes of reduction such as shortening, deletion of consonants in coda position, and vowel centralisation (see e.g. Jurafsky, Bell, Gregory and Raymond, 2001a; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory and Gildea, 2003; Bell et al., 2009). Several studies have investigated the deletion of word-final /t/ and /d/ and have found a strong correlation with word-frequency: /t/ and /d/ deletion occurs more often in high-frequency words (see Bybee, 2002; Jurafsky et al., 2001a; Gregory, Raymond, Bell, Fosler-Lussier and Jurafsky, 1999). Others have found a frequency effect on vowel

---

[1] As we will see later, besides the word category, other factors contribute to the higher degree of reduction of *him* compared to *hymn*, such as: word frequency, predictability and the systems of contrast the two words belong to.

reduction and elision (see e.g. Bybee, 2002) and on word duration (see e.g. Pluymaekers, Ernestus and Baayen, 2005a, 2005b).

The frequency effect on word production can be explained with reference to the automation of speech production. As Bybee states (2002: 268):

> "repeated neuromotor patterns become more efficient as they are practiced; transitions are smoothed by the anticipatory overlap of gestures, and unnecessary or extreme gestures decrease in magnitude or are omitted."

High-frequency words are more practiced than low-frequency words, which leads to a higher degree of gestural overlap and a lower degree of gesture magnitude – characteristic features of reduced speech.

1.1.2.3. **Predictability**

Predictability – the likelihood of a word given its lexical context – affects the phonetic realisation of words too: more predictable words have a higher degree of phonetic reduction (Jurafsky et al., 2001a; Bell et al., 2003, 2009).

The predictability of a word can be measured by taking into account either its preceding lexical context (the word immediately before it), or the following lexical context, or both preceding and following contexts (Jurafsky, Bell, Gregory and Raymond, 2001b). Jurafsky et al. (2001b) found that words with high predictability (given either their preceding or following context) are shorter, have reduced vowels and are more likely to show coda consonant deletion. Word duration is the parameter used by Bell et al. (2009) to establish the degree of word reduction in relation to the words' frequency and predictability in a large corpus of spontaneous speech. Interestingly, they found that frequency is the factor that most affects content words, while the duration of function words is more affected by contextual predictability (Bell et al., 2009). However, Gregory et al. claimed that frequency and predictability are two aspects of the same factor: "the informativeness of a word as measured by its probability" (1999: 151). They suggested that word production is adapted by speakers according to the probability of a word: more probable words have a lower informational load and can be produced with less articulatory effort, and thus undergo reduction (Gregory et al., 1999). This theory is strictly related to the idea that speakers fine-tune their productions to the needs of listeners depending (also) on the informativeness of the context (Lieberman, 1963; Lindblom, 1990). According to Lindblom (1990), speakers adapt their speech to the information available to listeners through "signal-complementary processes", such as word frequency and predictability. Assuming that information is transmitted in two forms, "signal-driven" and "signal-independent" processes, with the more information being given by the latter, the least information needs to be provided by the former (Lindblom, 1990). As stated also by Gregory et al., the "knowledge of the likelihood of words in contexts is used by speakers and affects their pronunciation" (Gregory et al., 1999: 163). In other words, speakers can allow or inhibit the activation of reduction processes on the basis of their intuition about the words' predictability (Bybee, 2002).

1.1.2.4. **Grammaticalisation**

The frequent use of combinations of words leads to their grammaticalisation (also grammaticisation) – a process whereby "a frequently repeated stretch of speech becomes automated as a processing unit" (Bybee and Scheibman, 1999: 577). During grammaticalisation, a sequence of words acquires a new grammatical (functional) or pragmatic meaning (Bybee and Scheibman, 1999) and loses part of its semantic meaning

(Heine, 2003). Grammaticalisation is strictly linked to lexical frequency and predictability. A word undergoing grammaticalisation is used in new contexts in which it was not used before, thus becoming more frequent and predictable. As a consequence, it undergoes "erosion (or 'phonetic reduction'), that is, loss in phonetic substance" (Heine, 2003: 579). A clear example is that of *be going to*, in which the original semantic meaning of motion verb is lost and the word sequence has acquired the new function of expressing tense and aspect of a verb, thus becoming extremely reduced (Heine, 2003). Other grammaticalised word combinations include *I don't know*, *why don't you*, *supposed to* (Bybee and Scheibman, 1999), *kind of*, *sort of*, *rid of*, *ought to* (Gregory et al., 1999), *used to* (Heine, 2003).

Bybee and Scheibman (1999), in a study of the phonetic realisation of *don't*, found that the most reduced form of *don't* occurs when it is preceded by the personal pronoun *I*, and that the item that occurs most frequently before *don't* is *I* (Bybee and Scheibman, 1999: 580). In other words, the most frequent combination in which *don't* occurs is also the most reduced one. As they point out, "the pronoun and *don't* constitute a storage and processing unit that is gradually undergoing reduction due to frequency of use" (1999: 581).

1.1.2.5. **Rhythmic and prosodic structure**

Prosodic factors such as stress, rhythm, and position in the prosodic structure play a crucial role in patterns of reduction. It is known that the stress placement affects the phonetic realisation of sounds and syllables, and that reduction processes are more likely to occur in weak and unstressed positions that do not bear word and phrasal stress (Shockey, 2003).

A position that is unstressed by definition is the so-called anacrusis. The concept of anacrusis was first employed in poetry and music to refer to the unstressed syllables or notes at the beginning of a verse or before the first downbeat of a piece of music (Noel, 2010). In linguistics, anacrusis is a rhythmic category that corresponds to “the sequence of weak syllables preceding the first accented syllable in the intonation group” (Cruz-Ferreira, 1998: 175). However, anacrusis is an under-investigated topic in linguistics. The only reference to anacrusis in the linguistic literature is found in studies on English isochrony (see e.g. Hill, Jassem and Witten, 1979; Bouzon and Hirst, 2004). In models of English isochrony, it is claimed that the durations of the rhythmic units (e.g. feet) in an utterance are approximately the same regardless of the number of subcomponents (e.g. syllables) within them. The only exception is anacrusis, which is “uttered as quickly as physiologically possible” (Jassem, 1952, in Bouzon and Hirst, 2004: 224). Unfortunately, in these studies, anacrusis is not investigated in detail. The only characteristics that emerge are that speech in anacrusis is less isochronous than the speech in other rhythmic units, and that there is some degree of compression in anacrusis – e.g. syllables are shorter as a function of the number of syllables per anacrusis (Bouzon and Hirst, 2004).

The spectrogram and waveform in Figure 2 provide an example of anacrusis. The sentence *they’ve got a day left* is segmented into two parts: the first part is the anacrusis (*they’ve got a*), followed by the rest of the sentence (*day left*).

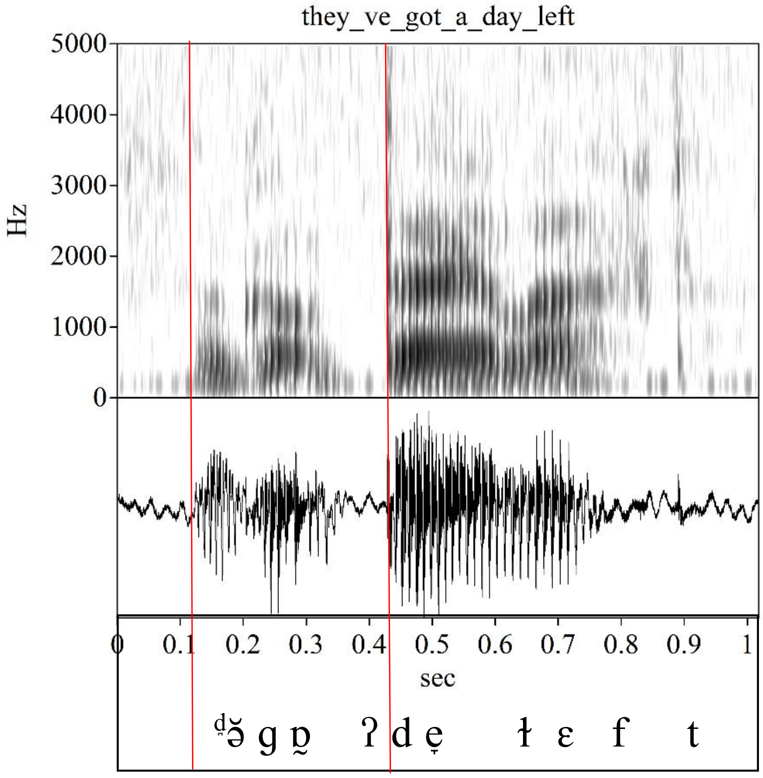


Figure 2. Spectrogram and waveform of *they've got a day left*.

In this example, the duration of the anacrusis (*they've got a*) is 306 ms (including the hold phase of the following plosive, as it is not possible to separate it from the preceding glottal stop). The duration of the following two syllables is 469 ms (starting from the burst of the initial plosive, thus excluding its hold phase). Particularly interesting is the temporal compression of the pronoun and auxiliary at the beginning of the sentence – *they've* is reduced to a short vowel, although a hint of dentality can be perceived at the beginning of the vowel, but is not identifiable in the spectrogram.

A well-studied phenomenon that seems to contrast with the idea that speech in anacrusis is highly reduced is domain-initial strengthening. Several studies have investigated the effects that the position of a sound in the prosodic structure has on the phonetic realisation of the sound itself (see e.g. Pierrehumbert and Talkin, 1992; Dilley, Shattuck-Hufnagel and Ostendorf, 1996; Fougeron and Keating, 1997). As explained by Cho, McQueen and Cox

(2007: 211), “prosodic strengthening […] can be defined as a spatio-temporal expansion of articulation and a coarticulatory resistance”. Research has shown that speech sounds are more forcefully articulated (Fougeron and Keating, 1997) at the beginning of a word, e.g. /t/ in English is more heavily aspirated in word-initial position (Byrd, 1996); at the beginning of a phrase, e.g. /n/ is articulated with more linguo-palatal contact in American English (Fougeron and Keating, 1997) and in French (Fougeron, 2001); and at the beginning of an utterance, e.g. rounded vowels are produced with more lip-rounding in Dutch (van Lieshout, Starkweather, Hulstijn and Peters, 1995). Adding to this, Fougeron and Keating, among others, claim that “consonant initial strengthening is generally cumulative, i.e., the higher the prosodic domain, the more linguopalatal contact the consonant has” (1997: 3728). Lastly, one more point in favour of domain-initial strengthening is the role of this position in spoken word recognition and speech understanding (see e.g. Nooteboom, 1981; Cho et al., 2007). As claimed by Kohler (1991: 189), “[t]he syllable or word-initial position has a higher signalling value for a listener and must therefore be given a more precise articulation by a speaker”. The question being raised here is how the two aspects of reduction in anacrusis and domain-initial strengthening can co-occur. A possible explanation lies in the type of experiments carried out on speech material occurring in initial positions. First, most of the studies on domain-initial strengthening focussed on the articulatory and acoustic properties of single segments rather than on words or syllables; therefore, they analysed the first consonant (or consonant cluster) and only sometimes did they go so far as to analyse the first vowel. Second, most studies looked at content words rather than function words; however, the anacrusis is more likely to be occupied by function words or fillers. Third, most – but not all – studies of initial positions analysed stressed syllables rather than unstressed ones, revealing that “boundary effects are crucially enhanced by the presence of pitch-accent" and "strengthening [is] due to accentuation or focus" (Cho et al., 2007: 227). Fourth, some studies of domain-initial strengthening did not analyse the beginning of an utterance, but looked instead at the beginning of other prosodic domains, such as the intonational phrase. For example, Cho et al. (2007: 236) compared the beginning of *ticket* in the sentences: *When you get on the bus, tickets should be shown to the driver* and *John bought several bus tickets for his family*. Fifth, Cho et al.'s experiments on lexical access and segmentation did not provide evidence that the domain-initial strengthening affects word recognition. In fact, they concluded that "[l]exical segmentation depends on a multitude of factors, including the effective set of lexical competitors" (Cho et al., 2007: 228).

To summarise, it has been shown that several factors play a role in domain-initial strengthening and that this process is probably not applicable to all items occurring in initial position. However, further research is needed before a satisfactory explanation of the co-occurrence of reduction in anacrusis and domain-initial strengthening can be found. The present research will contribute by providing an insight into the nature of speech in anacrusis.

1.1.2.6. **Summary of factors influencing reduction**

This chapter so far has explored the lexical, morpho-syntactic and prosodic factors that influence patterns of phonetic reduction in connected speech. It was shown that function words exhibit a higher degree of reduction and a wider range of phonetic realisations than content words (Section 1.1.2.1). It was explained that lexical frequency (Section 1.1.2.2) and predictability (Section 1.1.2.3) play a crucial role in the process of grammaticalisation, in which a stretch of speech is reinterpreted as a single phonological unit; this is the case, for example, in sequences of a personal pronoun and an auxiliary verb such as *I don't* (Sections 1.1.2.4). Finally, it was shown that words in unstressed positions are more reduced than words in stressed positions or carrying pitch accent (Section 1.1.2.5).

In order to investigate phonetic reduction, a set of words that correspond to all the factors just mentioned was selected: personal pronouns and contracted auxiliary verbs combinations, such as *she's*, *I'll*, *we'd*, etc. English personal pronouns and auxiliary verbs belong to the class of function words, are very frequent and predictable given the context, and being frequently used in combinations they become grammaticalised. Moreover, combinations of personal pronouns and auxiliary verbs are usually unstressed and do not receive phrasal stress, unless in contrastive position. To ensure that the personal pronoun and auxiliary verbs under examination occur in unstressed position, they were analysed in anacrusis position.

The next section explores the relevant literature on English auxiliary verbs, describing their main phonetic and phonological features (Section 1.2.1), and the relationship between their weak and strong forms (Section 1.2.2).

1.2. English auxiliary verbs

This section describes the characteristics of English Auxiliary Verbs (hereafter EAVs) and some of the theories on the relationship between their strong and weak forms. Two relevant issues are also discussed: the combination of personal pronouns and auxiliaries and the minimal phonetic features of auxiliaries that convey grammatical information.

1.2.1. Weak and strong forms of auxiliary verbs

English auxiliary verbs, like several (but not all) monosyllabic function words, have two kinds of forms: weak forms and strong forms (Selkirk, 1996). In traditional accounts, the weak forms are described as reduced versions of the strong forms along several parameters. Specifically, weak forms are characterised by shortening (reduction in duration), vowel centralisation (reduction in magnitude of the articulatory gestures) and sound elision (segmental reduction) (Cruttenden, 2008).

On a broader linguistic level, the main differences between EAV weak and strong forms are:

- the position they can occur in (their distribution);
- whether they can be stressed;
- the set of vowels that can occur in the nucleus;
- the permitted syllable structures;
- whether they can form the negative forms.

(from Kaisse, 1985; Simpson, 1992; Selkirk, 1996; Ogden, 1999)

Syntactically, weak and strong forms are not interchangeable as they do not occur in the same contexts (Kaisse, 1985). It can be said that weak and strong forms are in complementary distribution: in isolation, in contrastive position and in sentence-final position only strong forms can occur, while in all the other environments only weak forms can occur (Selkirk, 1996).

Prosodically, weak forms are “stressless and reduced”, while strong forms are “stressed and unreduced” (Selkirk, 1996: 200) (but see Palmer, 1965, for a different view). As a consequence, strong forms have full vowels, while weak forms can only have the vowel /ə/ and in a few cases /ɪ/ (Palmer, 1965; Ogden, 1999; however, Cruttenden, 2008, includes also /ʊ/).

Morphologically, the negative forms with the contraction <n't>, such as *aren't*, *don't*, *mustn't*, can be formed only by the strong forms, and not by the weak forms of auxiliaries (Ogden, 1999). The formation of some negative forms is more complex than just the addition of the finite auxiliary and the contraction <n't> – as in forms such as *can't* and *won't* – but this aspect is not covered here as it is not investigated in the present research.

1.2.2. Relationship between weak and strong forms

The relationship between weak and strong forms is a matter of debate. One way to relate the two types of forms is to treat the weak forms as deriving from the strong forms (Zwicky, 1970). The derivation can be explained with reference to the Auxiliary Reduction (AR) process (Zwicky, 1970; Kaisse, 1983), which involves both phonological and syntactic rules (Lakoff, 1970) and results in the contracted forms of auxiliaries, such as *'m*, *'ve*, *'ll*. According to Zwicky (1970), the auxiliaries that undergo AR – *will*, *would* and the finite forms of BE and HAVE – can be derived from their full forms by applying an ordered set of phonological rules: glide deletion and word-initial unstressed vowel deletion. For example, applying these rules to *will*, it first becomes **ill*, and subsequently *'ll*; the latter form must then be attached to a pronoun, to obtain a form such as *they'll*.

One of the main arguments against a derivational account is that there is no rule in English that deletes word-initial /w/ in all items in the lexicon (Kaisse, 1983, 1985). Even considering only a restricted word class, such as that of function words, *w*-Deletion never affects *what*, *why*, *with*, *was* and *were* (Kaisse, 1983; Ogden, 1999; Simpson, 1992). Kaisse also rejects the idea that AR is a phonological rule, stating that AR is rather a set of rules that indicates in which contexts the contracted forms can be used (Kaisse, 1983, 1985).

Kaisse treats Auxiliary Reduction as a case of simple cliticisation (Zwicky, 1977 in Kaisse, 1985): "a syntactic or morphological adjunction of some grammatical word to a 'host'" (Kaisse, 1985: 39). In her view, the contracted forms of auxiliaries are clitics: grammatical elements that are not independent but need to be attached to a word or phrase. A consequence of the process of cliticisation is the realisation of the "host-clitic group" as a "phonological unit" (Kaisse, 1985: 39). Moreover, according to Kaisse, the contracted auxiliary forms do not derive from their strong counterparts, but are "suppletive allomorphs, to be listed in the lexicon alongside the full forms" (Kaisse, 1983: 95). However, as stated by Ogden (1999), treating the weak and strong forms as suppletive allomorphs fails to acknowledge the connections and similarities between them.

Ogden (1999) provides an account of EAVs that highlights the phonological relationship between the various weak and strong forms. Using a polysystemic approach, he suggests dividing EAVs into four subsystems, depending on which forms each auxiliary can take. Each subsystem has its own phonology which accounts for the various forms in it. The four forms an auxiliary can take are (exemplified using *have*):

- strong forms – /CVV(C)/, e.g. /hav/
- (weak) syllabic forms with onset – /Cə(C(C))/, e.g. /həv/
- (weak) syllabic forms without onset – /əC/, e.g. /əv/
- (weak) non-syllabic forms – /C/, e.g. /v/

(Ogden, 1999: 67)

The distribution of syllabic and non-syllabic clitics is different: while syllabic clitics can be attached to non-pronoun hosts, non-syllabic clitics must be attached to a pronoun (with the exception of /s, z/) (Ogden, 1999: 70). Ogden, like Kaisse, supports the view that the pronoun and the auxiliary function as host and clitic and become "phonologically fused"

(Ogden, 1999: 75). Once again the tight relationship between pronouns and auxiliaries emerges.

This close relationship is not restricted to the contracted forms such as *I'm*, *you've*, *they'll*. In their study of *don't*, Bybee and Scheibman (1999) analysed the correlation between frequency of occurrence and phonetic realisation of *don't*, and found that the auxiliary is more tightly connected to the pronoun that precedes it than to the verb that follows it. They related this phenomenon to the process of grammaticalisation (see Section 1.1.2.4). However, it can be explained also with reference to the syntagmatic relations between words in a sequence. In the case of *don't* – whose context is highly constrained – Bybee and Scheibman (1999) observed that in their data, the range of lexical items occurring after *don't* is more than double the items that occur before it. As they stated, the "use of elements in sequence strengthens their syntagmatic relations" and the "positions with the fewest options are the most fused phonologically […]" (1999: 578). This is in accordance with the claim that there is a "general tendency in English for auxiliaries to lean on the subject, as evidenced by contractions such as *'ll*, *'s*, *'ve*, *'d*, *'re*". (Bybee and Scheibman, 1999: 591).

1.2.3. Phonetic features and grammatical information

One more factor that contributes to the wide range of variation in the realisation of EAVs is their belonging to a small paradigmatic system: a restricted set of items in paradigmatic relation, which means that the items that can occupy the auxiliary position are only a few. Considering again the example of word-final /m/ in *I'm* and *lime*, Local observes that *I'm* contrasts only with other "pronoun + nonpast forms of the verb *be*" and none of them contains a nasal consonant (Local, 2003: 334). Even if we extend the paradigmatic system of items that can occur after the pronoun *I* to include the auxiliary HAVE, only a few items can occur in that position. Table 1 gives an example of the small number of items in the paradigmatic system of auxiliaries that can occur after *I*.

I	’m	told
	’ve	
	’d	
	Ø	

Table 1. Paradigmatic system of contracted auxiliaries that can occur between the pronoun *I* and a past participle.

Table 1 shows that only three auxiliaries (or no auxiliary) can occupy the position between the personal pronoun *I* and a past participle. Of these three auxiliaries, only one contains a nasal consonant. This means that the place of articulation of the nasal is not critical to the identity of the auxiliary *’m* in this system. As the only nasal, nasality alone is enough to maintain the identity of the auxiliary, and changing its place of articulation does not change the meaning of the word. In the class of content words, however, a word such as *lime* contrasts, for example, with *line*. This means that in the case of content words, changing the place of articulation of the nasal consonant from bilabial to alveolar changes the meaning of the word.

This is due not only to the restricted paradigmatic system of auxiliaries, but also to the small set of sounds (or phonetic features) used in this word class. In fact, EAVs are “very economical in their exploitation” of the sound system and phonotactic possibilities of English (Simpson, 1992: 211). In the consonant system of EAVs, in coda position, /m/ is the only nasal consonant; all the other consonants in coda position are oral and are articulated at the alveolar ridge (e.g. *are*[2], *is*, *could*, *must*), except *have*. In both onset and coda position, only singleton consonants occur, except the consonant cluster /st/ in *must*. This economical set of sounds (and sound combinations), however, is sufficient to express the grammatical information provided by the auxiliaries: aspect, tense, number, and person. Dividing the system of auxiliaries into subsystems, a clear picture of the correlations between phonetic features and grammatical information emerges. In the subsystem of finite forms of BE, which is the most complex subsystem (Simpson, 1992), three parameters correspond to the three present tense forms *am*, *is*, *are*. As stated by Local (2003: 334), this is "a three-term system of contrast where the contrasts involve parameters of nasality (as in *I'm*), centrality (as in *you're*) and friction (as in *s/he/it's*)". As for the relation between past and present, Simpson (1992: 213) states that it "can be seen as being a phonological one of $^{y\text{-}w}$", where ^{y} stands for "non-rounding" and ^{w} stands for "backness and rounding". The tense aspect is expressed also in the modal auxiliaries, which, for the majority, do not express person and number, except for the auxiliary DO. In the modal auxiliaries, the tense aspect is expressed by the "stoppedness" at the end of the past tense (*did*, *could*, *should*, *would*) and again by the opposition between ^{y} and ^{w}, so that if the present tense form has ^{y}, its past counterpart has ^{w}, and vice versa, as in *do-did* (but *can-could*). Following the Firthian tradition (see Section 1.3.1), Simpson (1992: 214) represents the auxiliary through formulae which aim to express the grammatical relation between the various verb forms. For example, he represents *will* and *would* as ^{y}WcL and ^{w}WcD[3], where the symbols represent "complexes of phonetic parameters having varying temporal extents" and the

---

[2] In contexts and accents in which /r/ is realised, such as in rhotic accents and in pre-vocalic position in non-rhotic accents.

[3] The symbol "c" stands for "close vocalicity" (Simpson, 1992: 215).

superscripts represent phonetic properties that extend over the whole syllable (Simpson, 1992: 213).

1.2.4. Summary of English auxiliary verbs features

To summarise, English auxiliary verbs have strong and weak forms. These two kinds of forms occur in complementary syntactic contexts and differ in prosodic and morphological features. For this reason an account that derives the weak forms from the strong forms is unsatisfactory. A more suitable approach is to treat the weak forms as belonging to different phonological systems according to their forms, and to treat non-syllabic forms, which can be attached only to pronouns, as cliticised forms. The relationship between a pronoun host and a cliticised form of auxiliary is a close one that leads to the phonological fusion and grammaticalisation of the pronoun and the cliticised form of the auxiliary. Finally, personal pronouns and auxiliary verbs belong to two small paradigmatic systems, which means that they contrast with very few items. For this reason, they exhibit a wider range of phonetic variation that content words and can undergo a high degree of reduction without undermining speech comprehension.

The next section describes the theoretical approach of this thesis.

1.3. Theoretical approach

Reduced speech is characterised by a high degree of apparent deletion of sounds and sound sequences. Most traditional phonemic accounts of reduction focus on processes of deletion and quantify the degree of reduction by counting missing segments or syllables (see e.g. Guy, 1991; Greenberg, 1999; Johnson, 2004; Pluymaekers et al., 2005a; Raymond, Dautrincourt and Hume, 2006; Schuppler, Ernestus, Scharenborg and Boves, 2011; Van de

Ven et al., 2012; Ernestus, 2014). However, as mentioned in Section 1.1, reduction is a gradient rather than a categorical phenomenon (Kohler, 1999; Ernestus and Warner, 2011), and even in cases of extreme reduction in which segments and syllables are apparently missing, speech remains intelligible (Ernestus and Warner, 2011). As stated by Ernestus and Smith (2018: 130) "in extreme cases of reduction it may be impossible to linearly segment the speech signal, yet sufficient phonetic residue of a word's form may remain as to make it fully identifiable". For this reason, an account of reduction that is exclusively segmental and quantitative leaves many questions unanswered. For example, it fails to explain why reduction does not hinder the intelligibility of speech. If something is 'missing' in the acoustic output, how does human communication remain intelligible? A non-segmental approach like the one adopted in this thesis is more suitable to answering this question.

The following sections explain the theoretical framework of this thesis, starting from a brief description of the core principles of Firthian Prosodic Analysis that are relevant for the present work, the importance of fine phonetic detail, and the role of articulatory prosodies in speech understanding.

1.3.1. Firthian Prosodic Analysis

This thesis draws heavily on theoretical concepts from Firthian Prosodic Analysis (FPA) (Firth, 1948; Palmer, 1968, 1970; Ogden and Local, 1994; Ogden, 2012). Firthian Prosodic Analysis (FPA) is a phonological approach to studying languages that is fundamentally different from mainstream approaches such as the ones of the generative tradition (e.g. Chomsky and Halle, 1968). One of the concepts of FPA that is crucial for the present research is that of polysystematicity. In FPA's view, language can be treated as a set of interacting systems that can be analysed independently. Each system has its own phonological categories. While phonemic approaches rely on the concept of allophones to explain phonetic variation, FPA explains systematic phonetic variation with reference to the system lexical (and morphological) items belong to. This means that the same sequence of sounds has different phonetic exponents in different places in the structure. For example, Smith, Baker and Hawkins (2012) showed that phonetic detail reflects the morphological structure of words. In their experiment, they showed that the same sequence of phonemes exhibits different phonetic features according to its morphological status. They compared the phonetic features of sequences of sounds such as /dɪs/ and /mɪs/ in true prefixed words such as *mistype* and *discolour*, against the same phonemic sequences in pseudo-prefixed words such as *mistake* and *discover*. Their acoustic analysis revealed that the duration and spectral qualities of the prefix and pseudo-prefix differ as a consequence of the morphological word type. In FPA, this can be explained with reference to the systems the two categories of words belong to. True prefixes belong to a system that has different phonetic features from those of the terms in the pseudo-prefix system.

The view of language as polysystemic is crucial for the present work as it makes it possible to treat the weak forms of auxiliary verbs as belonging to a separate system. Instead of treating the weak forms as subordinate to, and derived from, the strong forms, a polysystemic approach explains the variation between strong and weak forms (and between the various types of weak forms) as a function of the system they belong to. The two forms are still related, but one is not derived from the other. As explained in Section 1.2.2, strong and weak forms occur in complementary positions in the structure or fulfil a different linguistic function. Each system has its own phonological features leading to different phonetic realisations from those in other systems.

One of the core principles of FPA is the focus placed on syntagmatic relations (Firth, 1948). While in traditional phonemic theories prominence is given to paradigmatic relations between contrasting elements (e.g. phonemes), FPA claims that syntagmatic relations come first. In FPA terms, syntagmatic relations are related to the structure of a language, while paradigmatic relations are related to the systems. The structure must be established first, before the systems, which operate in the structure, can be analysed. The paradigmatic relations, on the other hand, establish the systems, which "are, essentially, permutations: closed sets of elements operating at places in structure" (Kelly and Local, 1989: 140).

Intrinsically linked to syntagmatic relations is the concept of prosodies, which gives the name to FPA. Prosodies are phonological elements or entities that characterise and/or delimit stretches of speech. Prosodies operate in specific places in the structure; they are "phonological units which handle syntagmatic relations" (Ogden, 2012: 201). An example of prosodies can be found in the phonetic detail of the lateral consonant in onset of *led* and *let*. Hawkins and Nguyen (2004) showed that the lateral consonants in *led* and *let* differ in duration and spectral qualities depending on whether the consonant in syllable coda is voiced or voiceless. While a segmental approach cannot account for the spectral and durational differences in the lateral consonant in onset, FPA treats the coda voicing as a rhyme-level contrast, which has exponents over the whole syllable (Ogden, Hawkins, House, Huckvale, Local, Carter, Dankovicová and Heid, 2000). That is, the phonetic detail of the syllable is the effect of the 'voicing' feature of the rhyme.

1.3.2. Articulatory Prosodies

Used with a subtly different meaning, the concept of prosodies is central also in the work of Kohler (1999) and colleagues (Niebuhr and Kohler, 2011; Kohler and Niebuhr, 2011) on the production and perception of reduced function words in German. In his work on reduction, Kohler uses the term ‘articulatory prosodies’ to refer to articulatory residues that remain in the signal in reduced word forms. In his view, articulatory prosodies are distributed, long-domain features that are not “tied to specific segmental units’’ (Kohler, 1999: 89). Kohler and Niebuhr investigate the reduction of German function word *Ihnen* (Kohler and Niebuhr, 2011) and the modal particle *eigentlich* (Niebuhr and Kohler, 2011) in spontaneous speech. They claim that reduced word forms are characterised by ‘articulatory prosodies’ which extend over stretches of speech longer than the segment. Such articulatory prosodies are features such as nasality, palatality, labiality and glottality. According to Niebuhr and Kohler (2011), one of the roles of the articulatory prosodies is to retain the ‘phonetic essence’ of the word. They describe the phonetic essence of function words as their identity. That is, when words are reduced and segments cannot be identified due to reduction, the identity of the word is maintained by their articulatory prosodies. These articulatory prosodies are not temporally delimited, and, crucially, they are available to perception even when the segments they are tied to are not identifiable. Niebuhr and Kohler (2011: 319) share the view that a segmental approach is unsuitable for the analysis of reduction as it

> “runs into conceptual problems when the distinctive features of vowels and consonants as well as their assimilation or elision are no longer linearly segmentable […], and when phoneme strings, which may extend beyond syllables to whole words, need to be marked as deleted qua segmental units although the signal portion is still recognized as containing the full lexical information in the utterance context.”

Last but not least, an essential element of the non-segmental approach adopted in this thesis is the importance of fine phonetic detail, as described in the next section.

1.3.3. Fine phonetic detail

In the literature, the term fine phonetic detail has been used to refer to two slightly different types of phonetic phenomena. In its broader use, fine phonetic detail refers to any, usually local, phonetic feature that is not involved in maintaining the lexical contrast and it is thus considered irrelevant in phonemic theories (Hawkins, 2010). The other meaning of the term is used to refer to the systematic, usually distributed, variations observed in speech that convey grammatical, linguistic and conversational meaning (Hawkins, 2010). As explained by Local (2003) and Hawkins (2003) among others, speech is rich in phonetic detail that conveys meaning beyond the lexical semantic meaning of phonemes. Phonetic detail provides information about grammatical and communicative aspects of speech, such as the prosodic structure of utterances (Fougeron and Keating, 1997; Cho, 2016), the lexical category of words (Local, 2003), their morphological structure (Smith et al., 2012; Clayards et al., 2021), their pragmatic function (Plug, 2005) and more. In this thesis, the term fine phonetic detail is used with this second meaning to refer to distributed phonetic features that convey grammatical information or function.

Fine phonetic detail, which is often overlooked by segmental approaches, has been shown to be available to perception and used for the correct interpretation of speech (Local, 2003; Hawkins, 2003a, 2003b). Traditional phonological theories focus on phonemic contrast and consider phonetic detail 'acoustic noise' (see e.g. Luce, Pisoni and Goldinger, 1990). However, it has been shown that this kind of phonetic detail that provides grammatical, lexical, and interactional information, is available and is used in speech understanding. Several studies have investigated the acoustic phonetic differences between stretches of speech that have the same sequence of phonemes, but signal information such as grammatical function or structure. For example, Davis, Marslen-Wilson and Gaskell (2002) provided evidence that phonetic detail is used in lexical access. They investigated the role of acoustic cues in spoken word recognition in the same sequence of phonemes in three different word combinations: onset-embedded words (e.g. *captain*), short words in matching context (e.g. *cap tucked*) and mismatching context (e.g. *cap looking*). They found that listeners use the phonetic detail available at the beginning of these words to disambiguate between them. Their research demonstrated that phonetic detail provides important linguistic information that is perceptually salient and helps word recognition.

To summarise, the approach of this thesis is non-segmental and attentive to fine phonetic detail. A phonological approach such as Firthian Prosodic Analysis can explain the variation in fine phonetic detail that is related to the linguistic structure and meaning. As shown by Local (2003) and Hawkins (2003) amongst others, phonetic detail is rich in information and conveys meaning beyond the contrastive lexical meaning assigned to phonemes. This type of phonetic detail cannot be explained with a segmental approach which overlooks it. As summarised by Hawkins (2003: 374):

> "some of the important details are not readily accommodated by standard phonological-linguistic units; yet when they are systematically reflected in the speech signal, they too can be crucially important to communication. […] Firthian prosodic analysis (FPA) has the potential to systematize fine (and not-so-fine) phonetic detail into richly specified linguistic structures that represent the salient contrasts of speech used interactively."

1.4. Research questions

One of the aims of this thesis is to investigate the phonetic detail that helps maintain the grammatical contrast between minimal pairs (e.g. *she'd* and *she'll*) in reduced speech. Although it might seem a contradiction to look at phonetic detail and prosodic features beyond the phonemes to investigate the contrast between minimal pairs, the rationale for looking at fine phonetic detail is that in reduced speech, segments are apparently deleted or might undergo modifications, as in the case of assimilation. It is precisely in these instances, when the segmental features that (are thought to) maintain the contrast are apparently missing, that fine phonetic detail plays a crucial role in the intelligibility of speech. This approach is based on two fundamental assumptions. Firstly, that reduced speech is intelligible and that reduction does not constitute loss of information and does not hinder communication. Secondly, that the grammatical, linguistic, and communicative information is conveyed by fine phonetic detail, rather than by segments which might be apparently missing, especially in reduced speech.

Without disregarding the importance of linguistic and extra-linguistic contexts for the correct interpretation of speech, this research focusses on the phonetic features of pronoun and auxiliary combinations that play a role in maintaining speech intelligibility. In particular, starting from the assumption that fine phonetic detail conveys crucial linguistic information, which is available to perception even in reduced speech, this research aims to address the following questions:

1. Are function words characterised by fine phonetic detail that remains in the signal in reduced speech?
   - What are the main phonetic features of function words that remain in the signal in highly reduced speech?
2. Is the paradigmatic contrast between function words maintained by fine phonetic detail?
3. Is the fine phonetic detail that maintains the contrast in reduced speech available to perception and sufficient for the correct identification of words in reduced speech?

At the same time, through a detailed qualitative analysis, the research reported in this thesis aims to describe the range of variation in the realisation of pronoun and auxiliary combinations.

By analysing the phonetic features of reduced function words and the perception of fine phonetic detail in the same words, this research combines two aspects that so far have been investigated only separately. As described in this chapter, research has focussed either on the phonetic variation and the fine phonetic detail of function words (e.g. Lavoie, 2002; Local, 2003), or on the perception of reduced speech (e.g. Ernestus and colleagues; Kohler and colleagues) and fine phonetic detail (e.g. Hawkins and colleagues). The present research starts with the auditory and acoustic analysis of original material; follows with the identification of the phonetic features that distinguish minimal pairs of reduced function words; and concludes with the investigation of the perceptual salience of these phonetic features and their role in spoken word recognition.

The thesis is divided into two parts. The first part reports on the auditory and acoustic analysis of the data collected in a production study (Chapters 2-6). Chapter 2 describes the methodology of the data collection and the analyses carried out. Chapters 3 and 4 report on the variation observed in the pronoun and auxiliary combinations analysed, with a focus on the fine phonetic detail and essential phonetic features that characterise them. Chapter 5 investigates the contrast between pairs of paradigms focussing on the acoustic features that maintain the contrast in reduced speech. Chapter 6 describes some aspects of the variability observed in reduced speech.

The second part of the thesis focusses on the perception of reduced pronoun and auxiliary combinations. Chapter 7 reports on a perception experiment that investigates the intelligibility of highly reduced pronoun and auxiliary combinations. Chapter 8 reports on a perception experiment that investigates the role of two selected acoustic parameters in the correct identification of pronoun and auxiliary paradigms.

2. Methodology

This chapter describes the methodology of the production experiment and the analysis of the data collected. It is divided into three parts. The first part describes the reasoning behind the type of material recorded (Section 2.1.1), the elicitation material created (Section 2.1.2), the procedures of the experiment, and the participants who took part (Section 2.1.3). The second part describes the type of qualitative analysis carried out on the data with reference to the theoretical approach described in the first chapter (Section 2.2.1). It explains the conventions of the phonetic transcription that accompanies the figures (Section 2.2.1.3.1) and the terminology used in this thesis (Section 2.2.1.3.2). The third part describes the quantitative analysis carried out on the data. It explains the annotation of the recorded material (Section 2.2.2.1), the acoustic analysis (Section 2.2.2.2), and the statistical analysis (Section 2.2.2.3).

2.1. Data collection

The production study had three main aims. The first aim was to identify and describe the range of variation in the realisation of pronoun and auxiliary combinations (hereafter pr + aux) in anacrusis. As explained in Chapter 1, function words exhibit a wider range of variation than content words. The second aim was to establish whether pr + aux are characterised by fine phonetic detail that remain in the signal when speech is highly reduced and 'segments' are apparently lost. That is, whether pr + aux are characterised by essential phonetic features that are always present even in highly reduced speech. The third aim was to determine whether this fine phonetic detail helps maintain the contrast between minimal pairs of paradigms in reduced speech. For this third aim, minimal pairs of paradigms had to be acoustically analysed and compared. In light of these three aims, a production experiment was carefully designed. The following sections explain the design of the experiment and the audio material recorded.

2.1.1. Experiment design

The aim of the research was to analyse reduced speech and to investigate the role of fine phonetic detail in the intelligibility of reduced speech. For this reason, as explained in Chapter 1, it was decided to analyse pronoun and auxiliary combinations in anacrusis.

In order to investigate the fine phonetic details of reduced word form, the data collected had to be of excellent acoustic quality. For this reason, a carefully-designed production experiment was carried out in which all the factors that are known to correlate with reduction were controlled for. These include preceding and following phonological context, stress and rhythmic structure of the sentences, influence from neighbouring sounds, and speech rate.

2.1.1.1. Spontaneous versus read speech

Initially, this project aimed at analysing pr + aux in spontaneous speech. The rationale behind analysing spontaneous speech was that it is known that the style of speech influences the amount of reduction (Browman and Goldstein, 1990; Shockey, 2003; Johnson, 2004; Dilley and Pitt, 2007), with spontaneous speech exhibiting the highest degree of reduction. At first, several types of already recorded audio material from corpora available were examined, including radio programmes, telephone conversations, and face-to-face conversations. However, despite the large variety of material considered, it was not possible to find all the pr + aux combinations needed in any one of the corpora examined. Moreover, for an accurate acoustic analysis and comparison of the phonetic detail of pr + aux, it was important to control for the phonological context (preceding and following environment) and the prosodic structure (stress and rhythm) of the sentences. This was not possible using spontaneous speech. Moreover, it has been shown that reduction occurs in all styles of speech, including formal read speech (Cole and Shattuck-Hufnagel, 2018), careful speech (see Warner and Tucker, 2011), and in situations "that do not involve a communicative partner" such as laboratory settings (Clopper and Turnbull, 2018: 35). For these reasons, it was decided to elicit the material in a laboratory setting under carefully controlled conditions. Although the degree of reduction cannot be expected to be as high in read laboratory speech as in spontaneous speech, several methods were used to elicit the most reduced speech possible. On the whole, it was thought that the benefits of recording all the paradigms with high acoustic quality outweighed the downsides of renouncing speech in its natural use and possibly a higher degree of reduction. Using scripted sentences, it was possible to elicit all the target items needed and to control for the phonological context, the prosodic structure, and the speech rate. The following sections describe the design and procedures of the elicitation task.

2.1.1.2. **Target items (pr + aux)**

The aim of the study was to collect all paradigms of pr + aux in unstressed position in declarative sentences. In order to investigate reduction, it was decided to leave out those modal auxiliaries that have only a strong form and no weak forms: *need*, *may*, *might*, *dare*, *used* and *ought* (Simpson, 1992: 218). Initially, the auxiliaries *must*, *should*, *can* and *could* were included in the data. However, it was noticed in the recordings of the pilot study that these auxiliaries were usually stressed. Even when the speakers were given clear instructions about which word had to be stressed in each sentence, this was often not achieved in the case of sentences such as *you must take a bath*. The auxiliaries *must* and *should* were therefore discarded. Table 2 shows all the pr + aux that were recorded.

	BE **present**	BE **past**	HAVE **present**	HAVE **past**	WILL **(present)**	WOULD **(past)**
I	I'm	I was	I've	I'd	I'll	I'd
she	she's	she was	she's	she'd	she'll	she'd
he	he's	he was	he's	he'd	he'll	he'd
it	it's	it was	it's	it'd	it'll	it'd
you	you're	you were	you've	you'd	you'll	you'd
we	we're	we were	we've	we'd	we'll	we'd
they	they're	they were	they've	they'd	they'll	they'd

Table 2. Paradigms included in the production study and recorded.

For the acoustic comparison of the pr + aux, all the target items had to be in the same phonological context and the sentences uttered using the same rhythmic pattern. The next section explains how this was achieved.

2.1.1.3. Elicitation material

There were two main challenges in the experimental design of the production study. The first challenge was to record all the pr + aux combinations in the same phonological context and with the same prosodic structure to make them acoustically comparable. To overcome this challenge, all the pr + aux were placed in the same sentence frame, and clear instructions about the rhythm of the sentence were given to the speakers. The speakers were told that the focus of the sentence was on the last word and to place the main stress of the sentence on that word. The second challenge was to trigger an adequate degree of reduction. To overcome this challenge, the metrical structure was controlled for and the pr + aux placed in anacrusis. Furthermore, each sentence was repeated five times to trigger reduction. All these aspects are described in detail in the following sections.

2.1.1.4. Prosodic structure

To elicit reduced pronoun and auxiliary combinations, the position of the pr + aux in the sentences and the prosodic structure of the sentences were carefully constructed.

Articulatory realisation depends heavily on the position of an element in a phrase, so that elements such as function words that have more than one form (weak and strong forms) are produced differently according to the prosodic position they occupy in a sentence and the degree of stress they receive. Besides controlling for the position and stress of the target items, the rhythmic structure of the whole utterance had to be controlled for, as it can affect all the elements in the utterance.

All sentences were constructed to have a pronoun and auxiliary combination in anacrusis (initial unstressed position), followed by a monosyllabic verb, followed by an unstressed element such as a preposition, a pronoun, an article or an affix, and finally a monosyllabic noun. The desired rhythmic pattern is shown in Figure 3 (using Jill House' model of prosodic hierarchy, see Ogden et al., 2000).

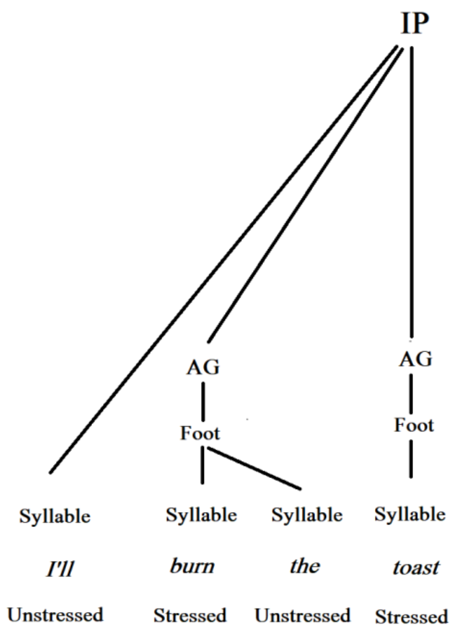


Figure 3. Desired rhythmic pattern of the sentences.

In this rhythmic pattern, the pr + aux is in the weakest position, it is unstressed, and can be highly reduced. The phrasal stress is on the last word of the sentence. In order to maintain this rhythmic structure, all verbs, nouns and adverbs were monosyllabic, with the exception of the inflected form of the present and past continuous, e.g. *burning*. Table 3 shows some of the sentences constructed according to this rhythmic pattern.

Weak position	**Strong position**	**Weak position**	**Strong position**
pr + aux	main verb	article	noun
she'll	*burn*	*the*	*steak*
pr + aux	verb stem	conjugation	noun (except *)
I'm	*burn-*	*-ing*	*sticks*
we were	*burn-*	*-ing*	*wood*
it's	*burn-*	*-ing*	*fast**

Table 3. Examples of sentences following the W-S-W-S rhythmic pattern. All sentences had a noun in coda position except for the sentences starting with *it's burning* (marked with an asterisk *).

2.1.1.5. **Phonological context**

In order to avoid any influence on the pr + aux from neighbouring sounds, the pr + aux were preceded by a pause and followed by the same verb in the appropriate form. The verb chosen was BURN. Several factors were taken into account in the choice of the verb, including phonetic, lexical and semantic factors. From an acoustic phonetic point of view, an initial plosive was considered the best sound to follow the pr + aux in that the plosive burst facilitates the measurement of duration. To avoid the influence of lingual articulations, the bilabial plosives /p/ and /b/ were considered the best choice, as the bilabial articulation does not interfere with the movement of the tongue. As for the vowel in the nucleus, the most neutral lingual articulation is that of a mid central vowel, which is similar to the neutral tongue configuration (Laver, 1994). The closest English vowel to the mid central position is /ɜː/. Verbs with /bɜː/ include *burn*, *burst*, *birth*, and *burp*. The verb *burn* is the only one that can be followed by a range of nouns both with and without an article, which is needed to maintain the metrical structure of the sentences with the various tense forms of the verb (e.g. *she's burning wood*, *she's burnt the soup*). It is also the verb with the highest lexical frequency of the four monosyllabic verbs with /bɜː/[4]. Moreover, the past participle form of the verb BURN is also monosyllabic and maintains the same vowel /ɜː/.

2.1.1.6. **Repetition**

---

[4] The lexical frequency of the verb *burn* by lemma in the BNC spoken and written sub-corpora (Leech, Rayson and Wilson, 2001) is 53/1,000,000, followed by *burst* with a lexical frequency of 25/1,000,000. In the spoken sub-corpus only (non-lemmatised), the lexical frequency of *burn* is 13/1,000,000.

Repetition is known to trigger reduction (e.g. Fowler and Housum, 1987; Aylett and Turk, 2004; Pluymaekers et al., 2005b). For this reason, each paradigm was repeated five times in a row in an almost identical sentence. The only difference between the five sentences containing the same paradigm was the noun at the end of the sentence. Table 4 shows an example of five sentences with the same pr + aux.

She's burnt the soup
She's burnt the bread
She's burnt the cake
She's burnt the toast
She's burnt the jam

Table 4. Example of repetition of a paradigm in five almost identical sentences except for the last word in each sentence.

Using repeated sentences but changing the final word meant that the noun at the end of the sentence was 'new information' in each repetition, while the rest of the sentence was 'old information'. New information is more likely to be the focus of the sentence and thus to carry the nuclear stress (Hawkins and Warren, 1994), while old information is more likely to be reduced. Moreover, each new noun had a contrastive function compared to the other nouns due to being placed in the same position in the same sentence. Contrastive items are more likely to be the focus of a sentence and to carry the phrasal stress.

2.1.2. Prompting material

A Microsoft PowerPoint presentation with a sentence on each slide was created. During pilot work, it emerged that, despite clear instructions on which word of the sentence had to receive the main stress, speakers would sometimes shift the stress from the noun to the verb. For this reason the noun of each sentence was written in bold characters and underlined. Two pictures were placed on each slide, one for the subject and one for the object of the sentence. The rationale behind adding pictures to the slides was twofold. On one hand, the pictures would make the reading task less boring for the participants. On the other hand, the pictures were meant to distract the participants so that they would not focus on their production and speak more naturally. Figure 4 shows an example of three consecutive slides.

Figure 4. Three slides of the PPT presentation used to elicit the spoken material.

To control for speech rate, the sentences were presented to the participants in a timed slide presentation. Each stimulus sentence was displayed for 2.45 seconds. The total number of sentences created to investigate reduction was 238. However, the total number of sentences in the elicitation task was 615.[5] Eight timed breaks were inserted in the slide presentation (approximately every 70 sentences). Each break lasted for 32 seconds.

During one of the pilot experiments it was noticed that sometimes speakers would change the rhythmic pattern of the sentences. Because of the importance of the constant rhythmic pattern for the acoustic analysis and comparison of the paradigms, the desired rhythmic pattern was repeatedly played to the participants, before the beginning of the recording

[5] During the elicitation task, 377 sentences were recorded to analyse another feature of speech that is not included in this thesis.

session and during each break. The rhythmic pattern was recorded by a male speaker of Standard British English using the words <ta-da, ta-da>.

2.1.3. Participants and recording procedures

Speakers were recruited at the University of York. Fifteen speakers were recorded, but four of them had to be discarded. One speaker was discarded because she had a cold on the day of recording and her speech was denasal. One speaker had a Northern accent. One speaker placed the nuclear stress on the incorrect word in more than 30% of sentences. Of the 12 remaining speakers, 11 were female and one male. The male speaker was discarded to make the acoustic analysis more consistent, especially the analysis of the formant dynamics. Although it was attempted to collect a balanced group of speakers, with equal numbers of male and female speakers, this was not possible as male informants proved more reluctant to participate in experiments than female informants. All the eleven informants were speakers of Standard Southern British English (SSBE). Their age ranged between 20 and 27 years (mean = 21.8, median = 21, SD = 2.2). None of the informants had any history of impaired speech or hearing. Before the recordings, participants were told the procedures of the experiment and how their data would be used for research. All participants gave their consent to participate in the recording and to the use of their recording for research purposes. All the recordings were carried out in a soundproof room of the Psycholinguistics Laboratory of the Department of Language and Linguistic Science at the University of York using a Behringer C1 Studio Condenser microphone, TAC Scorpion-Mixing desk, and Adobe Audition v1.0 on PC software with 16-bit PCM and 44.1 kHz sampling rate.

The informants were not told the aim of the study. They were told that the focus of the experiment was on the last word in each sentence and they were thus instructed to put the phrasal stress on that word. They were also asked to speak as naturally as possible, and they were advised to continue with the task in case of reading errors. The PPT presentation lasted a total of 32 minutes. The speakers were paid for their participation.

2.2. Data analyses

Most phonetic analyses of reduced speech found in the literature involve quantitative analyses of counts of missing segments or syllables (Ernestus and Smith, 2018). However, as explained in Section 1.3, a segmental approach that disregards phonetic detail is not suitable for the analysis of reduced data. For this reason, both qualitative and quantitative analyses were carried out on the data collected. Section 2.2.1 describes the methodology of the qualitative analysis. Section 2.2.2 describes the methodology of the quantitative analysis.

2.2.1. Qualitative analysis

Chapter 1 showed that speech is rich in phonetic detail that provides information beyond the lexical contrast of phonemes, and that this information is important for speech understanding (Hawkins, 2003; Local, 2003). In particular, reduced speech – in which segments can be difficult to identify – remains mostly intelligible and does not imply loss of information. For this reason, a solely quantitative analysis of reduced speech is likely to miss the fine phonetic detail that remains in the signal and maintains speech intelligible.

One of the assumptions of this analysis is that every detail can be relevant and therefore should be noted at first. This assumption follows the belief by Kelly and Local (1989: 26) that “at the beginning of work on language material we can’t, in any interesting sense, know beforehand what is going to be important”.

The qualitative analysis carried out in this research focusses on the phonetic detail that remains in the acoustic signal in reduced speech rather than on what is 'lost'. The idea that in reduced speech something is missing comes from the tradition of considering the citation form of words given in pronouncing dictionaries, or the full form of words uttered in isolation or in careful speech, as the reference form (see discussion in Johnson, 2004). Neither of the two options (citation form and full form of words) is satisfactory, as neither of them occurs in natural, everyday speech. When they occur in spontaneous speech, they have a specific function (e.g. contrastive function). For this reason, in the analysis that follows, the most unreduced forms of words found in the data are used as reference forms if the need of a reference form arises.

2.2.1.1. **Kinaesthetic sense**

The most important aspect of a qualitative analysis is the accurate and repeated listening of the data. Accompanied by spectral observation, the auditory impressionistic experience provides crucial insight into the production of speech. One of the goals of the analysis is to relate the auditory experience and the acoustic observations to the articulations that produced them. That is, to trace back from the sounds heard to the movements that produced them. No articulatory data such as palatography, EMA (Electromagnetic Articulography), or MRI (Magnetic Resonance Imaging) were collected. However, it is possible to acquire an insight into the positions and movements of the articulators through the "kinesthetic sense" (Pike, 1943: 14) or "kinaesthetic awareness" (Kelly and Local, 1989: 29) of the researcher. Through their kinaesthetic sense, the researcher is trained to replicate the movements and positions of the vocal tract of the speaker. By mimicking the sounds produced by the speaker, the phonetician can gain insight into the configurations of the vocal tract that produced the sounds. As claimed by Catford (2001: 14), "introspective analysis of the kinaesthetic sensations is by far the most powerful way of learning about articulation". Although the kinaesthetic awareness of the researcher is crucial in understanding vocal tract states and movements, it also has some limitations and cannot substitute articulatory data. The main limitation is that every human vocal tract differs in shape and size and the acoustic output of speech largely depends on these two aspects. The researcher can only approximate the configuration of a speaker's vocal tract. Despite this limitation and the fact that any information collected through the kinaesthetic sense of the researcher is subjective, the insight gained through the kinaesthetic sense is still valuable and can shed light on vocal tract movements and configurations in absence of articulatory data.

2.2.1.2. Parametric approach

Due to the nature of the data collected, and the aim of the analysis, rather than looking at segments, the analysis carried out in this research treats speech as a cluster of parameters and events. As stated by Kelly and Local (1989: 30), "[p]honetic observation begins by listening to speech in terms of independently varying auditory and movement parameters". The auditory parameters are the various events that characterise the acoustic output, such as friction, voicing, plosion, and so on. The parameters are also the gestures and states of the articulators in the vocal tract, such as lip-rounding, and the position of the velopharyngeal port. Multiple movements and states can occur simultaneously, e.g. the tip of the tongue moves towards the alveolar ridge while the velopharyngeal port opens. A parametric approach makes it possible to pay attention to the phasing of the movements and events, instead of assuming that they occur simultaneously. A parametric approach lifts the focus from preconceived phonetic categories (Kelly and Local, 1989). This approach was used also in the data annotation described in 2.2.2.1.

2.2.1.3. Notes on terminology, figures and transcriptions

The phonetic transcriptions that accompany the figures throughout the thesis, together with the observations reported in Chapters 3 and 4, are the output of the qualitative analysis described in this chapter. This section explains the conventions of the phonetic transcriptions (Section 2.2.1.3.1), some of the terminology adopted in the thesis (Section 2.2.1.3.2), and what is displayed in the figures in Chapters 3 and 4 (Section 2.2.1.3.3).

2.2.1.3.1. Transcriptions

The phonetic transcriptions accompanying the figures are impressionistic records of the auditory impression of the piece shown in the figures. The aim of the phonetic transcriptions is to provide the reader with information about the auditory qualities of the pieces. An attempt was made to record the sounds without preconceptions about the phonological categories they might belong to.

The symbols used are those of the IPA chart (IPA, 1999). In addition, some diacritics and conventions of the extIPA are also used (Ball, Howard and Miller, 2018), e.g. [z̥)] indicates that the alveolar friction is initially voiced and then becomes voiceless. A symbol in round brackets indicates that the sound is weakly articulated and barely audible, e.g. [(ḁ)a] indicates that the voiceless portion at the beginning of the vocalic articulation is hardly audible. Additionally, a superscript in round brackets is used to specify the quality of a sound (the main symbol) when there are no IPA symbols or diacritics for this purpose, e.g. [h$^{(ɪ)}$] indicates that the glottal friction has a near-close near-front unrounded quality.

The chapters on pronouns and auxiliaries (Chapters 3 and 4) report the phonemic transcriptions of each pronoun and cliticised form of auxiliary, and auxiliary, as found in pronouncing dictionaries. The Cambridge English Pronouncing Dictionary is used for reference (Roach, Hartman and Setter, 2006). Phonemic transcriptions are reported in slash brackets, e.g. /wiː/ for the strong form of *we*.

2.2.1.3.2. Terminology used

In any analysis of a language, “the starting point is the decision about what constitutes the 'piece', in other words which part of the language the analysis will concentrate on” (Ogden and Local, 1994: 483). The analysis reported in this thesis focuses on the combinations of a pronoun and an auxiliary verb in utterance-initial position. In the FPA tradition, the pr + aux combinations will be referred to as ‘piece’, which Firth defines as “combinations of words” (1948: 2).

The terms ‘vocoid’ and ‘contoid’ (Pike, 1943) are used in this thesis instead of the more ambiguous terms ‘vowel’ and ‘consonant’. The terms vowel and consonant are usually defined according to a mixture of phonetic features and phonological position and function (Laver, 1994). Due to the nature of the data analysed here, the terms vocoid and consonant are not appropriate as they assume the phonological status of a sound. The terms vocoid and contoid are more appropriate as they are “strictly delineated by the articulatory and acoustic nature of sounds, without reference to phonemic contextual function” (Pike, 1943: 78).

2.2.1.3.3. Figures

In the next chapters, the auditory and acoustic analyses are accompanied by figures of spectrograms and waveforms. All the figures in Chapter 3 and Chapter 4 display the entire pr + aux piece. In addition, they include the hold phase and the burst of the bilabial plosive in the verb *burn* and approximately 50 ms of pause before the beginning of the pr + aux. For consistency, the time marks below the waveforms are placed every 0.04 seconds in all figures, unless stated otherwise.

The production of all the speakers was used to illustrate the features observed in the data. Most features described in the next chapters are found across speakers. For this reason, an attempt was made to show tokens produced by a wide range of speakers. One speaker (S2) had a clearer and less reduced speech than the other speakers. In many cases, her production was used to illustrate unreduced instances of the pr + aux.

2.2.2. Quantitative analysis

This section describes the quantitative analysis carried out on the data. Section 2.2.2.1 describes the annotation of the data. Section 2.2.2.2 describes the parameters analysed in the acoustic analysis. Section 2.2.2.3 describes the statistical analysis.

2.2.2.1. Annotations

In order to carry out an acoustic analysis of the data, including measurements of duration, formants and spectral moments, all the audio files of the data were annotated using Praat (Boersma and Weenink, 2018). Due to the reduced nature of the data collected, the annotation was challenging. As pointed out in Section 2.2.1.2, a segmental approach was unsuitable and a parametric approach was used, focussing on phonetic events rather than segments. By events, it is meant the acoustic outputs of articulatory movements and states of the vocal tract. These events can be grouped into two categories: the events occurring in the larynx, and the events occurring in the supralaryngeal vocal tract. Examples of events are the periodic energy produced by the vibration of the vocal folds, or the friction produced by a narrow constriction in the oral cavity. This method had a few advantages. Firstly, it made it possible to measure acoustic features that were not tied to segments anymore. This was useful, for instance, when friction extended beyond the time-limited portion of a fricative. Secondly, by annotating the onset and offset of each event, it was possible to analyse the temporal reorganization of events. For example, it was possible to measure the friction overlapping with the formant structure of a sonorant. Thirdly, it made the comparison of different degrees of reduction more achievable. This was particularly important in the case of 'apparently missing' segments which were temporally realigned and thus not identifiable as temporally limited, but were still articulated.

Firstly, all the recordings were automatically divided into sentence-size sound files running a script in Praat with the appropriate instructions. This process resulted in 2618 sound files, each of them containing a sentence of the type *he's burnt the toast*. All the sound files were named with a number for the speaker and the entire sentence produced (e.g. "Speaker_11_She_d_burn_the_roast"). Throughout this thesis, the speakers will be referred to with a number preceded by S. The eleven speakers whose production was analysed are: S2, S3, S6, S7, S8, S9, S10, S11, S13, S14, and S15.

For each token, a Textgrid with two tiers was created. On the first tier (a point tier), the onset and offset of the various events were annotated. On the second tier (an interval tier), the onset and offset of events that could be acoustically analysed were labelled (e.g. the spectral properties of friction and the formant dynamics of sonorants. See section 2.2.2.2 for the detailed description of the parameters analysed).

All the annotation was performed manually. The option of an automatic annotation of the data was evaluated and discarded. Due to the large number of tokens, the automatic segmentation using a forced alignment software (HTK – Hidden Markov Model Toolkit) was trialled. However, as expected, the results were unsatisfactory. Due to the high degree of reduction of the data, the forced aligner missed short or ambiguous elements. The manual annotation was carried out as follows.

The onset and offset of the following events were manually annotated in the target items (pr + aux) of all sound files: voicing, visible formant structure corresponding to a sonorant articulation, friction, silence as in the closure of stops, and stop burst or onset of release. Where a period of friction or sonorant articulation exhibited a clear change in spectral quality, the onset or offset of the change were labelled too (e.g. in a sequence of two periods of friction produced at two different places of articulation, see Figure 11). During the annotation process, a large amount of notes was taken, describing details that could not be labelled or auditory impressions that would otherwise be difficult to capture from the acoustic analysis or spectral observation.

Although both waveforms and broadband spectrograms were used to identify the onset and offset of phonetic events, the annotation was carried out by referring to the waveforms. This is because waveforms are more temporally accurate than spectrograms, which sacrifice temporal accuracy to provide spectral information. As described in more detail below, a multitude of features were taken into account during the annotation process, along with repeated careful listening of portions of sound material.

The following sections describe the phonetic events that were annotated.

2.2.2.1.1. Voicing

The term voicing is used here to refer to the regular repeating cycles of vibration of the vocal folds that produce voicing. The most frequent type of phonation found in the data is modal voice. Where creaky voice occurred, the onset and offset of creaky voice were annotated in the same way but a note was added to indicate the type of phonation.

The onset of voicing was labelled at the first glottal cycle as observed in the waveform. This would normally occur before the beginning of the formant structure in vocoids. Annotating the offset of voicing was challenging. In several tokens the periodicity would slowly fade away diminishing in amplitude. When the offset of periodicity occurred during the closure, it was decided to mark the end of periodicity where the glottal cycles started to become irregular. When the offset of voicing could not be marked, the duration of voicing had to be discarded. Figure 5 shows an example of annotation of the onset and offset of voicing in an instance of *he'll* produced by S7. Note that the cycle after the label 'vcx' has a different shape from the previous ones.

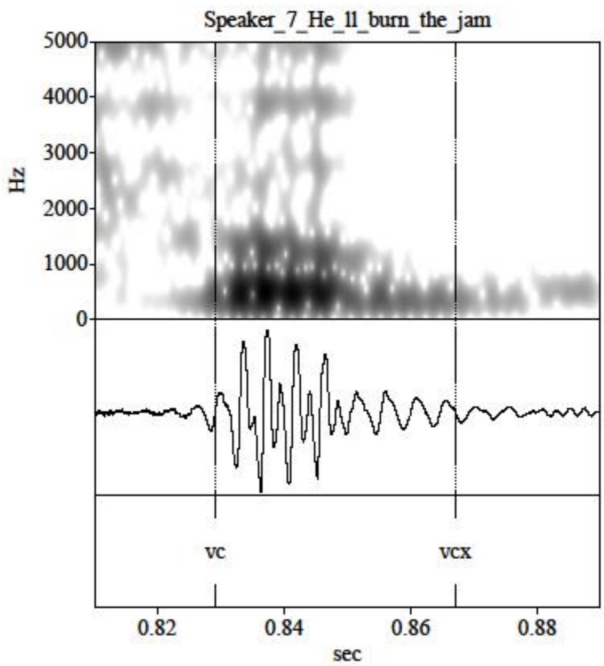


Figure 5. Annotation of the onset and offset of voicing in an instance of *he'll*. The label 'vc' indicates the onset; the label 'vcx' indicates the offset.

When the end of voicing occurred during a portion of friction, such as in pr+ *'s* tokens, it was more difficult to rely on the visible cycles on the waveform and a range of features were taken into account in annotating the offset. One of the features used was the voicing bar at the bottom of the frequency scale in the spectrograms. However, the degree of darkness of the voicing bar, depends on the amplitude of the fragment of speech with the highest amplitude displayed in the editor window in Praat. A change in the amplitude displayed in the editor window has a huge influence on how sounds are displayed. For this reason this method was not always reliable. Another (unreliable) method is to use the 'pulse' feature in Praat. Although useful at times, the pulses displayed also change when the time and amplitude in the editor window change. In summary, the challenge of annotating the offset of voicing was overcome by using multiple cues. Figure 6 shows an example of annotation of the onset and offset of voicing in an instance of *she's* produced by S2. In the waveform it can be noticed that there are still some cycles in the last 27 ms displayed. However, these cycles are very irregular and therefore were not included in the period of voicing.

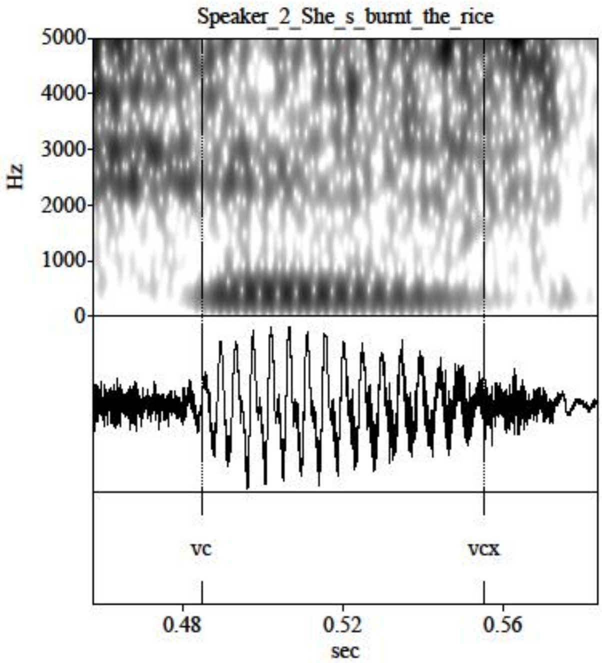


Figure 6. Annotation of the onset and offset of voicing in cases of simultaneous friction, such as in this instance of *she's*. The label 'vc' indicates the onset of voicing; the label 'vcx' indicates the offset of voicing.

2.2.2.1.2. Vocoids

Multiple factors were taken into account also in the annotation of the onset and offset of vocoids. The main features used were the start of the vertical striations, the start of the formant structure as visible in the spectrograms, the start of complex glottal cycles, and the higher amplitude of the glottal cycles compared to that of other sounds as visible in the waveform. Figure 7 shows an example of annotation of the onset and offset of the vocoid in an instance of *I'd* produced by S10.

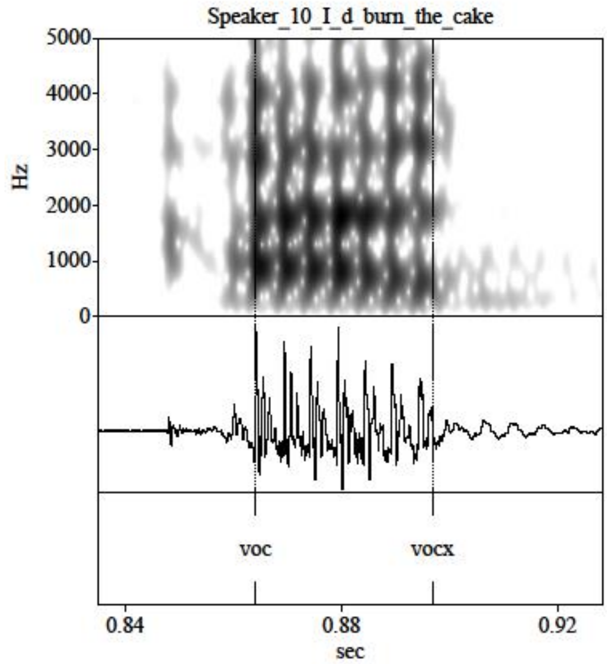


Figure 7. Annotation of the onset and offset of a vocoid in an instance of *I'd*. The label 'voc' indicates the onset; the label 'vocx' indicates the offset.

When friction is produced simultaneously to a vocoid, the striations of the formants become less clear, especially at high frequencies, and the aperiodic energy in the waveform makes it more difficult to observe the end of the complex cycles. Again a combination of features was taken into account to mark the beginning and end of the vocoid. Figure 8 shows an example of annotation of the onset and offset of the vocoid in an instance of *she's* produced by S11. In this case, F2 and F3 are clear well before the beginning of the vocoid, and the vertical striations are weak, so the onset of the vocoid was marked from the first complex glottal cycle. As for the offset of the vocoid, the properties of the glottal cycles are not visible in the waveform due to the aperiodic energy; however, there is an abrupt drop in amplitude of F2 and F3 in the spectrogram, which was used to mark the offset of the vocoid.

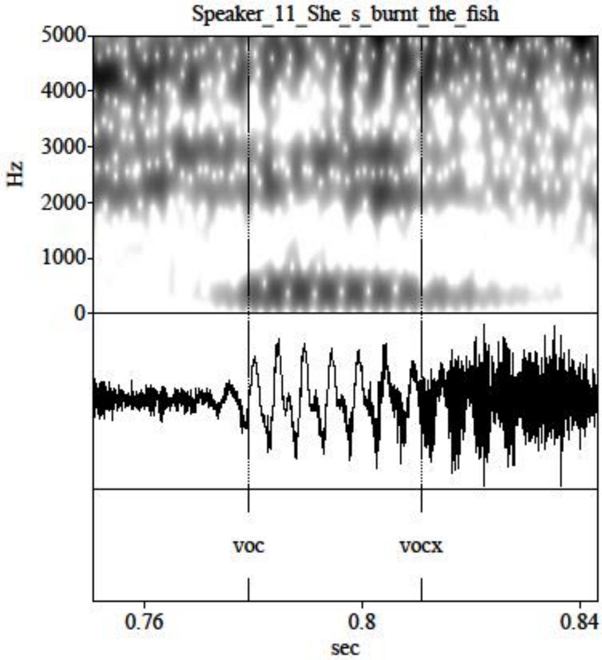


Figure 8. Annotation of the onset and offset of a vocoid in an instance of *she's*. The label 'voc' indicates the onset; the label 'vocx' indicates the offset.

2.2.2.1.3. Friction

Friction is produced when a narrow constriction is created in the vocal tract and air becomes turbulent. The acoustic correlate of friction is aperiodic energy. The onset of friction was marked at the first appearance of aperiodic noise in the spectrograms and waveforms. When the friction was in sentence-initial position, such as in *he*, *she* and *they*, there is often a slow increase of aperiodic energy which makes the task of defining its beginning difficult. It was decided to mark the onset of friction fairly early at the start of the visible friction. This method made the annotation consistent. The offset of friction before a vocoid was marked when the aperiodic energy decreased in amplitude in the waveform. Figure 9 shows an example of annotation of the onset and offset of friction in an instance of *he's* produced by S2.

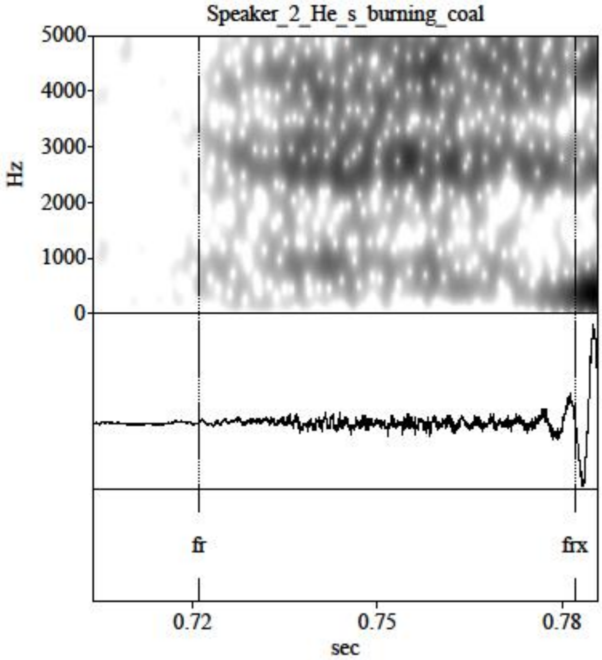


Figure 9. Annotation of the onset and offset of friction in sentence-initial position in an instance of *he's*. The label 'fr' indicates the onset; the label 'frx' indicates the offset.

The friction in coda position (in pr + *'s* paradigms), was characterised by a more abrupt onset and offset. Because friction is characterised by energy at high frequency, well above 5 kHz, sometimes it was necessary to display the frequency scale up to higher frequencies such as 8 kHz as in Figure 10. Figure 10 shows an example of annotation of the onset and offset of friction in coda position in an instance of *she's* produced by S14.

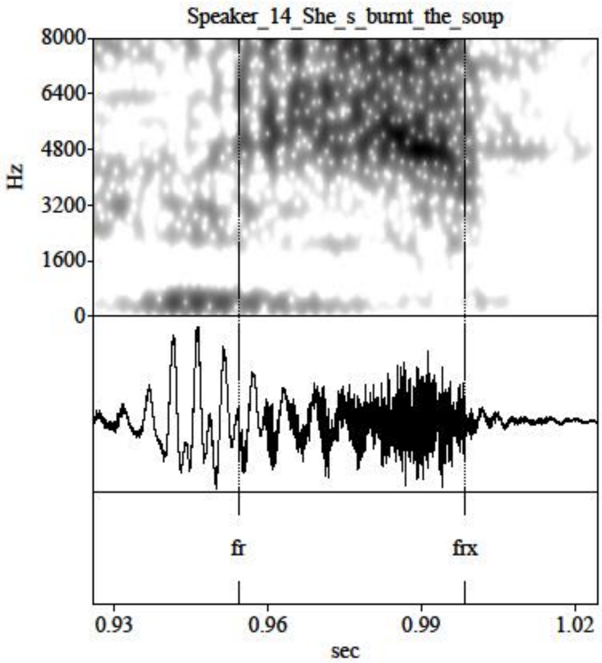


Figure 10. Annotation of the onset and offset of friction in coda of an instance of *she's*. The label 'fr' indicates the onset of friction; the label 'frx' indicates the offset. Notice that the frequency scale displays frequencies up to 8 kHz.

Due to the high degree of reduction, in some tokens, two portions of friction produced at different places of articulation were adjacent. In the majority of cases, it was possible to identify a spectral discontinuity between the two frictions. In the waveform, it would be displayed as a change in amplitude, and a decrease in amplitude towards the end of the first friction and a subsequent increase in amplitude at the beginning of the second friction. In the spectrogram, it would be displayed as a change in the distribution of energy along the frequency scale – which correlates with the place of articulation and the size and shape of the cavities of the vocal tract. Figure 11 shows an example of annotation of the onset and offset of two adjacent periods of friction in an instance of *she's* produced by S15.

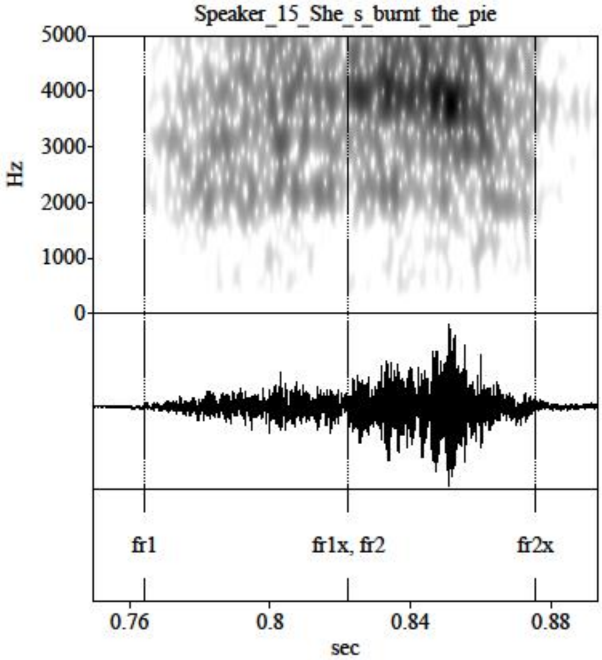


Figure 11. Annotation of the onset and offset of two adjacent periods of friction in an instance of *she's*. The label 'fr1' indicates the onset of the first portion of friction; the label 'fr1x, fr2' indicates both the offset of the first friction and the onset of the second portion of friction; the label 'fr2x' indicates the offset of the second portion of friction.

2.2.2.1.4. Oral closure

All pr + aux were followed by a phonological plosive in word-initial position of the verb BURN. This meant that, in most cases, the pr + aux were followed by a complete closure at the lips which resulted in a period of absence of acoustic energy at most frequencies. In some instances, there would still be voicing during the closure, displayed as a voicing bar at low frequencies (especially in pr + *'d* paradigms). The absence or presence of voicing, however, did not affect the annotation of the onset and offset of the oral closure. The two events were kept distinct in the annotation process. Figure 12 shows an example of annotation of the onset of closure in an instance of *you'd* produced by S9.

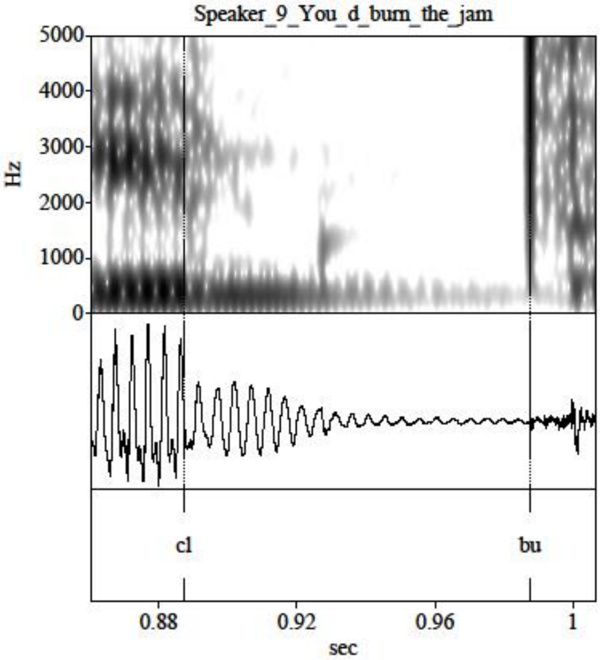


Figure 12. Annotation of the onset and offset of the closure in an instance of *you'd*. The label 'cl' indicates the onset; the label 'bu' indicates the offset of the closure and the burst of the release.

2.2.2.1.5. Release

During a period of complete oral closure there is usually a build-up of intra-oral air pressure, which is then released at the opening of the closure. The release can be abrupt in the form of a burst, or gradual, with a slow increase of aperiodic energy. Both these types of release were annotated at the first sign of energy in the waveform. Figure 12 above shows an example of annotation of the abrupt release of air pressure. The beginning of the release is characterised by a clear spike in both the spectrogram and waveform. Figure 13 shows an example of annotation of the gradual release of air-pressure in form of aperiodic energy after a period of closure in an instance of *we'd* produced by S8.

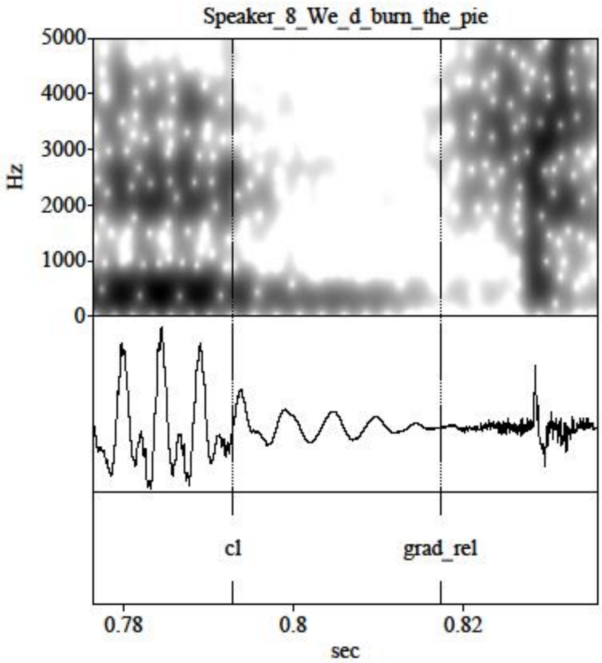


Figure 13. Annotation of a gradual release of turbulent air after a period of closure in an instance of *we'd*. The label 'cl' indicates the onset of the closure; the label 'grad_rel' indicates the onset of the gradual release of turbulent air.

2.2.2.1.6. Laterals

Depending on the context and position in the syllable, the onset and offset of lateral sounds can be characterised by a sudden change in spectral qualities compared to neighbouring vocoids. Acoustically, laterals are characterised by a low F2 frequency (especially dark, velarised laterals), a large gap between F2 and F3, a wide F1 bandwidth, and a general lower amplitude compared to neighbouring vocoids (Stevens, 1998). Although in the data collected very few laterals could be annotated, all these aspects were taken into account in the identification of the onset and offset of lateral sounds. Figure 14 shows one of the few instances of pr + *'ll* in which the onset of the lateral could be annotated.

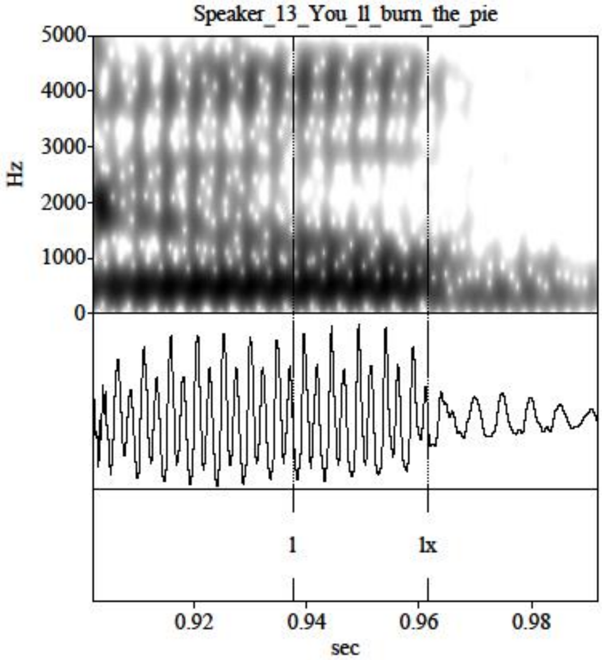


Figure 14. Annotation of the onset and offset of a lateral in an instance of *you'll*. The label 'l' indicates the onset; the label 'lx' indicates the offset.

The main features observed at the beginning of the lateral in Figure 14 (marked as 'l') are an abrupt change in F3, a wider F2 bandwidth close to F1, a change in overall amplitude (although it is not lower than the amplitude of the preceding vocoid), and a wide gap between F2 and F3.

2.2.2.1.7. Nasals

The onset and offset of nasal sounds, like those of laterals, are characterised by an abrupt change in spectral quality compared to neighbouring vocoids. Nasals are characterised by an overall low amplitude, weak energy at high frequencies, simpler glottal cycles than vocoids, and fainter formants compared to neighbouring vocoids. Figure 15 shows an example of annotation of the onset and offset of a bilabial nasal in an instance of *I'm* produced by S13.

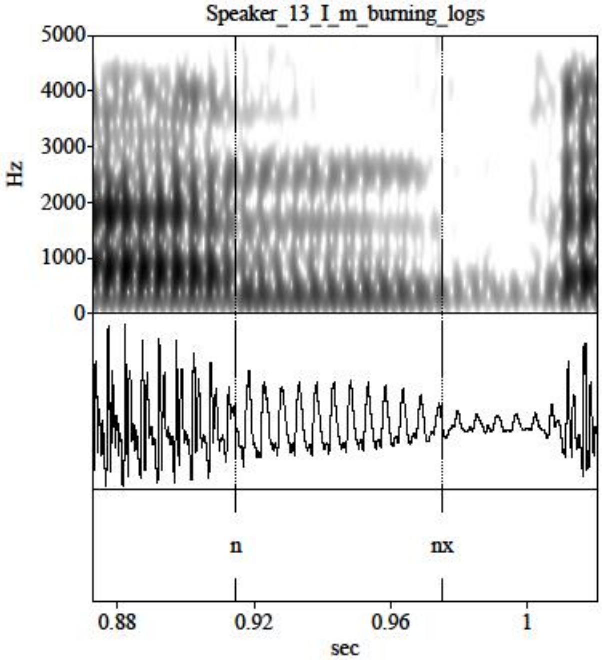


Figure 15. Annotation of the onset and offset of a nasal in an instance of *I'm*. The label 'n' indicates the onset; the label 'nx' indicates the offset.

The main acoustic features used to annotate the beginning of the nasal in Figure 15 were the evident change in frequency and amplitude of F1 and F2, the dip in overall amplitude between the vocoid and the nasal, and the lack of energy at frequencies above 3 kHz.

To summarise, several acoustic features were used to annotate the onset and offset of phonetic events such as voicing, friction, vocoids, closure, release, and lateral and nasal sounds. Repeated listening together with careful spectral observation were used. In particular, the main features taken into account were the spectral information displayed in the spectrograms, such as the frequency, bandwidth and amplitude of formants, the distribution of aperiodic energy along the frequency scale; as well as the information displayed in the waveforms, such as the periodicity of the glottal cycles, the overall amplitude of a sound, and the structure of the waveforms. The high degree of reduction of the data collected made the annotation particularly challenging. However, the challenge was overcome by choosing to annotate phonetic events rather than segments, and by looking at a wide range of acoustic features observable in the spectrograms and waveforms.

The next section describes the acoustic parameters analysed, and the statistical analysis performed.

2.2.2.2. **Acoustic analysis**

All the paradigms were uttered in the same rhythmic and prosodic structure in anacrusis and repeated five times by each speaker. Fifty-five repetitions were recorded for each paradigm across speakers. However, some tokens had to be discarded either because the incorrect pronoun or auxiliary had been uttered, or because the sentence had a different rhythmic structure, e.g. the pr + aux was stressed. The few cases of incorrect pr + aux and self-repair by the speaker were discarded. For this reason, when providing acoustic values, the count (N) of the tokens included in the acoustic analysis is always reported.

The acoustic parameters analysed were: duration, amplitude, spectral moments and formant dynamics of every event. All the measurements were obtained by running a script with the appropriate instructions in Praat. All Praat scripts can be found in Appendix D. This section describes in detail the analyses carried out.

2.2.2.2.1. Duration and amplitude

Duration is probably the parameter that is most frequently used to investigate the degree of reduction or strengthening. The overall duration of each piece was calculated from the onset of the first event, to the onset of the release of the plosive [b] in onset of the main verb BURN. Although the closure of the bilabial plosive in BURN is technically not part of the piece, the rationale for including it in the overall duration is that the release of the bilabial closure acts as a benchmark, a reference point that is present in all the sentences. In fact, the acoustic analysis was the main reason for choosing to have a plosive after the pr + aux: the release of the closure is a constant event that is always present at the end of each piece. Excluding the closure of [b] in the duration measurement would have made it impossible to measure the duration of the pr + *'d* paradigms, in that in most cases the offset of [d] cannot be identified and marked.

The duration of each phonetic event was calculated by subtracting the onset time point from the offset time point. For example, if the offset of friction was at the time point 0.463 seconds and the onset of friction was at 0.353 seconds, the duration of friction was 0.463 – 0.353 = 110 ms.

The mean amplitude of each sound was calculated in a temporal window of half the duration of the target sound centred at the mid-point (from 25% to 75% of the entire duration of the target sound).

It is good practice to normalise data for speech rate (by calculating the duration of the target item relative to that of the entire sentence) and amplitude (by calculating the amplitude of the target item against that of an adjacent stressed vowel). However, the experimental design of the data collection made it possible to control for speech rate and amplitude in the recording phase.

2.2.2.2.2. Spectral analysis

The parameters that are most commonly used to investigate the spectral properties of fricatives are the first four spectral moments. The first four spectral moments are statistical analysis of the distribution of energy along the frequency scale. They are: the spectral centre of gravity, the standard deviation, skewness, and kurtosis.

The centre of gravity (hereafter CoG) and the standard deviation (hereafter SD) are statistical analyses of the distribution of energy along the frequency scale. The CoG gives an amplitude-weighted mean frequency of the distribution of energy, while the SD indicates the spread of the distribution of energy around the mean (Jones and McDougall, 2009). It has been claimed that the CoG fails to capture information about the properties of fricatives beyond the "dominant front cavity" (Wrench, 1995). That is, fricatives are characterised by more than one peak in energy, but the CoG measures only the main concentration of energy missing to capture information about the other peaks in the spectrum. In an attempt to give a better account of fricatives' spectral properties, Wrench (1995) tested a Multiple Centroids Analysis (MCA). Based on an algorithm that Crowe and Jack (1987) developed to gain a better picture of vowel formants, MCA attempts to identify multiple peaks in fricative spectra. However, the fact that fricatives, in contrast with vowels, do not have a fixed number of formants (or peaks) per frequency range makes the MCA unreliable in the analysis of fricatives, as it struggles to identify any relevant peaks besides the main peak. Wrench (1995) concluded that the CoG combined with the other three spectral moments gives enough information for the analysis of fricatives and allows to identify their place of articulation. It must be pointed out that the spectral properties of sounds are linked to the size and shape of the vocal tract, therefore they are highly variable between speakers and genders.

Similarly to CoG and SD, Skewness and kurtosis are statistical measures which treat the spectrum of fricatives as a statistical distribution curve (Jongman, Wayland and Wong, 2000; Jones and McDougall, 2009). Skewness indicates to what extent the distribution of energy is (a)symmetrical around the mean. Positive skewness indicates that the aperiodic energy is concentrated in the lower frequencies, while negative skewness indicates that the aperiodic energy is concentrated in the higher frequencies. Kurtosis indicates to what extent the spectrum is flat or dominated by a peak. Positive kurtosis indicates a peaked distribution with a more clearly defined spectrum, while negative kurtosis indicates a flat distribution.

The four spectral moments were measured in a temporal window of half the friction noise centred at the mid-point (from 25% to 75% of the duration of friction).

2.2.2.2.3. Formant dynamics

Formants are resonant frequencies that reflect the size and shape of the vocal tract. They are high amplitude peaks of energy that are visible in spectrograms as horizontal dark bands. Each dark band is a peak of energy. The formants are the acoustic output of the configuration of the vocal tract (the filter) which shapes the sound (the vibration of the vocal folds) produced by the source (the glottis) (Fant, 1960). While higher formants seem to be more speaker-specific (McDougall, 2004), the first two formants can give an indication of the tongue position in the oral cavity. The first formant (F1) reflects the open-close (or high-low) dimension of the oral cavity. The second formant (F2) reflects the front-back dimension. As for the third formant (F3), it can reflect lip-rounding, rhoticity or a pharyngeal constriction (Johnson, 1997). Although formants are more well-defined in sonorant sounds, they can be visible and measured in obstruent sounds too.

Although steady vocoids characterised by a single quality that does not change in time can be analysed by calculating the mean frequency of the formants at mid-point or in a central portion of the vocoid, this analysis is unsatisfactory for vocoids that exhibit a variation in time. A more detailed and informative analysis can be obtained by measuring the trajectories, or formant dynamics, of the formants in time.

The formant dynamics were extracted at 9 equidistant points in time from the onset to the offset of each event. Throughout the thesis, the formant dynamics will be displayed in Hz. Although the Bark scale gives an idea of the perceived frequencies by the auditory system (Johnson, 2003), the Hertz scale was preferred. The rationale for choosing to display the formants in Hertz is that much of the work presented in this thesis focusses on the articulations that produced the sounds. The Bark scale was employed in the analysis of some data of a pilot study, but did not provide a useful insight into the quality of the sounds analysed. The difference between the Hertz scale and the Bark scale is most visible and useful at higher frequencies. However, in the present study, the focus is on the first three formants, that is, on the lower frequencies.

Throughout the thesis, the graphs of the formants display the mean frequency of each formant at each point in time (Time Point 1, Time Point 2, Time Point 3, …) calculated across speakers and repetition. Each Time Point (TP) is then connected to the preceding and following TP through a straight connecting line that does not necessarily represents the formant trajectory between those two TPs. It has to be considered that most of the vocoids are very short and it is thus unlikely that the formant trajectory moves extensively between two consecutive TPs. Where appropriate, the formant values obtained by the automatic measurements were manually checked.

2.2.2.3. **Statistical analysis**

For the analysis of the acoustic features of contrasting pairs of paradigms (Chapter 5), linear mixed-effects models were generated in R (R core team, 2016) using the *lmer()* function in the lme4 package (Bates, Maechler and Bolker, 2012) with the durations and/or the spectral moments of the phonetic events as dependent variables, the *Auxiliary* or *Pronoun* as a fixed effect, and *Speaker* as a random effect. P values were generated using likelihood ratio tests that compared the model with *Auxiliary* or *Pronoun* as a fixed effect against the null model without it. When multiple analyses were run with the same model, p values were adjusted using Bonferroni corrections. Otherwise, model outputs were considered to be significant when $p < .05$.

For the analysis of the relationship between reduction and repetition reported in Chapter 6 (Section 6.3), Spearman's Rank correlation tests were run. For each paradigm analysed, the test looked at the correlation between the duration of the vocoid or the duration of the piece and the number of repetitions of the paradigm.

2.3. Summary

This chapter has described the methodology of the data collection and the analyses carried out. Two aspects were important in the data collection: to trigger a high degree of reduction, while controlling for the phonological environment. In order to control for the phonological environment, speakers were asked to read sentences from a screen. The sentences were carefully constructed to influence the pr + aux as little as possible by choosing neutral articulations in the following phonological environment. Reduced speech was triggered by using several strategies. First, each paradigm was repeated five times in slightly different sentences which included a new noun ('new information') at the end of the sentence. Second, the speakers were instructed on where to place the phrasal stress and on the rhythm of the sentences. Third, the rhythmic structure of the sentences was played to the speakers during each break to remind them of the stress placements.

The methodology of the analyses carried out reflects the non-segmental approach of this thesis (see Section 1.3). In this chapter, the qualitative and quantitative analyses were described. Without disregarding the quantitative analysis, the importance of the qualitative analysis when dealing with reduced speech was emphasised. In particular, it was highlighted the role of the kinaesthetic sense of the researcher in linking the acoustic output to the states and movements of the vocal tract that produced it. Furthermore, it was explained that the movements and states of the vocal tract can be better analysed using a parametric approach, which allows the researcher to pay attention to the temporal organisation of the various phonetic events, instead of being fossilised on preconceived ideas on sound units.

Finally, the chapter described the acoustic and statistical analyses carried out throughout the research. The next two chapters report on the auditory analysis and spectral observations of the data collected. Chapter 3 describes the pronouns in pr + aux combinations, and Chapter 4 describes the auxiliaries in the pr + aux combinations.

3. Auditory and acoustic analysis of pronouns in pr + aux

This chapter and the following one (Chapters 3 and 4) report on an auditory and acoustic analysis of pronoun and auxiliary combinations (pr + aux) in reduced speech. The aim of these chapters is twofold. The first aim is to describe in detail the phonetic features of the pr + aux recorded and the wide range of variation found in the data collected. The second aim is to identify the fine phonetic detail or essential features that remain in the signal in reduced speech. That is, those phonetic features that constitute the identity of each paradigm, that convey the crucial grammatical information for the correct interpretation of the paradigms, and that for these reasons remain in the signal even in reduced speech.

Although the theory at the base of this research is that from a phonological perspective pr + aux combinations behave as a single unit and not as two separate items (a pronoun and an auxiliary), for practical reasons the analysis of the pronouns and auxiliaries are reported in two separate chapters. This chapter focusses on the description and analysis of the pronouns, and Chapter 4 focusses on the description and analysis of the auxiliaries. This structure has been chosen in order to avoid repeating the analysis of the features that are shared by more than one paradigm. For instance, in their phonological forms, the paradigms *I've*, *you've*, *we've*, *they've* share the labio-dental fricative in coda position. Bearing in mind the wide range of variation in the realisation of these paradigms, the descriptions of the various phonetic realisations of the labio-dental fricative of the cliticised form *'ve* have been grouped together.

In addition, the finite forms of the past tense of the auxiliary BE – *was* and *were* – were analysed too. The rationale for including *was* and *were* to the analysis is to investigate the contrast between the present and past tense of auxiliaries, which will be covered in Chapter 5. Table 5 summarises the combinations of pronoun and auxiliary that have been analysed.

Pronouns	Aux HAVE		Aux BE		Modal Aux	
	present	past	present	past	(present)	(past)
I	I've	I'd	I'm	I was	I'll	I'd
he	he's	he'd	he's	he was	he'll	he'd
she	she's	she'd	she's	she was	she'll	she'd
it	it's	it'd	it's	it was	it'll	it'd
we	we've	we'd	we're	we were	we'll	we'd
you	you've	you'd	you're	you were	you'll	you'd
they	they've	they'd	they're	they were	they'll	they'd

Table 5. The combinations of pronouns (rows) and auxiliaries (columns) analysed in this thesis.

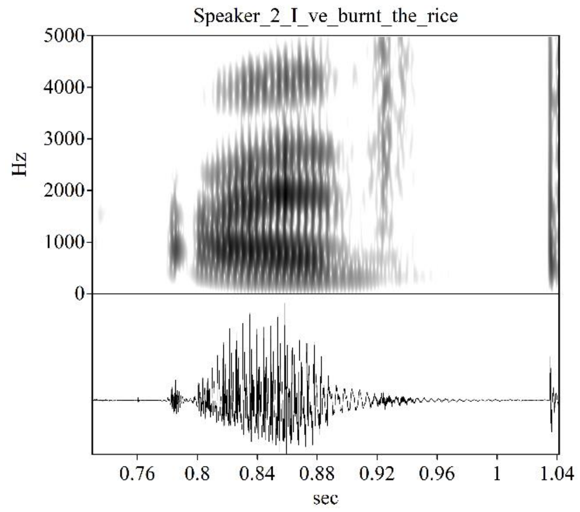
Where possible, the counts of the strategies observed for each paradigm are given. However, this is not always possible, as some instances proved difficult to be classified. The realisations of the pr + aux are better thought of as variations along a continuum between unreduced and reduced variants (or hyper- and hypo-speech in Lindblom's words), rather than distinct categories.

3.1. Pronouns

There are seven personal pronouns in English: *I*, *he*, *she*, *it*, *you*, *we*, *they*. The following seven sections describe the phonetic characteristics of each pronoun as found in the data collected. There is a wide range of variability in the realisation of each pronoun. To illustrate the degrees and types of variability found in the data, both unreduced and reduced variants will be described. In each section, the most unreduced realisations observed in the data are described first, followed by the most frequent variants, and the most reduced instances observed. The acoustic analysis is reported only where relevant to illustrate a point or feature.

3.1.1. Pronoun *I*

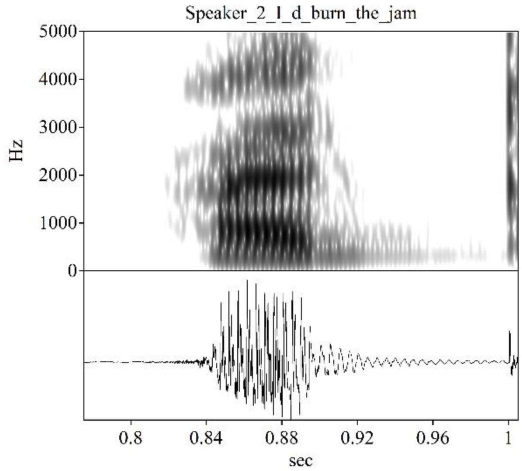
The phonemic transcription of the pronoun *I* as found in pronouncing dictionaries is /aɪ/. This pronoun has only a vocalic articulation. The sound source is solely in the glottis as the airstream flows from the glottis through the vocal tract unimpeded. The articulation of the vocoid changes in time. At the beginning of the vocoid the tongue is in an open front position and then moves to a close front position. Acoustically, the formants – the acoustic correlates of the tongue position in the vocal tract – reflect the change in the position of the tongue: at the beginning the frequency of F1 is higher and gradually lowers as the tongue moves to a higher position; while F2 starts at lower frequencies and gradually moves to higher frequencies as the tongue moves further front. Figure 16 shows a rare example of *I* in the dataset in which the formant movement is visible. In this instance, F1 starts at 822 Hz and ends at 614 Hz (range 208 Hz), while F2 starts at 1391 Hz and ends at 1855 Hz (range 464 Hz).



[ʔɐe̯v]

Figure 16. Spectrogram and waveform of an instance of *I've* produced by S2 in which there is a clear formant movement during the vocoid.

In the dataset collected, *I* is regularly produced with little movement of the tongue as reflected in little movement of the formants. In particular, the second formant (F2) – the acoustic correlate of the front-back dimension – exhibits little variation in the range of frequencies during the articulation of the vocoid, while the first formant (F1) – the acoustic correlate of the open-close dimension – moves along a wider frequency range. Figure 17 shows an instance of *I* in the pr + aux *I'd.* In this instance, F1 starts at 805 Hz and ends at 696 Hz (range 109 Hz), while F2 starts at 1857 Hz and ends at 1916 Hz (range 59 Hz).



[aẹd˺]

Figure 17. Spectrogram and waveform of an instance of *I'd* produced by S2 in which there is little formant movement during the vocoid.

The instances of *I* that do not display a noticeable formant movement can be treated as cases of decreased magnitude of gesture. The gesture of the tongue for the articulation of the vocoid is reduced in that the tongue makes a (spatially) smaller movement. Figure 18 shows the mean formant dynamics of the vocoid in *I* in three subsets calculated across speakers and repetitions (N = 165). The three subsets are *I'd burnt*, *I've burnt*, and *I'd burn*. The formant dynamics are calculated at nine equidistant points in time as explained in the methodology chapter (Section 2.2.2.2.3). The formants of the three subsets *I'm*, *I'll*, and *I was* are not included because the contoid in coda position of the pr + aux affects the formants of the vocoid.

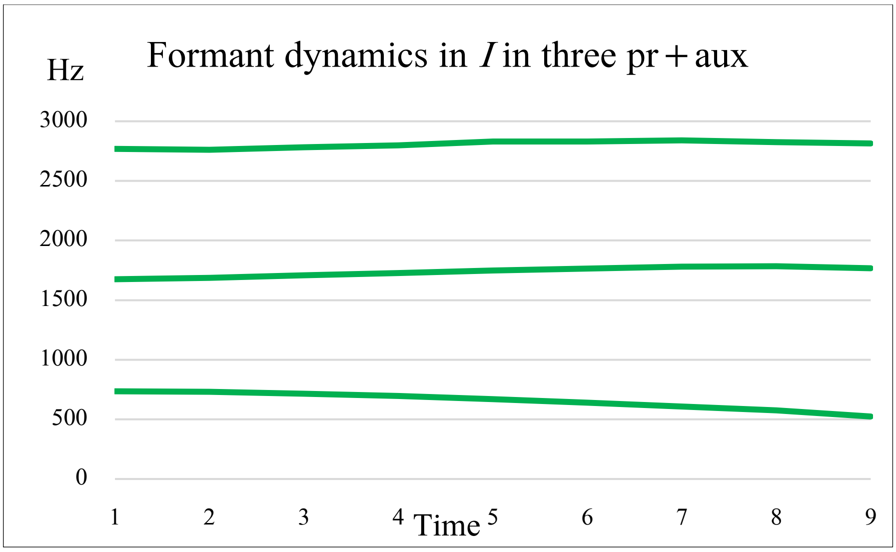


Figure 18. Mean formant dynamics of the vocoid *I* in three pr + aux: *I've*, *I'd burnt* and *I'd burn*, calculated at nine equidistant points in time.

It can be noticed in Figure 18 that the frequency of F2 increases slightly (from 1675 Hz at Time point 1 to 1767 Hz at Time point 9, range 92 Hz); while the frequency of F1 displays a larger movement (from 735 Hz at Time point 1 to 524 Hz at Time point 9, range 211 Hz). The formants indicate that the tongue moves from a lower to a higher position during the production of the vocoid and only slightly to a fronter position. The fact that the pronoun *I* can be produced as an almost steady (monophthongal) open vocoid is not surprising if we consider the paradigmatic system of pronouns. The pronoun system is a very small system of contrasts. Of the seven pronouns, *I* is the only one characterised by an open vocoid – all the other pronouns are characterised by a close or near-close vocoid /hi, ʃi, ɪt, wi, ju/ (*he, she, it, we, you*) or a close-mid vocoid /ðeɪ/ (*they*). Even if the articulation of the vocoid /aɪ/ is reduced to an open monophthongal vocoid, the pronoun cannot be confused with any other pronoun – its identity is maintained.

Besides the decrease in magnitude of gesture leading to a single-quality vocoid, the vocoid can be temporally reduced too. The vocoid duration across *I* + aux paradigms (with the exclusion of the pr + aux *I was*) across speakers and repetitions (N = 236) ranges between 6 ms and 96 ms (mean = 43 ms, median = 42 ms). Figure 19 shows the shortest instance of the vocoid in *I* found in the dataset.

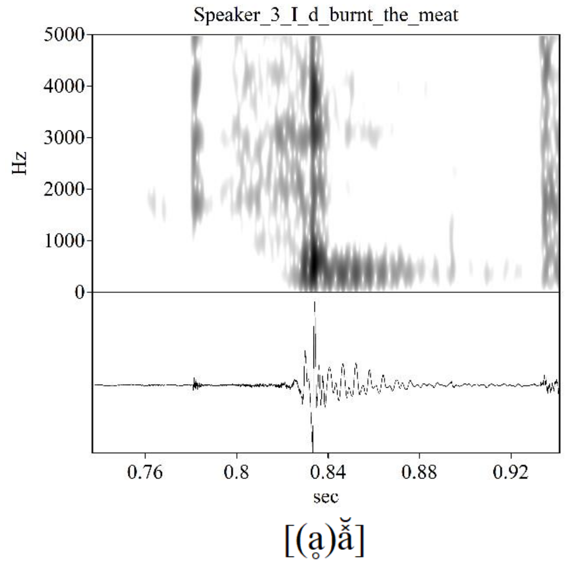


Figure 19. Spectrogram and waveform of the shortest instance (6 ms) of the vocoid in *I'd* produced by S3.

The vocoid in Figure 19 is characterised by only one glottal cycle of vibration of the vocal folds. However, it can be noticed in the spectrogram that there is a portion of friction before voicing starts and a portion of voicing after the end of the vocoid. The formant structure of the vocoid is visible during the friction that precedes the beginning of voicing. The duration of the friction from the spike to the beginning of the vocal folds vibration is 48 ms. The duration of voicing is 38 ms. Although the portion of friction before voicing starts is not audible, its presence suggests that the gesture for the articulation of the vocoid starts before the vibration of the vocal folds begins. Instead of occurring simultaneously, the articulation in the supralaryngeal vocal tract and the vibration of the vocal folds are temporally reorganised. The presence of friction with formant structure before the beginning of voicing was observed in several instances in the data and is covered separately in Chapter 6 (Section 6.1).

An even more reduced realisation of *I* occurs when the vocoid is voiceless. Figure 20 shows an instance of voiceless vocoid. Notice that in this case also the auxiliary is extremely reduced to the point that it is not identifiable in the spectrogram and waveform. The friction that is visible in the figure is produced at the glottis and it has a clear front open quality. There is a short and weak portion of vibration of the vocal folds at the end of the friction but it is hardly audible.

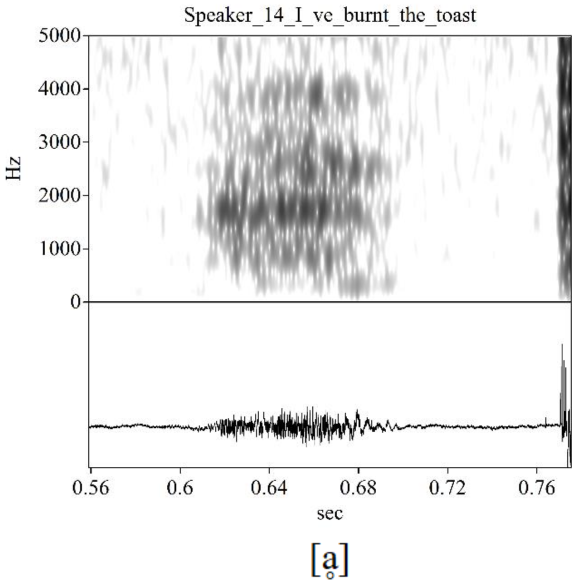


Figure 20. Spectrogram and waveform of a voiceless instance of *I've* produced by S14.

A common feature of any vocoid in sentence-initial position after a period of silence, is to be preceded by a glottal stop (Dilley et al., 1996). This occurs frequently at the beginning of *I*. In 38% of instances (N = 121/317), *I* begins with a glottal stop. Figure 21 shows an example of a short vocoid with two clear strong bursts at the beginning.

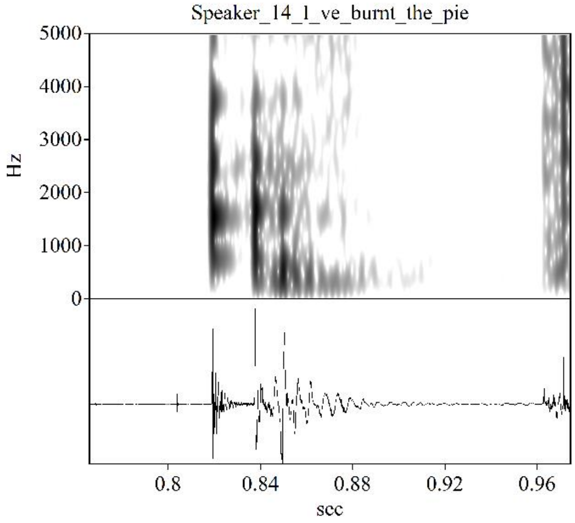


Figure 21. Spectrogram and waveform of an instance of *I* preceded by a glottal stop in *I've* produced by S14.

To summarise, the pronoun *I* in the dataset is regularly realised as a vocoid with an open front quality and little tongue movement. The duration of the vocoid is highly variable and can be very short (the shortest instance has a duration of 6 ms). The presence of friction before voicing starts highlights the articulation of the vocoid and suggests that the phonetic events involved in the production of the vocoid might be temporally reorganised. Specifically, the events occurring in the supralaryngeal vocal tract (articulation) and the events occurring in the laryngeal cavity (voicing) might be realigned. In the most extreme cases of reduction, the pronoun *I* can be realised as a single glottal cycle or even be entirely voiceless with weak glottal friction on which the oral articulatory gesture is superimposed. The quality of the vocoid depends considerably on the auxiliary it is in combination with. When in the pr + aux combinations *I'm*, *I'd* and *I've*, the essential phonetic feature of the pronoun *I* is the open front unrounded quality of the vocoid. However, when in the combinations *I'll* and *I was*, the quality of the vocoid is influenced by the nature of the sound that follows it (see Chapter 4).

3.1.2. Pronoun *he*

The phonemic transcriptions of the pronoun *he* found in pronouncing dictionaries are /hiː/ for the strong form, and /hi/ and /i/ for the weak forms. In all the items recorded, *he* is characterised by an initial period of friction produced in the glottis. The only part of the vocal tract involved in the production of the glottal friction is the glottis, which means that the rest of the vocal tract can adopt any configurations. This often results in the oral cavity adopting the position of neighbouring sounds during glottal friction. In *he*, the glottal friction is followed by a vocoid characterised by a near-close near-front unrounded articulation. The quality of the initial glottal friction is thus influenced by the articulation that follows it – the tongue is already in a near-close near-front position ready for the articulation of the vocoid. The acoustic result of this articulatory feature can be seen in the spectrograms in the forms of visible formants during the glottal friction. Figure 22 shows an unreduced instance of *he'd* with clear F2 and F3 during the initial glottal friction. The formants during the friction and the vocoid are at the same frequencies, indicating that the tongue and the lips do not change position or shape during the production of the piece.

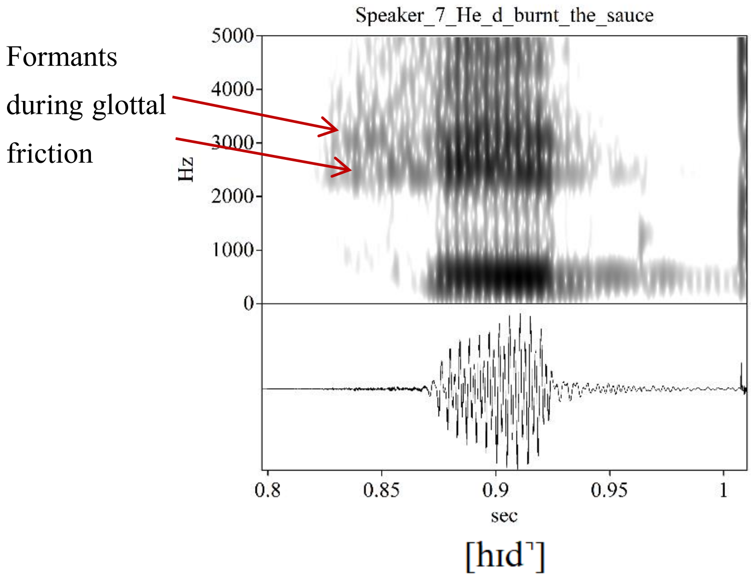


Figure 22. Spectrogram and waveform of an unreduced instance of *he'd* produced by S7 with clear formants during the glottal friction.

Although we would expect the second formant in *he* to be flat during the duration of the piece, as in the example in Figure 22, in several instances in the dataset the frequency of F2 decreases during the production of the pronoun. Figure 23 shows an example of sloping F2 from the beginning of friction to the end of the vocoid.

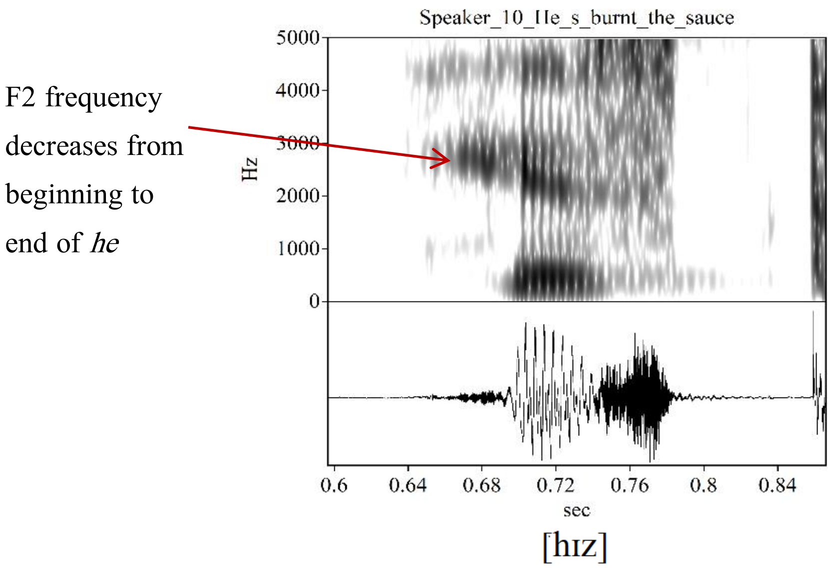


Figure 23. Spectrogram and waveform of an instance of *he's* produced by S10 in which F2 is noticeably sloping from the beginning to the end of the piece.

The decrease of F2 frequency in *he* is not unusual in the data collected. The mean F2 frequency calculated across speakers and repetitions exhibits a slope from the beginning to the end of *he*. Figure 24 shows the mean formant dynamics of the vocoid alone in two subsets of the data collected: *he's* on the left (N = 82), and *he'd* on the right (N = 96).

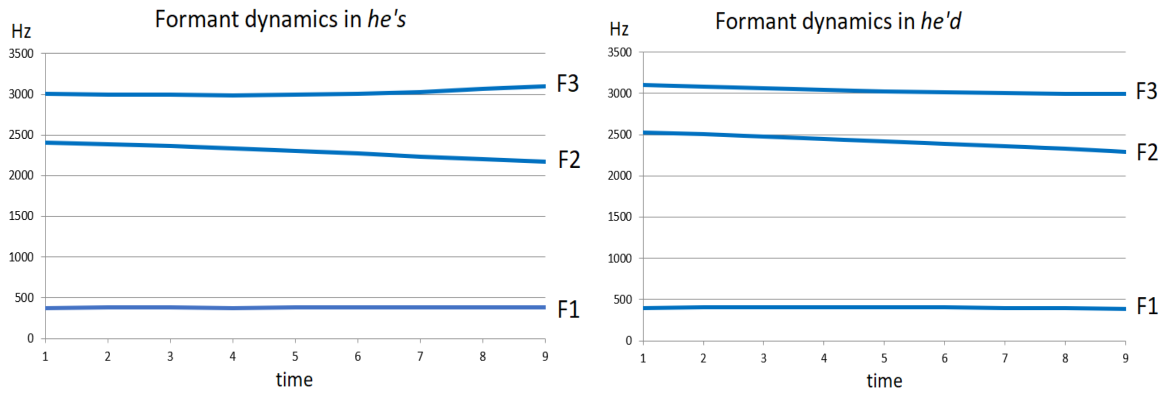


Figure 24. Formant dynamics of the vocoid in *he's* (left) and *he'd* (right) calculated across speakers and repetitions.

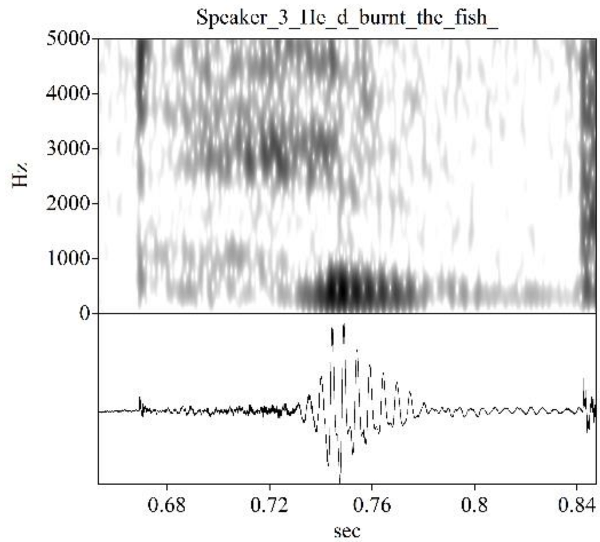
The mean frequency of F2 calculated across speakers and repetitions in all instances of *he's* in which the formants of the vocoid could be measured goes from 2436 Hz at Time Point 1 to 2157 Hz at Time Point 9 (range = 279 Hz). The mean frequency of F2 calculated across speakers and repetitions in all instances of *he'd* in which the formants of the vocoid could be measured goes from 2531 Hz at Time Point 1 to 2316 Hz at Time Point 9 (range = 215 Hz).

A few hypotheses can be formulated although none of them is entirely convincing. The downward movement of F2 suggests that the highest point of the tongue in the oral cavity is moving backward, albeit slightly. In both *he's* and *he'd* the vocoid is followed by an alveolar contoid. This means that the front of the tongue is moving closer to the alveolar ridge in a vertical 'closing' movement. However, this does not exclude that the direction of the tongue movement might be simultaneously upward and backward. The only element that contradicts this hypothesis is that F1 is stable throughout the duration of the vocoid. Another possible explanation for this feature is that the formants move towards the locus for the alveolar vocoid that follows. Although we would expect only the formant transitions – the formants in the last portion of the vocoid – to move towards the alveolar locus, the vocoids in the data are very short. Unfortunately, the pronoun *he* is followed only by alveolar sounds in *he*+aux combinations (*he's*, *he'd*, *he'll*) so this hypothesis cannot be tested using the data collected. Another hypothesis is that the vocoid becomes more central during its articulation. Vowel centralisation is a common feature of reduction (Lindblom, 1963). However, it usually affects the entire duration of the vocoid, in that the vocoid is still realised as a monophthong but with a more mid central quality than when it is unreduced. In the instances observed in the present dataset, the movement of the tongue occurs during the production of the piece. The movement towards a more central vocoid would explain the direction of F2.

The reduced instances of *he* in the dataset are characterised by a short vocoid. The duration of the vocoid in the subset *he's* ranges from 9 ms to 55 ms (mean = 29 ms, median = 28 ms, SD = 11 ms). The duration of the vocoid in the subset *he'd* ranges from 5 ms to 49 ms (mean = 25 ms, median = 24 ms, SD = 11 ms)[6]. The short duration of the vocoid can be explained with reference to two main articulatory phenomena. The first phenomenon is the temporal reorganisation of the phonetic events that occur in the vocal tract. As was observed in the pronoun *I*, the gesture in the oral cavity and the vibration of the vocal folds in the larynx can be temporally realigned so that they occur simultaneously only for a short

[6] As explained in Chapter 2 (methodology) the vocoid in *he'll* and *he was* cannot be segmented from the sounds that follow and therefore its duration cannot be measured.

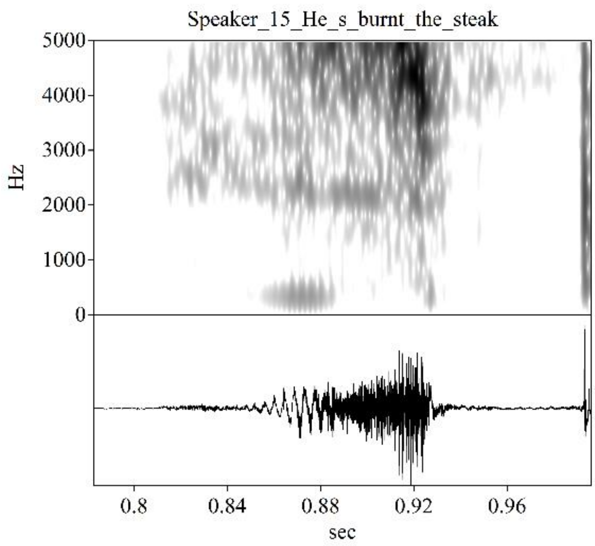
period of time. This is the case also in several instances of *he* in which the vocoid is short: the palatal articulation is still in place, as is the vibration of the vocal folds, but they are temporally reorganised rather than co-occurrent. Figure 25 shows an instance of *he'd* in which the palatal articulation can be seen (and heard) during the glottal friction, but not during voicing. In the portion of speech in which the vocal folds are vibrating, F2 and F3 are very weak or almost absent, suggesting that the alveolar closure might already be in place. Despite the presence of voicing, it is not possible to identify a vocoid in this instance of *he'd*.



[hĭ̬]

Figure 25. Spectrogram and waveform of an instance of *he'd* uttered by S3 in which a vocoid is not identifiable despite the presence of voicing.

The second phenomenon that leads to the realisation of a short vocoid in /hɪ/ is the short duration of the vibration of the vocal folds. Also in this case, the oral gesture tends to be in place, to last for a longer period of time than the actual vocoid, and to be visible in the spectrogram during the glottal friction. However, the period of time in which the vocal folds vibrate is shorter than the duration of the gesture of the oral articulation. Figure 26 shows an instance of *he's* in which F2 is visible from the start of the glottal friction to the end of the alveolar friction. This suggests that the gesture in the oral cavity is articulated over a longer period of time than the portion of voicing. In this instance, the duration of voicing is 29 ms. Notice that the frequency of F2 decreases throughout the duration of the piece, as highlighted above.



[hɪ̥s]

Figure 26. Spectrogram and waveform of an instance of *he's* produced by S15 with a short portion of voicing.

In the most extreme cases of reduction, the pronoun *he* is characterised by a long portion of friction with an underlying palatal quality and no voicing or visible vocoid. Figure 27 shows an instance of *he'd* that is entirely voiceless. The spikes observable in the spectrogram and waveform (just before the mark at 0.8 seconds) are due to saliva noise produced in the oral cavity.

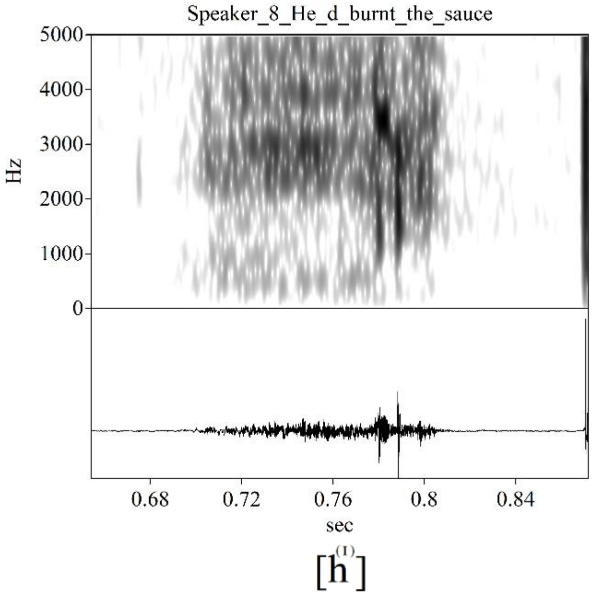


Figure 27. Spectrogram and waveform of an instance of *he'd* produced by S8 in which there is no voicing and no identifiable vocoid.

While in the literature the glottal friction in onset of *he* is considered optional in weak forms (see e.g. Ogden, 1999), in the data analysed it is always present. In a few instances it is weak and hardly audible. This feature seems to occur exclusively when *he* is in the pr + aux combinations *he's* or *he was*. Figure 28 shows an instance of *he's* in which the alveolar friction in coda is quite strong, but the glottal friction in onset is weak and hardly audible.

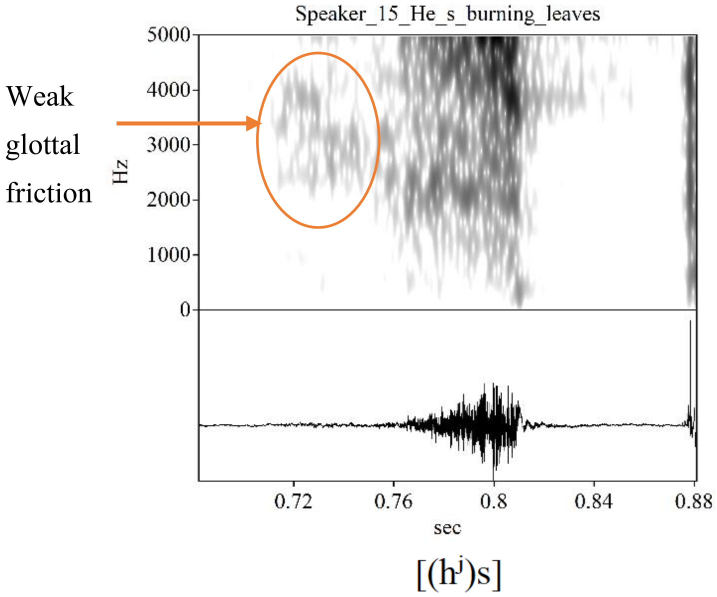


Figure 28. Spectrogram and waveform of an instance of *he's* produced by S15 in which the initial glottal friction is very weak.

To summarise, the pronoun *he* in the dataset is characterised by a palatal – or close front – tongue gesture that is audible throughout the piece and visible in the form of F2. Unexpectedly, the frequency of F2 decreases during the piece. Also *he* , like *I*, can be realised with a very short vocoid. The main reason for the realisation of a short vocoid is the temporal reorganisation of the events occurring in the laryngeal cavity and the supralaryngeal cavity. Another reason is the variable degree of voicing. The main phonetic features of the pronoun *he* are the palatal gesture and the glottal friction. A vocoid might be realised or not depending on the vibration of the vocal folds and the temporal alignment of voicing and the palatal articulation. The glottal friction can be very weak but is rarely absent. The essential phonetic features of the pronoun *he* are the close front articulation and to a lesser degree the glottal friction.

3.1.3. Pronoun *she*

The phonemic transcription of the strong form of the pronoun *she* found in pronouncing dictionaries is /ʃiː/, and the weak form is /ʃi/. The friction in *she* is produced by the tip or blade of the tongue in a position of close approximation to the post-alveolar region between the alveolar ridge and the hard palate. This palato-alveolar (or postalveolar) friction is described as being sibilant or strident because the turbulent air is produced when the high velocity airstream created in the narrow channel in front of the constriction hits the lower teeth (Ladefoged and Maddieson, 1996; Stevens, 1998). In unreduced instances, a vocoid with a near-close near-front quality follows the initial friction. Figure 29 shows an unreduced instance of *she'd* characterised by a strong initial friction and a long vocoid with steady formants.

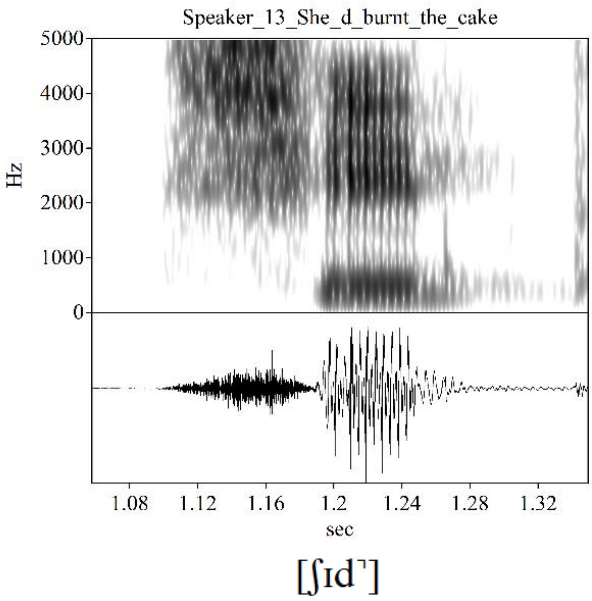


Figure 29. Spectrogram and waveform of an unreduced instance of *she'd* produced by S13.

Although to a lesser extent compared to the glottal friction at the beginning of *he*, the palato-alveolar friction at the beginning of *she* is also affected by the articulation of the vocoid that follows it. In *she*, the tongue is the active articulator involved in the production of the friction, which means that the shapes and positions that the tongue can adopt during the articulation of the friction are constrained by the production of the friction itself. However, the friction in *she* has an audible palatal quality. In some instances, the formants corresponding to the gesture of the vocoid are visible during the palato-alveolar friction. Figure 30 shows an instance in which F2 and F3 are visible from the start of the initial friction to the end of the vocoid.

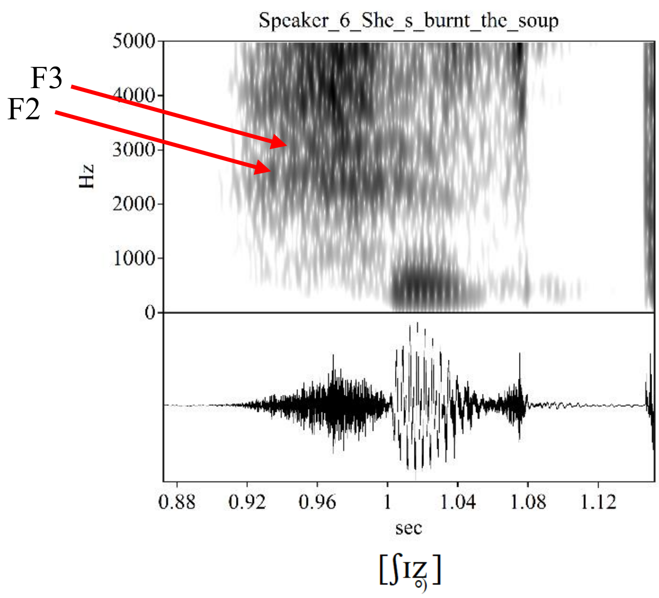


Figure 30. Spectrogram and waveform of an instance of *she's* produced by S6 in which F2 is visible from the beginning of the friction.

It can be noticed in Figure 30 that F2 decreases in frequency towards the end of the vocoid, similarly to what was observed in the pronoun *he*. The first three formants in the vocoid in *she* were measured and their means calculated across speakers and repetitions.

Figure 31 shows the formant dynamics of the vocoid in *she's* on the left (N = 78) and *she'd* on the right (N = 92).

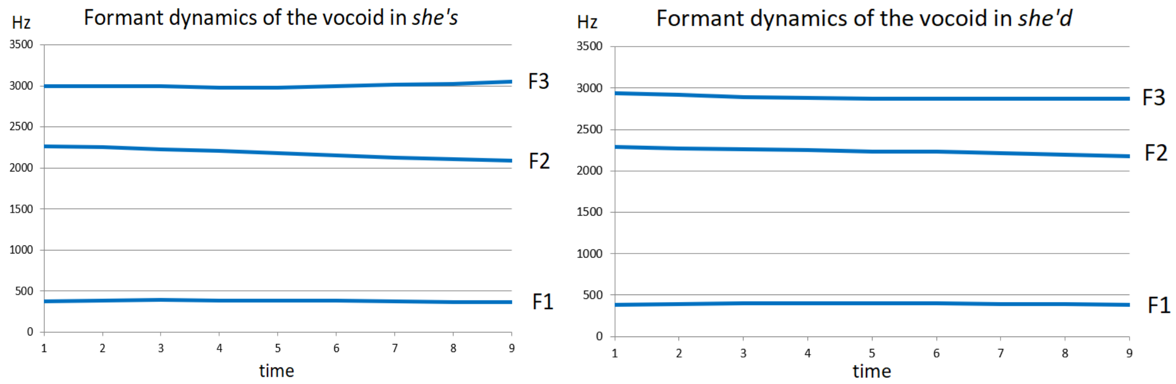
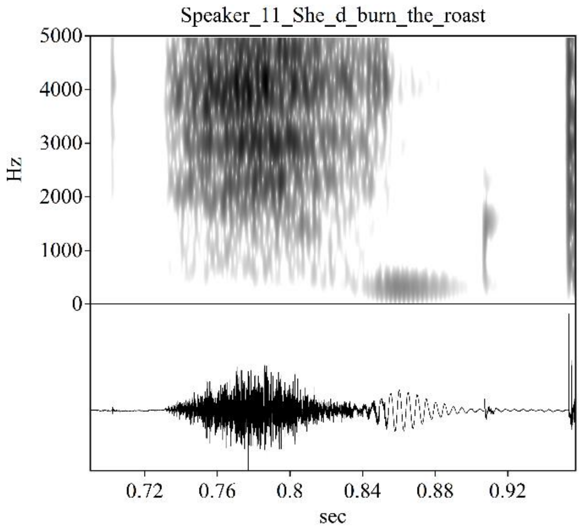


Figure 31. Formant dynamics of the vocoid in *she's* (left) and *she'd* (right).

Figure 31 shows that the slope of F2 in *she* is not as noticeable as it is for the pronoun *he*, especially in *she'd* (right). The mean frequency of F2 in *she's* goes from 2262 Hz at Time Point 1 to 2172 Hz at Time Point 9 (range = 155 Hz). The mean frequency of F2 in *she'd* goes from 2302 Hz at Time Point 1 to 2172 Hz at Time Point 9 (range = 130 Hz).

As for the pronoun *he*, most of the variation in the reduction of *she* can be classified as one of two main features: the temporal reorganisation of phonetic events, and the short duration or even absence of voicing. In the first case, the two main phonetic events are: the articulation occurring in the supralaryngeal cavity and the phonation occurring in the laryngeal cavity. Variations in the alignment of these two events result in different acoustic outputs. For example, Figure 32 shows an instance of *she'd* in which the vibration of the vocal folds occurs after the palatal articulation in the oral cavity (superimposed on the palato-alveolar friction) and is simultaneous to the period of alveolar closure.



(ɪ)
[ʃd̚]

Figure 32. Spectrogram and waveform of a reduced instance of *she'd* produced by S11 in which voicing occurs after the palatal articulation.

Figure 32 is a reduced instance of *she'd* in which the vibration of the vocal folds starts at the end of the palatal articulation and is simultaneous to the oral closure. The duration of the portion of voiced friction around the 0.84 mark is 15 ms.

Another feature of reduced instances of *she* is the short duration of voicing. Figure 33 shows an instance of *she's* in which the palatal articulation is visible throughout the piece in form of clear F2 and F3. Between the palato-alveolar friction and the alveolar friction there is a period of low-amplitude friction that corresponds to a more open constriction in the oral cavity in which the formants are more prominent in the spectrogram than the high-frequency friction. However, the vibration of the vocal folds starts only at the end of this more open articulation and lasts for a very short time. Although the short period of voicing could not be measured accurately in the waveform due to the friction, its duration is approximately 16 ms.

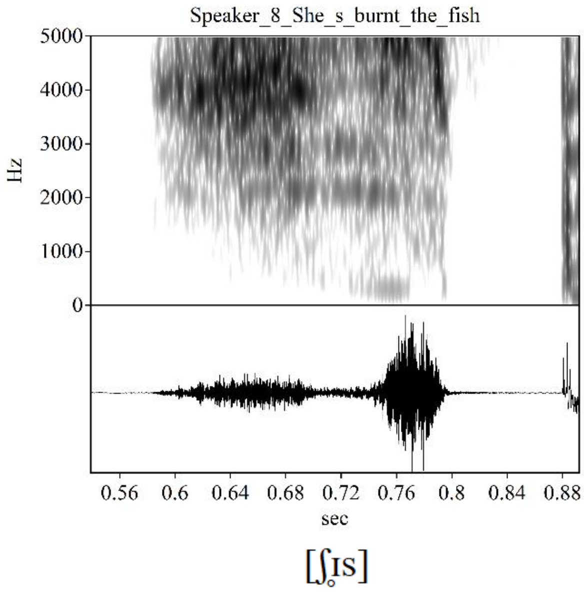
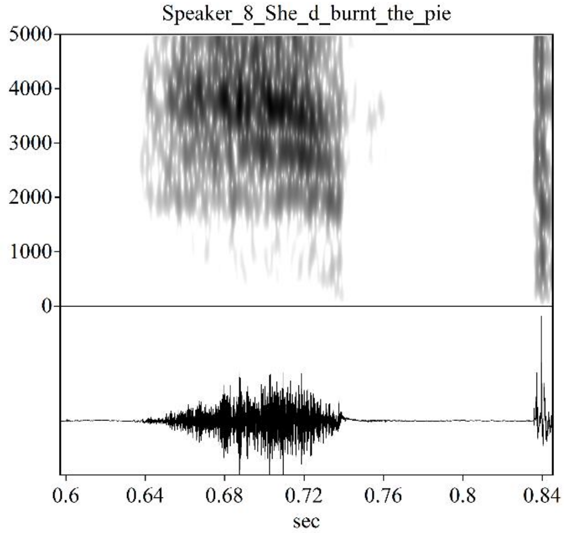


Figure 33. Spectrogram and waveform of an instance of *she's* with a short period of voicing produced by S8.

Figure 33 is an instance of *she's* in which, despite the audible palatal quality and the visible formants, a clear vocoid is not apparent or measurable due to lack of voicing in correspondence with the more open constriction in the oral cavity.

Although phonologically *he* and *she* are similar – they are characterised by initial friction and a close front vocoid – they seem to reduce in different ways. Firstly, the palato-alveolar friction is never absent in any of the *she*+ aux combinations. Its duration ranges from 58 ms to 160 ms (range = 102 ms, mean = 101 ms, median = 101 ms, SD = 16 ms, N = 278). The duration of the glottal friction in *he* ranges from 31 ms to 165 ms (range = 134 ms, mean = 66 ms, median = 65 ms, SD = 17 ms, N = 264).

In the most extreme cases of reduction of *she*, there is only palato-alveolar friction. Figure 34 shows a highly reduced instance of *she'd*.



[ʃ$^{w)}$]

Figure 34. Spectrogram and waveform of a highly reduced instance of *she'd* produced by S8.

Moreover, in the dataset collected, the instances of *she* characterised by absence of voicing (N = 13/215, 6%) are slightly more numerous than the instances of *he* characterised by absence of voicing (N = 7/215, 3%). However, voicing is also linked to the auxiliary of the piece. For both pronouns *he* and *she*, voicing is present in a higher number of instances in pr + *'d* than in pr + *'s*. There is no vibration of the vocal folds in 10% of instances of *she's* analysed (N = 11/108, 10%), while there is no vibration of the vocal folds in 2% of instances of *she'd* analysed (N = 2/107, 2%). The same pattern occurs in *he's* and *he'd*: 6% of instances of *he's* do not have any vibration of the vocal folds (N = 6/106, 6%), while only one instance of *he'd* does not have vibration of the vocal folds (N = 1/109, 1%). Table 6 summarises the percentage and number of instances of *he'd*, *he's*, *she'd* and *she's* in which there is no voicing.

	'd	*'s*
He	1% (1/109)	6% (6/106)
She	2% (2/107)	10% (11/108)

Table 6. Percentage and numbers of instances of *he'd*, *he's*, *she'd* and *she's* without voicing.

Although these are small differences between the vocoids in *he* and *she*, in the data analysed it seems that in the pronoun *he* the essential phonetic feature that constitutes and maintains the identity of the piece is the close front, or palatal, quality of the articulatory gesture, rather than the initial glottal friction; while in the pronoun *she* the essential feature is the initial palato-alveolar friction. From an acoustic phonetic point of view, this is not surprising. Sibilant fricatives such as [ʃ] inherently have higher amplitude and more defined spectral peaks than non-sibilant fricatives (Jongman et al., 2000). These acoustic features make them perceptually more salient than non-sibilant fricatives (Cho, Jongman, Wang and Sereno, 2020). Therefore, the most prominent acoustic and perceptual feature of *she* is likely to be the strident friction and for this reason it is more likely to be retained in reduced speech. Moreover, in the small paradigmatic system of English pronouns, *she* is the only pronoun to be characterised by strident friction. The other fricatives are [h] and [ð], which are non-strident. This means that no other acoustic features are needed for *she* to be correctly identified even in reduced speech.

To summarise, the pronoun *she* in the dataset is always characterised by a period of palato-alveolar friction. The friction can be followed by a vocoid of variable duration with a near-close near-front (or palatal) quality. If the vocoid is not audible and/or visible, the reason is most likely to be a temporal realignment of the phonetic events occurring in the oral cavity and in the laryngeal cavity. That is, the palatal gesture is always articulated and can be superimposed on the palato-alveolar friction in absence of voicing or when the vibration of the vocal folds lasts only for a short period of time. Even when the vocoid is apparently absent, the close front tongue gesture can be observed in the formant structure retained in the acoustic signal. The amount of vibration of the vocal folds is also highly variable and can be absent altogether. The only essential phonetic feature of the pronoun *she* is the palato-alveolar friction.

3.1.4. Pronoun *it*

The pronoun *it* shows the widest range of variation of all the pronouns. Its realisation largely depends on the auxiliary it is combined with. All the *it+* aux paradigms will be looked at in more detail in Chapter 4. In this section only the main variations and features of reduction in the realisation of *it* will be illustrated.

The reason for the high degree of variability in the realisations of *it* is twofold. Firstly, the voiceless alveolar plosive /t/ is known to be particularly prone to a wide range of variation in the accents of English (see e.g. Buizza, 2010; Harris, 1994; Lavoie, 2001). Secondly, the pronoun *it* is the only pronoun with a non-empty coda. All the other pronouns are characterised by an open syllable, which means that the cliticised form of the auxiliaries – a consonant – can occupy the coda position. In combination with the pronoun *it*, the cliticised forms of the auxiliaries create consonant clusters in coda position that are not allowed in English. As it will be described in the next section, there are various ways speakers tackle this issue with different outputs.

The phonemic transcription of *it* found in pronouncing dictionaries is /ɪt/. The initial vocoid is produced with the tongue in a near-close near-front position. In its unreduced realisation, the vocoid is followed by a complete closure at the alveolar ridge which creates a build-up of intra-oral air-pressure behind the closure. When the tongue moves away from the alveolar ridge, the air-pressure is released with a burst followed by a period of glottal friction. Figure 35 shows an example of unreduced realisation of *it's* produced by S2.

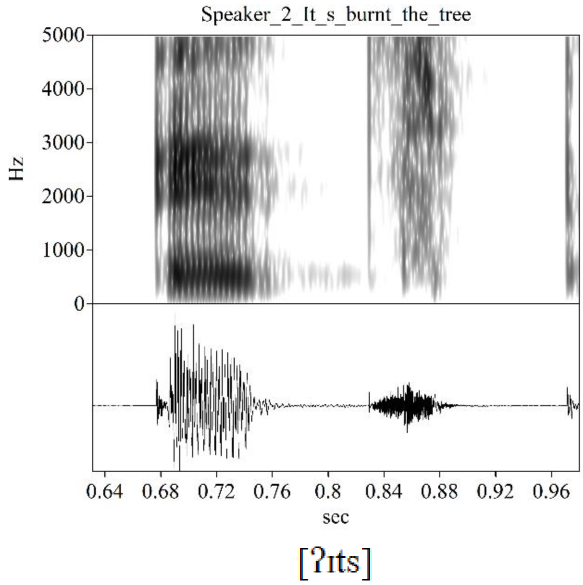
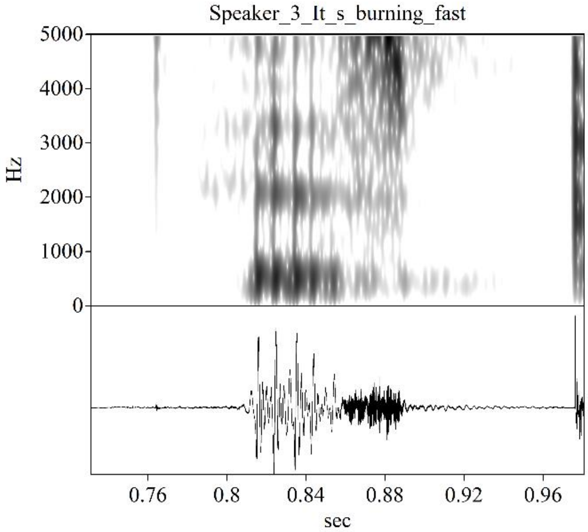


Figure 35. Spectrogram and waveform of an unreduced instance of *it's* produced by S2.

In the instance of *it's* in Figure 35, the burst is followed by a period of glottal friction (25 ms) followed by alveolar friction (33 ms). In the dataset, only 25% of instances of *it* exhibit a complete closure and release like the token in Figure 35.

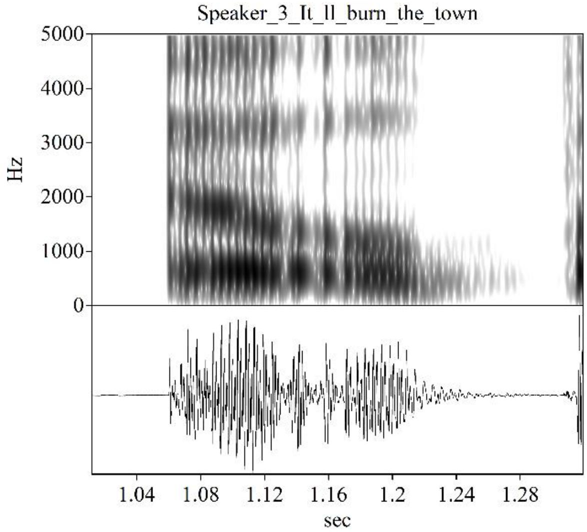
The range of variations in the realisation of *it* do not involve only the supralaryngeal articulation, but includes variation in the activity of the larynx too. In several instances in this dataset, the pronoun *it* is characterised by creaky voice during the production of the vocalic articulation. In these instances, creaky voice is the acoustic correlate of the phonological plosive in coda position. Figure 36 shows an instance of *it's* realised with creaky voice during the articulation of the vocoid, and with no closure and release for the plosive. The vocoid is followed by alveolar friction.



[ɪ̰s]

Figure 36. Spectrogram and waveform of an instance of *it's* produced by S3 with creaky voice and no complete closure and release.

Figure 36 is an instance of *it's* in which the vocoid is characterised by creaky voice throughout its duration. This is not always the case and there is a wide range of variation in the amount and type of laryngeal activity, which often seems to be speaker-specific. Figure 37 shows an instance of *it'll* in which three creaky glottal cycles are visible in the central portion of the vocoid.

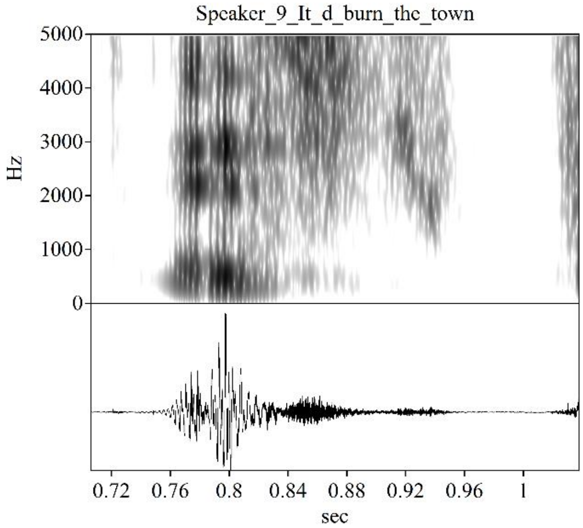


[ʔɪɪ̰o]

Figure 37. Spectrogram and waveform of an instance of *it'll* produced by S3 with a few irregular and longer glottal cycles in the middle.

Although most instances of *it* realised with creaky voice do not have a canonical plosive, in some instances, creaky voice is realised alongside a canonical plosive with a complete closure and a burst at release.

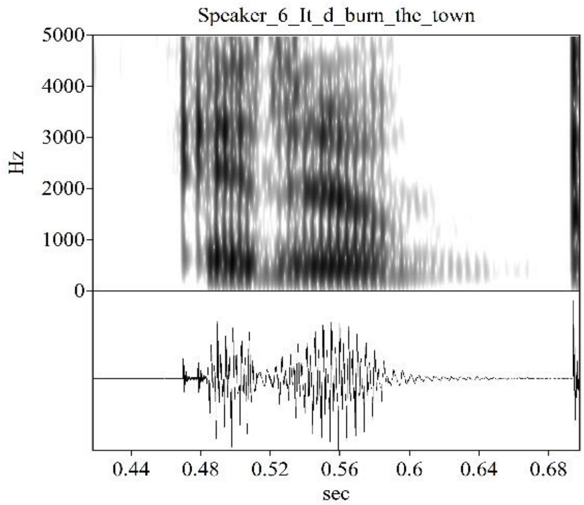
As mentioned above, /t/ is prone to a wide range of phonetic realisations. Figure 38 shows an instance of *it'd* in which both phonological plosives are realised without a complete closure but with friction at the place of articulation.



[ɪ̰t̞t̞ʷ]

Figure 38. Spectrogram and waveform of an instance of *it'd* produced by S9. Both plosives are realised with friction at the place of articulation.

In only one instance of *it'd*, /t/ is realised as a tap – a quick ballistic movement of the tongue against the alveolar ridge. The phonological representation of *it'd* is /ɪtəd/, which means that /t/ is in intervocalic position. Figure 39 shows the instance of *it'd* in which /t/ is realised as a very brief closure without build-up of air-pressure.



[ʔɪɾə]

Figure 39. Spectrogram and waveform of an instance of *it'd* produced by S6 in which the first plosive is realised as a tap.

Another feature of reduction of *it* is the apparent absence of the vocoid. Figure 40 is an instance of *it's* in which the vocoid is apparently absent. The release of the stop is weak, as it can be seen from the low amplitude of the burst. The auditory impression of the piece is that of an affricate. Since there are no other affricate sounds in the paradigmatic system of pronoun and auxiliaries, it is fair to assume that the piece *it's* can be highly reduced and realised as an affricate without becoming problematic for the correct interpretation of the pr + aux.

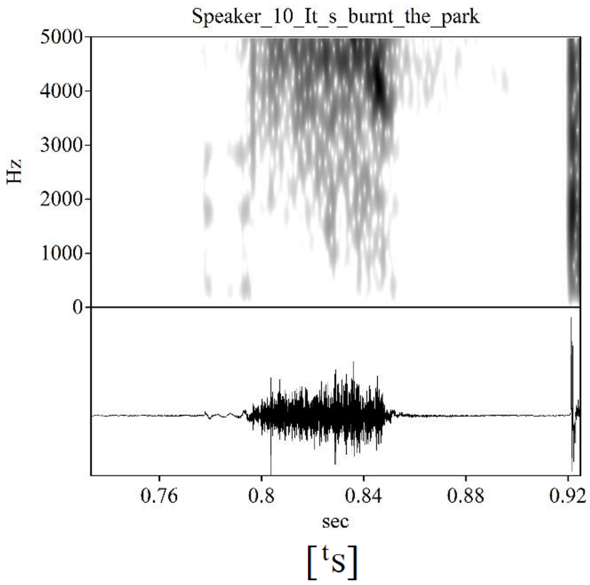


Figure 40. Spectrogram and waveform of an instance of *it's* produced by S10 characterised by an inaudible vocoid and a weak burst.

A more extreme case of reduction is shown in Figure 41 (note the change of the frequency scale to display frequencies up to 8 kHz). In this instance of *it's*, the pronoun seems to be absent altogether. In the spectrogram and waveform there is only a period of voiceless alveolar friction. Also in this instance the auditory impression is of an affricate rather than a fricative. This observation raises the question of how the auditory impression of an affricate is obtained in the absence of a burst.

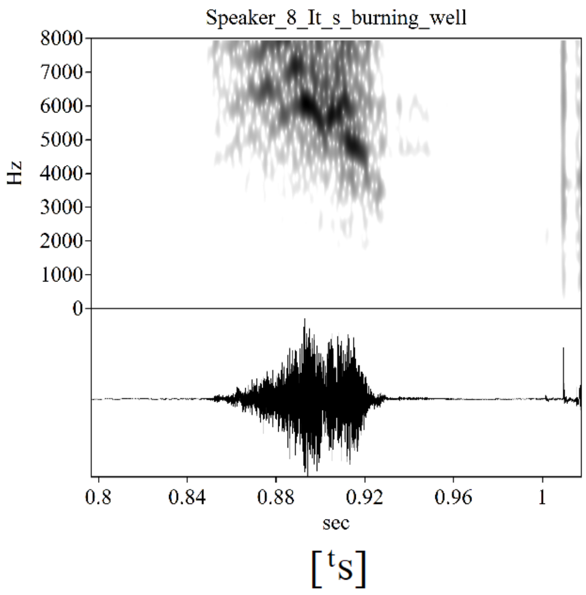


Figure 41. Spectrogram and waveform of a highly reduced instance of *it's* produced by S8. The frequency scale displays frequencies up to 8 kHz.

Two factors can explain the perceptual impression of an affricate in the instance in Figure 41. The first factor is the silence before the friction. Although we do not know whether there is an articulatory closure in the vocal tract, the period of silence could give the auditory impression of the hold phase of a plosive. The second factor is the sharp increase in amplitude of the friction noise and the high energy noise produced. The energy contour of the fricative as observed in the waveform is quite different from that of a phonological fricative.

To summarise, the pronoun *it* is characterised by a wider range of phonetic variation than any other pronoun. Its realisations largely depend on the auxiliary it is combined with. Some of the variations will be illustrated in more detail in Chapter 4. The main feature observed in the dataset is the use of laryngeal features such as creaky voice as acoustic correlates of the phonological plosive. When articulated in the supralaryngeal cavity, /t/ is characterised by a high degree of variation, including realisations such as friction throughout its duration, or a quick ballistic movement of the tongue. Since *it* is the only pronoun with a contoid in coda position, the vocoid can be highly reduced and even be absent altogether. Due to the wide range of variation in the realisation of *it*, it is difficult to establish the essential phonetic features of this pronoun. The feature most frequently observed is glottality. However, glottality is not always present. In its absence, alveolarity is present and can be considered the essential phonetic feature of *it*.

3.1.5. Pronoun *you*

The phonemic transcription of the strong form of the pronoun *you* is /juː/, while the weak forms are /ju/ or /jə/. *You* is characterised by a vocoid produced with two simultaneous movements. The tongue moves from an initial position in which the anterior part of the dorsum is in open approximation with the pre-palatal region producing a close front sound, to a position in which the posterior part of the dorsum is in open approximation with the velum producing a close back sound. At the same time, the lips change shape from being spread to being rounded. Figure 42 shows an instance of *you've* in which the movement of the formants reflects the movement of the tongue and the lips during the production of *you*. In particular, F2 – the acoustic correlate of the front-back tongue position – shows the largest movement from the frequency of 2650 Hz at the beginning of the vocoid to the frequency of 1950 Hz at the end of the vocoid (range 700 Hz). F1 – the acoustic correlate of the tongue height – moves from 350 Hz at the beginning of the vocoid to 410 Hz at the end of the vocoid (range 60 Hz). F3 – the acoustic correlate of lip shape and protrusion moves from 3030 Hz at the beginning of the vocoid to 2890 Hz at the end of the vocoid (range 140 ms).

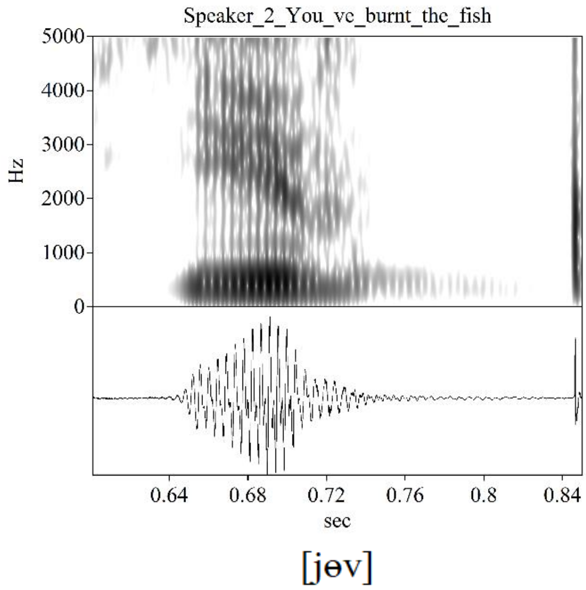
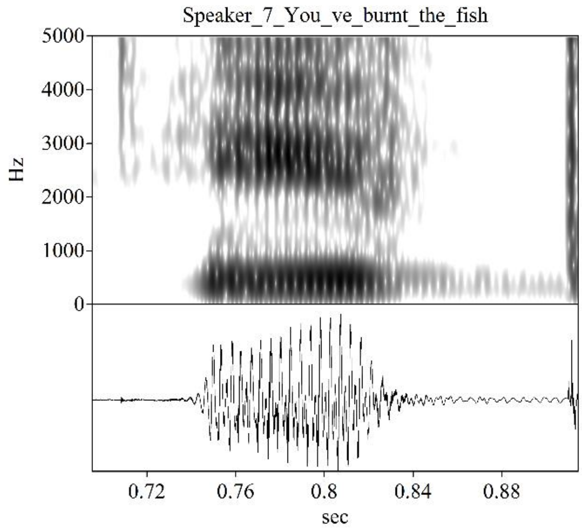


Figure 42. Spectrogram and waveform of an instance of *you've* produced by S2 in which the movement of the higher formants (F2, F3 and F4) is clearly visible.

Figure 42 is an instance of *you* in which the tongue is clearly moving during the production of the piece. However, most realisations of *you* in the dataset are characterised by flat (or almost flat) formants throughout the duration of the vocoid. From this, it can be inferred that the tongue does not move considerably during the production of the piece. This can be considered a case of reduced magnitude of the gesture. Although a reduction in magnitude of gesture is to be expected in reduced speech, in the dataset analysed the reduction of gesture magnitude is evident also in instances in which there is no apparent temporal reduction. Figure 43 shows an instance of *you've* in which the duration of the vocoid is 71 ms and the formants are almost steady, although they fall rather sharply at the end of the vocoid as the friction begins.



[i̹v]

Figure 43. Spectrogram and waveform of an instance of *you've* produced by S7 in which there is little formant movement. The duration of the vocoid is 71 ms.

Figure 43 shows an instance of *you've* in which the magnitude of the articulatory gesture is reduced but there is no evident temporal reduction. The only audible change in the vocoid is a hint of lip-rounding towards the end of the vocoid. This realisation characterised by a close front, or palatal, vocoid without a noticeable tongue movement is common across the dataset. However, like the pronouns described so far, also *you* is characterised by a wide range of realisations including the temporal reorganisation of phonetic events. Figure 44 shows an instance of *you'd* in which a long period of friction before the vibration of the vocal folds reveals that the articulation in the oral cavity started well before voicing. The duration of the friction up to the first glottal cycle is 107 ms.

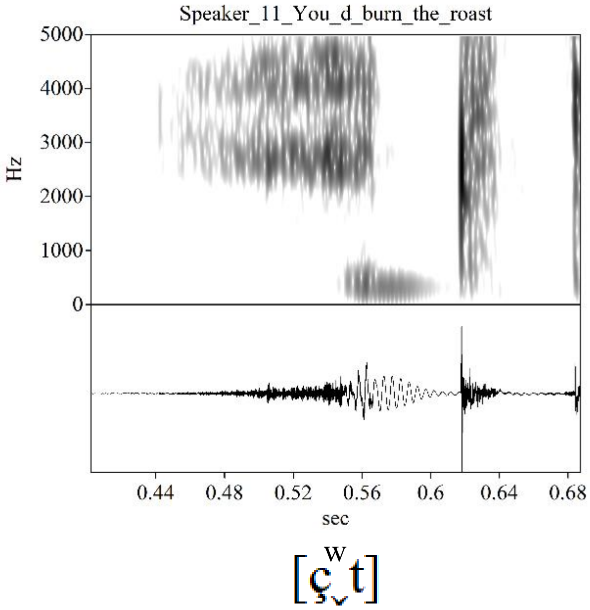


Figure 44. Spectrogram and waveform of an instance of *you'd* produced by S11 in which there is a long period of voiceless palatal friction (duration 107 ms) before the beginning of voicing.

The auditory impression is that of a palatal, or close front, sound with a trace of lip rounding throughout its duration. In fact, the palatal quality seems to be the constant feature of all instances of *you* analysed. The presence of friction at the beginning of the pronoun *you* has been observed in several instances in the dataset. This feature will be analysed in more detail in Section 6.1, Section 5.2.5.1, and Section 7.3.2.

The most extreme case of reduction of the pronoun *you* in the dataset is again an example of temporal reorganisation of laryngeal and supralaryngeal events. Figure 45 shows an instance of *you're* in which the vocoid is very short (5 ms). However, the articulatory gesture starts before the vibration of the vocal folds begin, and it can be seen in the spectrogram as a period of voiceless friction in which F2 and F3 are visible. Moreover, the vibration of the vocal folds continues after the closure of the oral cavity has been achieved.

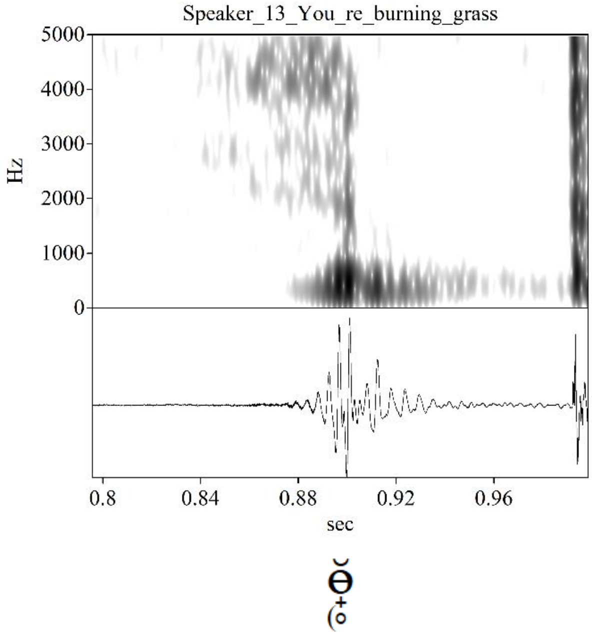


Figure 45. Spectrogram and waveform of a reduced instance of *you're* produced by S13 characterised by a short vocoid.

To summarise the phonetic characteristics of *you*, this pronoun is regularly produced with a reduced degree of magnitude of gesture as evident in the flat (or almost flat) formants even in instances in which there is no apparent temporal reduction. The most common realisation is that of a vocoid with variable duration and a close front quality, often with a hint of lip rounding. The essential phonetic features of the pronoun *you* are palatality and to a lesser degree labiality. While the palatal quality is the main feature in all instances of *you*, the labiality can be very weak in some of the most reduced instances of *you*.

3.1.6. Pronoun *we*

The phonemic transcription of the pronoun *we* found in pronouncing dictionaries is /wiː/ for the strong form and /wi/ for the weak form. In unreduced instances, *we* is articulated as a vocoid with two simultaneous movements. The tongue moves from an initial position in which the posterior part of the dorsum is in open approximation with the velum producing a close back sound, to a position in which the anterior part of the dorsum is in open approximation with the front of the palate producing a close front sound. At the same time, the lips change configuration from a rounded lip shape to a spread lip shape. While the first position is not held, the second position, and the associated lip shape, can have a steady state. Figure 46 shows an instance of *we'd* in which the movement of the formants is visible. F2 moves from a lower frequency (1241 Hz) at the beginning of the vocoid to a higher frequency (2318 Hz) towards the end of the vocoid. The movement of F2 indicates that the tongue moves from the back to the front of the oral cavity. F1 has a frequency of 428 Hz throughout the vocoid.

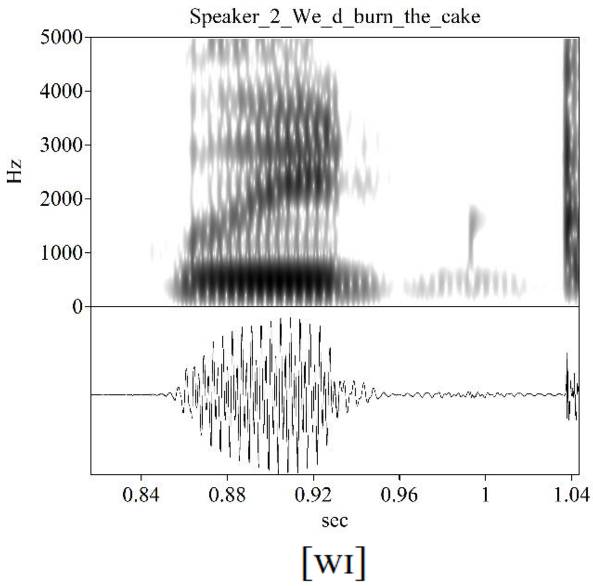
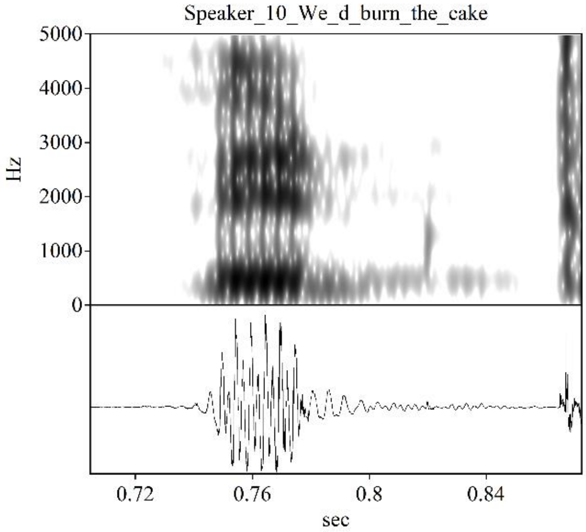


Figure 46. Spectrogram and waveform of an instance of *we'd* produced by S2 characterised by visible F2 movement.

Similarly to the pronoun *you*, also *we* is often realised with a steady or almost steady quality throughout its duration, suggesting that the tongue and lips do not move considerably during the production of the vocoid. In this case, the formants are flat instead of changing frequency in time. Figure 47 shows an instance of *we'd* in which the formants are steady throughout the duration of the piece.

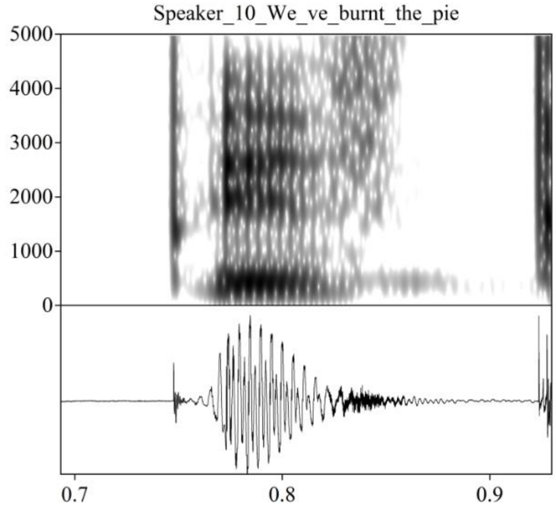


[ø]

Figure 47. Spectrogram and waveform of an instance of *we'd* produced by S10 characterised by flat formants.

The quality of the vocoid in Figure 47 is close-mid front and rounded, close to [ø].

In a few instances (N = 12/158, 8%), the labial-velar gesture at the beginning of *we* is articulated with a complete closure and burst instead of a stricture of open approximation. In these instances, a silent hold phase is followed by an abrupt release of air with a bilabial quality, which is visible in the waveform and spectrograms in form of a spike before the vocoid starts. Figure 48 shows an instance of *we've* with an initial bilabial burst.



[bɪf]

Figure 48. Spectrogram and waveform of an instance of *we've* produced by S10 that starts with a bilabial release.

Once again, the instances in which reduction manifests as a short vocoid, can be analysed as temporal realignment of phonetic events. Figure 49 shows an example in which the magnitude of the articulatory gesture is not reduced and can be observed in the spectrogram in form of weak formants starting before voicing begins. However, the vocoid is very short. While the vocoid is very short (13 ms) the gesture is visible for a duration of 72 ms. In this case, F2 moves considerably from 1603 Hz at the beginning of the weak friction to 2409 Hz just before the start of voicing (maximum frequency height) and then to 1805 Hz at the end of the vocoid.

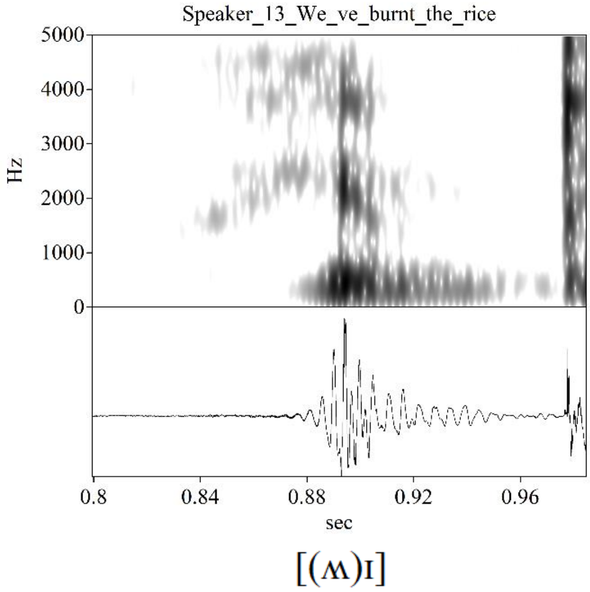
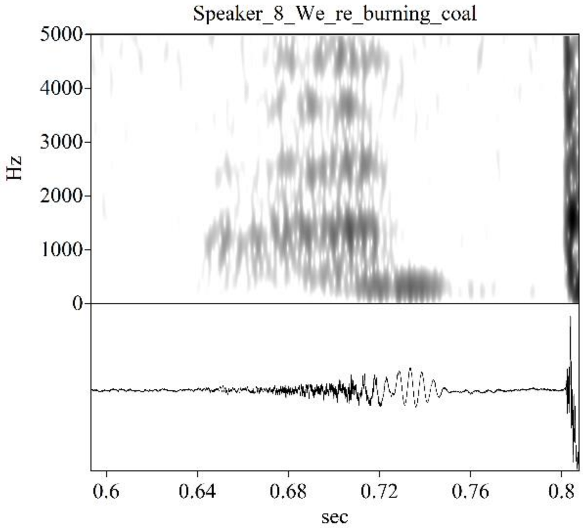


Figure 49. Spectrogram and waveform of an instance of *we've* produced by S13 with a short vocoid but a long articulatory gesture.

Instances such as the one shown in Figure 49 suggest that the realisation of the pronoun *we*, as well as other pronouns, as a short vocoid do not necessarily imply the temporal compression of the gesture, but rather a temporal reorganisation of the phonetic events involved in the production of the sounds.

In the most reduced instances, *we* is realised as a short vocoid with close-mid front rounded quality. Figure 50 shows an instance of *we're* in which there is a weak portion of friction characterised by low F2 displaying little movement followed by a period of voicing (33 ms).



[ʍ̬]

Figure 50. Spectrogram and waveform of a highly reduced instance of *we're* produced by S8.

To summarise, the reduction of *we* is characterised by an apparently short vocoid. The vocoid can have a steady quality or change quality in time. While labialisation, or lip rounding, is a constant feature of the pronoun *we*, the position of the tongue can be raised towards the back of the oral cavity, resulting in a sound similar to [w] or, more rarely, towards the front of the palate, resulting in a sound similar to [ɥ] or [ø]. Both vocoid qualities can be realised simultaneously to voicing or voicelessness. In the latter case, the production of weak friction is the only means to hear the articulated sound and to see it in the spectrograms. The essential phonetic features of the pronoun *we* are labiality and velarity. Although the tongue is usually raised towards the velum, it can have a more advanced position towards the hard palate.

3.1.7. Pronoun *they*

The phonemic transcription of the pronoun *they* is /ðeɪ/. In conventional descriptions, English /ð/ is described as a dental fricative (Jones, 1960; Roach, 2000; Cruttenden, 2008).

To produce dental friction, the tip of the tongue creates a narrow constriction behind the upper front teeth. Dental fricatives are described as non-sibilant sounds in which the turbulence is produced at the constriction (Ladefoged and Maddieson, 1996). However, the dental fricative in English is a peculiar sound from both a phonetic and a phonological point of view. From a phonological point of view, one of the peculiarities of /ð/ is its distribution. In word-initial position /ð/ is limited to function words (e.g. *they*, *the*, *this*, *that*, *though*) while in content words its voiceless counterpart occurs (e.g. *think*, *thick*, *thistle*). From a phonetic point of view, it has been observed that its phonetic realisation in connected speech is highly variable (see e.g. Manuel, Shattuck-Hufnagel, Huffman, Stevens, Carlson and Hunnicutt, 1992; Ogden 1998; Local, 2003; Ogden, 2012). Ogden (1998, 2009) states that the main phonetic features of /ð/ in the article *the* are voicing and dentality, and that the manner of articulation is variable but rarely exhibits strong friction noise.

In the pronoun *they*, the initial contoid is followed by a vocoid in which the tongue moves slightly from an initial mid-close front position to a near-close near-front position. The lips are spread throughout the pronoun. The movement of the tongue during the production of the vocoid is reflected in the movement of the formants in the spectrograms of unreduced instances of *they+* aux. Figure 51 shows an instance of *they're* produced by S2.

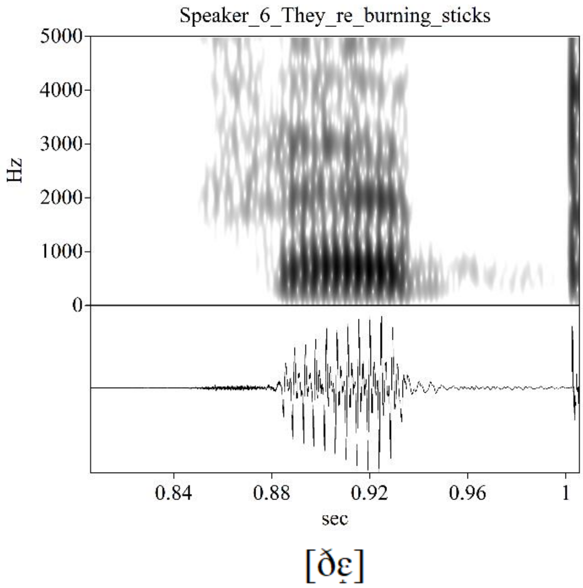


Figure 51. Spectrogram and waveform of an instance of *they're* produced by S6 with dental friction at the beginning of the piece.

In the instance in Figure 51 the piece begins with weak dental friction. In the dataset analysed, *they* is realised with initial dental friction in 19% of instances (N = 63/329). In most items collected the initial dental articulation does not create a stricture of close approximation. The most common realisation is a complete closure followed by a burst (N = 252/329, 77%). The period of silence created by the complete closure is not usually identifiable in the spectrogram or waveform because the pronoun is in sentence-initial position preceded by a pause. However, the presence of a burst indicates that during the closure there is a built-up of air-pressure that is then abruptly released. Figure 52 shows an example of a burst and a short release phase at the beginning of *they'd* produced by S10.

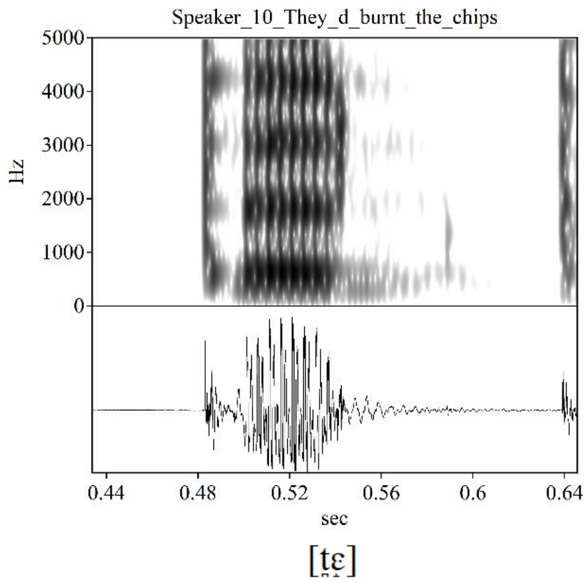
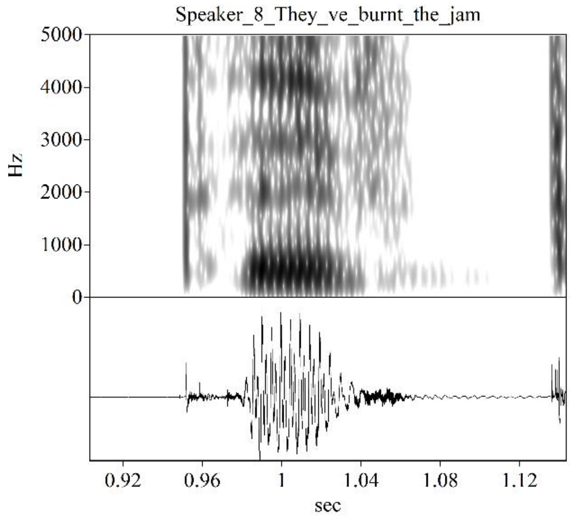


Figure 52. Spectrogram and waveform of an instance of *they'd* produced by S10 in which the initial contoid is produced with a complete closure and a burst.

Although the initial contoid in the instance shown in Figure 52 is not produced with friction, the articulation is audibly dental. In this case, the lag between the burst and the beginning of vibration of the vocal folds is 14 ms. The auditory impression is that of a voiceless unaspirated dental stop. However, there are instances in which the burst at the beginning of *they* is followed by a longer release phase. An example is shown in Figure 53. In this instance, the duration of the release phase is 30 ms. The auditory impression is that of a dental stop followed by a short portion of glottal friction.



[t̪ʰɛ̝]

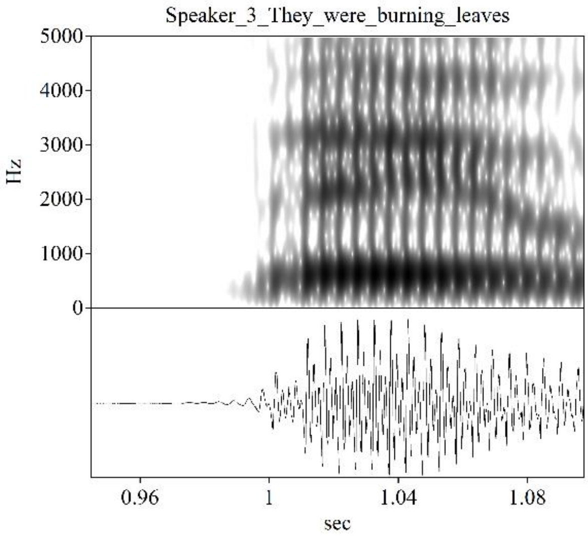
Figure 53. Spectrogram and waveform of an instance of *they've* produced by S8 in which the initial contoid is produced with a burst and release phase.

The duration of the initial contoid depends on its realisation. Table 7 reports the mean duration of the dental friction (when present) and the duration of the release phase of the burst (when present) in three pr + aux.

	Dental friction duration (ms)	N	Release after burst duration (ms)	N	Total N
They've	45	11	25	42	53
They'd	42	10	28	45	55
They'll	45	14	25	39	53

Table 7. Mean duration of the initial dental friction in the instances in which it is realised, and mean duration of the release phase in instances in which there is an initial dental burst.

In some instances, it is not possible to hear a burst or friction at the beginning of the piece or to see a burst or friction noise in the spectrogram and waveform. However, the dentality can still be perceived. This suggests that the speaker has moved the tip of the tongue towards the teeth but the stricture created is not narrow enough to cause the production of friction or a closure. It can be inferred that in these instances a dental articulation with open approximation is created, and that the auditory impression of dentality is retained in the formant transitions at the beginning of the vocoid even when a labiodental contoid cannot be identified. Figure 54 shows a zoomed-in token of *they* produced by S3 in which there is no visible burst or friction.



[ð̝̞̽ɛ̝]

Figure 54. Spectrogram and waveform of an instance of *they were* produced by S3 in which there is no visible burst or friction at the beginning, but dentality can still be heard.

As for the vocoid in *they*, in most instances collected, the vocoid is realised with a steady mid-close front or mid front unrounded quality rather than with a moving gesture. Figure 54 shows a clear example of flat formants, which suggest that the tongue is in a steady position throughout the duration of the vocoid. The variability in the realisation of the vocoid is due mainly to the amount of voicing and the temporal organisation of the supralaryngeal articulation and the vibration of the vocal folds. However, there are no cases in which *they* is completely voiceless. Figure 55 shows an instance of highly reduced *they've* uttered by S9. Although a short portion of weak voicing is visible in the waveform, the piece sounds voiceless [t̪ʰe̥]. It can be observed that there is a substantial amount of aperiodic energy after the burst and that it is characterised by a clear formant structure of a close-mid front articulation.

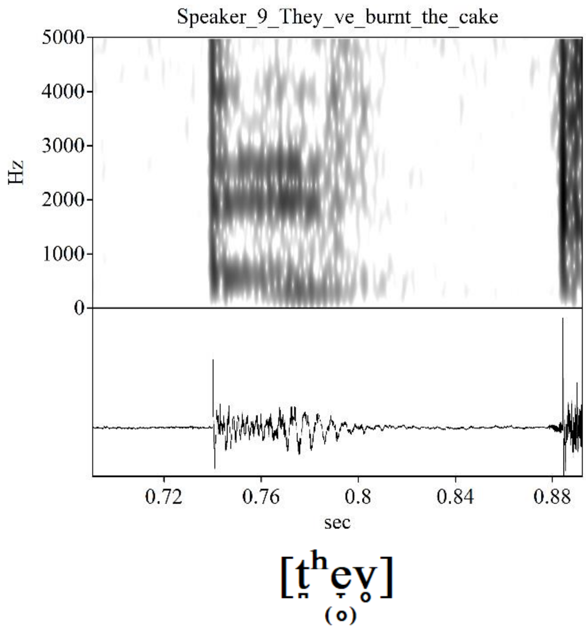


Figure 55. Spectrogram and waveform of an instance of *they've* produced by S9 in which there is a short period of voicing but the vocoid sounds voiceless.

Figure 56 shows another example of reduced instance of *they* in which the vocoid is realised with only two cycles of vocal folds vibration giving the auditory impression of creaky voice. The glottal cycles are preceded by a weak portion of dental friction.

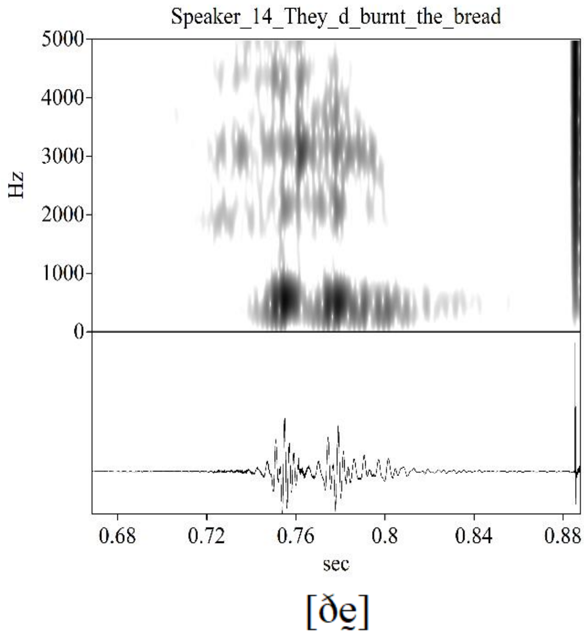


Figure 56. Spectrogram and waveform of an instance of *they'd* produced by S14 in which creaky voice is produced after a period of weak dental friction.

To summarise, the pronoun *they* is characterised by initial dentality and a mid or mid-close front spread vocoid. The variations in the realisation of *they* are mainly in the realisation of the initial contoid. Although the place of articulation – dental – is constant, the manner of articulation varies hugely, from complete closure, to friction, to open approximation. The vocoid is regularly articulated as a steady vocoid without change in quality. Like in the other pronouns, the temporal reorganisation of the phonetic events in the supralaryngeal cavity and in the laryngeal cavity dictate the duration of the vocoid. Voicing is also variable, but never completely absent. The essential phonetic features of the pronoun *they* are dentality and a mid front vocoid articulation.

3.2. Summary of observations

The aim of this chapter was twofold. The first aim was to describe the wide range of variation observed in the realisation of pronouns. The second aim was to identify the phonetic features that characterise reduced pronouns and remain in the signal in reduced speech.

This chapter illustrated some of the most common realisations of the English pronouns. For each pronoun, the most frequent realisations were described, together with some of the most extreme cases of unreduced and reduced realisations. Throughout the chapter it was shown that the data collected exhibits a wide range of variation.

The qualitative analysis reported in this chapter showed that all pronouns are characterised by one or two main phonetic features that are always present even in highly reduced tokens. A straightforward example is the dentality in the pronoun *they*. It was observed that the phonological fricative /ð/ in onset of *they* can have a wide range of realisations – it can be realised with friction, with a complete closure and a burst release, with open approximation, or even be absent as a distinct sound. However, the auditory impression of dentality is always present even when a time-limited dental sound cannot be identified. The pronoun *they* is also characterised by a mid front unrounded vocoid.

All the pronouns are characterised by a few phonetic features. Crucially, when the pronouns are reduced, these phonetic features can be reduced too, but they are always present. The pronoun *I* is characterised by an open front vocoid. It can be short or voiceless, but the gesture for the open front vocoid is always articulated. The pronoun *she* is characterised by the palato-alveolar friction, which is always produced although it can be shortened. The pronoun *he* is characterised by a close front articulation and glottal friction, although the latter can be very weak and almost inaudible. Surprisingly, in the data analysed the glottal friction is never completely absent as described in the literature on connected speech processes. The pronoun *it* exhibits the widest range of variation and its realisation largely depends on the auxiliary it is combined with. The most frequent acoustic correlate of /t/ is as creaky voice during the vocoid. That is, most instances of *it* are characterised by glottality in the form of creaky voice or a glottal stop. However, there are instances in which neither of them is present, as was shown in Figure 40 and Figure 41. In these cases, alveolarity is present.

The pronoun *you* is characterised by palatality and labiality, although the latter can be highly reduced and hardly audible. The pronoun *we* is characterised by velarity and labiality. The pronoun *they* is characterised by dentality and a mid front vocoid. The former can be auditorily weak but it is always present. Table 8 summarises the main phonetic features that characterise the pronouns in the data collected.

Pronouns	Essential phonetic features
I	Open front articulation
she	Palato-alveolar friction
he	Close front articulation Glottal friction (which can be very weak)
it	Glottality (but not always) If glottality is not present, alveolarity
you	Palatality and labiality
we	Labiality and velarity
they	Dentality and a mid front articulation

Table 8. Summary of the main phonetic features of the pronouns.

Table 8 reports the phonetic features that are always present in the pronouns even when they are highly reduced. Several of them are resonances that characterise the piece and cannot be identified as distinct sounds. This is the case of palatality, velarity, labiality, and glottality. As mentioned in Chapter 1 (Section 1.3.1), Niebuhr and Kohler (2011) use the term articulatory prosodies to refer to “articulatory residues in the reduction of function words […] which persist as non-linear suprasegmental features of syllables, reflecting, e.g. nasality or labiality that is no longer tied to specific segmental units” (Niebuhr and Kohler, 2011: 320). They propose the concept of essential phonetic elements that constitute the "phonetic essence" of words. The analysis of the English pronouns reported in this chapter fits well with the idea of articulatory prosodies that constitute the identity of function words. Features such as glottality in *it*, palatality in *you*, and velarity in *we* are long-domain resonances that characterise the whole piece, rather than time-delimited local features. Importantly, these phonetic features are retained in reduced speech.

Besides illustrating the range of variation in the realisation of pronouns, the analysis presented in this chapter provided some useful insight into the features of reduction. The most salient aspects of reduction that were observed are the reduction in the magnitude of gesture, temporal reduction, and the temporal reorganisation of phonetic events.

The reduction in the magnitude of gestures is visible in pronouns that have a vocoid that is canonically described as changing quality in time. Pronouns such as *I* and *they* are commonly produced without apparent tongue movement – the quality of the first part of the vocoid is maintained as a steady state or changes slightly in the expected direction. For example, Section 3.1.1 reported that the pronoun *I* is in most instances realised as a vocoid with an open front quality. Section 3.1.7 showed that the vocoid in *they* is articulated as a close-mid front unrounded vocoid without change in quality.

The instances of reduction are also characterised by the temporal reorganisation of phonetic events. The state of the glottis and the states and movements of the organs in the supralaryngeal cavity are independent from each other. Therefore, the phonetic events occurring in the larynx and the events occurring in the supralaryngeal vocal tract can be realigned. This means that, for example, in the production of a vocoid, the phasing of voicing and the oral articulation can be simultaneous or occur at different times. In this case, the resulting vocoid can be very short or even voiceless or apparently deleted.

However, it was observed that, in most instances, both phonetic events are still articulated. The presence of weak friction at the beginning of several pieces led to the observation of visible formants in the spectrograms. This feature suggests that the gestures in the oral cavity are articulated before the onset of voicing. That is, when voicing and oral articulation do not co-occur, the articulation in the oral cavity occurs first and the vibration of the vocal folds starts later. While the articulation of the vocoid might not be audible without voicing, the presence of the formants indicates that the gesture in the oral cavity is articulated. The temporal realignment of the vibration of the vocal folds and the articulation in the oral cavity is the main factor in the production of reduced instances of pr + aux.

A question that remains to be answered is where in the vocal tract the weak friction in onset is produced. Two main hypotheses can be formulated. The first hypothesis is that the voiceless friction is produced in the glottis, and then shaped by the positions and movements of the supralaryngeal vocal tract. The second hypothesis is that the friction is produced at the place of articulation in the oral cavity. For example, the friction can be produced near the hard palate at the beginning of the pronoun *you* and near the soft palate at the beginning of the pronoun *we*. This issue is investigated in detail in Chapter 6 (Section 6.1).

To conclude this chapter, the auditory analysis and spectral observation of the data collected showed a wide range of variation in the production of the pronouns. This chapter described in detail the most frequent realisations of each pronoun, as well as the most extreme cases of reduction. This analysis has highlighted the phonetic features of each pronoun that remain in the signal in reduced speech. The next chapter reports the analysis of the cliticised forms of the auxiliaries.

4. Auditory and acoustic analysis of auxiliaries in pr + aux

This chapter reports on an auditory and acoustic analysis of the cliticised forms of the present tense of the auxiliary verb BE (*'m*, *'s*, *'re*), the cliticised forms of the present and past tense of the auxiliary verb HAVE (*'s*, *'ve*, *'d*), the cliticised forms of the modal auxiliaries *will* (*'ll*) and *would* (*'d*), and the finite forms of the past tense of the verb BE (*was* and *were*).

4.1. Cliticised forms of the auxiliaries

As explained in Section 1.2.2, using a polysystemic approach, the various forms of English auxiliary verbs can be divided into four systems (Ogden, 1999). The focus of this thesis is the system of auxiliaries that includes the (weak) non-syllabic forms /C/, where C represents a contoid, such as /d/ (as in *we'd*), /z/ (as in *she's*), /s/ (as in *it's*), /v/ (as in *they've*), /m/ (as in *I'm*), /l/ (as in *you'll*). These weak forms of the auxiliaries can be attached only to a pronoun host. In addition, the finite forms of the past tense of the auxiliary BE – *was* and *were* – were analysed too. The rationale for including *was* and *were* in the analysis is to investigate the contrast between the present and the past tense of auxiliaries. Table 9 (which was also reported at the beginning of the previous chapter) summarises the combinations of pronoun and auxiliary that have been analysed.

Pronouns	Aux HAVE		Aux BE		Modal Aux	
	present	past	present	past	(present)	(past)
I	I've	I'd	I'm	I was	I'll	I'd
he	he's	he'd	he's	he was	he'll	he'd
she	she's	she'd	she's	she was	she'll	she'd
it	it's	it'd	it's	it was	it'll	it'd
we	we've	we'd	we're	we were	we'll	we'd
you	you've	you'd	you're	you were	you'll	you'd
they	they've	they'd	they're	they were	they'll	they'd

Table 9. The combinations of pronouns (rows) and auxiliaries (columns) analysed in this thesis.

All the auxiliaries are described in combination with a pronoun. The combinations of the auxiliaries with the pronoun *it* will be described separately. As mentioned in the previous chapter, the pronoun *it* differs from the other pronouns in that it has a closed syllable with a contoid in coda position. All the other pronouns have an open syllable with an empty coda. When combined with an auxiliary, the empty coda of the pronoun is filled by the cliticised form of the auxiliary. However, some of the consonant clusters formed in the pr + aux combinations with *it* are not allowed by the phonotactics of English. For example, the consonant sequence /td/ in *it'd* is not allowed in coda position in English, so an extra vowel has to be added and the piece becomes bisyllabic. The strategies used to overcome the phonotactic constraints of coda consonant clusters are described for each cliticised form of the auxiliaries in combination with *it*. Following the approach used in Chapter 3, where possible, unreduced realisations are described first.

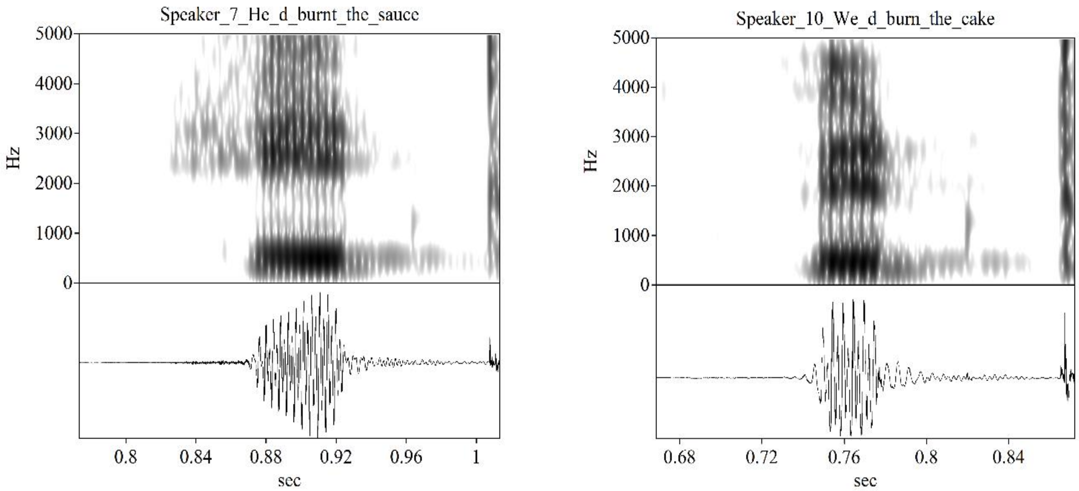
4.1.1. Auxiliaries *had* and *would*, cliticised form *'d*

The cliticised form *'d* represents either *had* or *would*. Initially, the two auxiliaries *had* and *would* were analysed separately. However, the acoustic analysis did not reveal any notable difference between the realisations of the cliticised forms of the two auxiliaries. For this reason, *had* and *would* are treated together here.[7] The cliticised forms of the auxiliaries *had* and *would* are disambiguated retrospectively by the finite form of the main verb (*burnt* or *burn*).

The phonemic transcription of the cliticised form *'d* is /d/. In the traditional phonetic description, /d/ is described as being produced by a complete closure in the oral cavity followed by an abrupt release of the air-pressure built up behind the closure. However, in English connected speech when there are two plosives in a row, the release of the first one is typically inaudible (Roach, 2000) or assimilated to the place of articulation of the following obstruent (Gimson, 1988). In the data collected, all pr + aux are followed by the verb *burn* in the appropriate form, which means that /d/ in all pr + *'d* is followed by a bilabial plosive. The most common realisation of /d/ across paradigms, speakers and repetitions in pr + *'d* is unreleased (or masked released) (78%, N = 420/537). Figure 57 shows two examples of unreleased /d/, in *he'd burnt* on the left, and *we'd burn* on the right.

---

[7] The acoustic analysis and comparison of *he'd* in the subsets *he'd burnt* and *he'd burn* and *we'd* in the subsets *we'd burnt* and *we'd burn* did not reveal any notable difference between the two groups.



[hɪd̚]   [ød̚]

Figure 57. Spectrogram and waveform of an instance of *he'd* produced by S7 on the left, and an instance of *we'd* produced by S10 on the right. In both instances the release of the first phonological plosive /d/ is masked by the closure of the second phonological plosive /b/.

In Figure 57 it can be noticed that the vocoid ends abruptly. This is because in the instances in which /d/ is unreleased, the complete closure of the plosive is still articulated, but the release is 'masked' by the bilabial closure for the bilabial plosive in onset of the following word. Perceptually, the closing gesture of the tongue moving towards the alveolar ridge is often audible. However, in some instances, the closure at the end of the vocoid sounds bilabial rather than alveolar. From the data available, it is not possible to know whether, in these cases, /d/ assimilates to the place of articulation of the following sound /b/, or whether the closure for /d/ is masked by the closure for /b/. A portion of voicing continues during the closure, as can be seen in the figure above.

In several instances of pr + *'d*, a weak spike in the middle of the closure can be observed at low frequencies. This spike can be seen also in the two figures above near the time mark 0.96 in the figure on the left, and between the time marks 0.8 and 0.84 in the figure on the right. This spike is likely to be the noise produced by the closure of the lips for the production of the bilabial plosive at the beginning of *burn/burnt*. It is visible only when /d/ is unreleased. In the whole dataset of pr + *’d* paradigms, it is present in 43% of instances of unreleased pr + *’d* across paradigms, speakers and repetitions (N = 180/420).

In several instances across pr + *’d*, /d/ is produced with friction at the place of articulation instead of a complete closure and a burst. To produce friction, the tongue moves towards the alveolar ridge but, instead of creating a complete closure, creates a stricture of close approximation. When the pulmonic air reaches the stricture, turbulent air is produced. Figure 58 shows an example of friction noise that starts before the end of the vocoid and increases in energy before the closure for the bilabial stop.

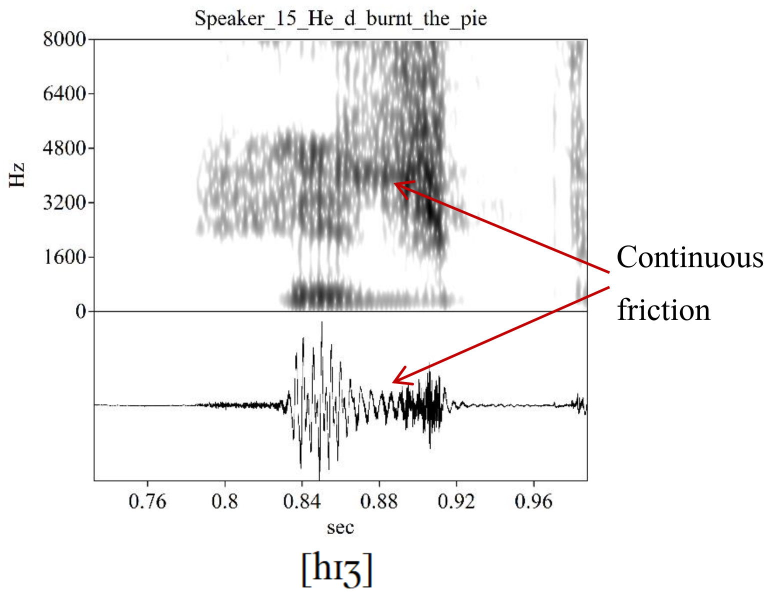


Figure 58. Spectrogram and waveform of alveolar friction in an instance of *he'd* produced by S15. The frequency scale displays frequencies up to 8 kHz.

Continuous friction noise instead of a complete closure for /d/ occurs in 8% of instances analysed (N = 42/537). It can be noticed in Figure 58 that the aperiodic energy is initially weaker, lower in amplitude and that the main concentration of energy is in a restricted range of frequencies between 3500 and 4700 Hz. Then the intensity of the aperiodic energy increases, as well as the range of frequencies at which it is found, from 1600 Hz to 10 kHz. The acoustic features of this friction suggest that the gesture of the tongue tip towards the alveolar ridge is in place, and that the tongue has reached the alveolar ridge and created a stricture. The stricture is initially narrower, then becomes wider and lets more turbulent air flow out through the constriction. This can be described as a case of reduced degree of gesture magnitude. The gesture is articulated, but its magnitude is reduced and instead of a complete closure, a narrow constriction is articulated. The constriction is then widened slowly rather than abruptly released. The auditory impression is that of a 'weak' affricate articulated with the blade of the tongue.

The realisation of /d/ with friction at the place of articulation raises the question of whether the instances of pr + *'d* in which /d/ is produced with friction differ from the pr + *'s* combinations. In traditional phonemic terms, the contoid /d/ and /z/ in coda position of /ʃɪd/ and /ʃɪz/ are responsible for the contrast between the minimal pairs *she'd* and *she's*. However, if the plosive /d/ is realised with continuous friction at the same place of articulation, how is the contrast with the phonological voiced alveolar fricative /z/ maintained? This issue is investigated in detail in Chapter 5 (Section 5.2.4.1).

In some cases, /d/ is articulated with a complete closure, after which turbulent air is gradually released. Figure 59 shows an example of *you'd* in which there is a gradual increase of aperiodic energy at the end of a complete closure.

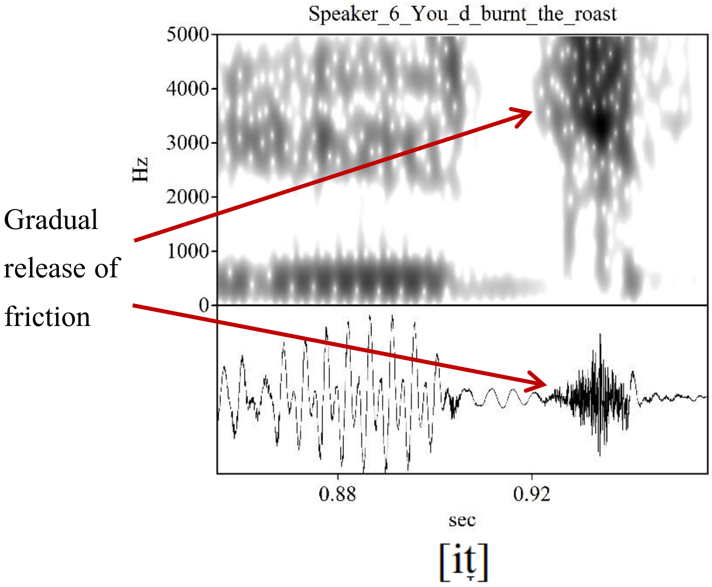


Figure 59. Spectrogram and waveform of an instance of gradual release of friction after a complete closure in *you'd* produced by S6.

Figure 59 shows an example of gradual release. It can be seen in the spectrogram and waveform that a complete closure is followed by a gradual increase of aperiodic energy instead of a burst as in the case of an abrupt release. From an articulatory point of view, this suggests that the tongue makes a complete closure at the place of articulation, but after the air-pressure builds up behind the closure, the tongue moves away gradually from the alveolar ridge creating the friction. A gradual release occurs in 5% of instances of pr + *'d* (N = 27/537) across speakers and repetitions. Also in this case the auditory impression is that of an affricate sound.

In 9% of the dataset, the plosive in pr + *'d* paradigms is realised with a closure, a burst, and a release phase. Compared to the most frequent variant in the dataset (unreleased plosive), the variant in which the air pressure is released with a burst cannot be considered an instance of reduction. On the contrary, it can be considered a case of hyper-articulation.

Figure 60 shows an instance of *you'd* in which the plosive is released with a burst followed by a short period of friction.

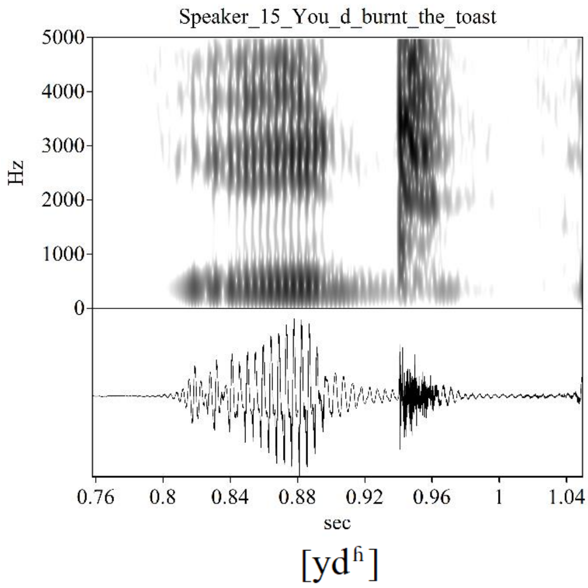


Figure 60. Spectrogram and waveform of an instance of *you'd* produced by S15 in which /d/ is realised with a complete closure and burst.

To summarise, the most common realisation of /d/ in pr + *'d* paradigms is unreleased, as we would expect in English connected speech when a plosive is followed by another plosive (in this case /b/ in onset of *burn*). Other realisations of /d/ include continuous friction at the place of articulation; a complete closure followed by a gradual increase of aperiodic energy; or abruptly released with a burst. Table 10 reports the number and percentage of tokens for each realisation in the data analysed. The pronoun *it* has been excluded from this table and will be treated separately in the next section.

Realisations of /d/ in pr + *'d* except *it'd*	**N**	**%**
Unreleased	420	78
Continuous fricated release	47	9
Gradual fricated release	43	8
Burst release	27	5
Total	**537**	

Table 10. Count of realisations of the alveolar plosive in pr + *'d*.

Another important aspect of pr + *'d* is the closure between the vocoid and the burst of /b/ in onset of *burn(t)*. The duration of the closure measured from the end of the vocoid to the /b/ burst in *burn(t)* includes both the closure for /d/ and for /b/. For this reason, the closure in pr + *'d* is longer than the closure in all the other pr + aux (as measured from the end of the contoid in the case of pr + *'ve* and pr + *'s*). Table 11 shows the mean duration of the closure and the mean duration of voicing after the end of the vocoid in five subsets with the pronoun *we*.

Piece	**Mean duration of closure in ms**	**Mean duration of voicing in closure in ms**	**N instances**
We'd	97	31	92
We've	67	(39)*	41
We're	71	33	43
We'll	70	29	49
We were	67	31	45

Table 11. Mean duration in ms of the closure before /b/ and voicing after the end of the vocoid in five subsets with the pronoun *we* and different auxiliaries. The asterisk (*) indicates that voicing in *we've* is not actually during the closure because the vocoid is followed by labio-dental friction.

Table 11 shows the difference between the mean duration calculated across speakers and repetitions of the closure in *we'd* compared to the mean duration of the closure in the other four *we*+aux. The reason for the longer duration of the closure in pr+ *'d* is the articulation of two adjacent closures for /d/ and /b/. The table also reports the duration of voicing after the end of the vocoid. In four of the five paradigms, voicing continues during the closure, while in *we've*, the vocoid is followed by friction, so voicing in this case is simultaneous to the labio-dental friction, rather than during the closure. While there is a substantial difference between the duration of the closure in *we'd* compared to the other paradigms, none of the paradigms exhibits a noticeable difference in the duration of voicing during the closure.

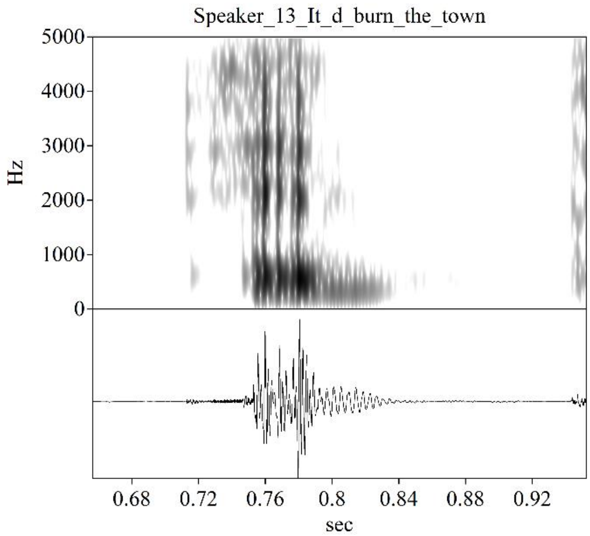
4.1.1.1. *It'd*

As mentioned in Section 3.1.4, the syllable structure of the pronoun *it* affects the realisation of the *it*+aux combinations. Not only does the same cliticised form of an auxiliary combined with the pronoun *it* exhibit different characteristics than when it is combined with all the other pronouns, but it also affects the realisation of the pronoun *it* itself.

In the data collected, the paradigm *it'd* is followed by the verb *burn* in the appropriate form (either *burn* or *burnt*). There are several variants of *it'd* and they can be grouped according to the realisation of the plosives. By far the most common realisation of the phonological /d/ in *it'd* is unreleased (98%, N=50/51). Depending on the realisation of /t/, the piece can have two forms. If /t/ is unreleased also, the piece is realised as a short vocoid followed by a long closure. The short vocoid is characterised by creaky voice. If /t/ is not unreleased, a short vocoid is produced after *it* between the two phonological plosives /t/ and /d/ and the piece takes the form of two vocoids with a variable contoid between them. The contoid between the two vocoids can be realised either with friction, a canonical burst, as a tap, or a dip in amplitude. In summary, the realisations of /ɪtəd/ observed in the dataset collected are:

- The two phonological plosives /t/ and /d/ are both unreleased. In this case, the vocoid is realised with creaky voice.
- The second plosive /d/ is unreleased, but the first plosive /t/ is not unreleased. The first plosive /t/ can be realised:
    - with a closure and a burst;
    - with friction;
    - as a short ballistic movement (tap);
    - as a dip in amplitude of the vocoid.

The most common realisation of /ɪtəd/ is as a short vocoid with creaky voice followed by a long hold phase. Figure 61 shows an instance of *it'd* produced by S13.



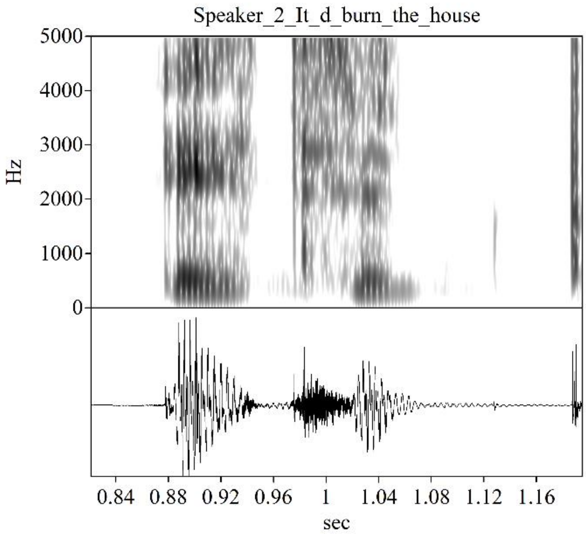
[ʔɪ̰̆]

Figure 61. Spectrogram and waveform of an instance of *it'd* produced by S13 in which the short vocoid is characterised by creaky voice and the two plosives are unreleased.

Figure 61 shows an instance of *it'd* realised as a short vocoid with creaky voice.

When /t/ is not unreleased, a vocoid or a period of voicing is produced after /t/ and the piece is thus realised with two vocoids. The following figures show the variants of the phonological plosive /t/: a complete closure followed by a burst (Figure 62), friction at the place of articulation (Figure 63), a short ballistic movement of the tip of the tongue such as a tap (Figure 64), or a dip in amplitude during the vocoid (Figure 65).

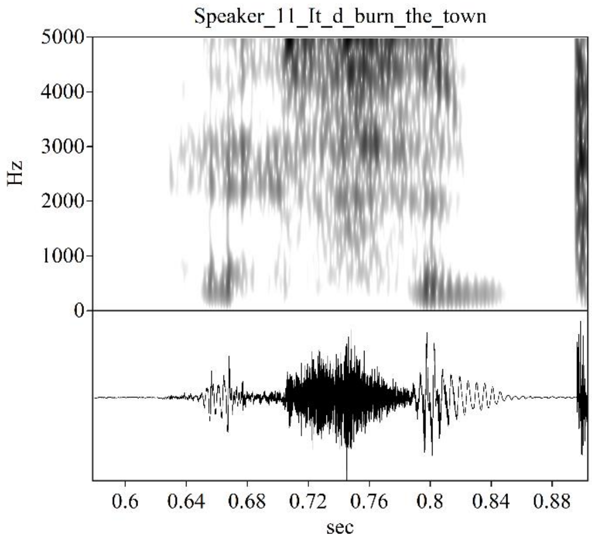
Figure 62 shows an instance in which /t/ is realised with a complete closure and an abrupt release of energy. The plosive is then followed by a short vocoid before the gesture for the closure of the second plosive /d/ is articulated.



[ʔɪtʰə]

Figure 62. Spectrogram and waveform of an instance of *it'd* produced by S2.

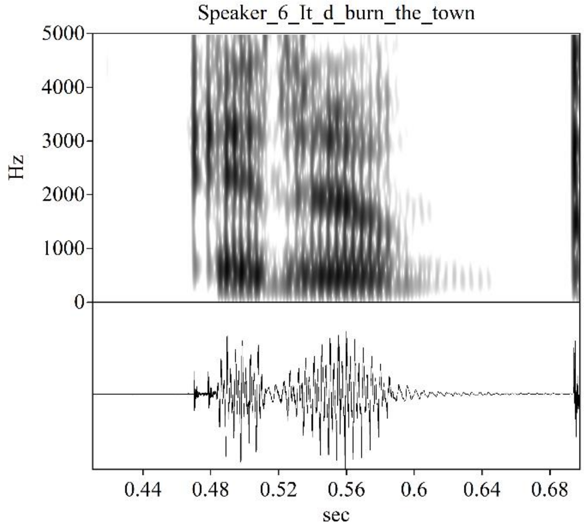
In Figure 62, /t/ is released with a burst after a short hold phase. However, in most instances in which it is not unreleased, /t/ is realised with continuous friction at the place of articulation, as shown in Figure 63. Here again the phonological plosive /t/ is followed by a short vocoid and a period of voicing continuing after the end of the vocoid.



[ɪ̰t̝ə]

Figure 63. Spectrogram and waveform of an instance of *it'd* produced by S11 in which /t/ is realised with continuous friction.

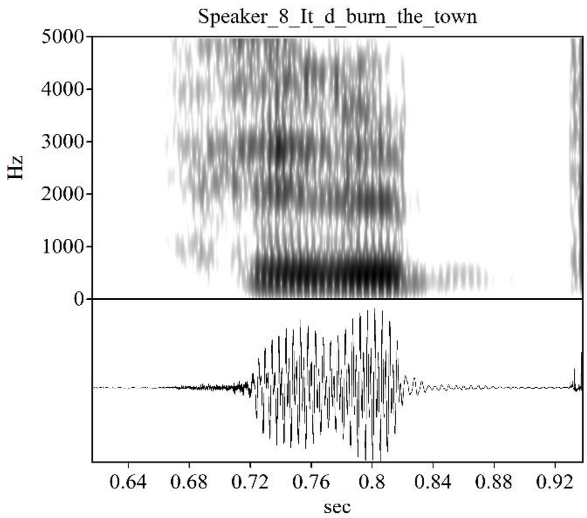
The phonological plosive /t/ can also be articulated with a short ballistic movement of the tip of the tongue against the alveolar ridge. This gesture produces a short closure in the vocal tract that is not long enough to allow a build-up of air pressure behind the closure. Figure 64 shows an instance of this articulation. In this case, the second vocoid is quite long and exhibits a falling F2. The closure at the end of the vocoid is clearly bilabial indicating that in this case the second plosive is not articulated.



[ʔɪɾɵ]

Figure 64. Spectrogram and waveform of an instance of *it'd* produced by S6 in which /t/ is realised as a tap.

In some instances, *it'd* is realised as a vocoid with only a dip in overall amplitude as observed in the waveform. In the spectrogram, the vocoid exhibits a weakening of F3. Figure 65 shows an instance of *it'd* in which the duration of the vocoid is quite long (94 ms). The variation in amplitude gives the auditory impression of two syllables.



[ɪ̥ɪˀə]

Figure 65. Spectrogram and waveform of an instance of *it'd* produced by S8 displaying only a brief decrease in amplitude during the vocoid.

From the area in which F3 exhibits lower amplitude in Figure 65, it can be inferred that the tongue has moved slightly towards the passive articulator, lessening the output of energy without creating a considerable constriction.

So far, all the instances of *it'd* described are characterised by an unreleased voiced alveolar plosive /d/. In one instance (N = 1/51, 2%) /d/ is realised as friction at the alveolar ridge as shown in Figure 66.

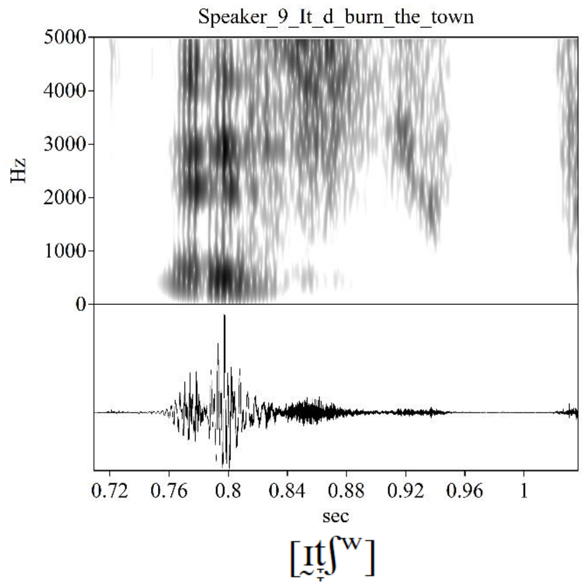
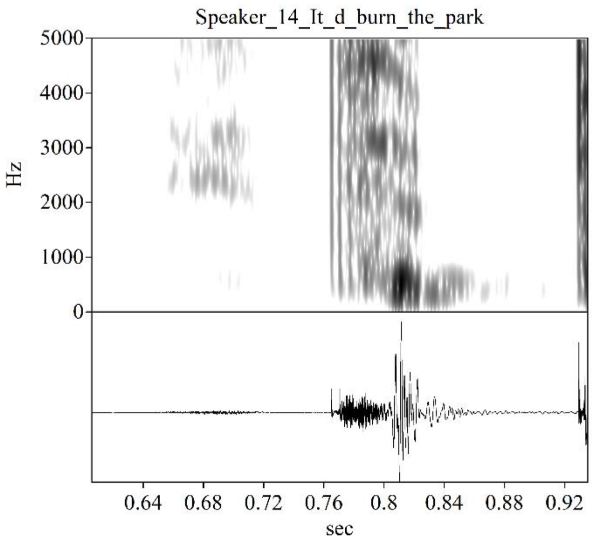


Figure 66. Spectrogram and waveform of an instance of *it'd* produced by S9 in which both phonological plosives are realised with friction.

Figure 66 shows an instance of *it'd* in which both phonological plosives /t/ and /d/ are realised with friction at the place of articulation. The first portion of friction is characterised by higher overall amplitude and aperiodic energy at higher frequencies. The second portion of friction is characterised by lower overall amplitude and aperiodic energy at lower frequencies. Auditorily, the second portion of friction sounds rounded or dark. This suggests that the labial gesture for the articulation of the following /b/ is already being articulated. It can be noticed in the spectrogram that /b/ too is realised with a gradual increase of friction rather than a sharp burst.

The first vocoid in *it'd* can also be reduced. In the dataset, there is only one instance in which the vocoid is voiceless. Figure 67 shows this instance.



[ɪ̥tʰə]

Figure 67. Spectrogram and waveform of an instance of *it'd* with a voiceless vocoid produced by S14.

Figure 67 shows an instance of *it'd* in which the vocoid is voiceless. However, the first plosive of *it'd* is realised with a complete closure and a burst, and it is followed by a short vocoid before the complete closure for the second plosive starts.

In the dataset, there is also an instance of *it'd* without the first vocoid. The piece starts with an abrupt release of air pressure at the alveolar ridge. The long portion of friction (duration 102 ms) exhibits a clear formant structure, which suggests that the quality of the friction is palatal.

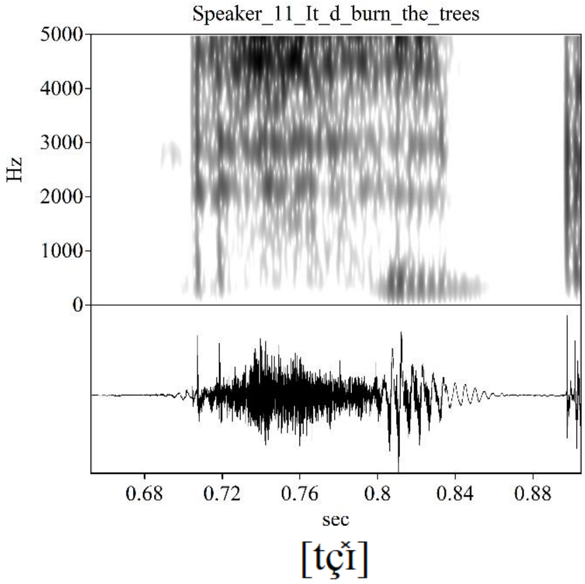


Figure 68. Spectrogram and waveform of an instance of *it'd* produced by S11 characterised by an initial burst followed by a long period of friction.

To summarise, the most frequent realisation of *it'd* is as a short vocoid with creaky voice followed by a long closure. When the phonological /t/ is released, it is followed by a mid central vocoid. The phonological plosive /t/ exhibits a high degree of variability, including friction, short ballistic movement, or complete closure and burst release.

4.1.1.2. **Summary of observations of the cliticised form *'d***

To summarise, the combinations of pr + *'d* can be divided into two groups according to the syllable structure of the pronoun. In combination with the pronouns with an open syllable, the most frequent realisation of /d/ is unreleased (78%). In reduced instances it can be realised as friction at the place of articulation without a complete closure, or as a gradual release of friction after a complete closure at the alveolar ridge. Both variants can be accompanied by a range of degrees of voicing, but voicing is always present. In combination with the only pronoun with a closed syllable (*it*), /d/ is realised as unreleased in 98% of instances. Depending on the realisation of the adjacent phonological plosive /t/, a vocoid can be produced between the two phonological plosives. The piece can thus be realised as a short creaky vocoid, or as a long vocoid with an intervening gesture in the middle. This gesture can create a complete closure, a narrow constriction resulting in alveolar friction, or briefly reduce the amount of output energy of the vocoid. An important aspect of all pr + *'d* is the duration of the closure after the end of the vocoid, which is longer than in all the other pr + aux. Moreover, the beginning of the closure is characterised by the abrupt end of the vocoid, and by a portion of voicing which continues during the initial portion of the closure. A quantitative analysis of these features is reported in Chapter 5.

4.1.2. Auxiliaries *has* and *is*, cliticised form *'s*

The cliticised form *'s* is the present tense of the third person singular of either the auxiliary HAVE (*has*) or BE (*is*). The phonemic transcription of *'s* is /z/ in *she's* and *he's*, and /s/ in *it's*. The forms *has* and *is* can occur only in combination with the third person singular pronouns *he*, *she* and *it*. In the traditional phonetic description, the alveolar fricative is described as a sibilant fricative, in which the friction noise is produced when the airstream strikes the teeth rather than at the constriction articulated at the alveolar ridge (Ladefoged and Maddieson, 1996). The tip or blade of the tongue produces a narrow constriction near the alveolar ridge which channels the airstream towards the teeth where the air becomes turbulent producing the friction noise.

Similarly to the pr + *'d* combinations, the realisations of the pr + *'s* combinations largely depend on the syllable structure of the pronoun. Although all paradigms (*she's*, *he's* and *it's*) have friction in coda position, the phonetic form of the piece differs between the paradigms that have a pronoun with an open syllable and the paradigms that have a pronoun with a closed syllable. For this reason, the description of the pr + *'s* realisations is divided into two parts. The first section describes the paradigms *she's* and *he's*. The second section describes the paradigm *it's*.

4.1.2.1. *She's* and *he's*

In combination with the pronouns *she* and *he*, the phonemic transcription of *'s* is /z/. The main variations in the realisations of /z/ are related to the amount of friction produced at the alveolar ridge and the amount of voicing. Figure 69 shows an unreduced instance of *he's* produced by S10.

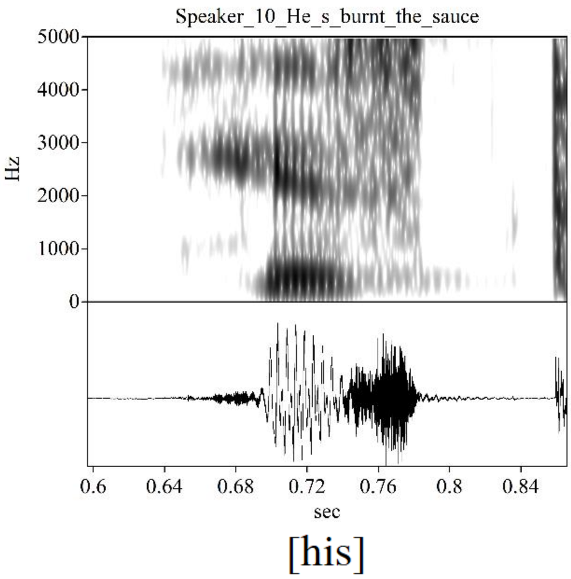
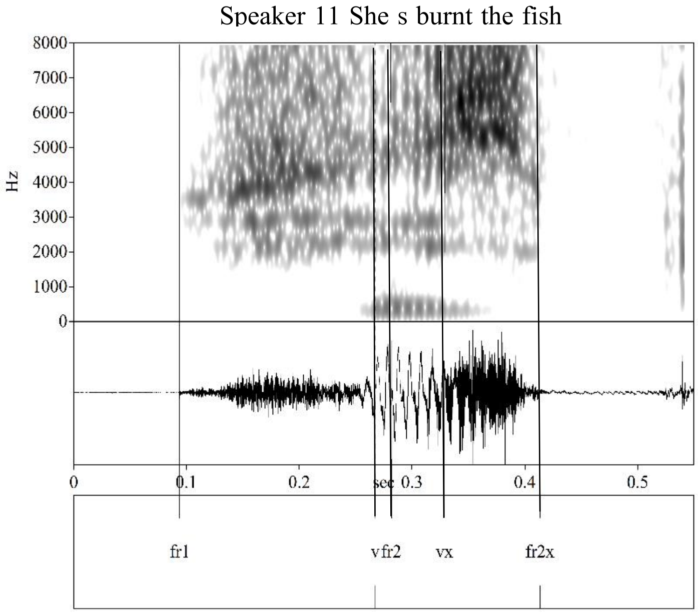


Figure 69. Spectrogram and waveform of an unreduced instance of *he's*.

In 70% of tokens of *she's* and *he's* across speakers and repetitions, the alveolar friction starts before the end of the vocoid and the two phonetic events overlap for a portion of time creating a 'fricated vocoid'. In the waveforms and spectrograms, fricated vocoids are characterised by weak and blurred F2 and F3 at lower frequencies (below 3000 Hz) with simultaneous friction noise at higher frequencies (above 3000 Hz). The mean duration of this overlap in the dataset *she's* and *he's* across speakers and repetitions is 11 ms (N = 196), while the mean duration of the non-overlapping vocoid in the same dataset is 33 ms and the mean duration of the non-overlapping alveolar friction is 46 ms. Figure 70 shows an example of overlapping friction and vocoid in *she's*.



[ʃɪz]

Figure 70. Spectrogram, waveform and annotation of an instance of *she's* produced by S11 in which friction overlaps with the vocoid. The frequency scale displays frequencies up to 8 kHz.

In Figure 70 the formant structure is visible throughout the piece, although F3 is faint in the final period of friction. However, during the period of vibration of the vocal folds, friction noise is visible at higher frequencies (above 4500 Hz). In this instance, the friction overlaps with the vocoid for its entire duration. The annotation shows the start (marked 'v') and end ('vx') of the vocoid, as well as the start ('fr2') and end ('fr2x') of the alveolar friction in coda. The palato-alveolar friction in onset ends with the start of the alveolar friction (marked 'fr2').

The paradigms *she's* and *he's* exhibit the same types of variation and reduction. In particular, the vocoid can be highly reduced because the initial friction that characterises the pronouns *she* and *he* is always produced (although in *he* the glottal friction can be very weak). The vocoid can be temporally reduced, simultaneous to friction (as above), voiceless or apparently absent. The most reduced realisations of *she's* and *he's* are characterised by one or two portions of friction produced at different places of articulation without voicing and without an apparent vocoid throughout the piece. Figure 71 shows an instance of *he's* in which glottal friction is followed by alveolar friction.

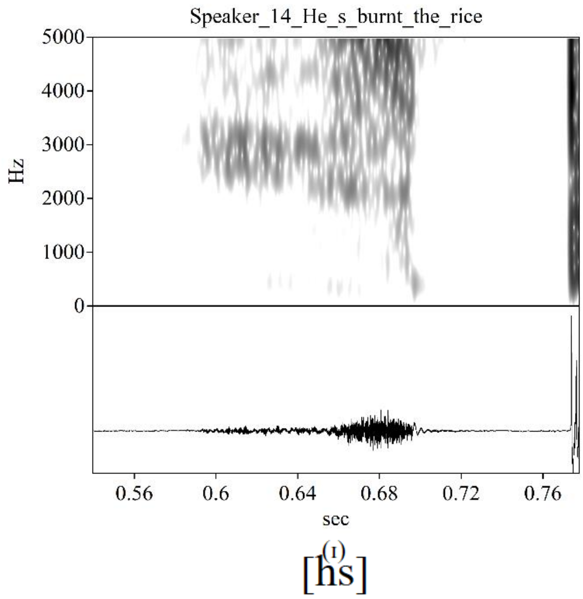


Figure 71. Spectrogram and waveform of a reduced instance of *he's* produced by S14 characterised only by friction at two places of articulation.

In the instance of *he's* in Figure 71, the two periods of friction differ both in the distribution of energy along the frequency scale and in the overall amplitude as shown in the waveform. The two portions of friction are articulated at different places of articulation: the first is articulated in the glottis, the second at the alveolar ridge. Throughout the piece, F2 is visible and it has the typical descending characteristic observed throughout the dataset in *he* and to a lesser extent *she*. A very short and weak portion of voicing is visible in the waveform between 0.64 and 0.68 seconds, but it is too weak to be audible.

She's can be further reduced to friction produced only at one place of articulation. Figure 72 shows one of three instances of *she's* in which friction is produced only in the postalveolar region.

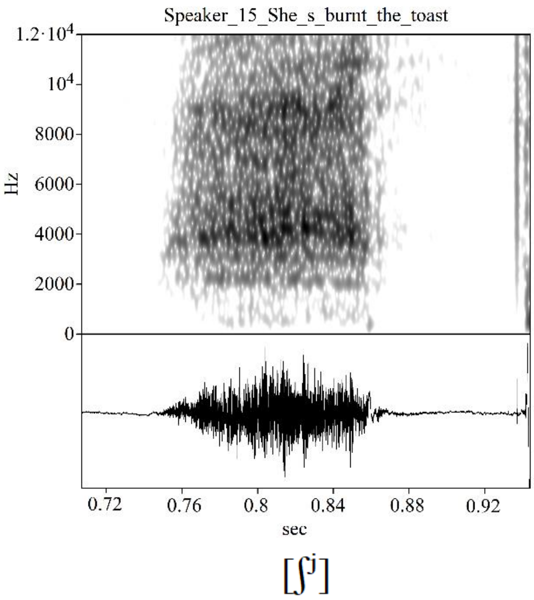


Figure 72. Spectrogram and waveform of a reduced instance of *she's* produced by S15 realised as friction only. The frequency scale displays frequencies up to 12 kHz.

Figure 72 shows an instance of *she's* that is highly reduced. In this case, the friction is produced only near the postalveolar region. Also in this case, F2 is visible throughout the piece.

To summarise, the reduction features that have been observed in *she's* and *he's* include a variable degree of overlap of phonetic events, absence of voicing and in extreme cases absence of alveolar friction in *she's* but not in *he's*. These reduction features can be combined to give highly reduced pieces characterised by voicelessness and continuous friction throughout the piece with or without obvious spectral variations. Although a vocoid is not always identifiable in the spectrograms, the articulatory gesture of a close front articulation seems to be in place as evident from the formant structure visible during friction.

4.1.2.2. *It's*

The case of *it's* is treated separately here because *it* is the only pronoun with a closed syllable, and also the only third person singular pronoun without friction in onset. As was reported in the section on *it'd*, the pronoun *it* is more prone to a wide range of variations.

The phonemic transcription of *it's* is /ɪts/. In its canonical form, *it's* is characterised by a near-close near-front vocoid followed by a complete closure produced by the front of the tongue against the alveolar ridge. The closure is then released with a burst which is followed by homorganic friction. Figure 73 shows an unreduced instance of *it's*.

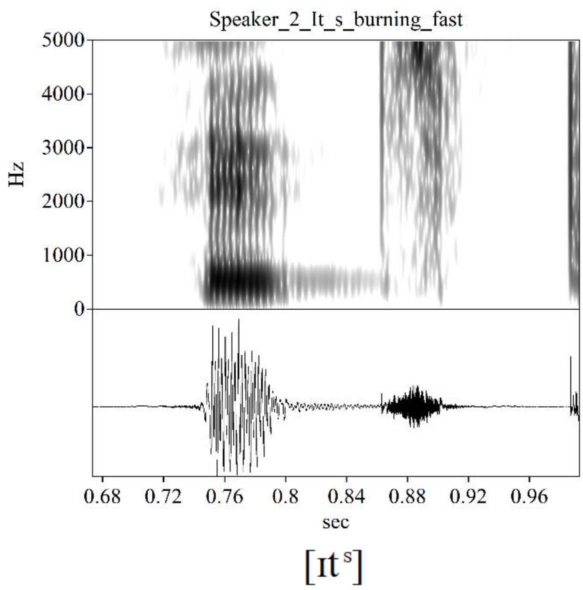


Figure 73. Spectrogram and waveform of an unreduced instance of *it's* produced by S2.

Figure 73 shows an instance of *it's* in which /t/ is realised with a complete closure in the oral cavity followed by an abrupt release and a period of alveolar friction. During the closure the vibration of the vocal folds continues but with diminishing intensity.

Most of the variations in the realisation of *it's* involve the phonological plosive /t/. As described in Section 3.1.4, creaky voice is one of the most frequent correlates of /t/. Figure 74 shows an example of *it's* produced by S3 in which /t/ is realised without a complete closure and the vocoid is characterised by creakiness.

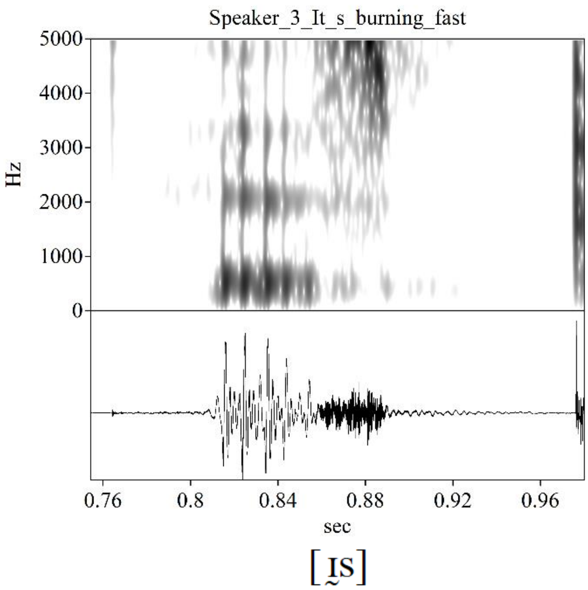
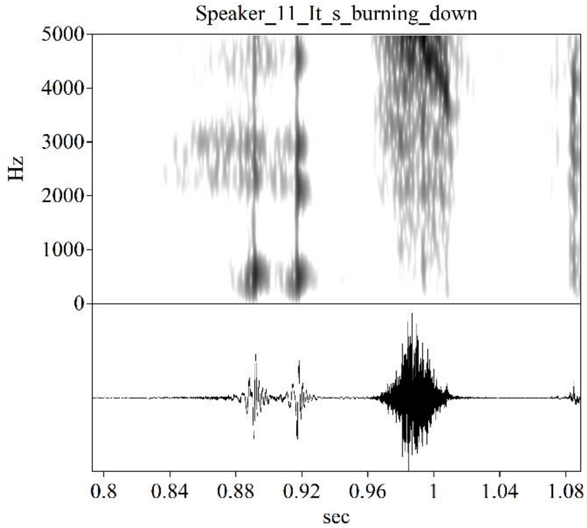


Figure 74. Spectrogram and waveform of an instance of *it's* produced by S3 in which the vocoid is realised with creaky voice and there is no visible closure for /t/.

Figure 74 shows an instance of *it's* in which the creaky vocoid is followed by alveolar friction without any intervening closure. In some instances, there is a period of silence between the vocoid and the alveolar friction, indicating that a complete closure is articulated in the oral cavity. Figure 75 shows an example of a reduced vocoid followed by a period of silence and alveolar friction.



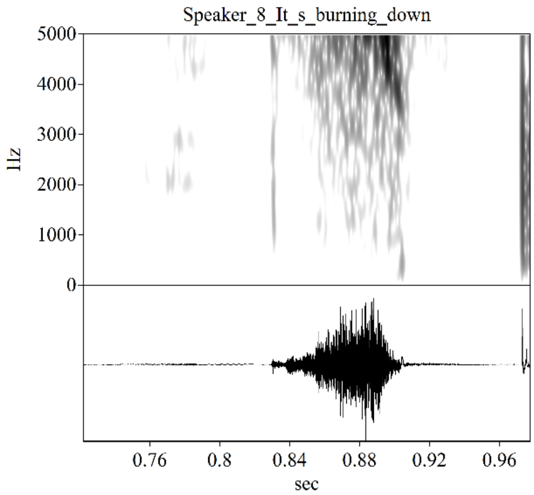
[ɪ̰ts]

Figure 75. Spectrogram and waveform of a reduced instance of *it's* produced by S11 in which the pronoun is realised as two cycles of vocal fold vibration.

The period of silence in the instance of *it's* in Figure 75 is followed by a gradual increase of aperiodic energy. The presence of a complete closure before the release of the air pressure gives the auditory impression of an affricate even though the release is not abrupt. The vocoid is realised as two irregular cycles of glottal folds vibration. Except for this glottal activity, the piece is voiceless. The vocoid can be further reduced to a barely audible portion of weak friction and even deletion. Figure 76 shows an instance of *it's* in which the vocoid is voiceless and barely audible.[8] The presence of the vocoid can be seen in the weak friction with visible formants before the burst of the release.

---

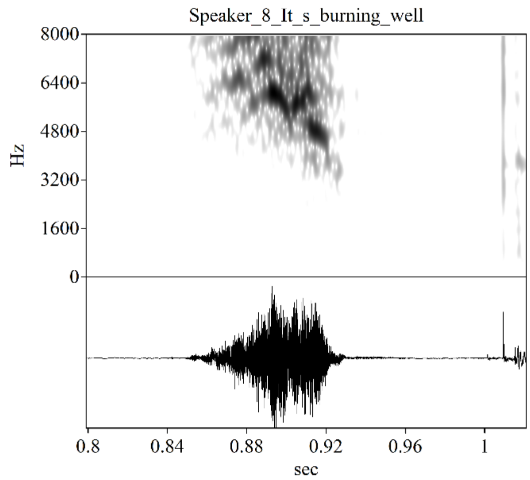
[8] In this instance of *it's*, the vocoid is audible only by listening to the audio file wearing headphones and in a quiet environment.



[(ɪ̥)ts]

Figure 76. Spectrogram and waveform of an instance of *it’s* produced by S8 in which the vocoid is articulated but it is voiceless and barely audible.

Finally, in the most reduced instances of *it’s*, only a period of voiceless alveolar friction is audible and observable in the spectrograms and waveforms. Figure 77 shows an instance in which only alveolar friction can be identified. The frequency scale has been adjusted to display frequencies up to 8 kHz.



[ts]

Figure 77. Spectrogram and waveform of a highly reduced instance of *it's* produced by S8 characterised by only a portion of voiceless alveolar friction. The frequency scale displays frequencies up to 8 kHz.

Despite the absence of a transient burst before the friction in the instance of *it's* in Figure 77, the auditory impression is that of a homorganic affricate [ts].

To summarise, the paradigm *it's* exhibits a wide range of realisations due mainly to the syllable structure of the pronoun and the presence of a phonological /t/. The reduced instances of *it's* are characterised by voiceless alveolar friction, which is always articulated and audible. The main source of variability is in the pronoun which can be realised as a short creaky vocoid or be apparently missing.

4.1.2.3. **Summary of observations of the cliticised form *'s***

As described in this section, the alveolar friction in coda position of the three paradigms analysed, *she's*, *he's*, and *it's*, is always articulated and audible in the pr + aux *it's* and *he's*, while it can be highly reduced and even absent in *she's*. This is not surprising if we consider the small paradigmatic system of contrast these paradigms belong to. To maintain their contrast in this small pr + aux system, *she's*, *he's* and *it's* need only to retain their fricative element(s). The reduced instances of *she's* retain the voiceless palato-alveolar friction; the reduced instance of *he's* retain both the voiceless glottal friction and the alveolar friction; the reduced instances of *it's* retain the alveolar friction alone. In all three paradigms, the piece can be produced without voicing. This means that the voicing contrast between *she's* and *he's* (described as having a voiced fricative /z/) on one side and *it's* (described as having a voiceless fricative /s/) on the other side is not maintained. Despite the absence of noticeable alveolar friction in a few instances of *she's* (N = 6/106, 6%), it can be said that the essential phonetic feature of the cliticised form *'s* is alveolar friction.

4.1.3. **Auxiliary *have*, cliticised form *'ve***

The cliticised form *’ve* is the present tense of the auxiliary verb HAVE in the first and second person singular and the plural. It can occur in combination with the pronouns *I*, *we*, *you* and *they*. The phonemic transcription of *’ve* is /v/. In the traditional phonetic description, the fricative /v/ is described as a non-sibilant voiced labio-dental fricative. To produce it, the lower lip is raised and retracted so that it forms a narrow constriction with the upper teeth. The turbulence is produced when the pulmonic airstream reaches the narrow constriction (Ladefoged and Maddieson, 1996). Simultaneously, the vocal folds vibrate to produce voicing. In the data analysed, the main variations in the realisation of the labiodental friction involve the amount of turbulent energy generated at the constriction and the temporal alignment of friction and voicing. Figure 78 shows an instance of *they’ve* with a typical amount of friction in instances of pr + *’ve* in the dataset.

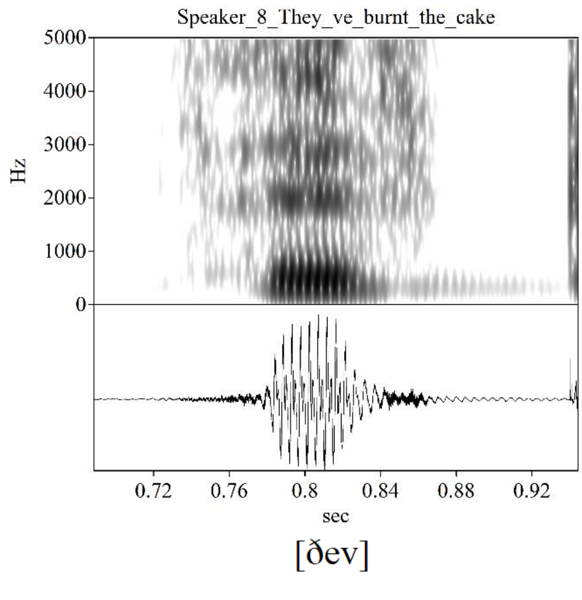
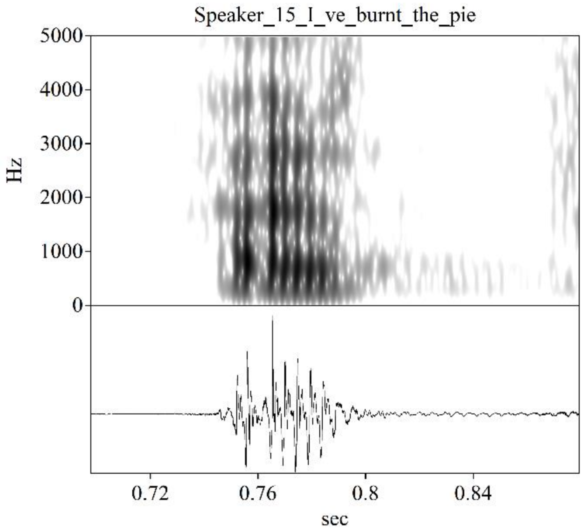


Figure 78. Spectrogram and waveform of an instance of *they’ve* produced by S8.

Figure 78 shows the typical friction produced in instances of pr + *’ve* across the dataset. The friction has low amplitude and energy distributed along the frequency scale rather than concentrated at specific frequencies. Voicing is usually present.

Most of the variations in the labio-dental friction of the auxiliary HAVE are in the duration and amplitude of the friction and the temporal alignment with the end of the vocoid. Figure 79 shows an instance of *I've* in which the friction is shorter and weaker than in the example shown above and starts before the end of the vocoid.



[ʔa̠$^{(v)}$]

Figure 79. Spectrogram and waveform of an instance of *I've* produced by S15 characterised by short and weak friction.

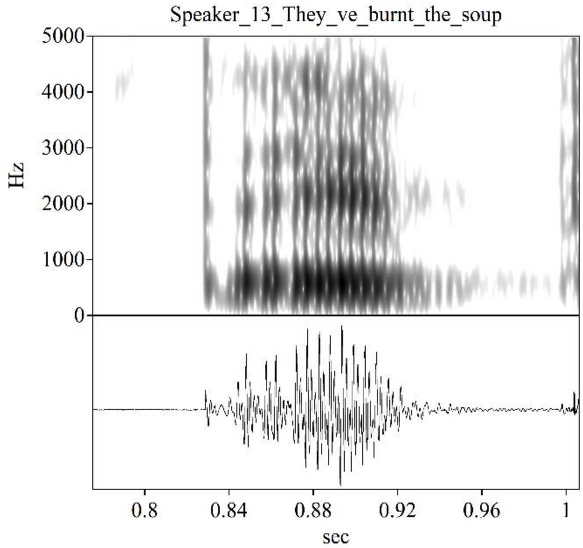
The weak realisation of the labio-dental friction might also be due to the accompanying vibration of the vocal folds. Because of voicing, the amount of unphonated air that reaches the front of the oral cavity is limited, which means that there is less air available to produce the turbulence. The labio-dental friction tends to be partially voiced. Table 12 shows the mean duration calculated across speakers and repetitions of the vocoid, the labiodental friction, the closure and the portion of voicing during friction in the four pr + *'ve* datasets. The last column on the right indicates what percentage of the labio-dental friction is voiced.

	Vocoid duration (ms)	Friction duration (ms)	Closure duration (ms)	Voicing during friction (ms)	Piece duration (ms)	Portion of voiced friction (%)
I've	39	30	73	22	140	73%
We've	45	32	65	26	143	80%
You've	43	31	65	25	135	81%
They've	46	31	72	26	149	84%

Table 12. Mean durations of the vocoid, the labiodental friction, the closure, the portion of voicing during the friction, and the mean duration of the entire piece in the four pr+ *'ve* datasets. The last column on the right indicates the percentage of labio-dental friction that is voiced.

Table 12 shows that the portion of friction that is voiced ranges between 73% and 84% of the duration of the friction.

In some instances (N=33/218, 15%) the labiodental friction is neither audible nor visible. Figure 80 shows an example of *they've* without a clear labiodental period of friction. In this instance, the vocoid is quite long (77 ms; the mean duration of the vocoid in *they've* calculated across speakers and repetitions is 46 ms).



[d̪e̞̰]

Figure 80. Spectrogram and waveform of an instance of *they've* produced by S13 without an apparent labiodental friction.

In the instance shown in Figure 80, the labiodental gesture is not audible. This feature raises the question of how the contrast with other pr + aux such as pr + *'d* is maintained. In order to investigate this issue further, a subset of data in which the labiodental friction could not be heard or identified in the spectrogram was selected. The results of the analysis of the *I've* subset is reported here. In 7 out of 53 (N = 7/53, 13%) instances of *I've* the labiodental friction is either not audible or it cannot be identified in the spectrogram. In some of these instances, the friction is not visible in the spectrogram but the labiodentality can be heard. Only in 3 instances the labiodentality cannot be heard at all (N = 3/53, 6%). The 7 instances were produced by three speakers only – S13, S14, S15. To identify the features that characterise *I've* in the absence of friction and/or labiodentality, the instances of *I've* produced by these three speakers (N = 15) were compared to the instances of *I'd* produced by the same three speakers (N = 15). The acoustic analysis of the two subsets revealed that the vocoids in *I've* and *I'd* differ in duration and F2. The mean duration of the vocoid in

I've in the S13-S15 subset is 37 ms (N = 15, the mean duration of the vocoid across the *I've* dataset is 39 ms, N = 53). The mean duration of the vocoid in *I'd* in the S13-S15 subset is 25 ms (N = 15, the mean duration of the vocoid across the dataset is 32 ms, N = 49). The formant dynamics of the two subsets are displayed in Figure 81.

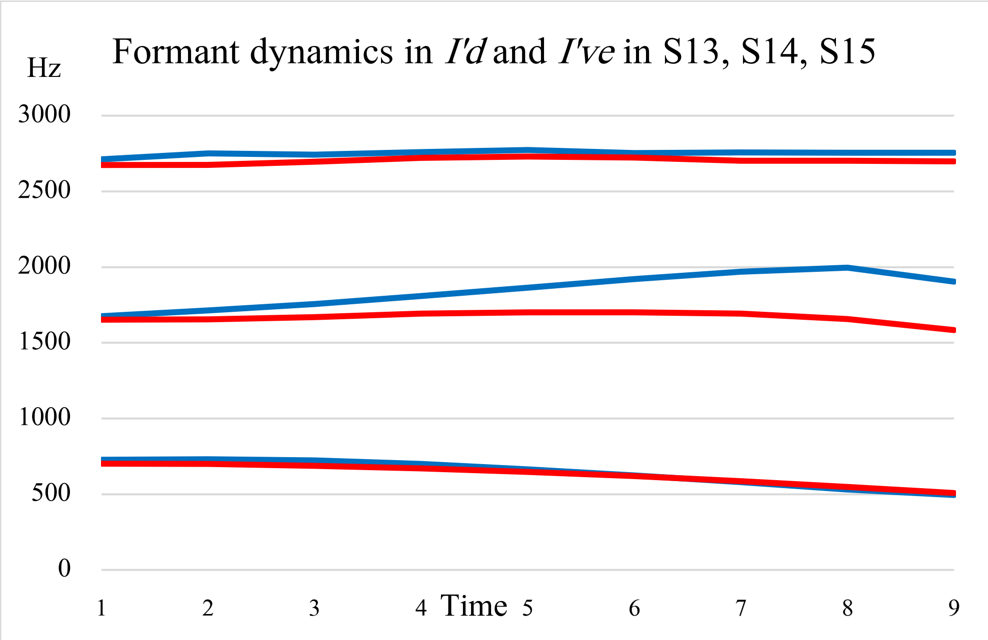


Figure 81. Formant dynamics of the vocoid in *I'd* (blue) and *I've* (red) in two subsets of the data produced by S13, S14, and S15, who produce very weak or absent labiodental friction in coda of *I've* (N = 15).

The same analysis and comparison were carried out on the *we've* and *we'd* datasets, with parallel results. The comparison between the acoustic features of pr + *'ve* and pr + *'d* revealed that the two paradigms differ along a set of parameters. Firstly, the duration of the vocoid is longer in instances of pr + *'ve* that do not exhibit an evident labiodental friction. Secondly, the frequency of F2 at the end of the vocoid is lower in pr + *'ve* than pr + *'d*, but not at the beginning of the vocoid. An additional feature that was observed in the data is the difference in the way the vocoid ends. The vocoid in pr + *'d* pieces ends abruptly with the articulation of the complete closure for /d/, while the vocoid in pr + *'ve* ends smoothly with energy diminishing at higher frequencies first, and then at lower frequencies. Figure 82 shows an instance of *they've* without audible labiodental friction on the left, and an instance of *they'd* produced by the same speaker (S13) on the right. Both tokens are displayed in the same time frame (230 ms).

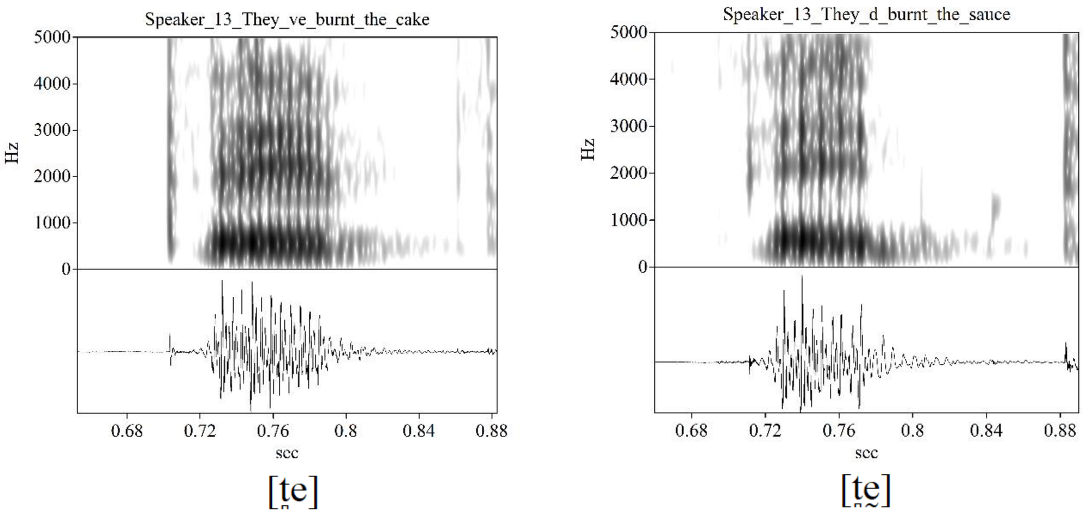


Figure 82. Spectrogram and waveform of an instance of *they've* without audible labiodental friction on the left, and an instance of *they'd* on the right. Both tokens are produced by S13 and are displayed in the same time frame of 230 ms.

Figure 82 shows an instance of *they've* on the left and an instance of *they'd* on the right uttered by the same speaker. The noticeable differences are the duration of the vocoid, which is longer in *they've* than *they'd*; the duration of the closure, which is longer in *they'd* than *they've*; and the abrupt offset of the vocoid in *they'd* compared to the smooth offset of the vocoid in *they've*.

There are also instances of pr + *'ve* that display a higher degree of turbulent energy compared to the ones shown so far. In these cases, the friction has a higher amplitude and sometimes a longer duration. Figure 83 shows an instance of *we've* in which the labiodental friction is characterised by high energy amplitude relative to the amplitude of the vocoid.

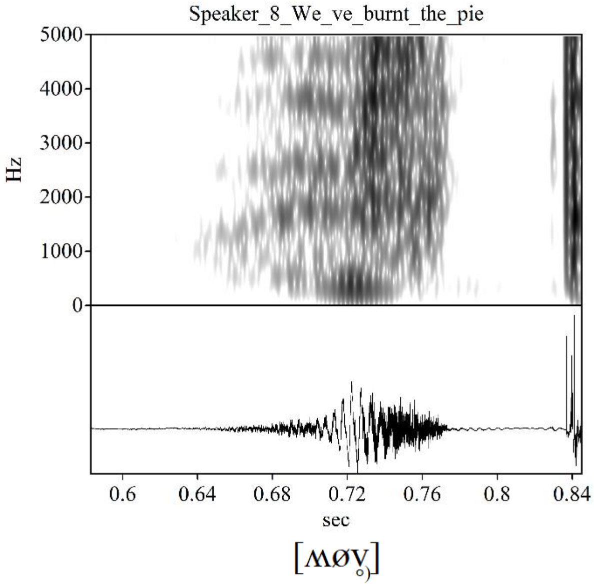


Figure 83. Spectrogram and waveform of an instance of *we've* produced by S8, characterised by a reduced vocoid and strong friction.

To summarise, the main variations in the realisations of the pr + *'ve* paradigms are related to the duration of the friction, to its amplitude, and to the temporal alignment with the end of the vocoid. In some instances the labiodental friction starts before the end of the vocoid. An unequivocal characteristic of the labiodental friction is to be accompanied by voicing. When the friction is apparently missing, the labiodentality can still be heard in some instances. In the few instances in which the labiodentality is not audible, the longer duration of the vocoid, the lower F2 frequency at the end of the vocoid compared to instances of pr + *'d*, and the smooth end of the vocoid are all exponents of the labiodentality. The main phonetic feature of the cliticised form '*ve* is the labiodentality. However, when the labiodentality is not audible, voicing seems to maintain the identity of the auxiliary '*ve*.

4.1.4. Auxiliary *am*, cliticised form *'m*

The cliticised form *'m* is the present tense of the auxiliary verb BE in the first person singular and it can occur only with the pronoun *I*. The phonemic transcription given by pronouncing dictionaries is /æm/ for the strong syllabic form and /əm/ and /m/ for the weak forms. In the traditional phonetic description, /m/ is produced with a complete closure at the lips and by opening the velo-pharyngeal port to let the pulmonic airstream flow out from the nasal cavities.

Nasal consonants are known to be prone to assimilation to the place of articulation of the following sound. Although bilabial nasals tend to be less prone to assimilation than alveolar nasals (Jones, 1960; Gimson, 1989), Ogden (1999) and Local (2003) claim that the bilabial nasal /m/ in the grammatical item *I'm* exhibits more phonetic variation than when it is found in coda position of lexical items such as *lime* or *time* (Local, 2003). For example, in a phrase such as *I'm going*, the bilabial nasal can assimilate to the place of articulation of the following velar plosive and be articulated at the velum. The assimilation of place of articulation does not occur in a sequence such as *lime green*, in which the bilabial nasal has to be articulated at the lips. This can be explained with reference to the polysystematicity of language: the coda position of function words and that of content words belong to two different systems which assimilate differently (Ogden, 1999).

The bilabial nasal in the dataset *I'm* analysed does not exhibit a wide range of variation. This is due to the following context, which is articulated at the same place of articulation. Most of the variations observed in *I'm* are in the realisation of the vocoid. Figure 84 shows an unreduced realisation *I'm* produced by S2.

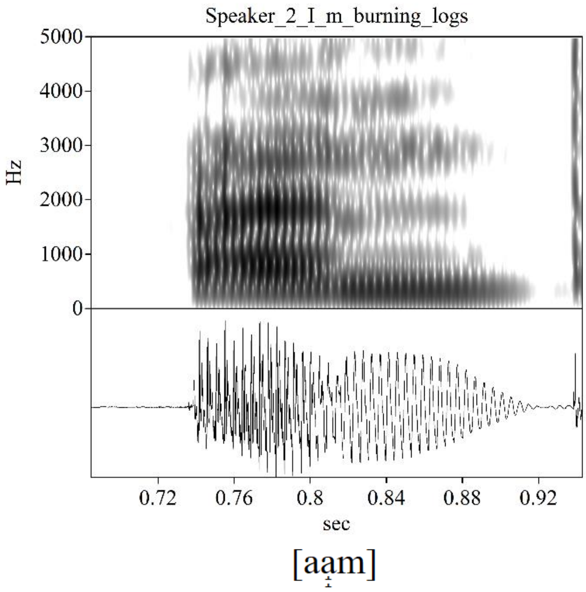


Figure 84. Spectrogram and waveform of an unreduced instance of *I'm* produced by S2.

There is not as much variation in the realisation of *I'm* as there is in other pr + aux pieces. This is surprising because it is the only pr + aux produced with an open vocoid and a nasal contoid, and thus it does not contrast with any other pr + aux combinations.

The main variation observed is in the temporal organisation of the various phonetic events. During the nasal contoid, there are three articulatory components: the bilabial articulation, the opening of the velo-pharyngeal port, and the vibration of the vocal folds. At the end of the nasal contoid, the bilabial closure can be maintained because *I'm*, like all pr + aux in this data, is followed by a bilabial plosive. Therefore, the closure at the lips is not released at the end of the nasality, but is maintained to the following sound. In 34 out of 55 tokens of *I'm* (61.8%), the hold phase of /b/ is completely voiced – the vocal folds vibrated throughout the hold phase up to the bilabial burst. In three out of 55 items (5%), also the nasality continues until the release of the bilabial plosive. The hold phase of the bilabial plosive is thus 'masked' by the nasal sound. Figure 85 shows an instance of masked hold phase as suggested by the formant structure which is visible throughout the hold phase, although the amplitude of the formants decreases in time. Even though the velo-pharyngeal port remains open, there is still build-up of air-pressure in the vocal tract that is then released with a burst.

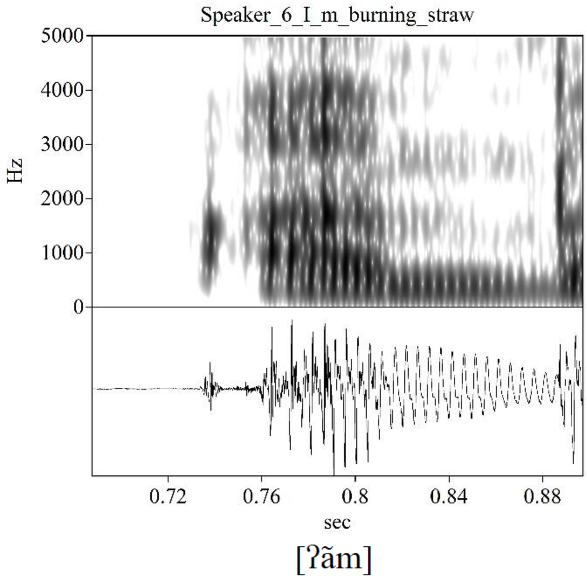
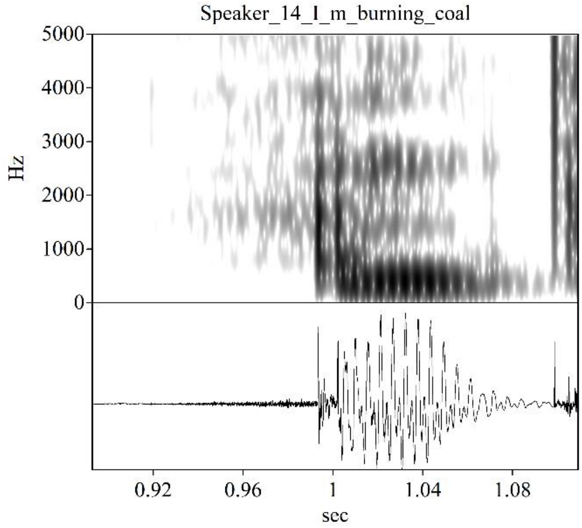


Figure 85. Spectrogram and waveform of an instance of *I'm* produced by S6, with visible formants up to (almost) the burst.

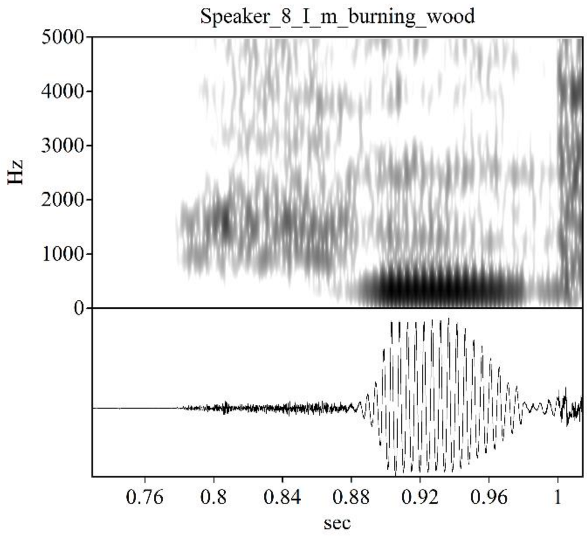
Despite being the only nasal sound in the small pr + aux system, in reduced instances of *I'm*, it is the oral vocoid that is reduced rather than the nasal contoid. On the one hand, the lack of reduction of the nasal sound can be explained by the fact that nasality is the main feature in *I'm*, as there are no other nasal sounds in the system of pr + aux. On the other hand, because it is the only nasal sound in the system, it could easily be reduced to a nasalised vocoid and still maintain the contrast with the other pr + aux. However, this is not the case, and the nasal contoid is always articulated. The vocoid is always articulated too, although it can be very short. In its most reduced form, its duration is 9.4 ms (see Figure 86). However, notice that there is a period of voiceless friction before the glottal stop and beginning of the vocoid.



[ʔăm]

Figure 86. Spectrogram and waveform of a reduced instance of *I'm* produced by S14.

The vocoid can also be entirely voiceless, as in Figure 87. The articulation of the vocoid is in place even though the vocal folds start vibrating at the beginning of the bilabial closure for the nasal contoid. The voiceless vocoid is audible.



[ḁm]

Figure 87. Spectrogram and waveform of an instance of *I'm* with a voiceless vocoid produced by S8.

Figure 87 shows the most reduced instance of the piece *I'm* observed in the data collected. The vocoid is voiceless, while the nasal contoid maintains its voicing.

To summarise the reduction features found in *I'm*, only the vocoid is affected by reduction, but not the nasal contoid. The vocoid can be very short and also voiceless, while the bilabial nasal is always articulated and it can be maintained up to the burst of the bilabial plosive in *burning*. In this data, the essential phonetic feature of the cliticised form *'m* is the nasal contoid.

4.1.5. Auxiliary *will,* cliticised form *'ll*

The cliticised form of *will* is *'ll*. The phonemic transcription of *'ll* is /əl/ or /ᵊl/. In all the pr + *'ll* combinations except *it'll*, /l/ occurs in syllable coda position. In Standard British English, the lateral approximant is realised differently depending on the syllable position it occurs in. When it occurs in syllable coda, the lateral is velarised, and it is called dark-L; while when it occurs in syllable onset, the lateral is not velarised and it is called clear-L (Jones, 1960; Abercrombie, 1967; Gimson, 1989). Both clear and dark alveolar lateral approximants in English have the same primary articulation, but they differ in their secondary articulation. The primary articulation is the articulation made with the tip or blade of the tongue against the alveolar ridge. One or both sides of the tongue are lowered so that they do not touch the molar teeth and the egressive pulmonic air flows out from the oral cavity from the sides of the tongue without producing friction. The secondary articulation is the articulation made by the body of the tongue. In velarised laterals, the back of the tongue is raised towards the velum. A feature of the secondary articulations is their resonances. Resonances can be defined as the "quasi-vocalic quality audible in consonantal production" (Carter, 2002: 22) resulting from a particular configuration of the oral cavity.

The resonances of a sound depend on the articulatory configuration of the vocal tract, such as the shape and size of the cavities, and the shape and position of the tongue and the lips, that is not directly related to the primary articulation (Kelly and Local, 1989). The resonances are the auditory correlates of the secondary articulation and are independent from the primary articulation in that they can occur on any consonantal articulation. Although any consonant can have resonances, several studies have shown that liquids such as laterals and rhotics have long-domain resonances that affect neighbouring sounds (Kelly and Local, 1989; Hawkins and Slater, 1994; West, 1999a, 1999b; Heid and Hawkins, 2000).

In the data analysed, the lateral is in coda position, therefore it is expected to be characterised by dark resonances. Although the tip tongue gesture is considered the primary articulation, and the velarisation is the secondary articulation, in the data analysed, the velarisation, and the presence of dark resonances throughout the piece, are the most perceptually prominent feature of pr + *'ll*. In this section, the descriptions of the pr + *'ll* paradigms have been divided into three groups according to the syllable structure of the pronoun. The three groups are: pronouns that have only a vocoid (*I*, *you*, *we*), pronouns that have a contoid followed by a vocoid (*she*, *he*, *they*), and the only pronoun that has a contoid in coda position and an empty onset (*it*).

4.1.5.1. *I'll*, *you'll*, *we'll*

The realisations of the pr + aux combinations *I'll*, *you'll* and *we'll* largely depend on the articulation of the lateral in coda position. As mentioned above, the lateral in coda position in English is characterised by a secondary articulation in which the body of the tongue is raised towards the velum. The resonances associated with the secondary articulation can affect neighbouring sounds, especially vocoids in unstressed syllables (Kelly and Local, 1989, Hawkins and Slater, 1994). The acoustic correlates of the dark resonances are lowered F2 and F3 (Heid and Hawkins, 2000).

In the subset *I'll*, *you'll*, and *we'll*, the lateral can be identified and measured as a discrete element only in nine tokens (N = 9/147, 6%). Figure 88 shows an instance of *I'll* in which the onset of the lateral can be identified.

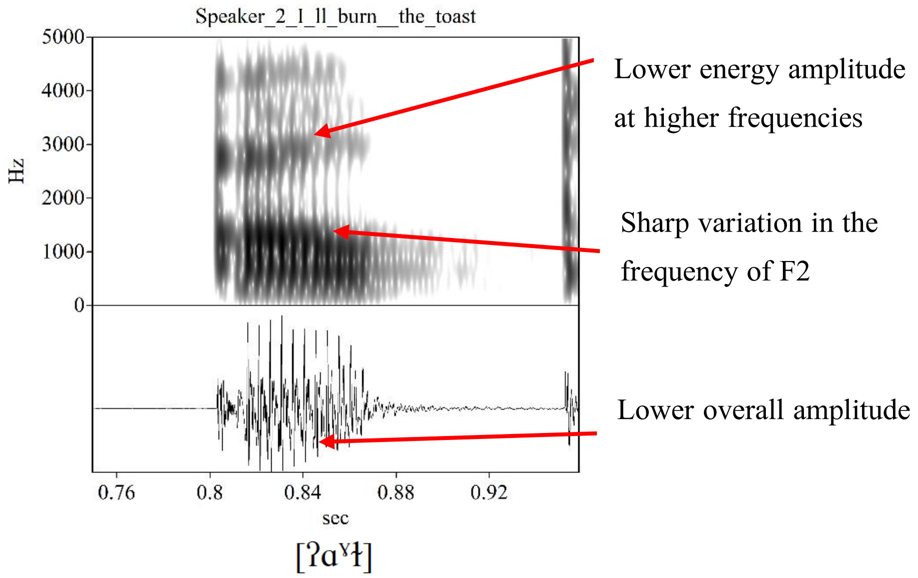


Figure 88. Spectrogram and waveform of an instance of *I'll* uttered by S2 in which a distinct /l/ sound can be identified in the spectrogram and waveform.

Figure 88 shows an instance of *I'll* in which the lateral is clearly identifiable. In the absence of articulatory data, it can be assumed that the abrupt change in the spectral quality of the vocalic sound corresponds to the beginning of the lingual contact of the lateral articulation. This is visible only in two tokens of *I'll* produced by S2 whose production, as mentioned in Chapter 2, has been used as reference point of unreduced realisations of the pr + aux paradigms for her slower and more careful articulation compared to all the other speakers. Although in all the other tokens it is not possible to identify through spectrographic inspection the beginning of the lingual articulation, it does not mean that the contact is not in place. In several instances the impression of a lateral sound is audible but not visible.[9] In these cases it is difficult to determine whether the lingual contact is reached or whether these are cases of phoneme restoration (Kemps, Wurm, Ernestus, Schreuder, and Baayen, 2005). Figure 89 shows an instance of *I'll* in which the perceptual impression of a lateral is present but a lateral sound is not delimitable from spectrographic observation.

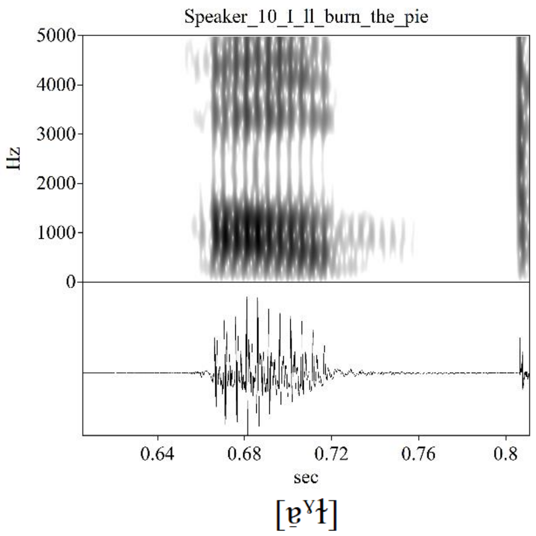
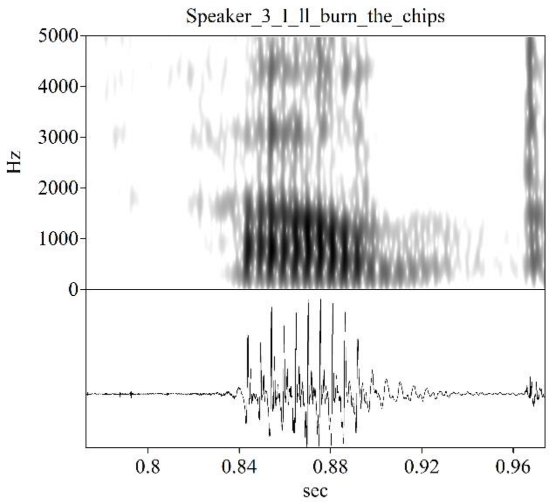


Figure 89. Spectrogram and waveform of an instance of *I'll* uttered by S10 in which a lateral sound can be heard but not delimited in the spectrogram.

In several instances of pr + *'ll*, the lateral sound not only is not identifiable as a distinct sound in the spectrogram and waveform but the laterality is also not audible from a perceptual point of view. Although there is a range of realisations that are not easy to categorise, it seems that most realisations without a lingual contact can be placed on a continuum between two types of realisation. In the first type /l/ is realised as a vocoid. This is a known phenomenon in some varieties of English and it is referred to as L-vocalisation

[9] By 'not visible' it is meant that although some spectral changes can be observed, it is not possible to identify the onset of the lateral at a precise point in time. A distinct sound cannot be identified and, therefore, its duration cannot be measured.

(Wells, 1982). To produce a vocalised L, the gesture of the primary articulation – the lingual contact with the alveolar ridge – is not fully articulated. The secondary articulation – the raising of the body of the tongue towards the velum – takes centre stage. The resulting sound is a vocoid with a back rounded quality. Figure 90 shows an instance of *I'll* in which the lateral is produced as a back rounded vocoid.



[ɐʏo̟]

Figure 90. Spectrogram and waveform of an instance of *I'll* uttered by S3 in which the lateral is realised as a back rounded vocoid.

Figure 90 shows an example in which the lateral is realised as a vocoid. It can be seen that both F1 and F2 slope to lower frequencies, F2 bandwidth becomes shallower, and F3 becomes fainter. Rather than an abrupt change (indicating that a lingual contact has occurred) the change in the formants is gradual. In this instance, the frequency of F1 is 769 Hz at the beginning of the vocoid, and 536 Hz at the end of the vocoid (range 233 Hz). The frequency of F2 is 1477 Hz at the beginning of the vocoid, and 1081 Hz at the end of the vocoid (range 396 Hz). The variation in the formant frequencies suggests that the tongue moves from a front and open position to a close and further back position. Moreover, the perceptual impression is that the lips move from a neutral to a rounded position. L-vocalisation is not a feature of reduction but occurs also in careful speech in some accents of English, including SSBE (Wells, 1982; Johnson and Britain, 2007).

The second type of realisation of pr + *'ll* occurs when the lateral sound is realised only as dark resonances audible throughout the piece. In this case, the lateral is not realised as a vocoid. The secondary articulation, or velarisation, is not subsequent to the preceding sounds, but it is articulated simultaneously. The distinction between these two types of lateral realisation is subtle and sometimes difficult to determine. In the case of the vocalised lateral, there is a vocalic period characterised by a back rounded quality at the end of the piece. In the resonance-only articulation, the secondary articulation is simultaneous to the preceding vocoid. In both cases, the gesture of the back of the tongue that produces the dark resonances is present. The perceptual impression of the piece is that of a 'dark' quality from the beginning. Figure 91 shows an instance of *I'll* realised as a dark vocoid.

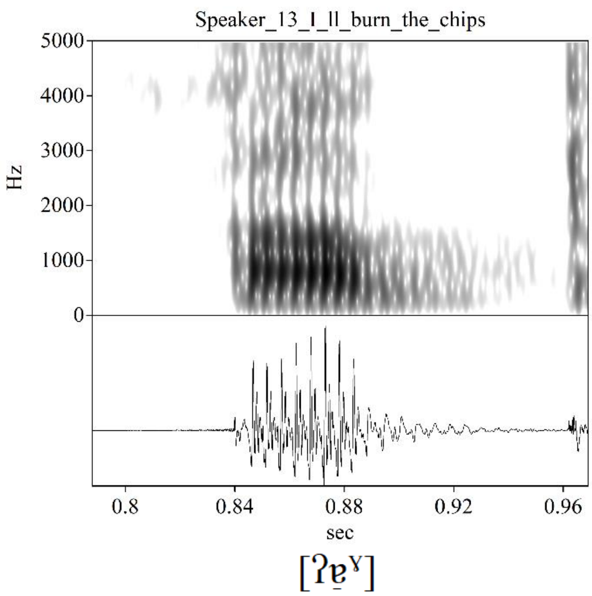


Figure 91. Spectrogram and waveform of an instance of *I'll* uttered by S13 in which the lateral is realised as dark resonances superimposed on the preceding sounds.

Figure 91 shows an instance of *I'll* in which the piece is characterised by a dark quality and no considerable formant movement. There is little change in the formant structure of the vocoid compared to the change found in the instance in Figure 90. F2 slopes slightly and its bandwidth becomes shallower. The frequency of F1 is 728 Hz at the beginning of the vocoid, and 666 Hz at the end of the vocoid (range 62 Hz). The frequency of F2 is 1348 Hz at the beginning of the vocoid, and 1224 Hz at the end of the vocoid (range 124 Hz).

More than other features of speech found in the data, the type of lateral realisation seems to be speaker specific. No single speaker produces all the main variants described. On the contrary, each speaker seems to use subtle variations of the same variant.

In reduced instances, the paradigm pr + *'ll* is realised as a short vocoid. Figure 92 shows an instance of *I'll* in which the vocoid is very short (duration 12 ms) and has a steady quality.

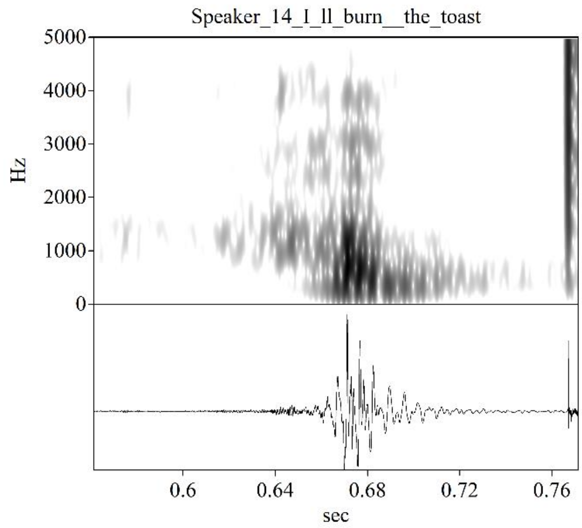


Figure 92. Spectrogram and waveform of a reduced instance of *I'll* uttered by S14 characterised by a short vocoid.

Figure 92 shows an instance of *I'll* in which the vocoid is very short. The vocoid has a dark quality due to the position of the back of the tongue raised towards the velum throughout the production of the piece. The vocoid is preceded by a portion of voiceless friction of the same quality.

As for the pr+aux *we'll*, by far the most common realisation is a back rounded vocoid. The main variation observed in *we'll* is the duration of the vocoid. Figure 93 shows two instances of *we'll* produced by the same speaker (S7).

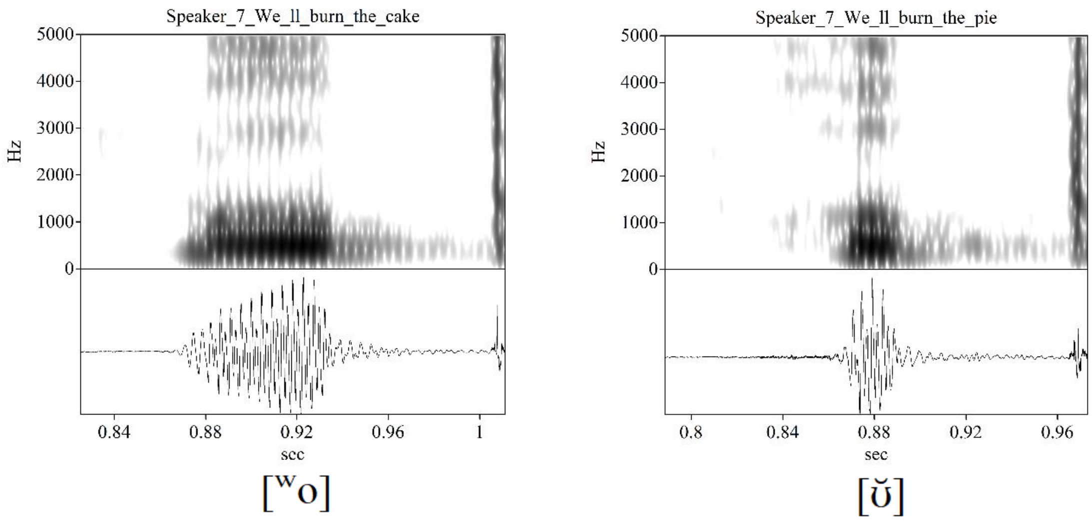
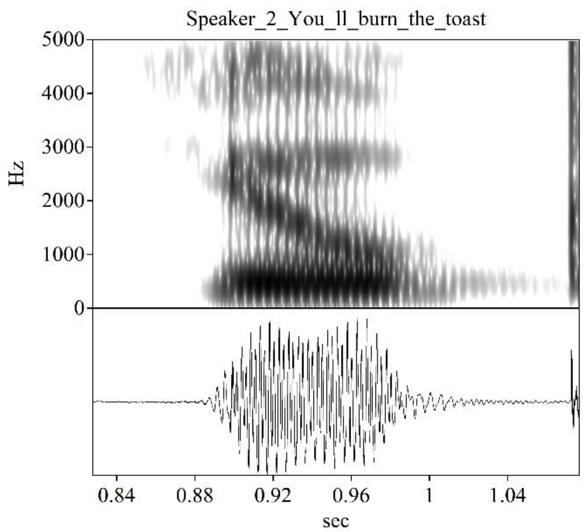


Figure 93. Spectrogram and waveform of two instances of *we'll* produced by S7. On the left, an instance of *we'll* in which the duration of the vocoid is 53 ms. On the right, an instance of *we'll* in which the duration of the vocoid is 18 ms. Both tokens were produced by the same speaker. The time windows display the same time frame.

Figure 93 shows two instances of *we'll* produced by the same speaker (S7) and displayed in the same time window. In both cases, like in the majority of instances of *we'll*, the piece is realised as a vocoid with a steady quality of a back rounded vocoid. The formant structure exhibits little variation in the frequencies of the first two formants, suggesting that the articulation does not change substantially during the production of the piece. The energy is concentrated at low frequencies (up to 1200 Hz) while the higher frequencies display only weak energy, e.g. F3 is visible but faint. The frequency of F1 is around 440 Hz in the instance on the left and 430 Hz in the instance on the right. The frequency of F2 is between

926 and 999 Hz in the instance on the left (range 73 Hz) and between 996 and 1026 Hz in the instance on the right (range 30 Hz). In both instances the frequency of F2 is higher in the central portion of the vocoid and lower at the beginning and end, displaying a domed shape. Although the duration of the vocoid in the instance on the right is 18 ms, the vocoid is preceded by a portion of weak voiceless friction of the duration of 35 ms.

While the vocoid in *we'll* is characterised by almost flat formants throughout its duration, the paradigm *you'll* is characterised by a falling F2 even in the most reduced instances. The realisations of *you'll* also exhibit more variation than *we'll*. In several cases of *you'll*, the lateral sound is realised as a vocoid (N = 12/47, 26%), but the vocoid can have either a back rounded quality or (less frequently) a front rounded quality. The range of realisations varies between [jọ ~ ʲə̹ ~ jø]. Figure 94 shows an unreduced instance of *you'll* in which the lateral is realised as a vocoid even in unreduced speech.



[jọ]

Figure 94. Spectrogram and waveform of an instance of *you'll* uttered by S2 in which the lateral is realised as a vocoid.

Figure 94 shows an instance of *you'll* that, like the great majority of tokens in this subset, is characterised by an initial close front unrounded vocoid and a tongue and lips movement to a mid-close back rounded vocoid. All instances of *you'll* show some degree of F2 movement, even the most reduced ones. Figure 95 shows a reduced instance of *you'll* in which the vocoid is very short (11 ms).

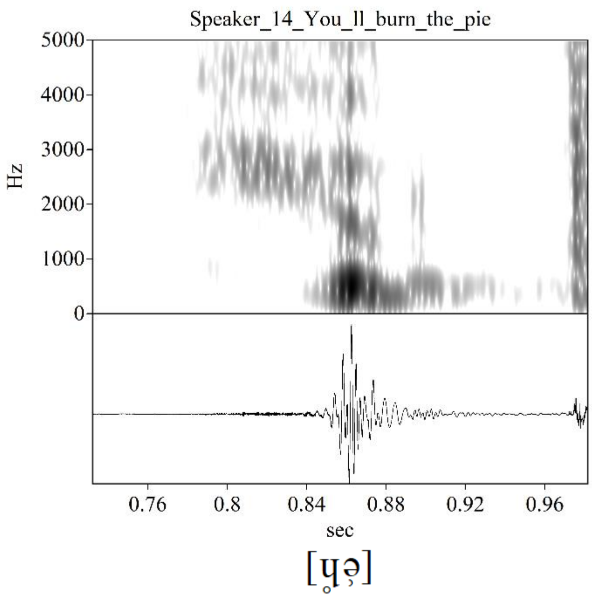


Figure 95. Spectrogram and waveform of a reduced instance of *you'll* produced by S14 in which the vocoid is very short (duration 11 ms).

Figure 95 shows a reduced instance of *you'll* in which, despite the temporal reduction, the movement of the second formant is visible in the vocoid as well as in the voiceless friction preceding it.

To investigate the degree of reduction of the magnitude of the tongue gesture from front to back, the formant dynamics of the first three formants in reduced and unreduced instances of *you'll* across speakers and repetitions were analysed. To identify the reduced and unreduced instances of *you'll*, the duration of the vocoid was taken into account together with the spectral observation of each token. The formant of the 15 tokens with the shortest vocoid duration (vocoid duration from 9 ms to 45 ms, mean = 34 ms, median = 39 ms) were compared to the 15 tokens with the longest vocoid duration (vocoid duration from 74 ms to 116 ms, mean = 92 ms, median = 87 ms). Figure 96 shows the formant dynamics of the reduced and unreduced instances of *you'll* measured at 9 equidistant points in time.

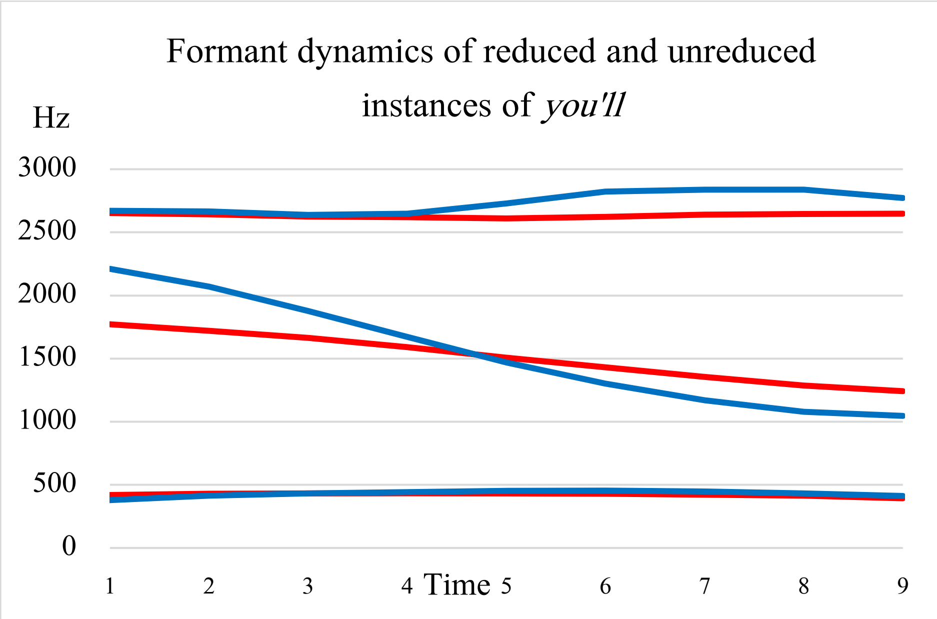


Figure 96. Formant dynamics of reduced (red) and unreduced (blue) tokens of *you'll* measured at 9 points in time. Although both F2 and F3 exhibit less variation in the reduced tokens, the movement from front to back (indicated by the falling F2) is present also in reduced tokens.

Figure 96 shows the first three formants in unreduced instances of *you'll* (blue) and reduced instances of *you'll* (red). Although all three formants differ between the two groups, F2 (the acoustic correlate of the front-back dimension) exhibits the greatest difference. At the beginning of the vocoid (Time point 1), F2 already starts at lower frequencies in the reduced tokens (1771 Hz in reduced instances, 2210 Hz in unreduced instances). At the end of the vocoid (Time point 9), F2 ends at higher frequencies in the reduced tokens (1241 Hz in reduced instances, 1046 Hz in unreduced instances). The frequency range of F2 in reduced tokens is 539 Hz, while the frequency range of F2 in unreduced tokens is 1164 Hz.

Nevertheless, despite the decrease in the magnitude of the gesture, the tongue movement from front to back – as shown by the movement of F2 from higher to lower frequencies – is always articulated even in reduced instances of *you'll*.

4.1.5.2. Summary of observations of *I'll, we'll, you'll*

To summarise the observations of the realisation of the paradigm *I'll*, *you'll* and *we'll*, the variants observed can be divided into four main groups, although the differences between them are subtle and not always easily identifiable.

The most common realisation of the lateral is as resonance only that is audible throughout the piece and that is not time delimited. This resonance has a dark quality and it is produced by the secondary articulation: the raising of the back of the tongue towards the velum and by lip rounding (or labialisation). Although the dark resonance is audible in all the tokens analysed and affects the entire piece, in the most common realisation, the resonance is the only acoustic cue that indicates the presence of a phonological lateral in the citation form of the paradigms. The second most common realisation of /l/ is as a (usually mid) back rounded vocoid. The primary articulation, if it is in place, does not create contact between the front of the tongue and the alveolar ridge. However, the secondary articulation, or velarisation, is articulated and produces a vocoid with a back rounded quality. This realisation has been observed in 15% of the data analysed (N = 22/147). The least frequent realisation of the lateral is with a lingual contact between the front of the tongue and the alveolar ridge. The beginning of the contact is identifiable in the spectrogram and waveform and the lateral sound can be temporally measured. This realisation is rare but it does occur in some unreduced instances in the dataset (6%, N = 9/147). Some instances of pr + *'ll* have been classified as a fourth type of realisation. These instances are characterised by a seemingly audible lateral sound that nevertheless cannot be identified in the spectrogram and waveform. These realisations are ambiguous and not well defined. However, since several instances were found to give the auditory impression of a lateral gesture being articulated, despite the absence of acoustic cues in the spectrograms, these instances were grouped together in a separate group. The differences between the variants are subtle and sometimes difficult to determine. Despite this, a count of the four main variants was attempted and it is shown in Table 13.

Pr + aux	Audible lateral	Audible but not visible lateral	Back rounded vocoid	Resonance only	Unsure	Total N
I'll	4 (9%)	11 (23%)	2 (4%)	29 (62%)	1 (2%)	47
You'll	4 (9%)	7 (15%)	12 (25%)	24 (51%)	0 (0%)	47
We'll	1 (2%)	5 (9%)	8 (15%)	38 (72%)	1 (2%)	53
	9 (6%)	23 (16%)	22 (15%)	91 (62%)	2 (1%)	147

Table 13. Count and percentage of four main variants of /l/ realisations in the paradigms *I'll*, *you'll* and *we'll*.

Table 13 shows the classification of the realisations of /l/ in four main variants. The boundaries between the four groups are not always clear cut and some instances are difficult to classify. Of the three paradigms (*I'll*, *you'll* and *we'll*), *we'll* shows less variation in the magnitude of gesture. The large majority of instances of *we'll* is realised with dark resonances only. A possible explanation is that the piece begins with a back rounded vocoid similar in quality to a vocalised-L. From an articulatory point of view, it is more economical to maintain the tongue position throughout the piece than to change it. This might explain why the first two formants of the vocoid in *we'll* are regularly flat and at low frequencies (below 1200 Hz) whether the piece is reduced or not.

In all instances of *you'll* analysed, the movement of the second formant from higher frequencies to lower frequencies is always visible, even in highly reduced tokens.

The paradigm *I'll* presents the largest range of variation of the three paradigms. One reason for this feature might be that the open front unrounded vocoid in *I* is unique in the small paradigmatic system of pronouns and therefore can undergo a wider range of variations without becoming problematic for speech perception. The shared feature between all the paradigms analysed in this section is the dark resonance throughout the duration of the piece, regardless of the actual tongue tip articulation of the lateral.

4.1.5.3. *She'll, he'll, they'll*

The paradigms *she'll*, *he'll* and *they'll* have the same syllable structure of a contoid in onset, a vocoid and an empty coda. The paradigm *she'll* is used in this section to illustrate the realisations of these three paradigms. All pieces are realised along a continuum between two main variants. One variant is characterised by a tongue movement during the production of the vocoid, from a front articulation to a back articulation. The other variant is more temporally compressed: the vocoid has a shorter duration and a steady quality of a front rounded vocoid. Figure 97 shows two instances of *she'll* that exemplify these two variants. In the instance on the left, there is a clear F2 movement during the vocoid. In the instance on the right, F2 is flat.

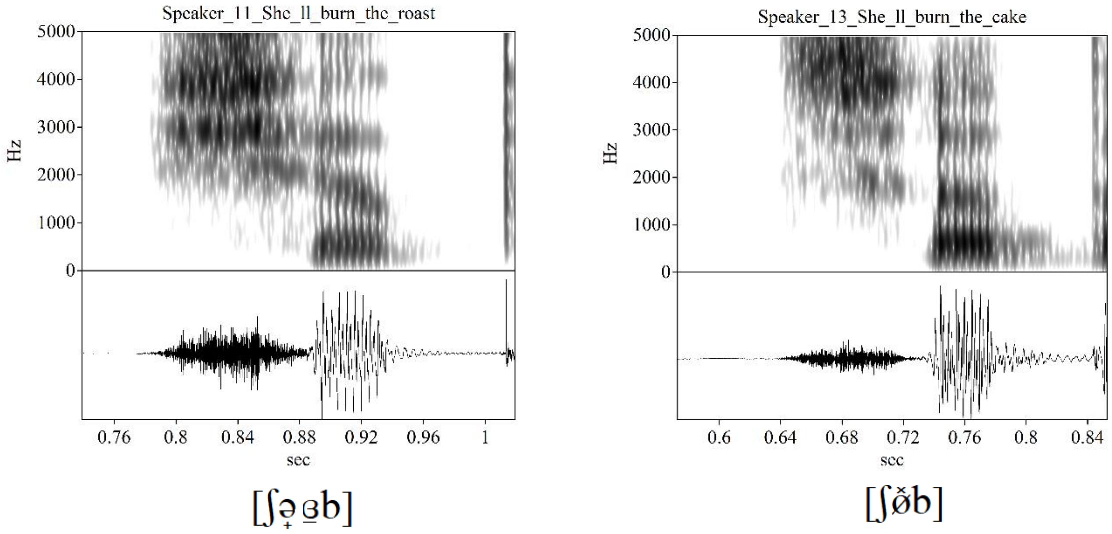
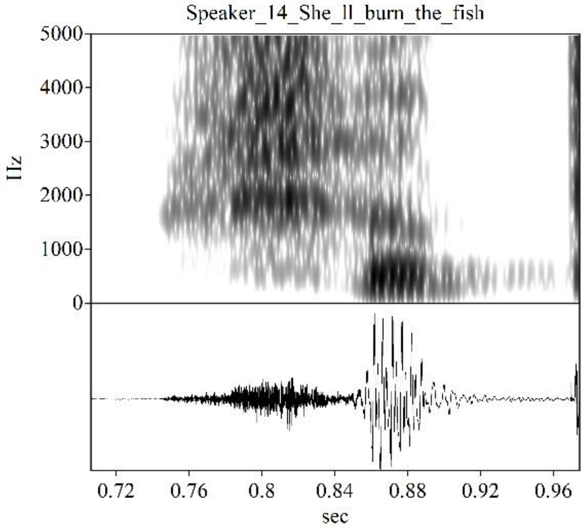


Figure 97. Spectrogram and waveform of two instances of *she'll*. On the left, an instance of *she'll* produced by S11 in which there is tongue movement from a close front position to a back one. On the right, an instance of *she'll* produced by S13 in which there is little tongue movement and the vocoid has a back rounded quality.

Figure 97 on the left shows an instance of *she'll* in which F2 moves from higher to lower frequencies during the vocoid, indicating that there is a tongue movement from front to back. In Figure 97 (right), however, the variation in the frequency of F2 is much smaller and perceptually the vocoid has a back rounded quality.

In all pr + *'ll*, the dark resonances are audible throughout the piece. Figure 98 shows an instance of *she'll* in which the palato-alveolar friction in onset is characterised by a particularly dark quality.



[ʃʷə̹]

Figure 98. Spectrogram and waveform of an instance of *she'll* uttered by S14 in which the initial friction has a dark quality.

In Figure 98, F2 is clearly visible in the palato-alveolar friction around 2000 Hz. In order to investigate the acoustic correlates of the dark resonances during the friction, the spectral moments of the palato-alveolar friction in *she'll* were compared to those of the palato-alveolar friction in *she'd*, which is characterised by clear resonances. Table 14 shows the spectral moments measured in the central portion of the friction, as explained in Section 2.2.2.2.2.

Paradigm	**Amp (dB)**	**CoG (Hz)**	**SD (Hz)**	**Skewness**	**Kurtosis**
she'll	57.7	3777	1266	1.89	8.24
she'd	57.3	4050	1364	1.55	5.70

Table 14. Mean values of amplitude and first four spectral moments measured in a central portion of the palato-alveolar friction in *she'll* and *she'd*, calculated across speakers and repetitions.

Similar results were obtained from the spectral moments of the glottal friction in *he'll* and *he'd*, as shown in Table 15.

Paradigm	Amp (dB)	CoG (Hz)	SD (Hz)	Skewness	Kurtosis
he'll	46.9	2314	1997	2.35	9.69
he'd	46.2	2795	2066	1.42	4.27

Table 15. Mean values of amplitude and first four spectral moments measured in a central portion of the glottal friction in *he'll* and *he'd*, calculated across speakers and repetitions.

The main feature to notice in Table 14 and Table 15 is that the mean CoG is lower in pr + *'ll* than pr + *'d*. A lower CoG is compatible with dark resonances. This feature is investigated further in Section 5.2.2.3.

The most reduced instances of *she'll*, *he'll* and *they'll* are characterised by a very short vocoid or a temporal reorganisation of the phonetic events in the oral cavity and the vibration of the vocal folds in the glottis. The feature of *she'll* that is always present even in most reduced instances is the palato-alveolar friction in onset. In the *she'll* subset there are no instances of complete voicelessness (which occurs in other *she* + aux subsets). Figure 99 shows an instance of reduced *she'll* in which the vocoid is short (duration 7 ms) but F2 is visible during the palato-alveolar friction and voicing continues after the end of the vocoid.

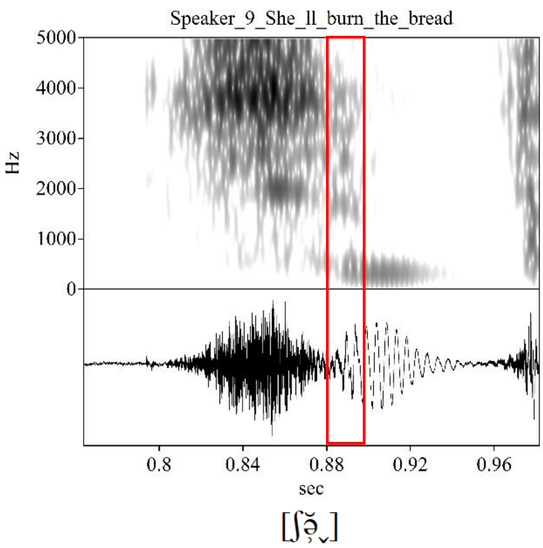


Figure 99. Spectrogram and waveform of a reduced instance of *she'll* with a very short vocoid produced by S9.

4.1.5.4. Summary of observations of *she'll*, *he'll* and *they'll*

To summarise the observations of the realisations of the phonological lateral approximant in the paradigms *she'll*, *he'll* and *they'll*, it was reported that the two main types of variants are articulated without lingual contact at the front of the oral cavity. In the first type (or group of subtly different instances) /l/ is realised as a vocoid with a back rounded quality: the first articulation is lost, and the secondary articulation becomes more prominent. Before the back rounded vocoid, a short onglide with a close front quality is audible and usually visible in the spectrograms. In the second type (or group of variants), /l/ is realised as dark resonance only. The dark resonances are audible throughout the piece in all the variants observed, but in the case of the second type of variant, the dark resonances are the most prominent acoustic correlate of /l/.

In the most reduced instances of *she'll*, *he'll*, and *they'll*, the vocoid is very short and can even be apparently absent. However, the dark resonances, as well as voicing, are always present.

4.1.5.5. *It'll*

The paradigm *it'll* behaves differently from all the other pronouns because of its syllable structure. Like the paradigm *it'd*, *it'll* is characterised by a second vocoid between the contoids as the consonant cluster /tl/ in coda position breaks the rules of English phonotactics (while /ts/ in *it's* is allowed).

In the paradigm *it'll*, there is a wide range of variation in the realisation of the phonological plosive in coda of *it*, but the lateral approximant exhibits very little variation compared to the variability observed in the other pr + *'ll* paradigms. The lateral approximant is realised as a vocoid in the great majority of instances (N = 38/48, 79%). In the remaining 10 tokens

(N = 10/48, 21%), the lateral is articulated with lingual contact at the alveolar ridge. In three of these instances, the preceding stop is laterally released (N = 3/48, 6%).

Figure 100 shows an instance of *it'll* that is realised as a vocoid changing from a close front unrounded quality to a mid back rounded quality. The absence of a visible intervening contoid allows a clear observation of the formant structure of the piece.

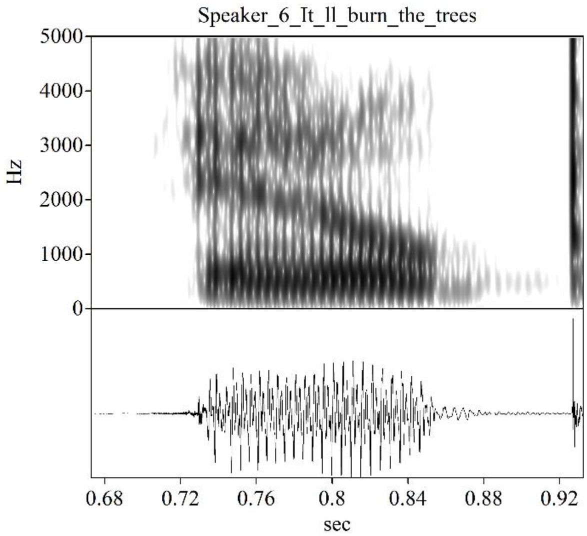
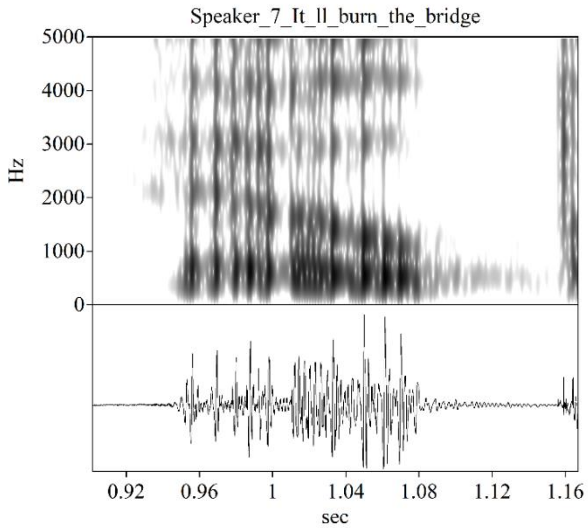


Figure 100. Spectrogram and waveform of an instance of *it'll* produced by S6 and realised as a vocoid changing in quality and no apparent intervening contoid.

Figure 100 shows an instance of *it'll* uttered by S6 in which the piece is realised as a continuous vocoid. This allows a clear observation of F2 which moves from the frequency of 2321 Hz at the beginning of the vocoid to the frequency of 960 Hz at the end of the vocoid. F1 also moves from 575 Hz at the beginning of the vocoid to 437 Hz at the end of the vocoid. The movement of the first two formants suggest that the tongue moves from a mid front position to a back position. Perceptually, the contoid can be heard as a very weak tap. It can be assumed that the front of the tongue moved towards the alveolar ridge but stopped far before reaching it. The small dip in amplitude and simultaneous slightly fainter

F2 are the only visible acoustic cues that indicate a change (albeit small) in the configuration of the oral cavity.

The same F2 movement can be observed in instances in which the phonation mode is creaky voice. Figure 101 shows an instance of *it'll* in which the piece is produced with creaky voice throughout its duration.

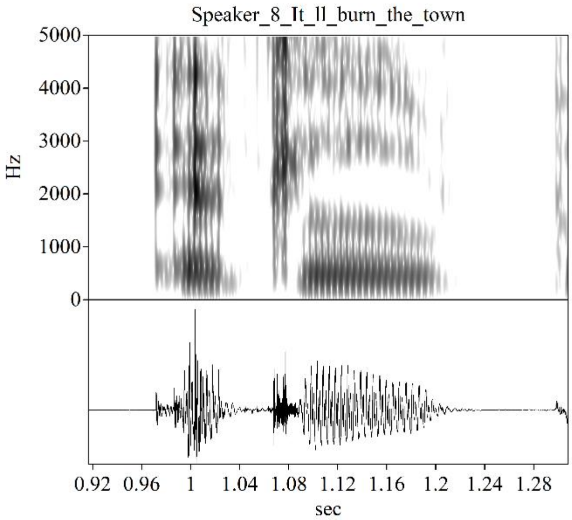


[ḛɔ̰̽]

Figure 101. Spectrogram and waveform of an instance of *it'll* produced by S7 and realised as a vocoid with creaky voice.

Figure 101 shows an instance of *it'll* in which the movement of the formants is visible and similar to the movement shown in Figure 100. In this case, F2 starts at the frequency of 2073 Hz and ends at the frequency of 1018 Hz. F1 starts at the frequency of 628 Hz and ends at the frequency of 444 Hz. As mentioned in Section 3.1.4, creaky voice is the acoustic correlate of /t/ in the absence of a contoid realisation.

In three instances, /t/ is realised as a plosive with a lateral release. Figure 102 shows an instance of *it'll* in which /t/ is laterally released.



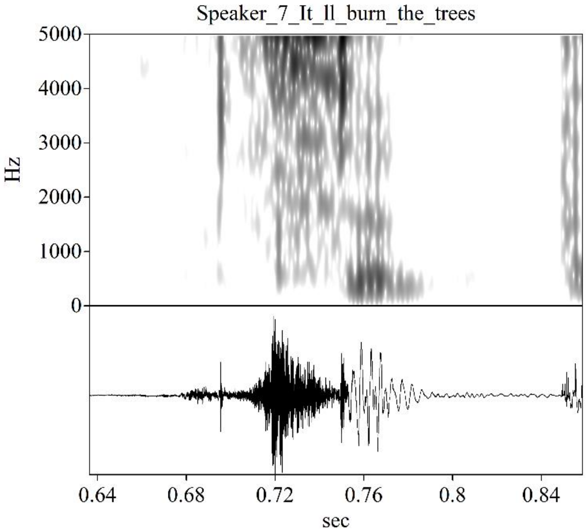
[ʔɪtˡɫ]

Figure 102. Spectrogram and waveform of an instance of *it'll* produced by S8 in which /t/ is laterally released.

Figure 102 shows an instance of *it'll* in which /t/ is realised with a hold phase with a complete closure, a burst, and a release phase. The build-up of air pressure in the oral cavity is released from the sides of the tongue.

While in all the other pr + *'ll*, there was a type of realisation of /l/ that was 'resonance only', this realisation has not been observed in the paradigm *it'll*. The reason for this can be ascribed once again to the syllable structure of the pronoun *it*.

The most reduced instances of *it'll* are characterised by the apparent absence of the vocoid in *it*. Figure 103 shows an instance of *it'll* in which the vocoid in *it* cannot be identified.



[tˢə̹]

Figure 103. Spectrogram and waveform of a reduced instance of *it'll* produced by S7 in which the vocoid in /ɪt/ is apparently absent.

Figure 103 shows an instance of *it'll* in which the vocoid at the beginning of *it* is apparently absent. The piece starts with some friction and a burst. These are followed by a long release phase (58 ms from the spike in the waveform to the first cycle of vocal fold vibration) characterised by glottal friction (aspiration) and followed by a back rounded vocoid.

To summarise the observation of the paradigm *it'll*, the phonological lateral /l/ is realised as a back rounded vocoid in the majority of instances (79%) and as a lateral with lingual contact in 21% of instances. The main source of variation in the realisation of the paradigm *it'll* is the phonological plosive, which can take many forms including being laterally released. Reduction is more apparent in the absence of a vocoid before /t/.

4.1.6. Auxiliary *are,* cliticised form *'re*

All the pr + *'re* combinations in this data are followed by the contoid /b/. SSB English (the variety of English recorded) is a non-rhotic variety – /r/ is not pronounced in pre-consonantal position. Therefore, in the dataset collected, the three paradigms *we're*, *you're* and *they're*, do not have a coda. However, the peculiarity of the cliticised form *'re* is that it changes the quality of the vocoid of the pronoun it is combined with. The following sections describe the three pr + aux *we're*, *you're*, and *they're*.

4.1.6.1. *We're*

The phonemic transcription of *we're* found in pronouncing dictionaries is /wɪə/. In the dataset analysed, *we're* is realised as a vocoid with changing quality in time only in unreduced instances. Figure 104 shows an unreduced instance of *we're* in which the formant movement is clearly visible: F2 moves from 946 Hz at the beginning of the vocoid to 1460 Hz 16 ms before the end of the vocoid.

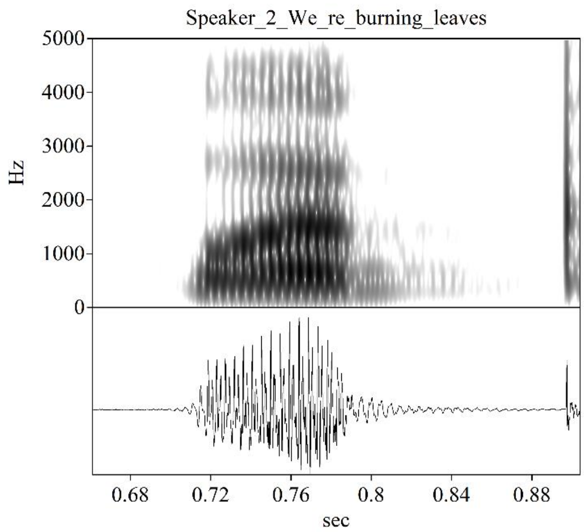


Figure 104. Spectrogram and waveform of an unreduced instance of *we're* produced by S2.

Figure 104 shows an unreduced instance of *we're*. To investigate the quality of the piece, the first three formants of *we're* were compared with the first three formants of the vocoid in *we've*. The vocoid in *we've* is used as a benchmark because the vocoid in *we've* is unlikely to be influenced by the contoid in coda position, as it is not articulated with the tongue. Figure 105 shows the formant dynamics of the vocoids in *we've* (blue lines) and *we're* (red lines), calculated across speakers and repetitions.

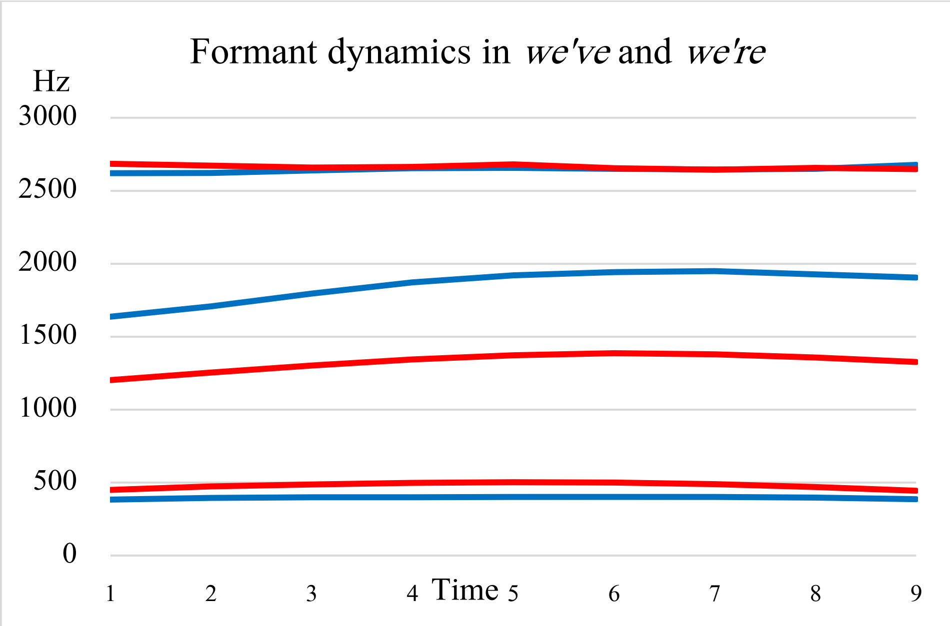
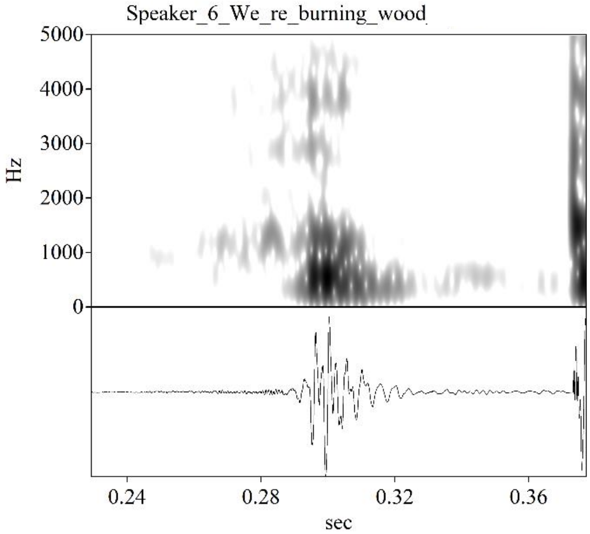


Figure 105. Formant dynamics of the vocoid in *we've* (blue) and *we're* (red).

As it can be seen in Figure 105, the formant dynamics, especially F1 and F2, of *we're* and *we've*, differ throughout the piece. F1 in *we're* is higher than in *we've*, indicating that the vocoid is produced with a lower tongue position. F2 in *we're* is much lower than in *we've*. The low frequencies of F2 suggest that the vocoid is articulated further back in the oral cavity. However, a lowered F2 can also be the acoustic correlate of lip-rounding and rhoticity. Auditorily, *we're* has a dark, or rounded quality throughout the piece compared to the more palatal quality of *we've*.

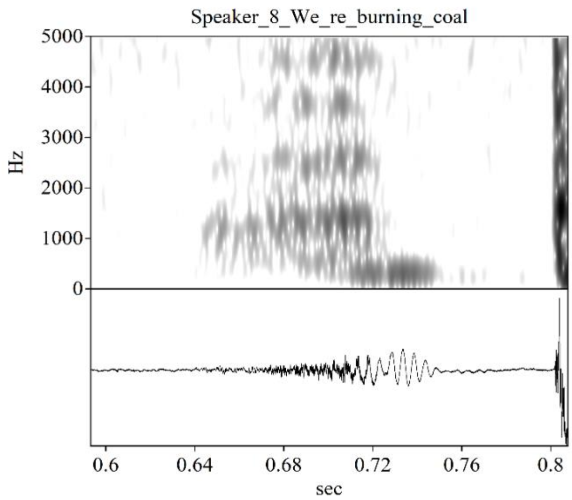
In its most reduced instances, *we're* is produced as a very short vocoid. Figure 106 shows a reduced instance of *we're* in which the duration of the vocoid is 10 ms.



[ɔ̆̽]

Figure 106. Spectrogram and waveform of a reduced instance of *we're* produced by S6. The duration of the vocoid is 10 ms.

As in the other pr + aux, also in *we're* there are instances that exhibit a realignment of the phonetic events occurring in the larynx and in the supralaryngeal vocal tract. Figure 107 shows an instance of *we're* in which most of the articulation in the oral cavity occurs before the vocal folds start vibrating.



[ʌ̬]

Figure 107. Spectrogram and waveform of a reduced instance of *we're* produced by S8 in which the articulation of the vocoid occurs before voicing starts.

Figure 107 shows an instance of *we're* characterised by an initial portion of friction (duration 76 ms) in which the formants are visible. It can be observed that the articulation in the oral cavity largely occurs before the vocal folds start vibrating. Voicing continues for 28 ms after the end of the friction.

4.1.6.2. *You're*

In the dataset analysed, the vocoid in *you're* changes quality in time, from a front close to a mid-central or back quality. Auditorily, the piece often sounds rounded from the beginning. Figure 108 shows a near-canonical realisation of *you're*. The formant movement is clear: F2 moves from 2641 Hz at the beginning of the vocoid to 1260 Hz at the end of the vocoid.

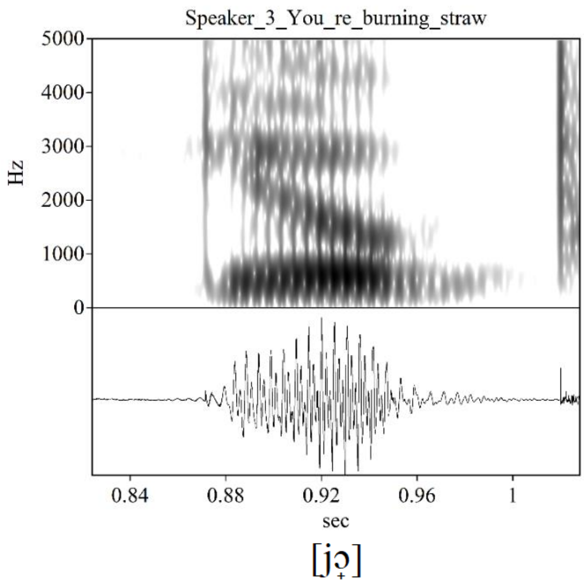


Figure 108. Spectrogram and waveform of an unreduced instance of *you're* produced by S3 with well-defined F2 movement.

The formant dynamics of the vocoid in *you're* were compared to the formant dynamics of the vocoid in *you've*. As mentioned above, the paradigm pr + *'ve* was chosen as reference because the contoid /v/ is less likely to affect the oral articulation of the preceding vocoid.

The mean frequencies of the first three formants in *you're* and *you've*, calculated across speakers and repetitions are shown in Figure 109.

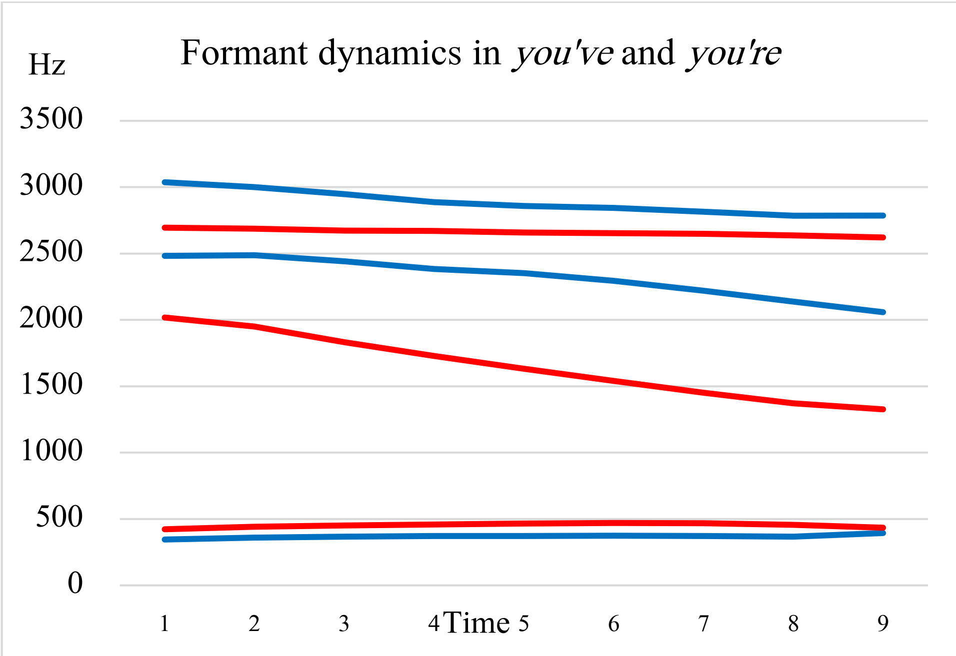


Figure 109. Comparison of the formant dynamics of the vocoid in *you've* (blue) and *you're* (red).

Similarly to F1 in *we*, also in *you*, F1 is higher in *you're* than *you've*, suggesting that the vocoid is articulated at a lower position in the oral cavity. In *you're* both F2 and F3 are lower than in *you've*, suggesting that the vocoid is articulated further back in the oral cavity, but it is possibly also more rounded.

Figure 110 shows a highly reduced realisation of *you're* in which the vocoid is only one glottal cycle long (the duration of the vocoid is 7 ms). However, before the vocal folds start vibrating, there is a period of palatal friction.

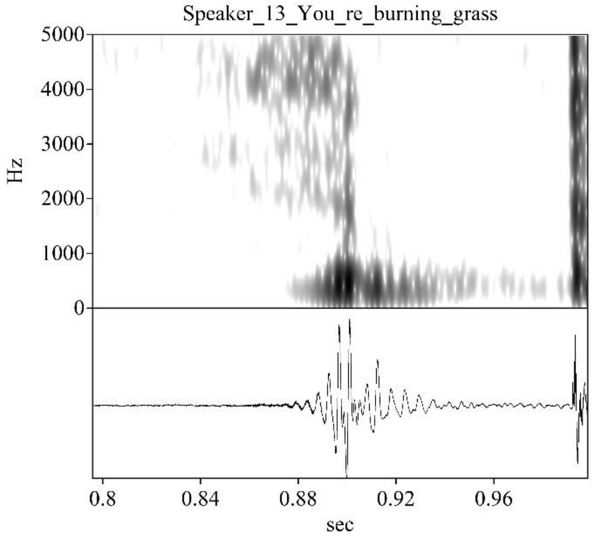


Figure 110. Spectrogram and waveform of a reduced instance of *you're* produced by S13 in which the short vocoid (duration 7 ms) is preceded by a portion of friction with palatal quality.

Figure 110 shows an instance of *you're* characterised by an initial portion of friction with visible formants reflecting the articulation of the vocoid. The friction ends abruptly when the bilabial closure is articulated.

4.1.6.3. *They're*

As described in Section 3.1.7, the vocoid in *they* is often produced with little movement of the tongue. This is the case also when *they* is in combination with the auxiliary *are*. Figure 111 shows an unreduced instance of *they're*.

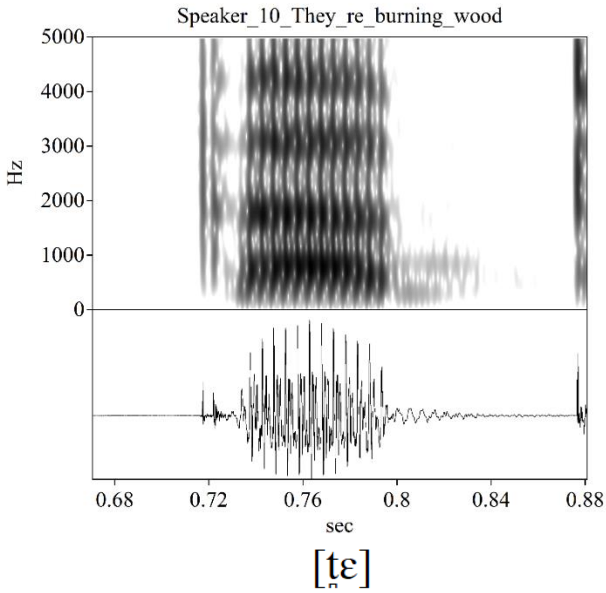


Figure 111. Spectrogram and waveform of an unreduced instance of *they're* produced by S10.

The first three formants of the vocoid in *they're* were compared to the formant dynamics of the vocoid in *they've*. Figure 112 shows the formant dynamics of the two paradigms.

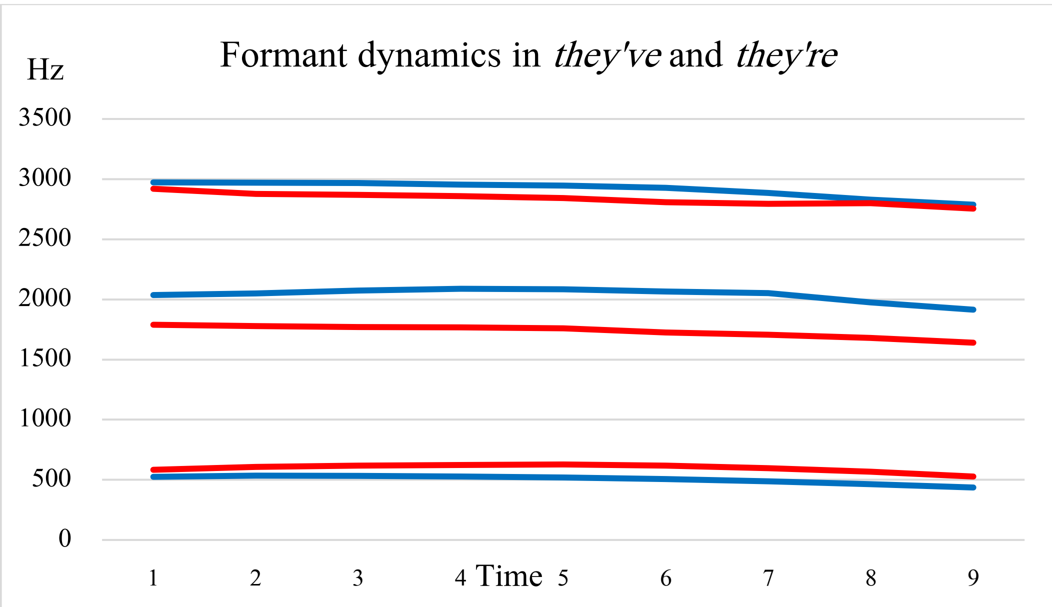


Figure 112. Formant dynamics of the vocoid in *they've* (blue) and *they're* (red).

The formants in the vocoid of *they're* show the same pattern observed in *we're* and *you're*. F1 is higher in *they're* than in *they've*, suggesting a more open oral articulation; while F2 is lower in *they're* suggesting a position of the tongue further back in the oral cavity than for the production of the vocoid in *they've*.

Reduced instances of *they're* exhibit temporal reduction. Figure 113 shows an instance of *they're* with a short vocoid (duration 37 ms) and irregular vibration of the vocal folds.

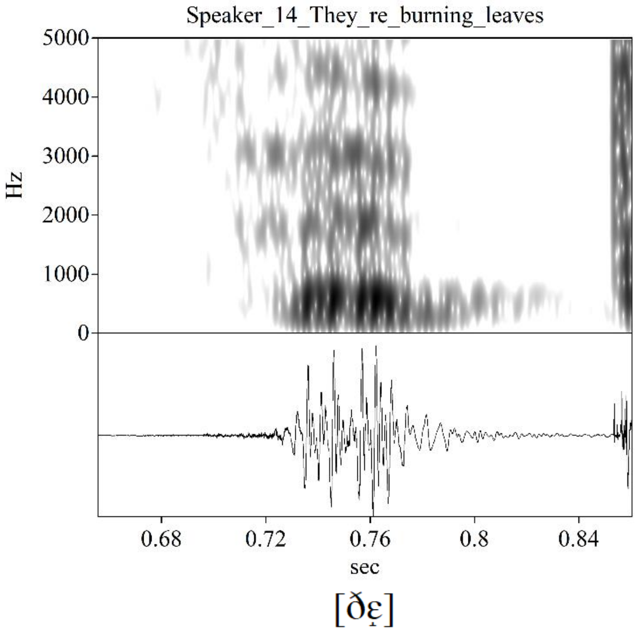


Figure 113. Spectrogram and waveform of a reduced instance of *they're* with a short vocoid produced by S14.

Figure 113 shows an instance of *they're* realised with creaky voice and a short vocoid, while the initial contoid is realised with dental friction. Compared to *we're* and *you're*, *they're* does not exhibit the temporal realignment of the events occurring in the supralaryngeal vocal tract and the larynx, although they could be masked by the friction of the contoid in onset.

4.1.6.4. **Summary of observations of the cliticised form *'re***

The three paradigms *we're*, *you're*, and *they're* exhibit some similarities. The comparison of their formant dynamics with the corresponding pr + aux *we've*, *you've*, and *they've* show that in combination with the auxiliary *are*, F1 is higher and F2 is lower throughout the piece. The durations of the various phonetic events of the three paradigms were compared and are reported in Table 16.

	friction	vocoid	voicing	closure	piece
we're	(42)	46	86	71	116
you're	(48)	47	85	68	116
they're	24	50	84	71	145

Table 16. Mean duration of the main phonetic events in *we're*, *you're*, and *they're*, calculated across speakers and repetitions.

Table 16 shows the mean duration of the main acoustic events of the three paradigms. The column 'friction' reports the duration of the phonological /ð/, and the friction produced at the beginning of instances of *we're* and *you're* as described in this and the previous chapter. The duration of the initial friction in *we're* and *you're* is not included in the duration of the piece. While *we're* and *you're* exhibit similar reduction patterns, *they're* differs in that a vocoid is always produced in all instances of *they're*. Moreover, in reduced speech, the phonetic events in *we're* and *you're* are temporally realigned. On the one hand, the articulation in the supralaryngeal vocal tract tends to occur before the vocal folds start vibrating. On the other hand, voicing tends to continue after the closure at the lips is articulated. This temporal reorganisation of the events results in a short vocoid (when present) and an initial portion of friction in which formants are visible. This is not the case in *they're*. The vocoid in *they're* is always produced and is simultaneous to the vibration of the vocal folds. If the gesture for the vocoid starts before voicing, it is masked by the contoid preceding it, regardless of its phonetic realisation.

4.2. Auxiliaries *was* and *were*

Although the main type of auxiliary analysed in this thesis is the cliticised form that must be attached to a host, the past tense forms of BE were analysed too. The rationale behind this decision is twofold. Firstly, the pr + aux paradigms analysed so far contrast in terms of present versus past tense: present and past tense of HAVE (*has*/*have* and *had*), and the pair *will* and *would* 10. Since the present tense of the auxiliary BE has been recorded (because it has the form of a weak non-syllabic clitic, as explained in 1.2.3) analysing the past tense of the auxiliary BE completes the set of present versus past tense contrasting pairs in which the present tense has a weak non-syllabic form. With the analysis of *was* and *were*, the present versus past tense contrasting pairs are:

present	past
has, have	*had*
will	*would*
is	*was*
are	*were*

Table 17. Pairs of paradigms that contrast in the tense aspect: present versus past.

The contrast between present and past tense pairs of auxiliaries is covered in Chapter 5.

Secondly, the contrast with the present tense of BE is conveyed by lip-rounding or labiovelarity (Ogden, 1999). This means that the contrast is conveyed by the resonances of the auxiliary. As stated by Ogden (1999: 72), "[i]n Firthian terms" the labiovelarity of *was* can be treated "as a prosody of the syllable, because /w/ has implications over the whole domain". Since one of the aims of this research is to investigate the role of the prosodies in

10 *Will* and *would* have been included following Ogden (1999) treatment based on their "behaviour in reported speech: I will do the cleaning. I said I would do the cleaning".

speech understanding, the contrast between the present and the past tense of BE can shed light on this issue.

The following sections report on the analysis of the paradigms *I was*, *she was*, *he was*, *it was*, *you were*, *we were*, *they were*. First the auxiliary *was* is described, then the auxiliary *were*. The section on *was* is further divided into two groups: *was* in combination with a pronoun with an open syllable (*I was*, *she was*, *he was*) and in combination with the only pronoun with a closed syllable (*it was*).

4.2.1. Auxiliary *was*

The phonemic transcription of the auxiliary *was* is /wɒz/ ~ /wəz/. Unreduced realisations of *was* exhibit a long vocoid with a noticeable dip in F2 and weaker energy at higher frequencies (from 2 kHz) in correspondence of the labial-velar articulation. Figure 114 shows an unreduced realisation of *she was*.

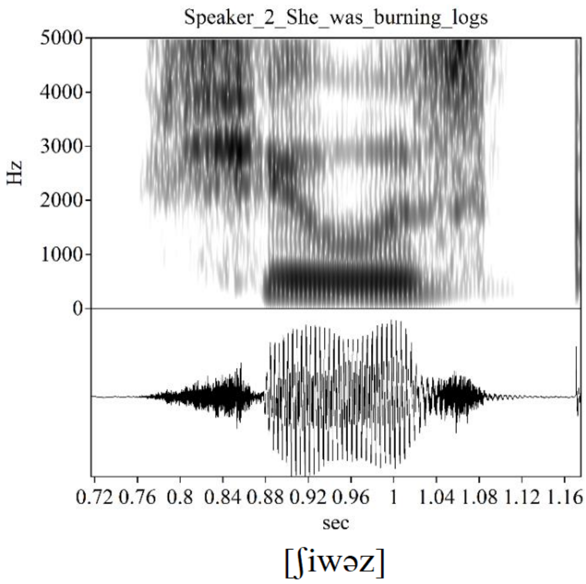


Figure 114. Spectrogram and waveform of an unreduced realisation of *she was* produced by S2.

Figure 114 shows an instance of *she was* in which F2 moves from higher frequencies (2352 Hz) at the beginning of the vocoid, to lower frequencies (1203 Hz at 0.96 sec), and up again to 1709 Hz at the end of the vocoid. Although F2 is considered the acoustic correlate of the front-back dimension of the vowel space, it is also affected by lip-rounding (Fant, 1992). The movement of F2 visible in Figure 114 is the acoustic correlate of the labiovelarity of /w/. Interestingly, F3 – which is affected by a range of articulatory parameters, including the shape of the tongue (Lindblom and Sundberg, 1971) – does not exhibit a noticeable variation along the frequency scale, although in other unreduced instances its amplitude is much lower than in the surrounding vocoids.

Most instances of *was* in the dataset exhibit flat F2 throughout the duration of the vocoid. Figure 115 shows an instance of *she was* in which F2 moves slightly downward but does not exhibit the dip shown in Figure 114.

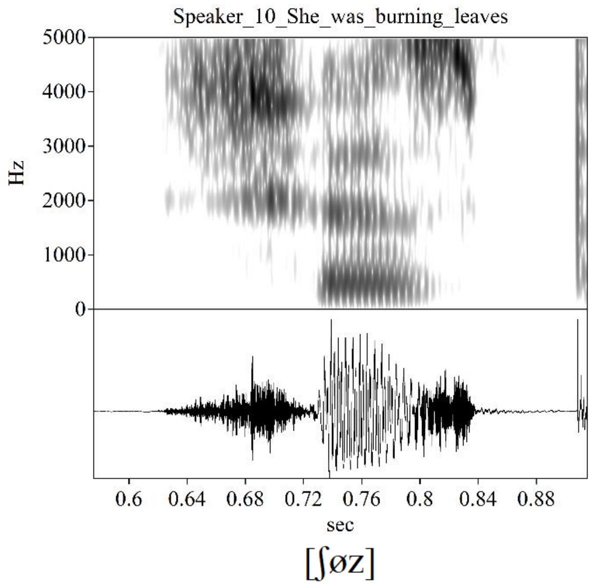
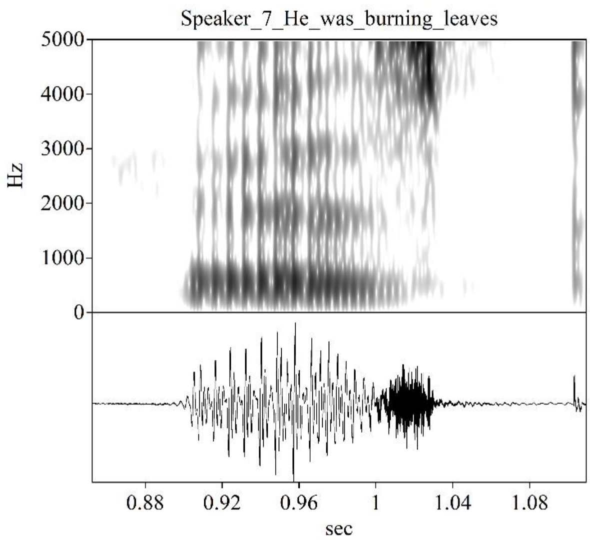


Figure 115. Spectrogram and waveform of a reduced realisation of *she was* produced by S10.

Figure 115 shows an instance of *she was* in which the vocoid is short (59 ms, mean duration of the vocoid in *she was* across speakers and repetitions is 85 ms) and has a steady quality. The quality of the vocoid is that of a near-close near-front rounded vocoid.

The paradigms *she was* and *he was* exhibit very similar patterns of reduction. The main features observed in reduced instances are a vocoid with short duration and flat formants, and the dark quality of the entire piece due to lip-rounding and possibly velarisation. The initial friction is always produced except in one of the rare instances in which the pronoun *he* is realised with hardly any glottal friction. In that token, the piece is characterised by creaky voice, as shown in Figure 116.



[ʏ̰ø̽z]

Figure 116. Spectrogram and waveform of a reduced realisation of *he was* produced by S7, and characterised by very weak initial glottal friction and creaky voice.

Reduced instances of *I was* exhibit the same short vocoid and steady formants without dip in F2 observed in *she was* and *he was*. However, auditorily, *I was* is characterised by an open central vocoid with little rounding. That is, *I was* does not exhibit the same dark resonances of other pr + *was*. To further investigate the quality of the vocoid in *I was*, two subsets of *I was* and *I've* with a vocoid duration shorter than the mean vocoid duration of the entire dataset were selected and their formant dynamics compared. The mean vocoid duration across speakers and repetitions of the entire *I was* dataset is 101 ms. Twenty-two tokens had a vocoid duration shorter than 101 ms and were thus selected. The same process was applied to the *I've* dataset. Twenty-three tokens had a shorter vocoid duration than the mean across speakers and repetitions (38 ms) and were thus selected. The rationale to compare the formant dynamics in *I was* to those in *I've* is that the vocoid in *I've* is unlikely to be affected by the articulation of the contoid in coda position since /v/ is not articulated with a lingual articulation. Therefore, bearing in mind that there is a durational difference between the two vocoids and that only the quality of the vocoid is being compared, comparing the formants of the vocoid in *I was* with the formants of the vocoid in *I've* can provide information about the articulation and the resonances of *I was*. Figure 117 shows the formant dynamics of *I was* and *I've*.

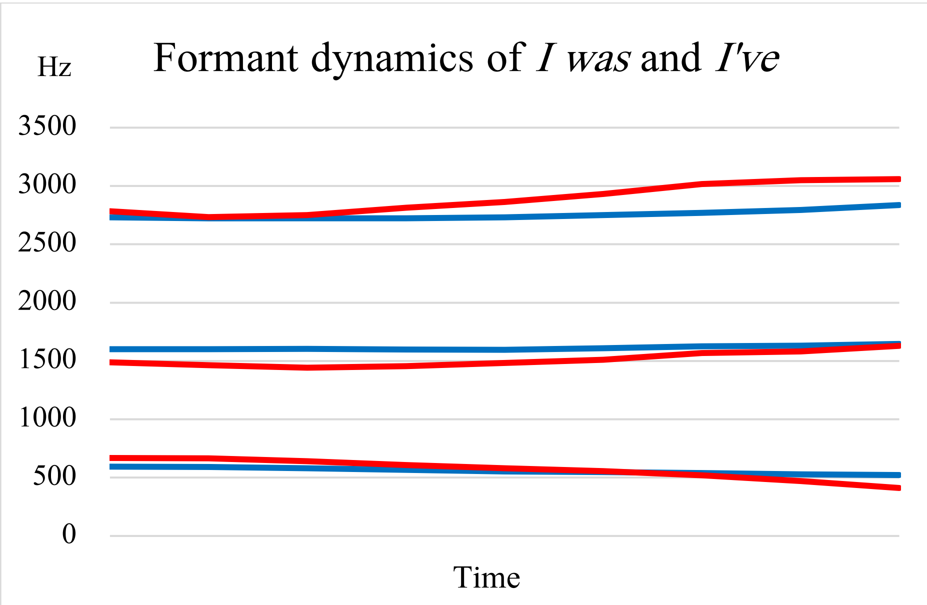
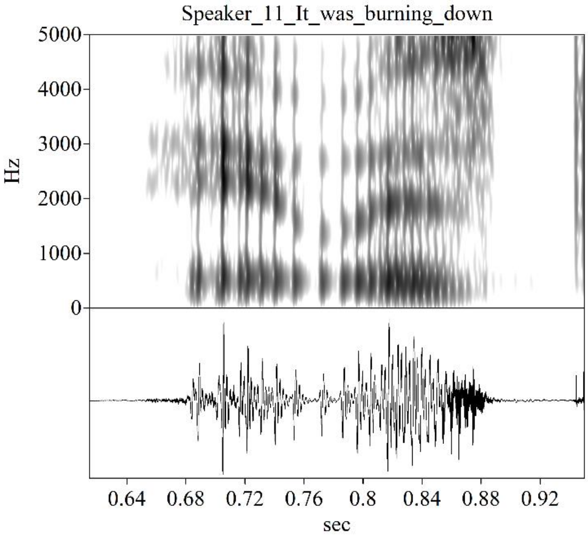


Figure 117. Formant dynamics of a subset of instances of *I was* (red) and *I've* (blue).

Figure 117 shows that there are several differences between the formant dynamics in reduced instances of *I was* and *I've*. Although F2 is slightly lower in *I was* than *I've*, it does not exhibit the expected difference. The auditory impression of reduced instances of *I was* is that they are less rounded than reduced instances of *she was* and *he was*. This might be due to the fact that *she was* and *he was* must maintain the contrast with other two paradigms that have voiced alveolar friction in coda position, namely *she's* and *he's*; while *I was* is the only *I*+ aux with voiced alveolar friction in coda. This means that *I was* does not have to maintain the contrast with a **I's* counterpart. The present counterpart is *I'm*, which always retains the nasality. Therefore, *I was* needs only to retain the friction in coda position to be correctly identified, and does not need a high degree of rounding like *she was* and *he was* do.

4.2.1.1. *It was*

The most common realisation of the paradigm *it was* is with creaky voice throughout the vocoid or in a central portion of the vocoid in correspondence of a dip in F2 as observed in unreduced instances of *was* mentioned in the previous section. Figure 118 shows an unreduced instance of *it was* produced by S11.

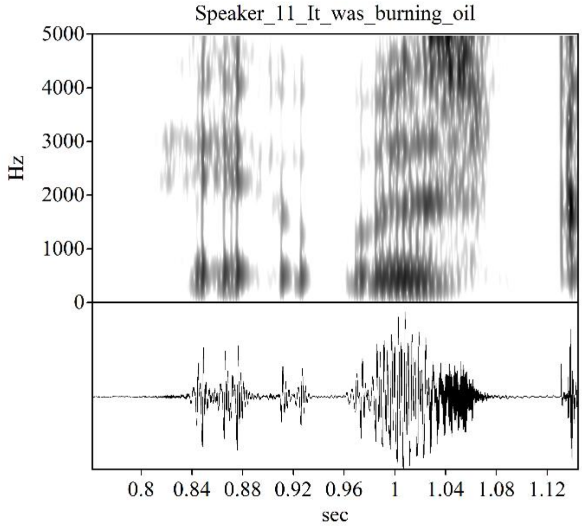


[ɪ̰w̰ə̦z]

Figure 118. Spectrogram and waveform of an unreduced instance of *it was* produced by S11.

Figure 118 shows an instance of *it was* characterised by creaky voice throughout the duration of the vocoid (except a short portion of modal voice at the end of the vocoid). The vocoid exhibits a clear dip in F2 in a central portion of the vocoid. F3 also exhibits a much smaller but still noticeable dip.

In some instances, an oral articulation and a glottal articulation occur simultaneously. Figure 119 shows an instance of *it was* in which the labial-velar articulation occurs simultaneously to a glottal closure.

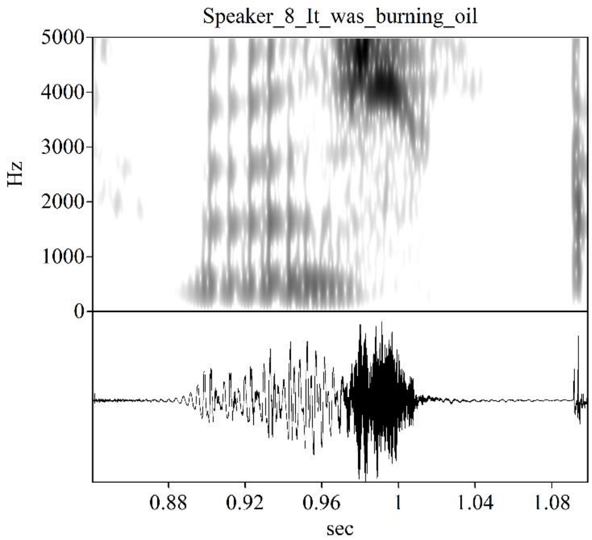


[ɪ̰ʔʷøz]

Figure 119. Spectrogram and waveform of an instance of *it was* produced by S11.

Figure 119 shows an instance of *it was* in which the formant movement (especially F2) indicates that the labial-velar gesture is simultaneous to the glottal articulation.

The paradigm *it was* can be reduced to a short creaky vocoid followed by alveolar friction, which can be partially voiced or voiceless. Figure 120 shows a reduced instance of *it was* uttered by S8.



[ø̽s]

Figure 120. Spectrogram and waveform of a reduced instance of *it was* produced by S8.

Figure 120 shows a reduced instance of *it was* realised with a short vocoid with initial creaky voice, followed by alveolar friction. The quality of the vocoid is that of a close-mid near-front rounded vocoid. This realisation is not dissimilar to some instances of *it's* realised with a short creaky vocoid and a portion of friction. However, the quality of the vocoid is markedly different: *it was* is characterised by a dark back, rounded quality, and *it's* is characterised by a clear front, unrounded quality.

4.2.2. Auxiliary *were*

The auxiliary *were* can be combined with the pronouns *you*, *we*, *they*. Its phonemic transcription is /wɜː/ ~ /wə/. The three paradigms *you were*, *we were* and *they were* exhibit the same features of reduction and are described together in this section. Unreduced instances of pr + *were* are characterised by the dip in F2 that was observed in *was*. Figure 121 shows an unreduced instance of *you were*.

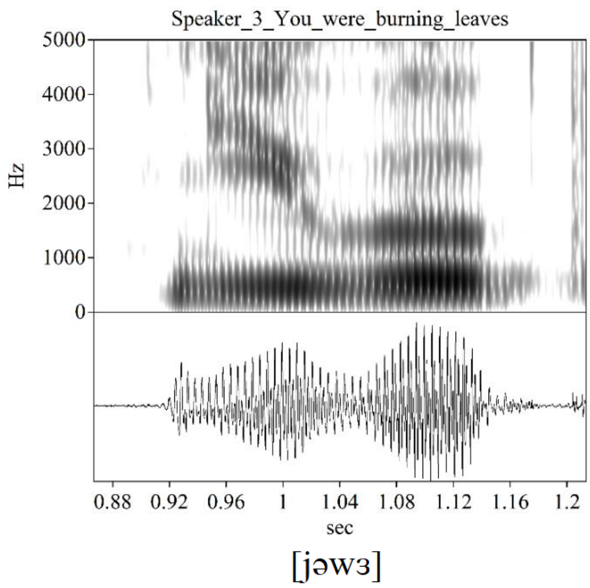


Figure 121. Spectrogram and waveform of an unreduced instance of *you were* produced by S3.

Figure 121 shows an unreduced instance of *you were*. The piece is produced with a long vocoid that starts with a high front palatal quality which changes into a back rounded quality. The spectrogram shows a clear F2 movement from higher frequencies to lower frequencies (from 2665 Hz to 1396 Hz in the central portion of the vocoid; range 1269 Hz). In the central portion of the vocoid there is a decrease in overall amplitude and energy at higher frequencies (above 2000 Hz).

Reduced instances of pr + *were* have shorter vocoids and less formant movement. Figure 122 shows a reduced instance of *you were*.

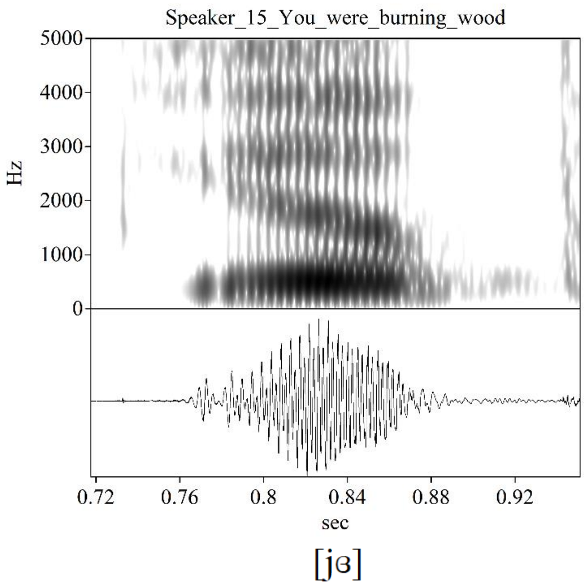
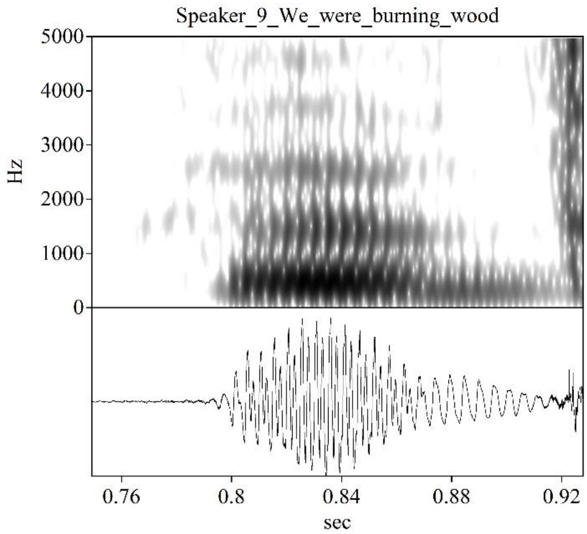


Figure 122. Spectrogram and waveform of a reduced instance of *you were* produced by S15.

Figure 122 shows a reduced instance of *you were*. The piece is characterised by a vocoid that starts with a close front rounded quality, and ends with an open-mid back rounded quality. Lip rounding is audible from the beginning of the piece.

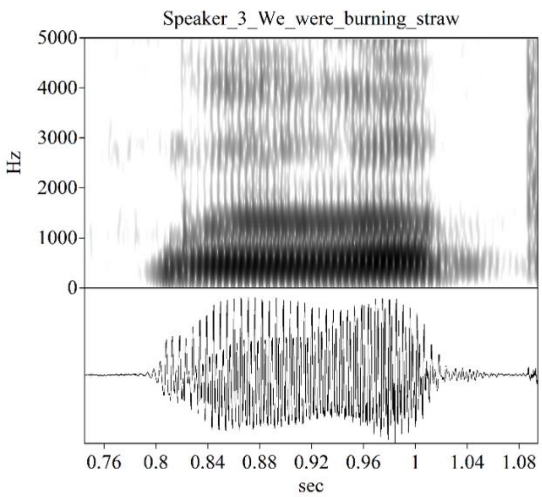
Reduced instances of *we were* are produced as a vocoid with little formant movement. Figure 123 shows a reduced instance of *we were*.



[ɵ]

Figure 123. Spectrogram and waveform of a reduced instance of *we were* produced by S9.

Figure 123 shows a reduced instance of *we were* in which the vocoid is quite short (63 ms; mean vocoid duration of *we were* across speakers and repetitions is 136 ms). The auditory impression is of a mid central rounded vocoid. In several instances of *we were*, the vocoid exhibits the same formant structure, but with a lower overall amplitude in a central portion of the vocoid. Figure 124 shows an example.



[ʷɵᶷɵ]

Figure 124. Spectrogram and waveform of an instance of *we were* with a dip in overall amplitude in a central portion of the vocoid.

Figure 124 shows an instance of *we were* in which there is a dip in overall amplitude starting between the marks 0.92 and 0.96 sec, which is reflected also in lower energy at higher frequencies (above 2000 Hz). The auditory impression is of a piece with two sonority peaks.

4.2.3. Summary of observations of the auxiliaries *was* and *were*

While unreduced realisations of pr + *was* and pr + *were* exhibit a dip in amplitude in the central portion of the vocoid which gives the auditory impression of two syllables, reduced instances are monosyllabic, or characterised by a single sonority peak. The vocoid can exhibit flat formants, as in reduced instances of *she was* and *we were*, or a falling F2 as in reduced instances of *you were*. In pr + *was* the alveolar friction in coda is always retained. However, voicing during the friction is variable, resulting in partially voiced or voiceless friction. Auditorily, both pr + *was* and pr + *were* are characterised by a dark quality throughout the piece. The dark resonance is due to labiovelarity. An exception is *I was*. In some reduced instances of *I was*, the labiovelarity – and therefore the dark quality of the vocoid – is very weak. A possible explanation for this lack of dark resonances in *I was* is in the system of paradigmatic contrasts of the pronoun *I*, in which *I was* is the only paradigm with alveolar friction in coda position. This means that the alveolar friction in coda is the only phonetic feature needed to maintain the identity of *I was*. This is not the case for the paradigms *she was* and *he was*, which must maintain the contrast with the present tense forms *she's* and *he's*. The contrast between the past and the present tense of BE in these two paradigms is conveyed by the labiovelarity, which seems to be the main phonetic feature of all pr + *was* and pr + *were* paradigms, except *I was*.

4.3. Summary of observations

This chapter, in conjunction with Chapter 3, aimed to describe the most common realisations, as well as the wide range of variation, in the English pronoun and auxiliary combinations. The focus of these two chapters was on the features of reduction that could be observed in the data, and in particular on the phonetic features that remain in the signal in reduced speech. That is, rather than focussing on what is deleted, or missing, in reduced speech, like most studies on reduction do, the aim was to identify what remains in the acoustic signal when speech is reduced. The assumption at the base of this analysis is that in reduced speech crucial information is retained by fine phonetic detail.

The phonetic features of the cliticised forms of the auxiliaries are not as straightforward to identify as those of the pronouns. In particular, /d/ does not exhibit only one feature that is always present in all instances. Its high degree of variability and range of possible realisations mean that it is difficult to pin down /d/ to just one or two phonetic features. The most common feature of /d/ is the closure. In 96% of instances in the dataset (N = 503/525), pr + *'d* paradigms are realised with a complete closure, regardless of the release. However, in 4% of instances (N = 22/525) /d/ is realised with friction throughout its duration. The realisation of /d/ with alveolar friction raises two issues. Firstly, it becomes problematic to treat 'closure' as the phonetic feature that characterises all instances of *'d*. To solve this issue it needs to be specified that pr + *'d* is not always characterised by a closure, but can also be characterised by alveolarity (with or without a closure). Alveolarity alone cannot be considered the main feature of pr + *'d* because the place of articulation of the closure can be assimilated to the sound that follows. In the data analysed, some instances of pr + *'d* were characterised by a bilabial closure due to the following /b/. Secondly, it raises the question of how the instances of pr + *'d* with friction differ from instances of pr + *'s*, such as *she's* and *he's*. This question will be addressed from an acoustic point of view in Chapter 5, and from a perceptual point of view in Chapter 7.

Similarly, the cliticised form of the auxiliary *'ve*, /v/ does not seem to be characterised by a single phonetic feature. Two main phonetic features were identified as properties of pr + *'ve*: labiodentality and voiced friction. However, in a small number of instances, neither labiodentality nor voiced friction can be heard. The analysis of these instances revealed that they are characterised by a falling F2 at the end of the vocoid, a smooth, gradual end of the vocoid, and longer vocoid duration and shorter closure duration compared to corresponding pr + *'d* paradigms. These features can all be treated as the exponents of labiodentality even when the labiodentality cannot be heard.

As for the other cliticised forms of the auxiliaries, *'s* is characterised by alveolar friction. The friction can be partially voiced, as in *she's* and *he's* or voiceless as in *it's*. The main variations observed in pr + *'s* are in the amount of friction produced, in its duration and amplitude, in the temporal onset of the friction relative to the vocoid and in the duration of voicing. The friction can be very weak and even absent in *she's*. In combination with the pronoun *it*, the friction is always produced, while the pronoun can be highly reduced and even apparently absent. This is not surprising considering the small paradigmatic system of contrast of the pronouns that can be combined with *'s*. The pronoun *it* occurs in the paradigms *it's*, *it'd*, *it'll* and *it was*. In this system, there are two paradigms in which *it* is combined with an auxiliary that contains friction, *it's* and *it was*. *It was* is characterised by labiality (rounding) and voiced friction. *It's* is characterised by voiceless friction. Therefore, *it's* and *it was* differ in resonance and in the voicing of the friction in coda.

The cliticised form *'ll* exhibits a wide range of realisations, but all of them are characterised by velarity, regardless of the articulation of the tip of the tongue. The primary articulation –

the alveolar contact and lateral airflow – can be lost, but the secondary articulation – the tongue dorsum gesture – is always articulated and its resonances can be heard throughout the piece.

The cliticised form *’m* is always realised as a bilabial nasal contoid in the data analysed. The nature of the data collected, in which all pr + aux are followed by /b/, means that the bilabial nasal does not exhibit much variation. In fact, the variations in the pr + aux *I’m* are observed mostly in the vocoid rather than in the nasal contoid. The vocoid can be highly reduced, but the nasal contoid is always produced.

As for the cliticised form *’re*, its peculiarity is that it changes the quality of the vocoid of the pronoun it is combined with. The comparison between the formant dynamics of the pr + aux *we’re* and *we’ve*, *you’re* and *you’ve*, *they’re* and *they’ve* revealed that in all three pairs, in pr + *’re* F1 is higher, F2 is lower, and F3 is lower than in pr + *’ve* (except *we’re*, in which F3 is slightly higher than in *we’ve* at the beginning of the piece). For this reason, the main phonetic feature that characterises all three pr + *’re* is the openness of the vocoid.

Finally, the auxiliaries *was* and *were* were analysed even though they do not have a cliticised form. The rationale for analysing them is to compare the present and past tense forms of the auxiliary BE. Even though *was* and *were* do not have a cliticised form in the orthography, the analysis reported in Section 4.2 shows that reduced instances of pr + *was* and pr + *were* can be realised as monosyllabic. The main phonetic feature of the auxiliaries *was* and *were* is the labiovelarity. The only exception is *I was*, in which the labiovelarity can be weak, but is always characterised by voiced friction in coda.

Table 18 summarises the main phonetic features that characterise the cliticised forms of the auxiliaries, and *was* and *were*.

Clitics	Essential phonetic features
'd	Closure and/or alveolarity
's	Alveolar friction
've	Labiodentality and/or voicing
'm	Nasal contoid
'll	Velarity
're	Openness
was	Labiovelarity and friction
were	Labiovelarity

Table 18. Summary of the main phonetic features that characterise the forms of the auxiliaries analysed in this chapter.

The qualitative analysis presented in this and the previous chapter indicates that all pronouns and auxiliaries are characterised by one or two main phonetic features that are always present even in highly reduced tokens. Returning to the concept of the essential phonetic components that constitute the identity of function words proposed by Kohler and Niebuhr (2011), and mentioned in Section 1.3.1 and 3.2, most of the data analysed in Chapters 3 and 4 support this concept. In particular, the data provide some support to the concept that the essential components are articulatory prosodies that are not tied to individual segments but distributed throughout the piece. That is, the identity of function words is not manifest in segmental units, but in long-domain resonances. A clear example is the dark resonance of pr + *'ll*. Regardless of the tip tongue articulation of the lateral, in pr + *'ll* combinations the entire piece is characterised by a dark quality triggered by the secondary articulation (velarity) of the lateral. The information about the auxiliary in pr + *'ll* paradigms is conveyed by the velarity alone, as demonstrated by the high percentage of instances of pr + *'ll* in which only the dark resonance is present (62%).

However, the wide range of variability in some pr + aux, makes it difficult to define exactly what is 'essential'. In some cases, such as for the cliticised forms *'d* and *'ve*, it seems that a bundle of subtle phonetic detail characterise these forms. For example, in the case of *'d*, in the great majority of cases, stoppedness, or closure, is the phonetic feature that characterise pr + *'d*. However, the presence of a few instances in which there is no closure, questions the validity of suggesting that stoppedness is essential. In the few cases without closure, alveolarity is present, in form of friction. It can be claimed that the identity of pr + aux is maintained by either or both closure and alveolarity, but none of them is essential – in the absence of stoppedness, alveolarity is present, and vice-versa. A more suitable approach for the data analysed is to identify a set of phonetic features that can occur variably in each piece. In the case of *'d*, the set of phonetic features that characterise pr + *'d* includes closure and alveolarity – they can occur together, but any one of them is sufficient for the identification of the pr + aux.

The essential phonetic features reported in Table 18 largely correspond to the phonological features of the auxiliaries described by Simpson (1992) and Ogden (1999). For example, in Simpson's terms, *are* is characterised by openness, *am* by nasality, *is* by sibilance, and *'d* by stoppedness. Moreover, both authors highlight the essential role of the labiovelarity – or rounding and backness (Simpson, 1992) – in *was* and *were*, as this is the feature that marks the past tense of the auxiliary BE. The main difference between the phonological accounts proposed by Simpson (1992) and Ogden (1999), and the observations reported here, is that the cliticised form *'ll* in the data analysed is characterised by velarity rather than laterality as in the phonological accounts. While Simpson (1992) and Ogden (1999) consider the primary articulation the most prominent feature of *'ll*, in the data analysed, the secondary articulation is the essential feature of all pr + *'ll*. As mentioned above, whether the primary articulation is realised or not, the velarity, or dark resonance, is always present and affects the entire piece.

The variability in the realisations of the pr + aux is also related to the paradigmatic system of contrast they belong to. In the case of pr + *'d*, depending on the form of the main verb, pr + *'d* contrasts with pr + *'ve* or pr + *'s*. Both pr + *'ve* or pr + *'s* are characterised by friction, which would contrast nicely with the stoppedness of pr + *'d*. However, the presence in the data of instances of pr + *'d* realised without closure, but with friction, confounds this hypothesis. In order to investigate this issue further, the acoustic features of instances of pr + *'d* in which /d/ is realised with friction were analysed and compared to the acoustic features of pr + *'s* in the dataset. The results of this analysis are reported in the next chapter (Section 5.2.4.1).

The reduced instances analysed in this chapter exhibit the same types and variations in reduction described in Chapter 3: variation in the magnitude of gestures, the temporal reorganisation of phonetic events, and the temporal variability of voicing. The main source of variation and reduction seems to be the phasing of the phonetics events that occur in the supralaryngeal tract and the larynx.

Some of the variation observed in the data raises the question of how the contrast between paradigms is maintained in reduced speech. For example, in reduced instances of *we'll*, and *we'd*, produced without an apparent contoid in coda, what acoustic features distinguish the two paradigms? The auditory analysis reported in this chapter highlighted the presence of phonetic features such as resonances that characterise each piece. The next chapter (Chapter 5) attempts to answer this question by comparing the acoustic properties that remain in the signal when pairs of paradigms are reduced.

5. Contrast in the paradigmatic system of pronouns and auxiliaries

From the auditory analysis of pr + aux reported in Chapters 3 and 4, it emerged that pr + aux can be highly reduced, so that pr + aux combinations such as *we'll* and *we'd* can both be produced as a very short vocoid, and that the only apparent difference between highly reduced pairs of paradigms is their resonances. In this case, *we'll* is characterised by dark resonances, while *we'd* is characterised by clear resonances. This feature raises the question of which acoustic features maintain the contrast between pairs of paradigms conveying grammatical information in reduced speech.

This chapter expands on the findings of the previous two chapters by reporting on a quantitative analysis of the acoustic features that maintain the contrast between paradigms. The pairs of paradigms that are compared in this chapter were analysed along a range of acoustic parameters, including duration, amplitude, spectral moments, and formant dynamics. The statistical analysis highlights the significant differences between the acoustic qualities of the paradigms analysed.

5.1. Background

The pr + aux analysed in this thesis are in paradigmatic contrast. They belong to a small system of contrasting items in which each pr + aux contrasts with only a small number of other pr + aux. By 'contrast' here it is meant that the difference between two items conveys a difference in meaning. It also means that each item in the system can occur in the same position in the utterance as all the other items in the same system. In a sentence like *she's burnt the toast*, *she's* can be replaced by any other pr + aux in which the auxiliary is a contracted form of the verb HAVE, for example *she'd* (past tense), *he's*, *you've*, *I've*, etc.

The difference between two items, e.g. *she's* and *she'd*, conveys a difference in grammatical meaning – in this case, present or past tense. Each of the items in the system of pr + aux provides grammatical information about the person and tense of the verb.

Traditional phonemic approaches base the inventory of a language exactly on this: contrasting sounds that when replaced by another sound in the same position bring about a change in meaning. In the example mentioned above, the symbols /z/ and /d/ represent phonemes because changing one for the other in *she's* and *she'd* changes the meaning. The question arises of how the contrast is maintained in highly reduced speech. For example, in Section 4.1.1, it was reported that /d/ in *she'd* (and all other pr + *'d*) can be realised with friction instead of a complete closure. This means that both paradigms *she's* and *she'd* can have voiced alveolar friction in coda position. If this is the case, how is the contrast between the two paradigms maintained in reduced speech? While a phonemic approach tends to consider reduced speech as loss of information (Niebuhr, 2016), there is little evidence that this is true. This chapter aims at investigating how the contrast is maintained – if it is maintained at all – in reduced speech, by comparing the acoustic features of pairs of paradigms.

5.1.1. Phonetic realisation of contrast in a small paradigmatic system

The fact that the paradigmatic system of pr + aux is a small system (it contains only a few items relative to other systems such as that of content words), means that the paradigms can exhibit a wider range of phonetic variation than paradigms in bigger systems. A typical example that explains this feature is the difference between the assimilation of coda /m/ in *I'm* (function word) and that in *lime* (content word) reported by Local (2003). The nasal in *I'm* can undergo assimilation to a following alveolar or velar sound because *I'm* belongs to a small paradigmatic system and contrasts only with other combinations of a pronoun and a contracted auxiliary. Such combinations are, for example, *he's*, *she's*, *we're*, *you're*, etc. None of the paradigms that can occur in the same position in the utterance contains a nasal sound. This means that the place of articulation of the nasal can vary without hindering word identification. This is not the case for a word such as *lime*, which contrasts with *line* (Local, 2003).

Two aspects of the paradigmatic system of pr + aux are important here. Firstly, the contrasting paradigms provide grammatical information that must be retained even in reduced speech for successful lexical access. Secondly, as just mentioned, because the paradigmatic system of pr + aux is small, the items in the system can be highly reduced. The question arises as to how these two aspects – high degree of reduction, and need to transmit grammatical information – are balanced. Two hypotheses can be made. The first hypothesis is that the information is conveyed by the context. Ernestus (2014) and colleagues (Ernestus et al., 2002) claimed that high degrees of reduction need their context to be correctly interpreted. The second hypothesis is that the fine phonetic detail that is crucial for the correct identification of words is present in the signal even in highly reduced speech and is available to perception. The type of phonetic detail that remains in the acoustic signal is the one described in Chapter 1 (Section 1.3) – systematic, distributed, long-domain features such as resonances (or articulatory prosodies) that are not tied to segmental units. This fine phonetic detail constitutes the identity of the piece in reduced speech and, for this reason, must always be present for the piece to be correctly identified. The aim of this chapter is to identify the acoustic features that maintain the contrast in pairs of contrasting paradigms. Later on in the thesis, Chapters 7 and 8 investigate the perceptual salience of these acoustic features and their role in word identification through two perception experiments. Crucially, the minimal context that surrounds the pr + aux in the present data can shed light on the role of the acoustic cues and the role of the context in the correct interpretation of reduced pr + aux.

5.1.2. Present versus past tense of auxiliaries

The contrast analysed in this chapter is that of present versus past tense of English auxiliaries (except Section 5.2.5.1 on *he'd* versus *you'd*). The pairs analysed are *we'll* versus *we'd*, *I'll* versus *I'd*, and *she'll* versus *she'd* for the contrast between *will* and *would*; *she's* versus *she was*, *you're* versus *you were*, and *we're* versus *we were* for the contrast between the present and past tense of BE; and *he's* versus *he'd* for the contrast between the present and past tense of HAVE.

The rationale for analysing the contrast between present and past tense, is two-fold. Firstly, the present and past tense forms of the auxiliary are followed by the same finite form of the main verb, (e.g. *she's* <u>*burnt*</u> vs *she'd* <u>*burnt*</u>). This means that the disambiguation of a sentence must occur at the level of the auxiliary and not at the level of the main verb. Therefore, the auxiliary, even in its most reduced forms, must convey the grammatical information of the tense of the verb. This aspect fits perfectly with the aim of this study to investigate the phonetic features that remain in the signal in reduced speech and that convey grammatical information. Secondly, as described by Simpson (1992) and Ogden (1999), the phonology of the present and past tense of English auxiliaries is characterised by systematic patterns. The various auxiliaries have different strategies to express the grammatical relationship between present and past tense. The distinct behaviours of the auxiliaries can be explained by using a polysystemic approach in which each auxiliary (or group of auxiliaries in the case of modal auxiliaries) belongs to a separate system. The present-past relationship in three systems is reported here: present and past tense of HAVE, present and past tense of BE, and *will* and *would*.

Starting with the modal auxiliaries, only *will* and *would* are covered in this thesis. Following Ogden (1999: 72) *will* and *would* can be treated as present and past tense "on the basis of their behaviour in reported speech:

> I will do the cleaning
> I said I would do the cleaning"

Like other modal auxiliaries such as *shall* and *can*, the past tense is expressed by the stoppedness in coda position. This is true also for the auxiliaries HAVE and DO.

present	past
will	woul<u>d</u>
can	coul<u>d</u>
shall	shoul<u>d</u>
have, has	ha<u>d</u>
do, does	di<u>d</u>

Table 19. Present and past tense of auxiliaries whose past form is marked by stoppedness.

Table 19 shows the present and past forms of the auxiliaries which have a stop in coda position in the past tense form.

Moreover, the modal auxiliaries *will*, *shall*, and *can*, share the opposition of lip-rounding: the present tense is characterised by non-rounding (*will*, *shall*, *can*), the past tense is characterised by lip-rounding (*would*, *should*, *could*). Rounding and non-rounding, which can be referred to also as labiality, are resonances that characterise the entire form. According to this account, *will* and *would* contrast on two parameters: resonances (lip-rounding/labiality) and coda stoppedness. This is true for the syllabic forms of *will* and *would*, but not for the non-syllabic ones.

The system of finite forms of HAVE do not exhibit the opposition between lip-rounding and non-rounding, but only the stoppedness in coda position (*has*/*have* versus *had*).

The patterns found in the system of finite forms of BE are slightly more complex (Simpson, 1992). The present versus past tense relationship is not conveyed by the coda stoppedness. The present tense is characterised by an empty onset, while the past tense is characterised by labiovelarity in onset, which is mandatory (Ogden, 1999). The labiovelarity of *was* and *were* is different from the labiovelarity of *will* and *would*, as demonstrated by the fact that *will* and *would* can be contracted (*'ll* and *'d*) losing the labiovelarity. The forms *was* and *were* cannot be contracted because the labiovelarity conveys the past tense in the system of BE. In the system of modal auxiliaries, the present and past tense of an auxiliary share the same onset (*will* ~ *would*, *shall* ~ *should*, *can* ~ *could*), and the present-past relationship is marked by the coda stoppedness (*will* ~ *would*, *shall* ~ *should*, *can* ~ *could*); in contrast, the present and past tense of BE share the coda (*is* ~ *was*, *are* ~ *were*), and the present-past relationship is conveyed by the labiovelarity in onset (*is* ~ *was*, *are* ~ *were*) (Ogden, 1999). Therefore, the past tense forms of BE retain the labiovelarity, as it marks the past tense. However, Ogden (1999: 73) points out that "non-syllabic forms of *was* can frequently be observed in normal speech. When such forms occur, they contain front rounded vowels:

ø̰z 'it was'
ʃyz 'she was'
aʉz 'I was'"

To summarise, the phonological accounts of the present-past relationship formulated by Simpson (1992) and Ogden (1999) highlight the unique ways in which auxiliaries belonging to distinct systems mark the present versus past contrast. In the present study, the present-past relationship of three systems of auxiliaries is analysed: present versus past forms of HAVE, present versus past forms of BE, and *will* versus *would*. From a phonological point of view, the finite forms of the auxiliary HAVE mark the past tense with coda stoppedness only. The finite forms of the auxiliary BE mark the past tense with labiovelarity. The pair *will* ~ *would* mark the past tense with both stoppedness and labiality in their syllabic forms. The present chapter reports on a quantitative analysis of the actual phonetic realisations of these three contrasts and the acoustic features that they retain in reduced speech.

In addition, the contrast between the pronouns *he* and *you* is investigated further. Several instances of *you* in the dataset (70%) exhibit a portion of voiceless friction with palatal quality before the onset of voicing. This feature raised the question of how the contrast with the pronoun *he* is maintained. Section 5.2.5.1 describes and compares the acoustic features of the two paradigms *he'd* and *you'd*.

5.2. Acoustic analysis of contrasting paradigms

The following sections describe the acoustic phonetic features of contrasting pairs of paradigms. Section 5.2.1 describes some analysis procedures including the statistical tests used in this chapter. Section 5.2.2 reports the analysis of combinations of pronouns with the cliticised forms of the auxiliaries *will* and *would*. Section 5.2.3 reports the analysis of combinations of pronouns with the cliticised forms of the auxiliary BE in the present and past forms: first *is* and *was*, then *are* and *were*. Section 5.2.4 reports the analysis of the present and past tense of the auxiliary HAVE in the pair *he's* and *he'd*. Finally, Section 5.2.5 reports the analysis of the contrast between the pronouns *he* and *you*.

5.2.1. Analysis procedures

For each contrast discussed in this chapter, mixed-effects models were run for each of several dependent variables. For the statistical analysis of duration, six mixed-effects models were run with the following dependent variables: *Silent_articulation* or *Friction1, Vocoid*, *Voicing*, *Closure*, *Voicing_in_closure* or *Friction2*, and *Piece.* For the statistical analysis of the spectral properties of fricatives, five mixed-effects models were run with the following dependent variables: *Amplitude*, *CoG*, *SD*, *Skewness*, and *Kurtosis*. The factor *Auxiliary* was included in each model as a fixed effect with two levels (*will~would*, *is~was*, *are~were*, *has~had*). For the contrast between *you* and *he* (Section 5.2.5.1) the factor *Pronoun* was included as a fixed effect with two levels (*you~he*) instead of *Auxiliary*. A random intercept for *Speaker* was also included in each model. P values were generated using likelihood ratio tests that compared the model with *Auxiliary* (or *Pronoun*) as a fixed effect against the null model without it. P values were adjusted using Bonferroni corrections throughout. The R model syntax and outputs can be found in Appendix C.

In some cases, to verify that two paradigms (e.g. *we'd* and *we'll*) retain their distinct resonances also in highly reduced speech, the temporally reduced tokens in each dataset were selected for comparison. For each paradigm, the instances in which the duration of the vocoid was shorter than the mean duration of the vocoid calculated across speakers and repetitions in each dataset were selected. Table 20 reports the mean vocoid duration calculated across speakers and repetitions in each dataset and the number of tokens with shorter vocoid duration that were selected for the comparison of their formant dynamics with the formant dynamics of all the tokens of the same paradigm.

Dataset	Mean vocoid duration in the dataset	N of tokens with a shorter vocoid duration
we'd	49 ms	23/46, 50%
we'll	46 ms	23/49, 47%
I'd	42 ms	25/39, 64%
I'll	53 ms	22/39, 56%
she's	36 ms	20/50, 40%
she was	86 ms	25/47, 53%

Table 20. Mean vocoid duration of all the tokens in each dataset, and number and percentage of tokens with a shorter vocoid duration than the mean vocoid duration in that dataset.

5.2.2. Contrast between *will* and *would*

As mentioned above, the present versus past contrast in the pair *will* and *would* is conveyed by the opposition of lip shape and the coda stoppedness in the past tense form. However, in the cliticised forms *'ll* and *'d*, the lip shape associated with *will* and *would* is lost. As reported in Section 4.1.1, being followed by another stop, the most frequent realisation of /d/ in *would* is unreleased, which means that the stoppedness in coda position of *would* is lost too. Moreover, as reported in Section 4.1.5, the most frequent realisation of *'ll* in the data is as dark resonances only without tongue tip contact. The combination of these features means that apparently both pr + *'ll* and pr + *'d* paradigms are realised without an actual contoid in coda position. This raises the question of how the grammatical information about the tense of the auxiliaries *will* and *would* is phonetically realised. The following sections describe the acoustic features of *we'd* and *we'll* (Section 5.2.2.1), *I'll* and *I'd* (Section 5.2.2.2), and *she'll* and *she'd* (Section 5.2.2.3).

5.2.2.1. *We'll* versus *we'd*

This section compares the acoustic properties of the paradigms *we'd* and *we'll*. In their phonological representation, the paradigms *we'd* and *we'll* differ in the contoid in coda position. As mentioned above, the stoppedness in coda of *would* also marks the past tense. In the data collected, in the large majority of instances of *we'd* and *we'll* there is no apparent contoid in coda position. Both pieces are realised as a short vocoid – in *we'd* the plosion is lost, while in *we'll* the laterality is lost. In *we'd*, the gesture for the alveolar closure of the stop can be articulated, but in most instances (N = 82/100, 82%) the release is masked by the lip closure for the bilabial stop that follows. In most instances of *we'll* (N = 50/53, 94%), the lateral approximant is realised as a back rounded vocoid or as dark resonances throughout the piece – the tip of the tongue does not make contact with the alveolar ridge (the instances of *we'll* in which the lateral can be identified are N = 3/53, 6%). However, the secondary articulation – the movement of the tongue dorsum towards the velum typical of dark-L in coda position in English – is articulated even when the primary articulation is not. Moreover, the opposition of lip shape in the vocoids of *will* (non-rounding) and *would* (rounding) is lost in the cliticised forms.

Figure 125 shows an instance of *we'd* on the left, and one of *we'll* on the right, produced by the same speaker (S6). In both instances, the pr + aux combination is realised as a short vocoid. In both instances, there is no noticeable formant movement during the vocoid, although there is a small increase in the frequency of F2 in the voiceless portion of friction at the beginning of *we'd* (on the left). The spike in the hold phase of *we'd burn* is likely to be the closure of the lips for the bilabial plosive in *burn*.

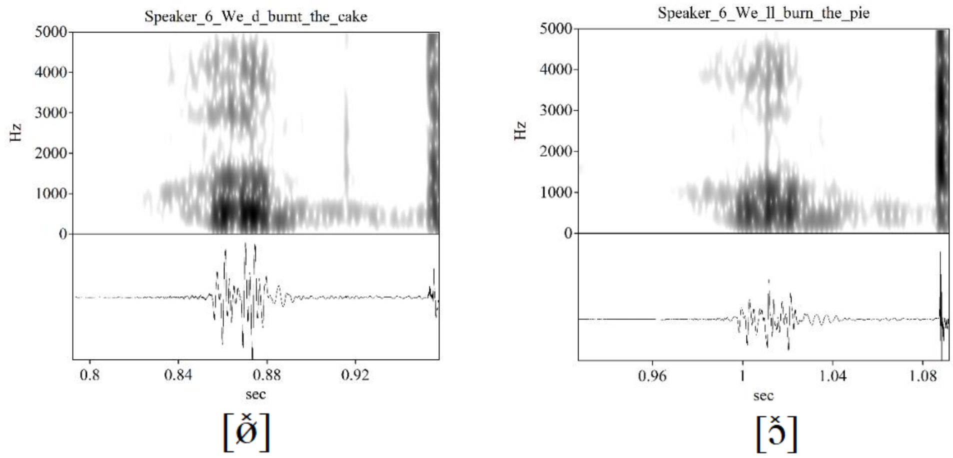


Figure 125. Spectrogram and waveform of an instance of *we'd* on the left, and an instance of *we'll* on the right, produced by the same speaker (S6).

The analysis of the two paradigms included the duration of various phonetic events. The results of the six mixed effects models reveal a main effect of *Auxiliary* for two dependent variables, *Closure* ($\chi^2(1) = 68.02$, $p < 0.001$), and *Piece* ($\chi^2(1) = 41.97$, $p < 0.001$). *Auxiliary* did not have a significant effect on the duration of the other intervals measured (that is, in each case, the p-value was > 0.001).

Figure 126 shows the mean durations of the five parameters tested in *we'd* and *we'll*, calculated across speakers and repetitions (*we'd* N = 93; *we'll* N = 49).

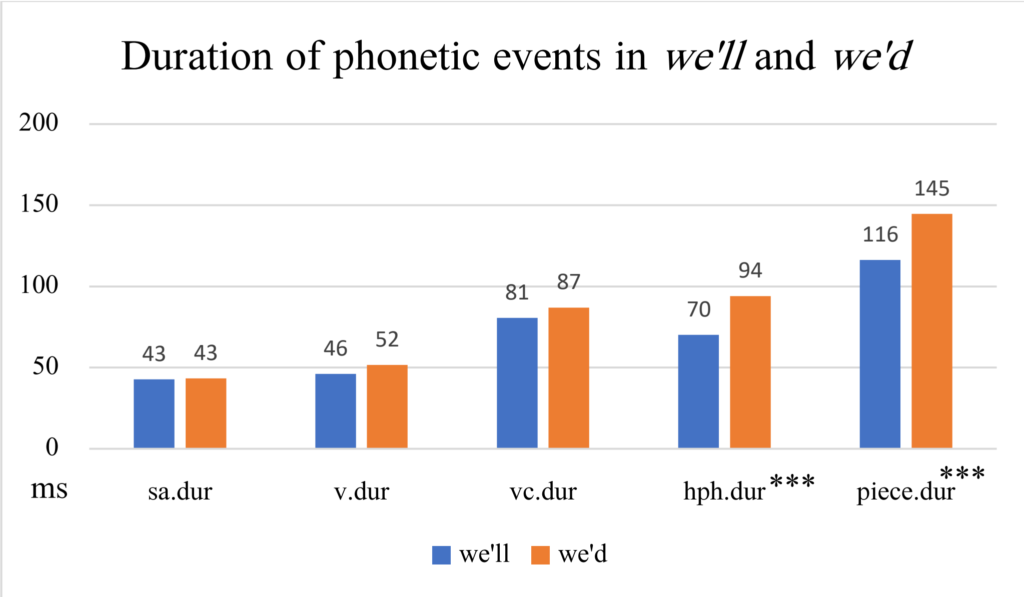


Figure 126. Mean duration (in ms) across speakers and repetitions of the silent articulation (sa), the vocoid (v), voicing (vc), the closure (hph), and the piece (piece) in *we'll* (blue) and *we'd* (orange). *** indicates that the duration of the two paradigms is significantly different.

The longer duration of the closure in *we'd* is due to the instances of *we'd* in which /d/ is articulated but not released (N = 42/47, 89%). In these instances, the closures of /d/ and /b/ merge in a single long closure.

The other parameter analysed is the formant dynamics, which are the acoustic correlates of the resonances. To compare the resonances, the formant dynamics have been measured at nine equidistant points in time from the onset to the offset of the vocoid in all instances of *we'd* and *we'll*. Figure 127 shows the mean frequencies of the first three formants calculated across speakers and repetitions (*we'd* N = 93, *we'll* N = 49).

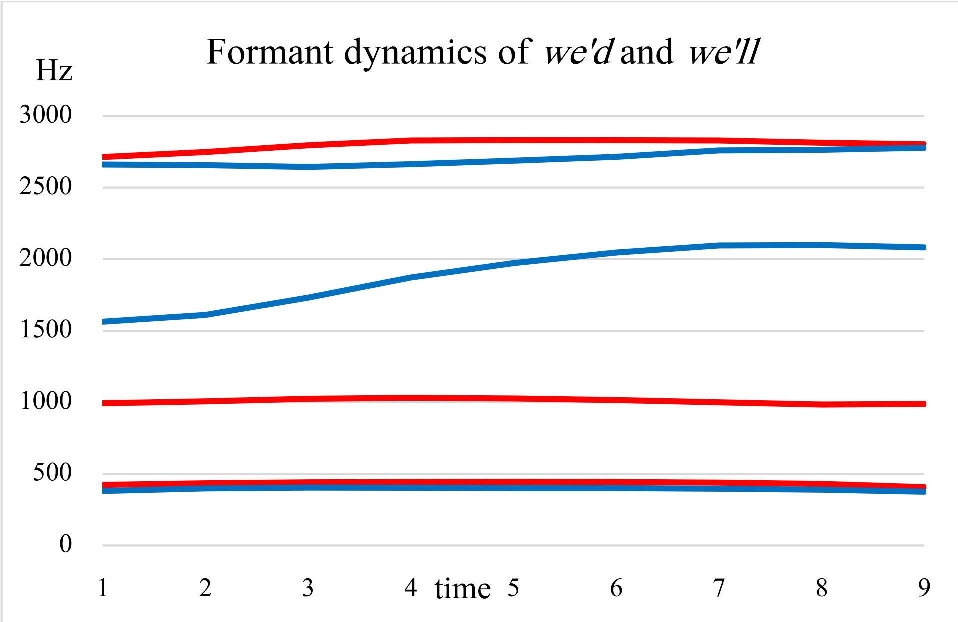


Figure 127. Formant dynamics of the vocoids in *we'll* (red) and *we'd* (blue) calculated across speakers and repetitions.

As it can be seen in Figure 127, the main difference is in the frequency of F2. A low F2 is typical of dark-L and back vocoids, which are articulated with the tongue dorsum further back (or raised) in the oral cavity. A high F2 suggests a front vocoid or palatal articulation, and clear resonances, which are articulated with the front of the tongue in a more advanced or raised position in the oral cavity. The frequencies of F2 in *we'd* and *we'll* suggest that *we'll* is characterised by dark resonances, while *we'd* is characterised by clear resonances. Note that the frequencies of all three formants differ from the start of the vocoids, indicating that the whole piece is characterised by either dark or clear resonances.

To verify that the two paradigms retain their distinct resonances also in highly reduced speech, the temporally reduced tokens of *we'd* and *we'll* were selected for comparison following the procedure explained in Section 5.2.1. Figure 128 shows the mean formant dynamics of the two subset of *we'd* and *we'll* with shorter vocoid durations.

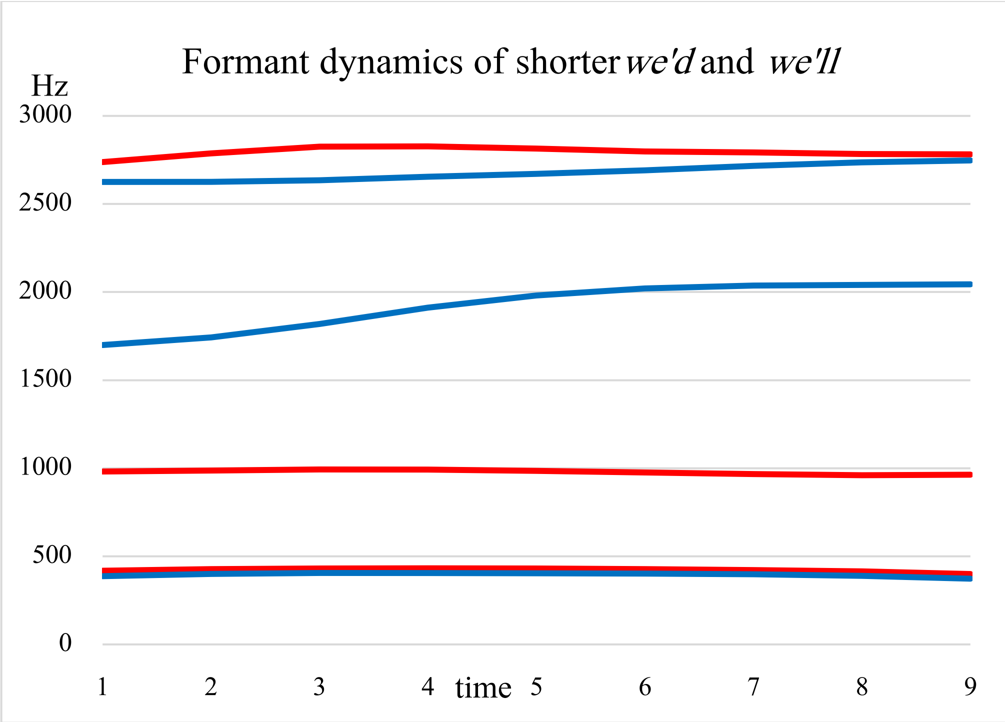


Figure 128. Formant dynamics of the vocoids in shorter instances of *we'll* (red) and *we'd* (blue).

In the two subsets of *we'd* and *we'll* with a shorter duration, the formant dynamics indicate that the resonances of the two paradigms are clearly distinct also in reduced tokens. The trajectory of F2 in Figure 127 and Figure 128 is very similar, although F2 in the reduced subset is slightly flatter. Figure 129 shows the formant dynamics of the vocoid in *we'd* and *we'll* in the entire datasets on the left, and in the two reduced subsets of *we'd* and *we'll* on the right for comparison.

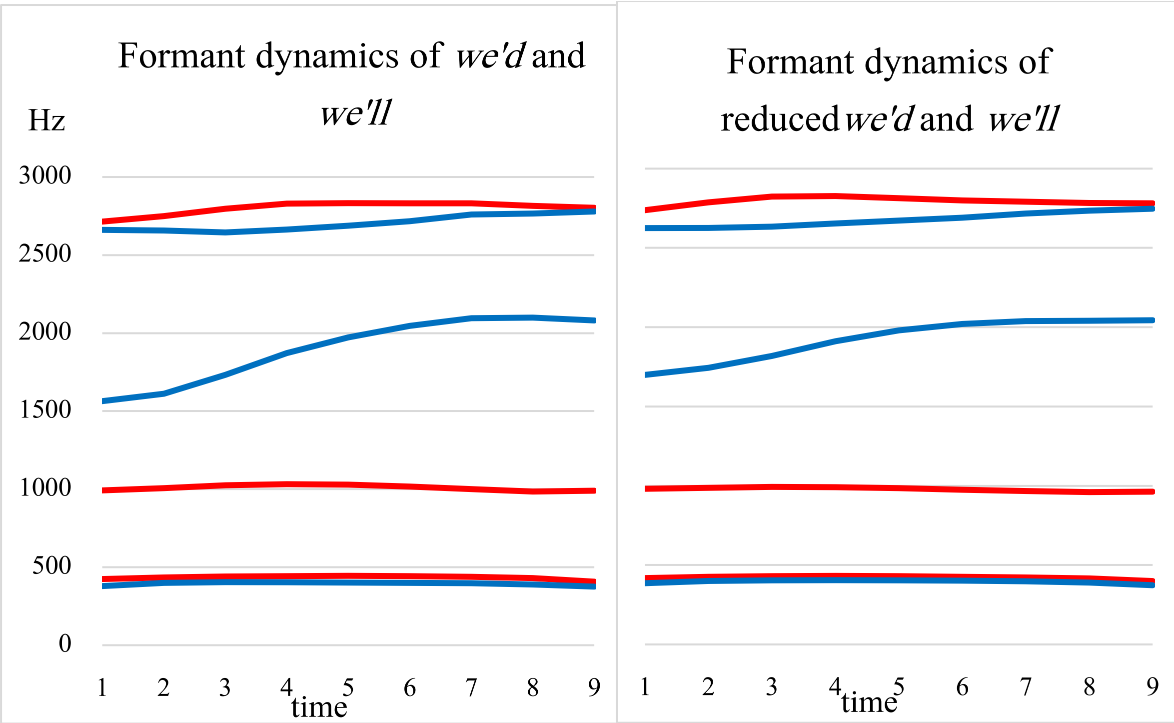


Figure 129. Formant dynamics of the vocoid in the entire *we'd* (blue) and *we'll* (red) datasets on the left, and the formant dynamics of the vocoid in the two reduced subsets of *we'd* and *we'll* on the right.

Three aspects emerge from the comparison between the formant dynamics of the whole datasets of *we'd* and *we'll* and the reduced subsets of *we'd* and *we'll*. Firstly, F2 in the reduced *we'd* subset is slightly flatter than F2 in the whole *we'd* dataset. While in the whole dataset F2 in *we'd* starts at 1564 Hz and ends at 2082 Hz, in the reduced subset F2 starts at 1699 Hz and ends at 2044 Hz. The frequency range is 518 Hz in the whole dataset, and 345 Hz in the reduced subset. This indicates that the magnitude of gesture is indeed reduced in reduced instances of *we'd*. Secondly, as mentioned above, the F2 difference between both the two datasets and the two subsets is visible from the beginning of the paradigms. This suggests that the clear and dark resonances are a feature of the entire *we'd* and *we'll* pieces rather than features of the contoid in coda position. The low frequencies of F2 in *we'll* indicate that the entire vocoid, and therefore the entire piece, is characterised by dark resonances. This contrasts with the higher F2 frequencies in *we'd*, which indicate clear resonances throughout the piece from the beginning.

Thirdly, comparing the whole datasets, the difference between the frequencies of F2 at the beginning of *we'd* and *we'll* is 571 Hz. However, comparing the reduced subsets, the difference between the frequencies of F2 at the beginning of *we'd* and *we'll* is 718 Hz. This suggests that the contrast between the two paradigms is even enhanced in the reduced tokens. This feature is apparent also in the frequencies of F3. Comparing the frequencies of F3 at the beginning of *we'd* and *we'll* in the whole datasets, the difference is 52 Hz. The difference between the frequencies of F3 at the beginning of *we'd* and *we'll* in the reduced subsets is 113 Hz. This feature is not apparent in the frequencies of F1. The difference between the frequencies of F1 at the beginning of *we'd* and *we'll* in the entire datasets is 44 Hz. The difference between the frequencies of F1 at the beginning of *we'd* and *we'll* in the reduced subsets is 32 Hz.[11]

To summarise, this section has shown that the paradigms *we'd* and *we'll* differ in two main parameters: the duration of the closure and the resonances of the piece. As for the duration, it has to be said that the closure in *we'll* belongs to the following sound, rather than to the pr + aux itself. But in this context, the long closure of *we'd* is a feature of the piece that distinguishes it from the other paradigms such as *we'll* and *we've*. The constant phonological context of a following plosive allowed the analysis of the closure in pr + *'d* which would not otherwise have been possible. As for the resonances, the formant dynamics of *we'd* and *we'll* indicate that the two paradigms differ in their resonances: *we'd* is

[11] The rationale for analysing the formant frequencies at the beginning of the vocoid rather than at the end of the vocoid is that at the end of the vocoid the formant frequencies might tend to converge towards the locus of the sound that follows (Sussman, McCaffrey and Matthews, 1991).

characterised by clear resonances and *we'll* is characterised by dark resonances. This means that the resonances of the cliticised forms of the auxiliaries *will* (pr + *'ll* dark) and *would* (pr + *'d* clear) are the opposite of the resonances of the syllabic forms in which *will* is clear and *would* is dark due to the opposition of lip-rounding of the vocoid. This is one more piece of evidence that the various forms of the auxiliaries belong to different systems and that a polysystemic approach is needed for their analysis. Finally, from the analysis of a subset of data characterised by a short vocoid, it emerged that despite the decreased magnitude of gesture (as observed in a flatter F2 in the reduced subset of *we'd*), the contrast between the resonances of *we'd* and *we'll* is enhanced rather than weakened, as suggested by the larger difference in the frequencies of F2 at the beginning of *we'd* and *we'll* in the reduced subsets compared to the whole datasets.

5.2.2.2. *I'll* versus *I'd*

The acoustic features of the paradigms *I'll* and *I'd* were compared using the same methodology described above. The results of the six mixed-effects models reveal a main effect of *Auxiliary* on *Closure* ($\chi^2(1) = 42.64$, $p < 0.001$), *Vocoid* ($\chi^2(1) = 8.82$, $p = 0.003$), and *Voicing* ($\chi^2(1) = 8.48$, $p = 0.004$). The other domains did not show a significant difference in duration for different auxiliaries (all have p-value > 0.001). Figure 130 shows the mean durations of the five parameter tested in *I'd* and *I'll* calculated across speakers and repetitions (*I'd* N = 39; *I'll* N = 39).[12]

[12] The low number of tokens in these two datasets is due to the fact that three speakers placed a stress on the pr + aux during the production study. If a speaker production of *I'll* had to be discarded, the production of *I'd* of the same speaker was discarded too.

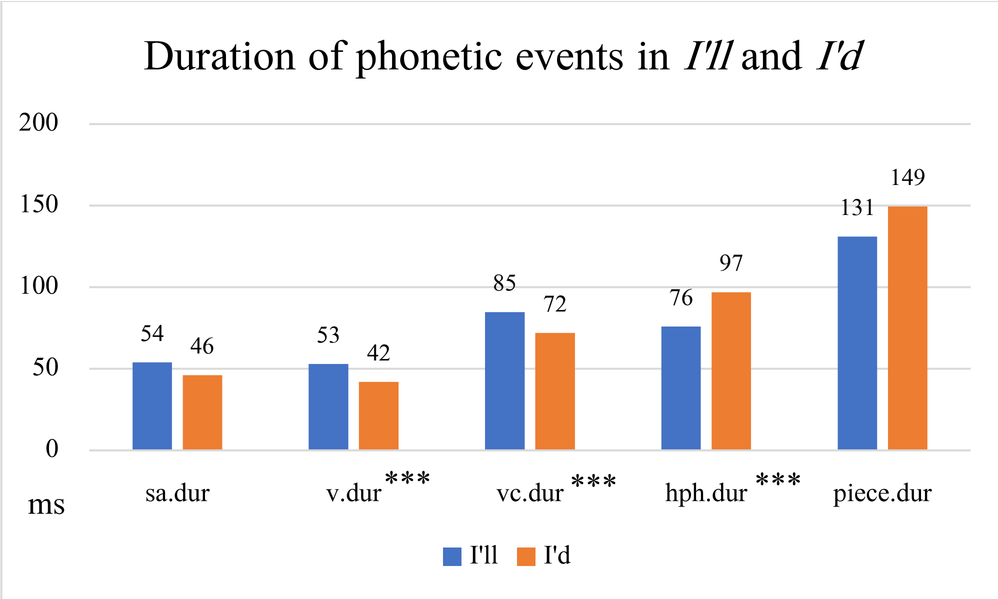


Figure 130. Mean duration (in ms) across speakers and repetitions of the silent articulation (sa), the vocoid (v), voicing (vc), the closure (hph), and the piece in *I'll* (blue) and *I'd* (orange). *** indicates that the duration of the two paradigms is significantly different.

The paradigms *I'd* and *I'll* contrast on the duration of the vocoid and the duration of the closure, as well as the duration of voicing, which is strictly connected to the duration of the vocoid. Compared to *we'll* and *we'd*, one additional parameter differs between the two paradigms: the duration of the vocoid. Figure 131 shows the distribution of the duration of the vocoid in *I'll* and *I'd*.

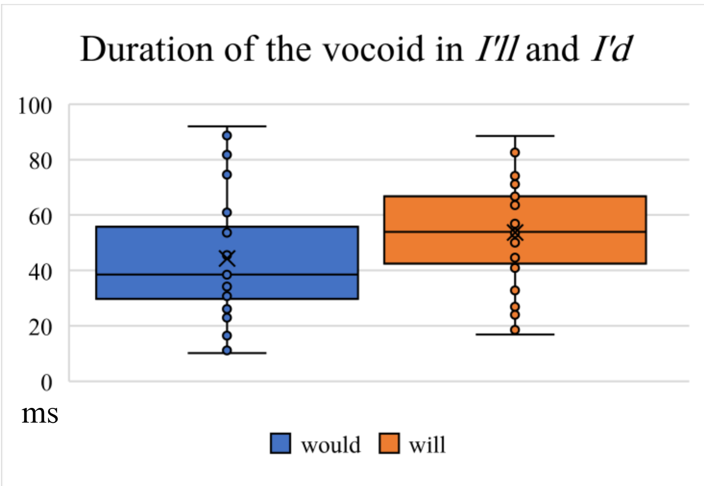


Figure 131. Distribution of the duration of the vocoid in *I'd* (blue) and *I'll* (orange).

As for the formants, Figure 132 shows the mean frequencies of the formant dynamics of the first three formants calculated across speakers and repetitions at nine equidistant points in time.

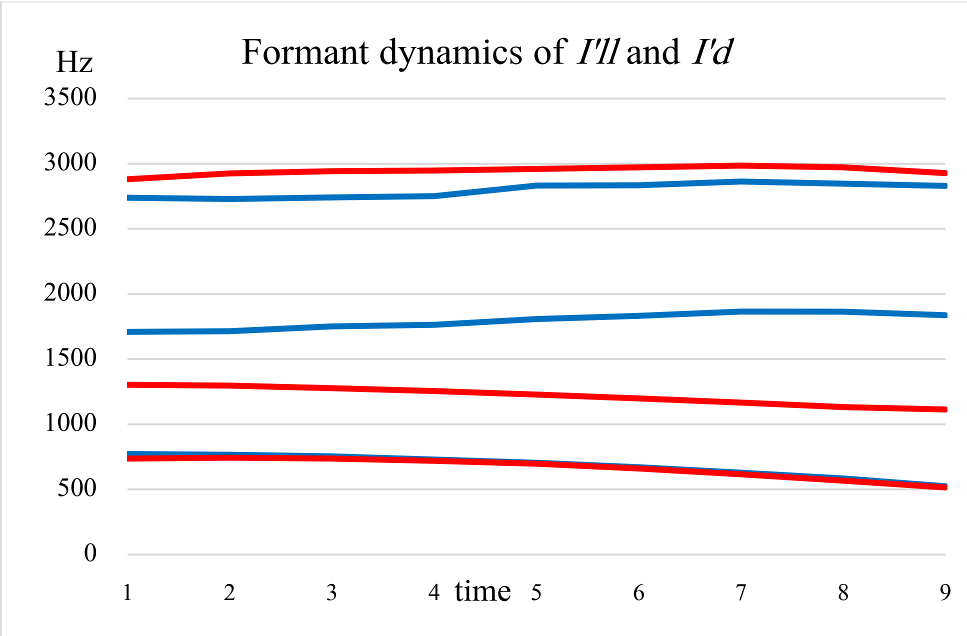


Figure 132. Formant dynamics of *I'll* (red) and *I'd* (blue).

Figure 132 shows the formant dynamics of the vocoid in *I'd* and *I'll*. F2 exhibits the largest difference between the two paradigms, with F2 in *I'll* being lower (406 Hz at onset and 730 Hz at Time point 8 where the largest difference is). A lower F2 in *I'll* confirms the dark resonances of the vocoid due to velarisation. The velarisation is the secondary articulation of the lateral in coda in *I'll*, which is realised without laterality, but maintains the velarity. F3 also differs between the two paradigms. However, it does not exhibit the predicted pattern of lower F3 in the dark paradigm.

To verify that *I'd* and *I'll* retain their distinct resonances in temporally reduced instances, a subset of each set of data was selected for comparison following the procedure explained in Section 5.2.1. Figure 133 shows the mean formant dynamics of the two subsets of *I'd* and *I'll* with shorter vocoid durations.

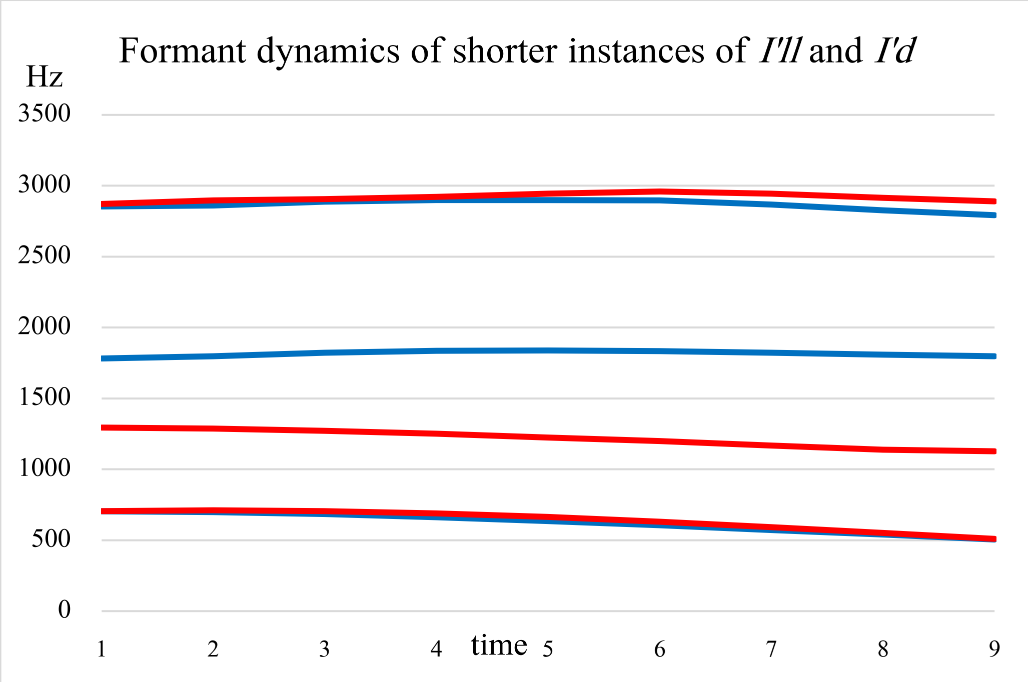


Figure 133. Formant dynamics of the vocoids in shorter instances of *I'll* (red) and *I'd* (blue).

In the two subsets of *I'd* and *I'll* with a shorter vocoid duration (Figure 133), F2 is clearly distinct and exhibits the same pattern observed in the entire datasets: F2 is lower in *I'll* than in *I'd*. This suggests that the two paradigms maintain their resonances in reduced instances. F2 is also flatter in the two shorter subsets, indicating a decrease in the magnitude of gestures. The difference in the frequencies of the first three formants between the entire datasets and the two subsets were compared. Table 21 shows the mean frequencies of the first three formants in the vocoid of *I'd* and *I'll* in the entire datasets, and the difference between each formant in the two paradigms calculated by subtracting the respective frequencies, e.g. F1 in *I'd* – F1 in *I'll*, F2 in *I'd* – F2 in *I'll* , and F3 in *I'd* – F3 in *I'll*. Table 22 shows the mean frequencies in the vocoid in *I'd* and *I'll* in the two subsets of instances with a shorter vocoid duration, and the difference in the frequencies of the first three formants between the two subsets.

Hz	1	2	3	4	5	6	7	8	9
F1 *I'd*	772	767	754	730	705	670	629	584	525
F2 *I'd*	1710	1716	1750	1764	1807	1833	1865	1864	1836
F3 *I'd*	2738	2729	2742	2752	2833	2835	2864	2848	2830
F1 *I'll*	738	744	737	720	697	661	618	568	516
F2 *I'll*	1304	1296	1278	1256	1228	1200	1167	1131	1114
F3 *I'll*	2881	2926	2943	2948	2961	2972	2986	2973	2929
F1 *I'd* – F1 *I'll*	34	23	17	10	7	8	11	17	9
F2 *I'd* – F2 *I'll*	406	419	472	508	579	633	698	733	722
F3 *I'd* – F3 *I'll*	-143	-196	-202	-196	-128	-137	-122	-125	-99

Table 21. Mean frequency in Hz of the first three formants of the vocoid in *I'd* (N = 39) and *I'll* (N = 39) calculated in the entire datasets across speakers and repetitions at nine equidistant points in time. The difference between the frequencies of F2 in the two pr + aux is in red.

Hz	1	2	3	4	5	6	7	8	9
F1 *I'd*	704	697	684	661	635	605	573	539	505
F2 *I'd*	1782	1797	1823	1836	1839	1834	1823	1808	1797
F3 *I'd*	2855	2860	2889	2898	2899	2896	2869	2827	2793
F1 *I'll*	705	711	705	688	663	630	592	551	510
F2 *I'll*	1295	1287	1272	1251	1225	1199	1169	1139	1127
F3 *I'll*	2872	2897	2906	2921	2945	2960	2944	2915	2890
F1 *I'd* – F1 *I'll*	-1	-14	-20	-27	-29	-25	-19	-12	-5
F2 *I'd* – F2 *I'll*	487	510	550	585	613	635	655	669	670
F3 *I'd* – F3 *I'll*	-17	-37	-17	-23	-46	-64	-75	-88	-97

Table 22. Mean frequency in Hz of the first three formants of the vocoid in two subsets of *I'd* (N = 25) and *I'll* (N = 22) with shorter vocoid durations. The mean frequencies were calculated at nine equidistant points in time. The difference between the frequencies of F2 in the two pr + aux is in red.

The comparison between the difference in the frequencies of F2 in the two shorter subsets and in the two entire datasets, indicates that the distance between F2 in the reduced subsets is larger than in the entire datasets. That is, the difference between the frequencies of F2 in

I'd and *I'll* at the beginning of the vocoid calculated in the entire dataset is 406 Hz. The difference between the frequencies of F2 in *I'd* and *I'll* at the beginning of the vocoid calculated in the reduced (shorter) subsets is 487 Hz. This means that in the reduced instances the difference between the mean frequency of F2 in *I'd* and *I'll* is 81 Hz larger than in the entire dataset. The difference becomes smaller in time and it is the same (633 Hz and 635 Hz) at Time point 6, after which the mean frequencies in the entire datasets diverge more than the mean frequencies in the shorter subsets. This comparison suggests that the contrast between the two paradigms is enhanced in the reduced tokens, at least in the first part of the vocoid. This pattern is not observed in F1 and F3. In fact, the difference between F3 in *I'd* and F3 in *I'll* is larger in the entire datasets than in the two reduced subsets.

5.2.2.3. *She'll* versus *she'd*

The same acoustic and statistical analyses were carried out for the contrasting pair *she'll* and *she'd*. The statistical analysis reveals a main effect of *Auxiliary* only on *Closure* ($\chi^2(1) = 61.63$, $p < 0.001$). The other domains did not show a significant difference in duration for different auxiliaries (all have p-value > 0.001). This means that *she'd* and *she'll* differ in closure duration, but not in vocoid duration. This result is similar to that of *we'd* and *we'll*, but not *I'd* and *I'll*, which are significantly different also in vocoid duration. Figure 134 shows the mean duration of the phonetic events in *she'll* (N = 49) and *she'd* (N = 45).

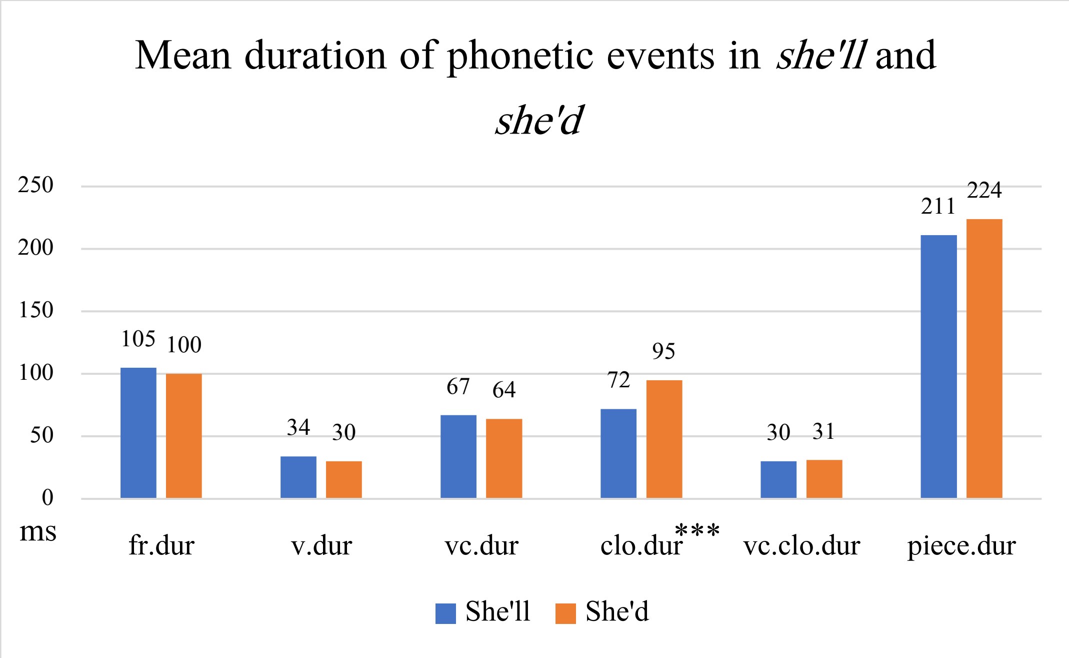


Figure 134. Mean duration in ms of the palato-alveolar friction (fr.dur), the vocoid (v.dur), voicing (vc.dur), the closure (clo.dur), voicing in closure (vc.clo.dur), and the entire piece (piece.dur) in *she'll* (blue) and *she'd* (red) calculated across speakers and repetitions.

The difference between the spectral properties of the palato-alveolar friction in onset of *she'll* and *she'd* were investigated by running five mixed effects models (see Section 5.2.1). The results reveal a main effect of *Auxiliary* on *CoG* ($\chi^2(1) = 17.81$, $p < 0.001$), *Skewness* ($\chi^2(1) = 7.05$, $p = 0.008$), and *Kurtosis* ($\chi^2(1) = 6.92$, $p = 0.008$). *Auxiliary* did not have a significant effect on *SD* and *Amplitude*. Table 23 shows the mean values of the four spectral moments of the palato-alveolar friction in *she'll* (N = 49) and *she'd* (N = 45).

	CoG (Hz)	SD (Hz)	Skewness	Kurtosis
She'll	3777	1266	1.89	8.24
She'd	4050	1364	1.55	5.70

Table 23. Mean CoG, SD, skewness and kurtosis in *she'll* and *she'd*, calculated across speakers and repetitions. In red the variables that exhibit a significant difference between the two paradigms.

The formant dynamics of the vocoid in *she'll* and *she'd* were compared too and are shown in Figure 135.

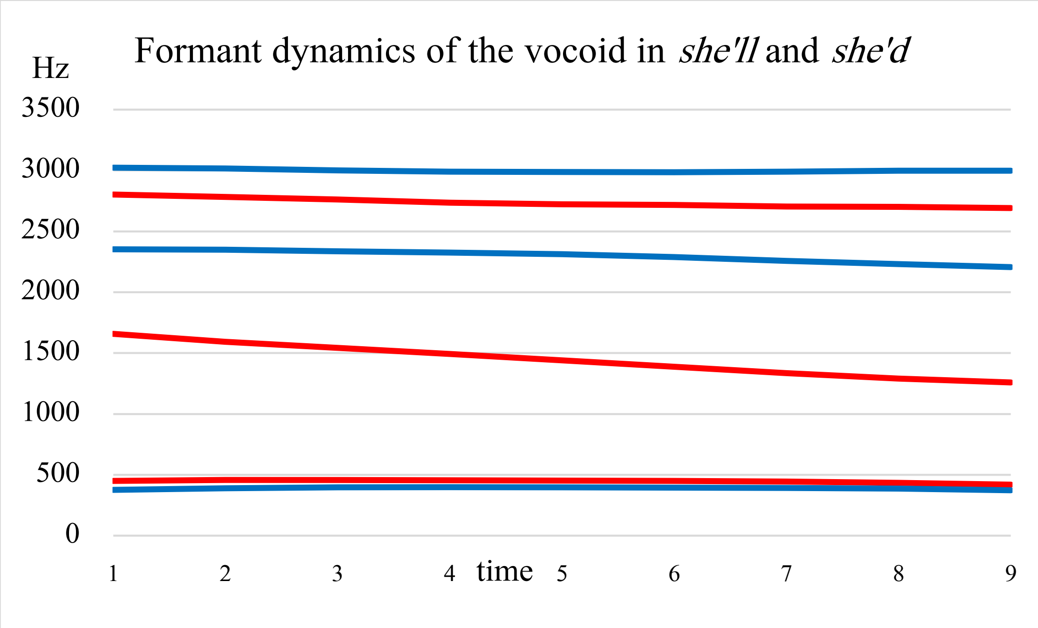


Figure 135. Formant dynamics of *she'll* (red) and *she'd* (blue).

Figure 135 shows the formant dynamics of the vocoid in *she'll* and *she'd*. F2 exhibits a similar pattern to that observed in the pairs *we'll* and *we'd*, and *I'll* and *I'd*. F2 is lower in *she'll* than *she'd*, confirming the dark quality of the vocoid in the pr + *'ll* paradigms compared to the pr + *'d* paradigms. F3 shows the same pattern observed in the pair *we'll* and *we'd* (but not *I'll* and *I'd*) in that it is lower in *she'll* than *she'd*. As for F1, the three pairs of paradigms exhibit different patterns.

5.2.2.4. Summary of results of *will* versus *would*

This chapter so far reported the acoustic analysis of the paradigms *we'd* and *we'll*, *I'd* and *I'll*, and *she'd* and *she'll*. The acoustic analysis of contrasting paradigms pr + *'d* and pr + *'ll* reveals that the duration of the closure is significantly different in all three pairs. In addition, the formant dynamics in all three pairs exhibit a similar pattern for F2, which is lower in pr + *'ll* than pr + *'d* paradigms. This suggests that the vocoid in pr + *'ll* is affected by the velarisation from the phonological lateral in coda position, regardless of the actual realisation of the lateral itself. The low F2 also suggests that the resonances of pr + *'ll* paradigms are dark, contrasting with the clear resonances of pr + *'d* paradigms, which are characterised by a higher F2.

In addition to the resonances and the duration of the closure, the paradigms *I'd* and *I'll* differ significantly also in the duration of the vocoid. The vocoid in *I'll* exhibits significantly longer duration than the vocoid in *I'd*. This means that this pair contrasts in three parameters rather than two. The analysis of the spectral moments of the palato-alveolar friction in *she'd* and *she'll* indicates that the two frictions differ along three parameters, including CoG, skewness, and kurtosis.

Finally, the analysis of the formant dynamics of a subset of tokens with a vocoid duration shorter than the mean vocoid duration of all tokens in each dataset, shows that F2 is flatter in the subset of the reduced tokens compared to the mean F2 calculated across the entire dataset. This feature suggests that in the reduced (shorter) tokens there is a decrease in the magnitude of gesture. From the comparison of the formant dynamics of the reduced tokens with the formant dynamics of the entire datasets, it also emerged that the difference in F2 between pr + *'ll* and pr + *'d* is larger in the reduced tokens. This feature suggests that a decrease in the magnitude of gestures does not necessarily weaken the contrast between the two paradigms. On the contrary, as far as the resonances of the two pr + aux are concerned, the contrast in the reduced tokens seems to be enhanced.

5.2.3. The present vs past tense contrast of the finite forms of the auxiliary BE

The present tense forms of BE are *am*, *is*, *are*. The past forms of BE are *was* and *were*. In this section, the contrasting pairs *is* versus *was*, and *are* versus *were* are analysed.

As mentioned in Section 5.1.2, the present versus past contrast in the finite forms of the auxiliary BE is expressed by the labiovelarity in the past tense, which is a prosody of the whole syllable (Simpson, 1992; Ogden, 1999). The labiovelarity of the past tense contrasts with the non-rounding of the present tense in the pair *are* versus *were*, and non-rounding and palatality in *is* versus *was*. The forms *was* and *were* can never lose the labiovelarity because it is the feature that marks the past tense. The question arises of whether the labiovelarity is retained in reduced speech or whether other phonetic features maintain the present versus past contrast. The quantitative analysis reported in the following sections will shed light on the phonetic realisations of the paradigms *she's* and *she was* (Section 5.2.3.1), *you're* and *you were* (Section 5.2.3.2.1), and *we're* and *we were* (Section 5.2.3.2.2).

5.2.3.1. *She's* versus *she was*

The rationale behind the choice of the pronoun *she* for the analysis of the contrast *is* versus *was* is that the labiovelarity of *was* and the non-rounding and palatality of *is* are prosodies of the entire piece and should be present in the palato-alveolar friction in onset of *she*. During the preliminary analysis of the data, it was noticed that the palato-alveolar friction is always articulated and is thus suitable for the analysis of the spectral moments. The glottal friction of *he* would also be suitable, since it is not articulated in the oral cavity, but in several instances of *he*+ aux, the glottal friction is very weak which might make the analysis of the spectral moments challenging.

The same acoustic and statistical analyses employed in the previous sections were carried out for the contrasting paradigms *she's* and *she was*. The results of the mixed-effects models reveal a main effect of *Auxiliary* on the duration of *Vocoid* ($\chi^2(1) = 83.22$, $p < 0.001$), *Voicing* ($\chi^2(1) = 83.97$, $p < 0.001$), and *Piece* ($\chi^2(1) = 20.54$, $p < 0.001$). The other domains did not show a significant difference in duration for different auxiliaries (all have p-value > 0.001). Figure 136 shows the mean duration across speakers and repetitions (*she's* N = 50; *she was* N = 47) of the palato-alveolar friction (fr1), the vocoid (v), the alveolar friction (fr2), the closure (hph), and the piece in *she's* (blue) and *she was* (orange).

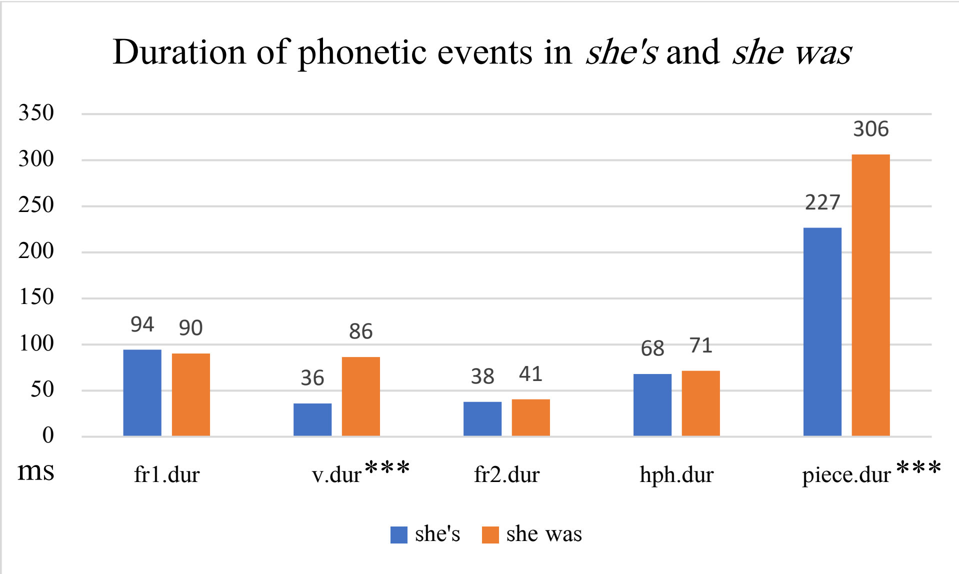


Figure 136. Mean duration (in ms) across speakers and repetitions of the palato-alveolar friction (fr1), the vocoid (v), the alveolar friction (fr2), the closure (hph), and the piece in *she's* (blue) and *she was* (orange). *** indicates that the duration of the two paradigms is significantly different.

Figure 136 shows the duration of a range of phonetic events in the paradigms *she's* and *she was*. The main feature to notice is that the duration of the vocoid in the pr + aux *she's* is significantly shorter than the duration of the vocoid in *she was*. The durations of voicing and the piece are positively correlated to the duration of the vocoid.

The spectral moments of the palato-alveolar friction in onset and the alveolar friction in coda in both paradigms were analysed too. The results of the statistical test reveal a main effect on *SD* ($\chi^2(1) = 9.48$, $p < 0.002$), *Skewness* ($\chi^2(1) = 19.78$, $p < 0.001$), and *Kurtosis* ($\chi^2(1) = 13.54$, $p < 0.001$), but not *CoG* ($\chi^2(1) = 0.59$, $p = 0.44$). Figure 137 shows the distribution of the four spectral moments in the palato-alveolar friction in onset of *she's* (blue) and *she was* (orange).

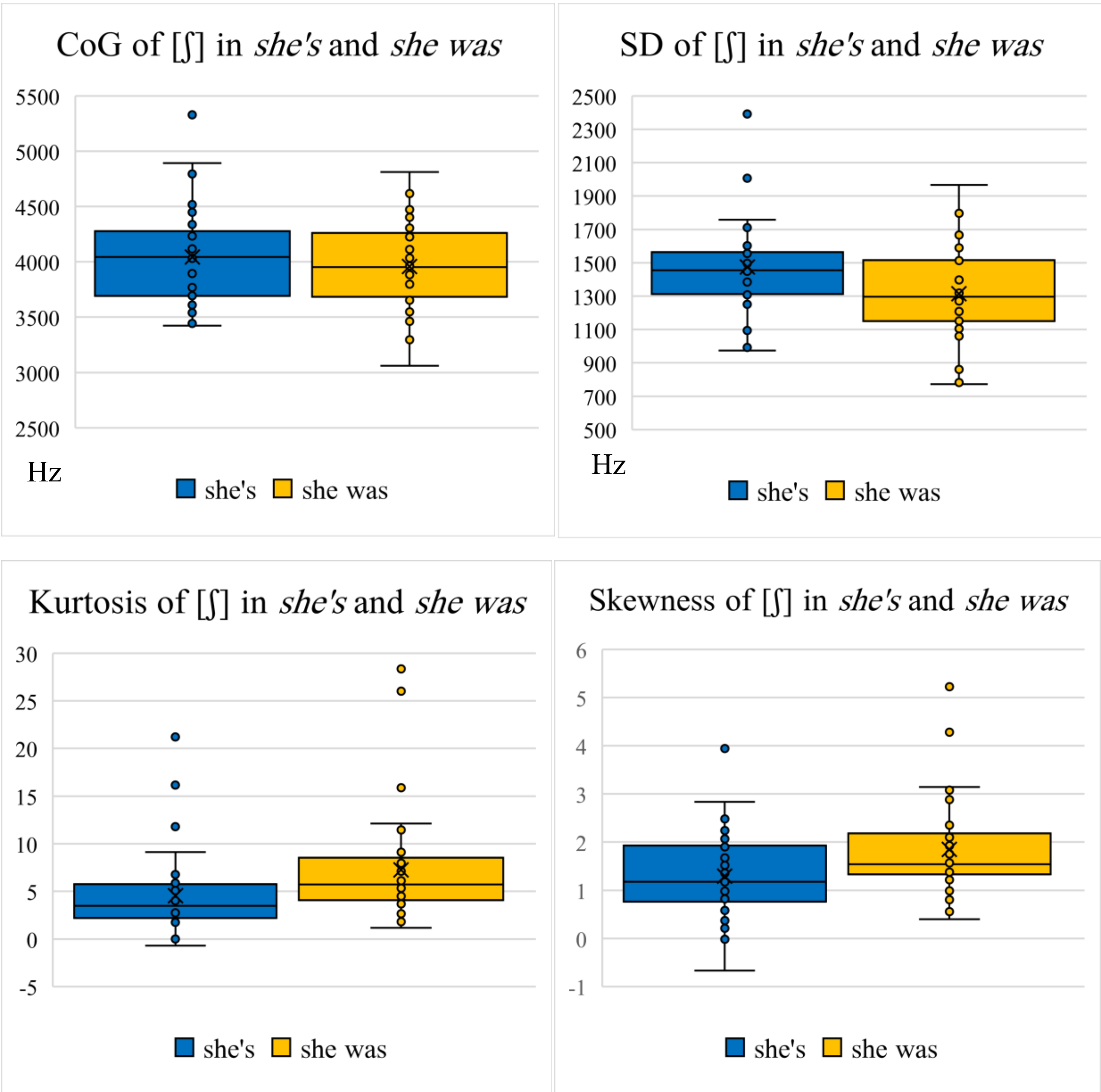


Figure 137. CoG (top left), SD (top right), kurtosis (bottom left), and skewness (bottom right) of the palato-alveolar friction in onset of *she's* (blue) and *she was* (orange).

Surprisingly, the CoG of the palato-alveolar friction in onset of *she's* and *she was* does not exhibit a significant difference. The mean CoG in *she's* is 4044 Hz, while the mean CoG in *she was* is 3957 Hz. Both kurtosis and skewness are positive in both *she's* and *she was*. However, the kurtosis is significantly higher in *she was*, indicating a more peaked distribution of aperiodic energy than in *she's*. The skewness is also significantly higher in *she was* than *she's*, indicating that the aperiodic energy in the palato-alveolar friction in *she was* is more asymmetrically distributed at lower frequencies than in *she's*. The spectral qualities of the alveolar friction in coda position do not exhibit any significant difference.

The formant dynamics of the vocoid in *she's* and *she was* were compared too. Although formants are usually measured when the source of noise is in the glottis, such as in sonorant sounds, they can be measured also in voiceless sounds. Since the CoG of the palato-alveolar friction did not exhibit the expected pattern, and the formants are usually visible during the friction, the formant dynamics of the palato-alveolar friction in *she's* and *she was* were analysed too. Figure 138 shows the formant dynamics of the palato-alveolar friction (left) and the vocoid (right) in *she's* (blue) and *she was* (red). The frequency scale is the same and displays frequencies up to 4 kHz.

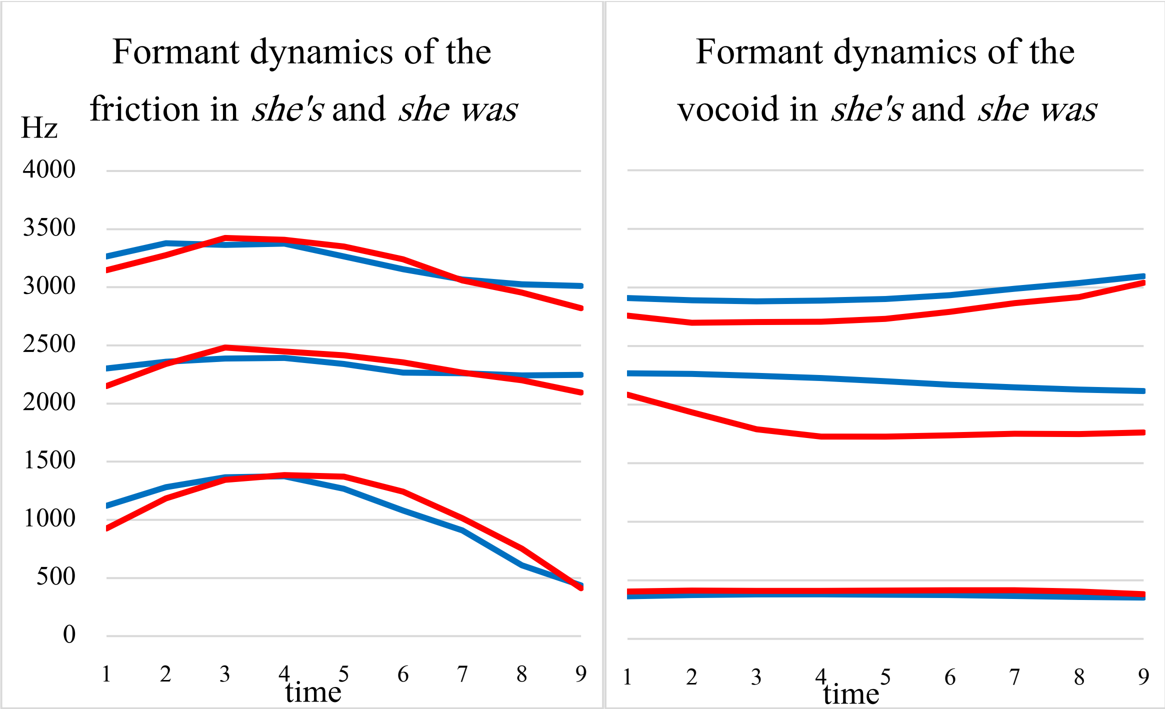


Figure 138. Formant dynamics of the palato-alveolar friction (on the left) and the vocoid (on the right) in *she's* (blue) and *she was* (red).

Figure 138 shows the formant dynamics of the palato-alveolar friction (on the left) and the vocoid (on the right) in *she's* and *she was*. The formant dynamics of the vocoid show that *she was* is characterised by lower F2 and F3 throughout the vocoid duration, suggesting that the vocoid is either articulated further back in the oral cavity or with lip-rounding or both. This suggests that the vocoid in *she was* is characterised by darker resonances than the vocoid in *she's*. The formant dynamics of the vocoid in *she's* suggest that the vocoid is articulated further forward and higher in the oral cavity than the vocoid in *she was*. Although all three formants differ from the beginning of the vocoids, F2 and F3 in *she was* exhibit a dip at Time points 3 and 4 typical of a labiovelar articulation. The formant dynamics of the palato-alveolar friction confirm the results of the CoG. The expected pattern of lower frequencies in *she was* is not present in the friction suggesting that the friction does not exhibit a darker quality in *she was* than in *she's*. A possible explanation is that [ʃ] in English is realised with lip-rounding (Ogden, 2009), therefore, the labiovelarity of *was* does not affect the articulation of [ʃ] which is already labialised. However, the high number of instances of *she's* realised without a vocoid raises the question of how the contrast between the paradigms *she's* and *she was* is maintained in reduced speech. So far, the main acoustic differences that emerged from the comparison of the two paradigms are: the vocoid duration and the vocoid's formants. If a piece is realised without an identifiable vocoid, other acoustic cues might be available to maintain the contrast. To investigate this issue further, and to test whether the two paradigms retain their distinct resonances also in highly reduced speech, a subset of *she was* and two subsets of *she's* were selected for comparison.

A high proportion of instances of *she's* (N = 16/50, 32%) are realised with only friction with no identifiable vocoid. This means that the instances of *she's* with a short vocoid are not the most reduced instances. If the most reduced instances of *she's* exhibit only friction and the spectral CoG of the friction in *she's* and *she was* do not differ significantly, the question is how the instances of *she's* without a vocoid maintain their distinction in reduced speech. Since *she was* is always realised with a vocoid, a pr + aux realisation without a vocoid is already an indication that it cannot be *she was*. However, there might also be other acoustic cues signalling the contrast. For this reason, the spectral qualities of the palato-alveolar friction in instances of *she's* without a vocoid, and the spectral qualities of the palato-alveolar friction in instances of *she was* with a vocoid duration shorter than the mean vocoid duration of entire dataset were compared. Sixteen instances of *she's* (32%) and 25 instances of *she was* (N = 25/47, 53%) were selected and analysed. The statistical analysis revealed a main effect on *CoG* ($\chi^2(1) = 9.9678$, $p = 0.001$), *Skewness* ($\chi^2(1) = 10.214$, $p = 0.001$), and *Kurtosis* ($\chi^2(1) = 6.4017$, $p = 0.011$), but not *SD* ($\chi^2(1) = 5.7692$, $p = 0.016$).

Figure 139 shows the mean CoG of the palato-alveolar friction in instances of *she's* without a vocoid (blue) and shorter instances of *she was* (orange).

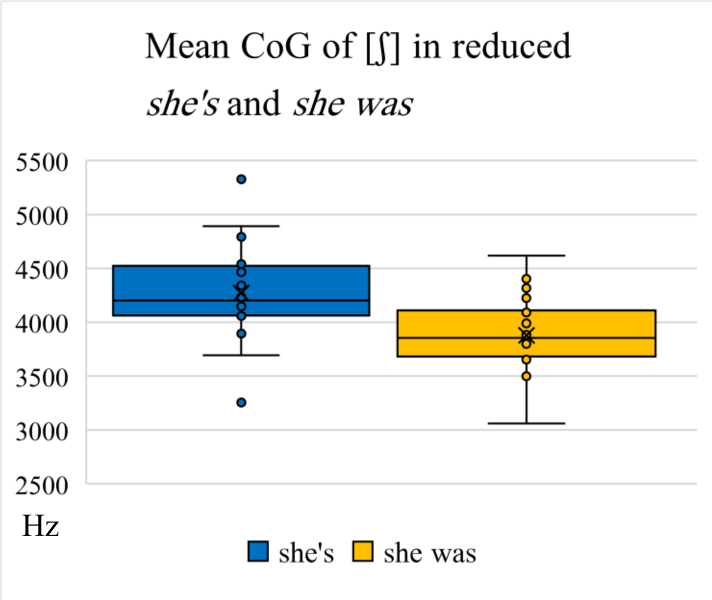


Figure 139. Mean CoG (in Hz) of the palato-alveolar friction in reduced instances of *she's* (blue) and *she was* (orange).

Figure 139 shows the CoG of instances of *she's* without a vocoid and shorter instances of *she was*. Compared with the boxplot at the top left of Figure 137, the difference between the two paradigms is noticeable. The mean CoG of the palato-alveolar friction in the subset of instances of *she's* without a vocoid (4270 Hz) is higher than the mean CoG calculated across the entire *she's* dataset (4044 Hz), and it is significantly different from the mean CoG of a subset of reduced tokens of *she was* (3878 Hz). This suggests that the contrast between the two paradigms is retained in reduced speech and it is conveyed by the palato-alveolar friction.

The instances with a vocoid with a shorter duration than the mean vocoid duration for each paradigm were selected too and their formants compared. Figure 140 shows the mean formant dynamics of the two subsets of *she's* and *she was* with shorter vocoid durations.

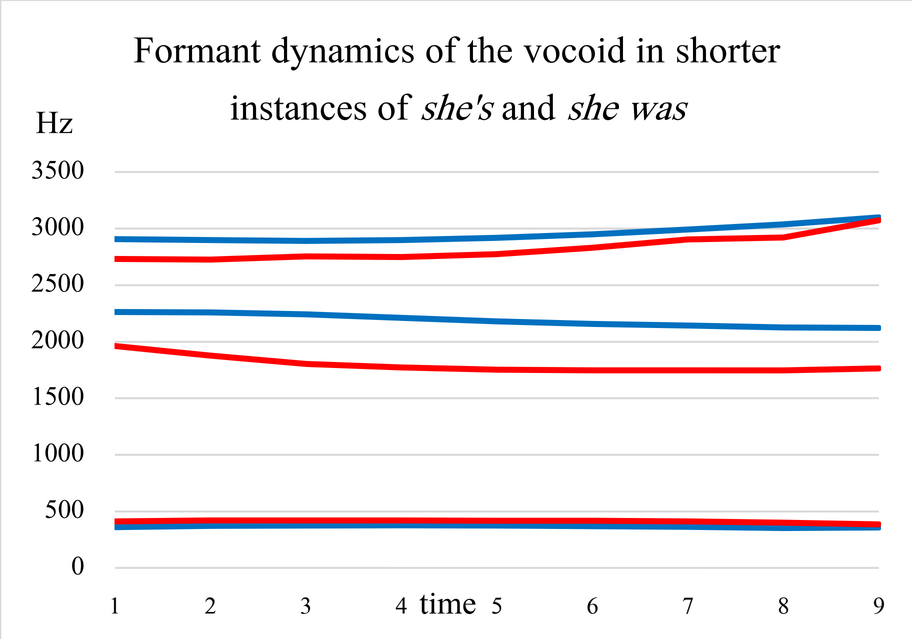


Figure 140. Formant dynamics of short instances of *she's* (blue) and *she was* (red).

Figure 140 shows the formant dynamics of shorter instances of *she's* and *she was*. All three formants show the same pattern exhibited in the entire datasets (F1 is lower in *she's*, while F2 and F3 are lower in *she was*). The formant dynamics suggest that *she was* is characterised by dark resonances, while *she's* is characterised by clear resonances even in reduced instances. In both paradigms, the formants are flatter in the reduced instances than in the whole dataset. However, they suggest that the pieces maintain their resonances also when reduced in duration.

To summarise, this section has shown that the paradigms *she's* and *she was* differ in two main parameters: the duration of the vocoid and the resonances. The vocoid in *she's* exhibits a significantly shorter duration than the vocoid in *she was*. However, 32% of instances of *she's* do not exhibit an identifiable vocoid, while it is present in all instances of *she was*. The absence of an identifiable vocoid can be a contrasting feature in itself.

As for the resonances, the formant dynamics of *she's* and *she was* indicate that the vocoids differ in their resonances: the vocoid in *she's* is characterised by clear (or palatal) resonances, while the vocoid in *she was* is characterised by dark (or labiovelar) resonances. Based on the assumption that the resonances are long-domain features and not locally delimited (Kelly and Local, 1989), we expected to find the effect of the resonances in the palato-alveolar friction in onset too. The comparison of the spectral properties of the palato-alveolar friction across the entire datasets of the two paradigms revealed that they differ significantly in SD, skewness and kurtosis, but not CoG. Although three of the spectral moments are significantly different, the CoG, which can be considered the acoustic correlate of the place of articulation, which is usually affected by the resonances, is not. This unexpected result raised the question of how the contrast between *she's* and *she was* is maintained in those instances of *she's* in which there is no vocoid, if the contrasting resonances are available only in the vocoid. The pr + aux *she's* and *she was* differ significantly in the duration of the vocoid. The question is whether the duration of the vocoid alone can maintain their distinction in reduced speech. A further investigation into the spectral properties of reduced instances of *she's* and *she was* revealed that in the instances of *she's* in which there is no vocoid, the palato-alveolar friction has a higher CoG (mean 4270 Hz) which is significantly different from the CoG of a subset of reduced instances of *she was* (mean 3878 Hz). The significant difference between the CoG of reduced instances compared to the non-significant difference across the whole datasets, suggests that in reduced speech the resonances in the friction are enhanced rather than weakened. The results of the analysis reported here suggest that in reduced speech the grammatical information about the present and past tense of BE is not lost but conveyed by a set of parameters depending also on the degree of reduction. In instances of *she's* that are realised with a vocoid, the contrast is maintained by the duration and the resonances of the vocoid. In instances of *she's* that are realised as friction only, the spectral properties of the palato-alveolar friction differ significantly between the two paradigms *she's* and *she was*, reflecting the dark and clear qualities of the resonances. In addition, the absence of a vocoid could be a discriminating feature too, since all instances of *she was* in the dataset are realised with a vocoid, while 32% of instances of *she's* are realised without a vocoid.

5.2.3.2. *Are* versus *were*

The finite forms of the present and past tense of BE, *are* and *were*, can be combined with the pronouns *we*, *you*, and *they*. In this section, only the analyses of the paradigms *you're* and *you were*, and *we're* and *we were* are reported. As described in Chapter 3, *we* is characterised by labiovelarity, while *you* is characterised by palatality and often lip-rounding too. Considering that the present versus past tense of BE is marked by lip-rounding, the question arises of how the present tense, which is marked by non-rounding is conveyed by the pronouns that have rounding (or labiality) as essential phonetic feature. The following sections describe the acoustic features that distinguish *you're* and *you were*, and *we're* and *we were*.

5.2.3.2.1. *You're* versus *you were*

This section reports on the analysis and comparison of the acoustic properties of the paradigms *you're* and *you were*. The same methodology used in the previous sections was applied. The acoustic features of all instances of *you're* (N = 55) across speakers and repetitions were analysed and compared to all instances of *you were* (N = 51).

The results of the six mixed-effects models reveal a main effect of *Auxiliary* on *Vocoid* ($\chi^2(1) = 136.36$, $p < 0.001$), *Voicing* ($\chi^2(1) = 148.98$, $p < 0.001$), and *Piece* ($\chi^2(1) = 126.02$, $p < 0.001$). The other domains did not show a significant difference in duration for different auxiliaries (all have p-value >0.001). Figure 141 shows the mean duration of the phonetic events in *you're* and *you were*.

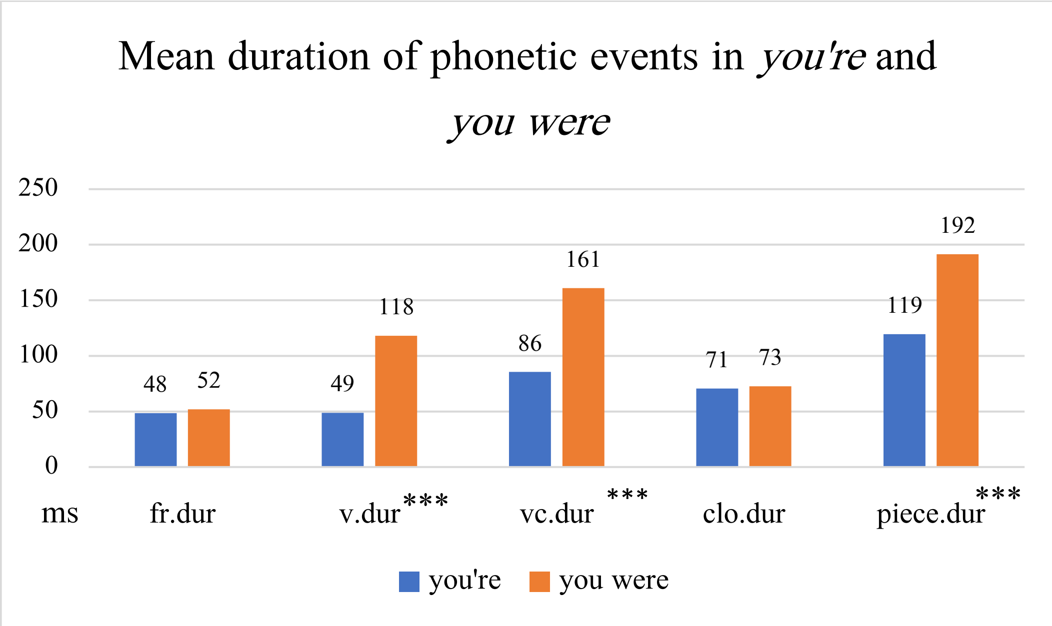


Figure 141. Mean duration (in ms) across speakers and repetitions of the silent articulation (fr), the vocoid (v), voicing (vc), the closure (clo), and the piece in *you're* (blue) and *you were* (orange). *** indicates that the duration of the two paradigms is significantly different.

As for the resonances, Figure 142 shows the formant dynamics of the vocoid in *you're* and *you were* calculated across speakers and repetitions at nine points in time.

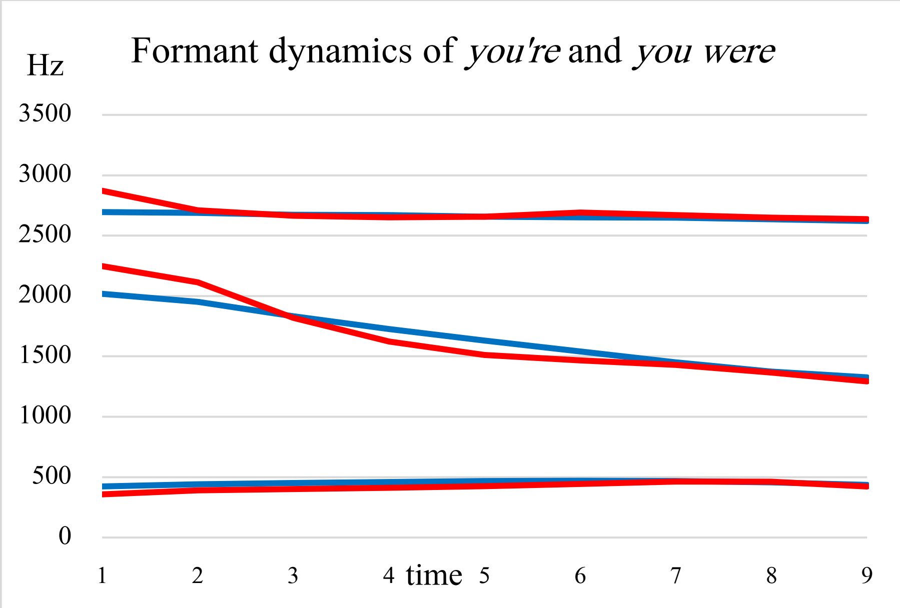


Figure 142. Formant dynamics of the vocoid in *you're* (blue) and *you were* (red).

The formant dynamics do not exhibit a clear pattern of contrast, or polarity. F2 in *you were* shows a higher degree of magnitude of gesture compared to *you're*. In particular, F2 in *you're* slopes steadily and gradually on the same trajectory as F2 in *you were*, but F2 in *you were* is at higher frequencies at Time points 1 and 2, but slopes to lower frequencies between Time points 4 and 6. All three formants exhibit a greater difference between paradigms at the beginning of the vocoid, suggesting a stronger palatality in onset of *you were* than in onset of *you're* (Time Points 1-2). Moreover, F2 in *you were* exhibits a dip in frequency in the central portion of the vocoid, suggesting a higher degree of lip-rounding or labiovelarity. Despite these small differences, a clear pattern of clear versus dark resonances is not visible. In this case, the duration of the vocoid might be the only parameter that maintains the distinction between *you're* and *you were* in reduced speech.

The formant dynamics of two subsets of data with only the instances of *you're* and *you were* that have a vocoid duration that is shorter than the respective mean vocoid durations across the datasets were analysed. Figure 143 shows the formant dynamics of the instances of *you're* (N = 27) and *you were* (N = 29) with a reduced duration.

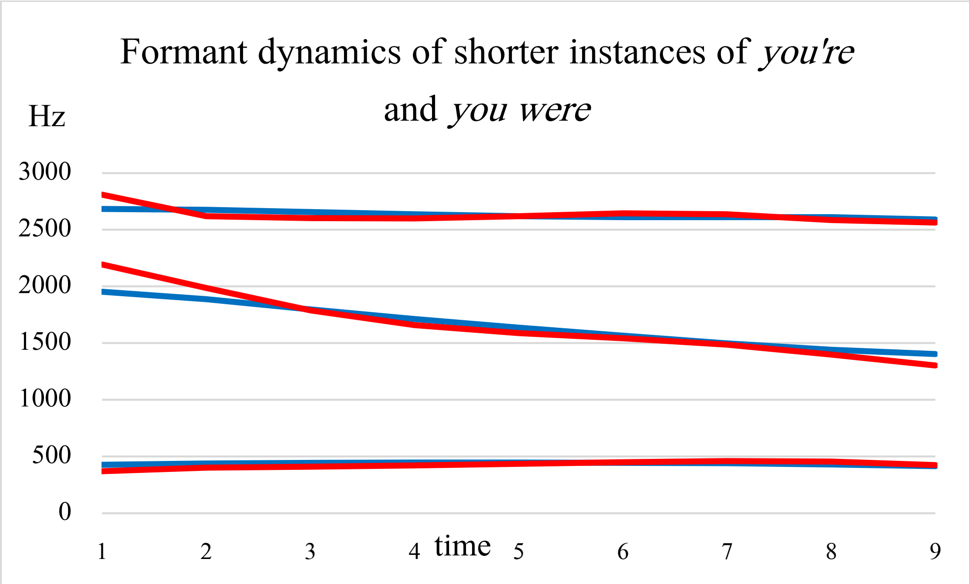


Figure 143. Formant dynamics of instances of a subset of *you're* (blue) and *you were* (red) with a shorter vocoid duration.

The analysis of the formant dynamics of two datasets of *you're* and *you were* characterised by a shorter duration than the mean vocoid duration shows that the formant dynamics are more similar than in the whole datasets. In particular, F2 in *you were* is flatter and exhibits less movement than in the entire dataset. This suggests that in shorter instances, the magnitude of the articulatory gesture is reduced. However, the formant dynamics in Figure 143 confirm that there is not a clear polarity in the resonances of *you're* and *you were*.

5.2.3.2.2. *We're* versus *we were*

The acoustic features of the contrasting pair of paradigms *we're* and *we were* exhibit the same patterns observed in *you're* and *you were*. The same methodology was applied.

The results of the statistical analysis reveal a main effect on *Vocoid* ($\chi^2(1) = 136.37$, $p < 0.001$), *Voicing* ($\chi^2(1) = 125.71$, $p < 0.001$), and *Piece* ($\chi^2(1) = 113.08$, $p < 0.001$). The other domains did not show a significant difference in duration for different auxiliaries (all have p-value >0.001). Figure 144 shows the distribution of the duration of the vocoid.

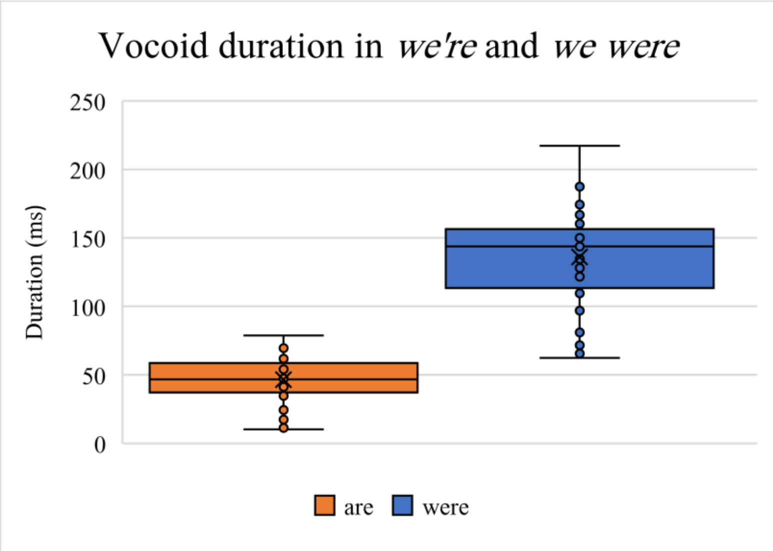


Figure 144. Distribution of the vocoid duration in *we're* (orange) and *we were* (blue).

Figure 144 shows that the duration of the vocoid in *we're* is shorter than the duration of the vocoid in *we were*. The formant dynamics were analysed too and are shown in Figure 145.

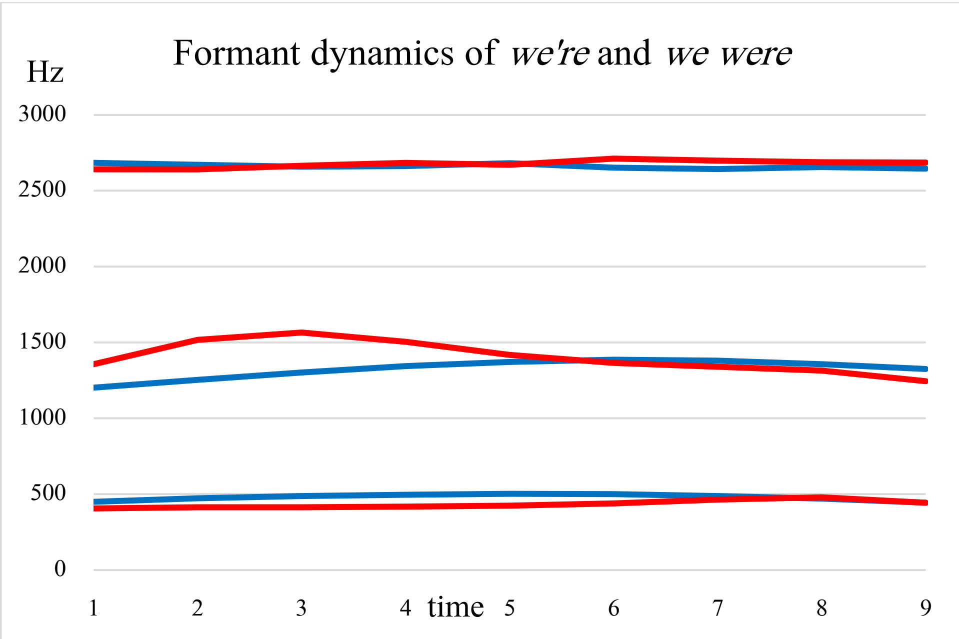


Figure 145. Formant dynamics of *we're* (blue) and *we were* (red).

The formant dynamics in *we're* and *we were* exhibit a similar pattern to the formant dynamics in *you're* and *you were*, in that in the past tense pr + *were* F2 is characterised by a larger movement and variation in frequency than F2 in the present tense pr + *'re*, which is flatter throughout. The absence of a clear polarity between clear and dark resonances observed in *you're* and *you were* is apparent also in the pair *we're* and *we were*.

To summarise, the contrast in the pairs of paradigms *you're* and *you were*, and *we're* and *we were* exhibits a similar pattern. The present tense *you're* and *we're* have a significantly shorter vocoid duration than their past tense counterparts *you were* and *we were*. As for the resonances, auditorily the pronoun *you* is characterised by palatality and the pronoun *we* by labiovelarity. However, it was shown in Chapter 4 that in combination with the auxiliary *are*, both *you* and *we* change quality, as indicated by the change in the frequency of F1 and F2 when *you* and *we* are combined with *are* in comparison to when they are combined with *have* (see Figure 105 in Chapter 4 comparing the formant dynamics of the vocoid in *we're* and *we've*). In combination with *were*, the quality of *we* does not change so dramatically. Only the first part of the vocoid exhibits higher F2. Although the similarity between the formant dynamics in pr + *'re* and pr + *were* raises the question of whether the resonances have a role in the perception of the contrast and the intelligibility of the paradigms, the lack of a contrast in the resonances might be balanced by the large difference in vocoid duration. While the durations of the vocoid in, for example, pr + *'ll* and pr + *'d* overlap greatly, as seen in Figure 131, the durations of the vocoid in pr + *'re* and pr + *were* exhibit a much larger, and nearly categorical, difference.

5.2.4. The present vs past tense contrast of the finite forms of the auxiliary HAVE

Phonemically, the contrast between the present tense and the past tense of HAVE is conveyed by friction in the present tense (pr + *'ve*, /v/, and pr + *'s*, /z/~/s/), and stoppedness in the past tense (pr + *'d*, /d/). The contrast between pr + *'ve* and pr + *'d* was briefly covered in Section 4.1.3, because it was pointed out that 15% of instances of pr + *'ve* across pronouns were realised without friction. The comparison between the acoustic features of pr + *'d* and the instances of pr + *'ve* without friction reported there revealed that they differ in vocoid duration, F2 frequency in the last portion of the vocoid, and the abrupt versus gradual end of the vocoid.

In the following section only a specific type of contrast is reported: the contrast between pr + *'s* and the instances of pr + *'d* in which /d/ is realised with friction instead of a complete closure.

5.2.4.1. *Has* versus *had*

This section reports on the analysis of the contrast between the cliticised forms of *has* and *had*. The rationale for analysing this contrast is that the realisation of alveolar friction at the end of pr + *'d* instead of a complete closure, means that the contrast with pr + *'s* could be neutralised. This feature occurs when *'d* is in combination with any of the three pronouns *he*, *she*, and *it*. In this section, the contrast between the present and past tense of HAVE is analysed in conjunction with the pronoun *he*.

5.2.4.1.1. *He's* versus *he'd*

In the data collected, 22% of instances of *he'd* (N = 24/110) were realised without a complete closure at the end of the vocoid, but with a portion of friction at the place of articulation. In these cases, both contoids in coda of *he's* and *he'd* are articulated at the alveolar ridge and can be partially voiced. The realisation of /d/ with alveolar friction raises the question of whether the two homorganic frictions differ, and whether the contrast between the two paradigms is maintained.

This section reports on the comparison of the acoustic features of *he's* and *he'd*. Only some speakers produced friction instead of a complete closure in *he'd*. These speakers are: S3, S6, S7, S8, S15. For the comparison between the acoustic properties of *he's* and *he'd*, it was decided to analyse only the tokens produced by these speakers. Therefore, 50 tokens of *he's* (including both *he's burnt* and *he's burning*), and 24 tokens of *he'd* (including both *he'd burnt* and *he'd burn* dataset) were analysed. The sub-datasets of *he'd* (*had*) and *he'd* (*would*) were collated together for this analysis. The rationale behind this decision is that the number of instances of friction instead of closure in *he'd* tokens is small (N = 24/110). A larger dataset makes the statistical analysis more robust. Moreover, the comparison between the acoustic features of *he'd* (*would*) and *he'd* (*had*) did not reveal any significant difference.

The parameters analysed are duration, amplitude, first four spectral moments, and formant dynamics of all sounds in each piece. The analysis included tokens in which there is friction from the end of the vocoid, as shown in Figure 146 on the left, and also tokens in which at the end of the vocoid there is a portion of weak friction which then gradually increases in energy, as shown in Figure 146 on the right. In the second type, it can be assumed that the tongue creates a narrower constriction at first and then a slightly wider constriction that allows more turbulent air to flow out from the oral cavity. The difference between the two types of friction is evident also in the waveform, in which the second type shows an increase in the overall amplitude of the friction.

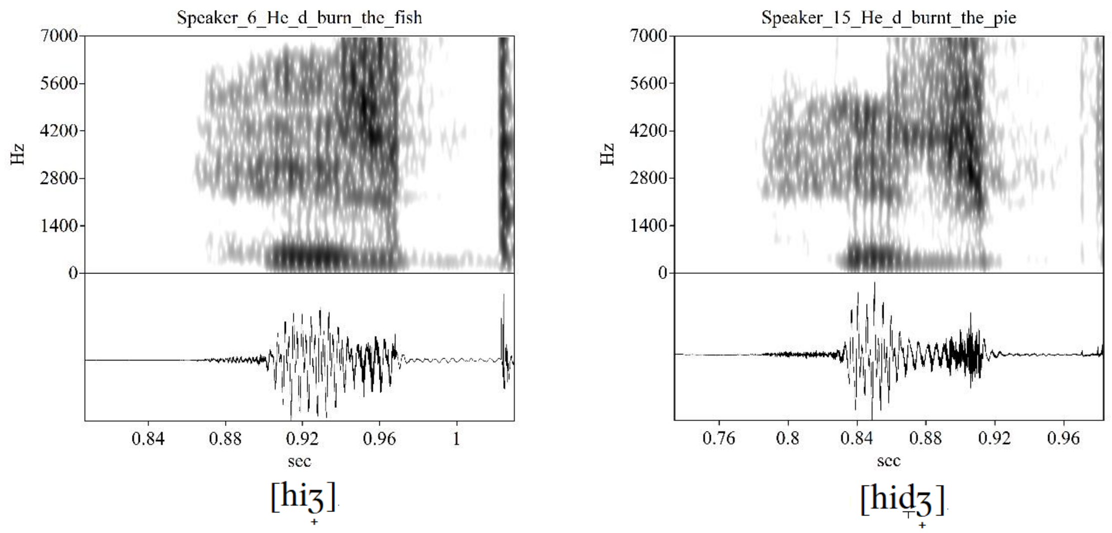


Figure 146. Spectrogram and waveform of two instances of *he'd* with friction in coda. On the left, an instance of constant friction from the end of the vocoid. On the right, an instance of friction that increases in time after a portion of weak friction. Both spectrograms show the frequency scale up to 7 kHz.

The statistical analysis revealed a main effect of *Auxiliary* on *Friction2* ($\chi^2(1) = 44.16$, $p < 0.001$). *Auxiliary* did not have a significant effect on the duration of the other intervals measured (all have p-value > 0.001). Figure 147 shows the mean duration of a range of phonetic events.

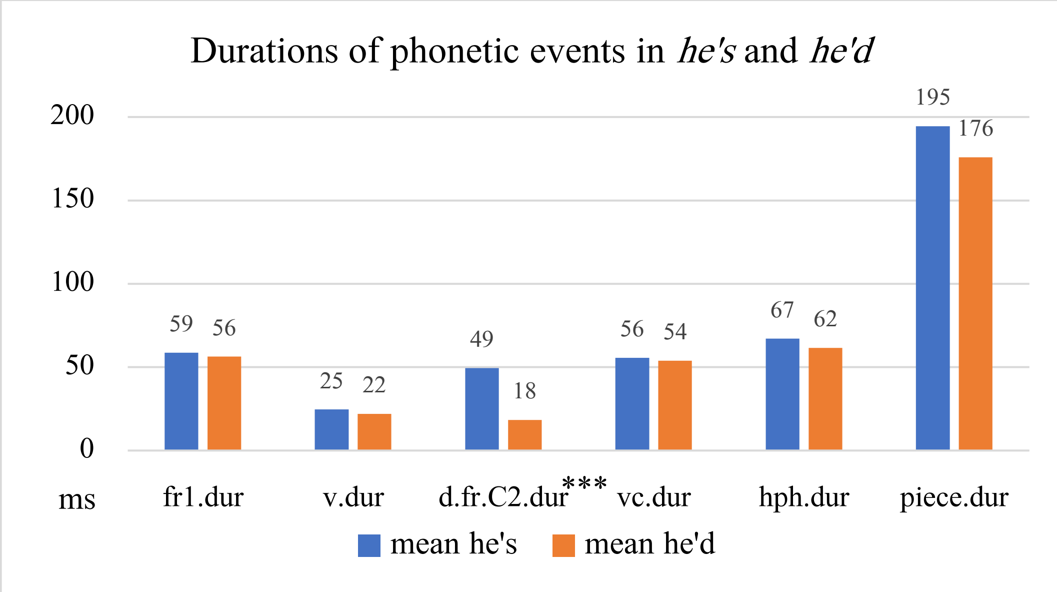


Figure 147. Mean durations of glottal friction (fr1), vocoid (v), alveolar friction (d.fr.C2), voicing (vc), closure (hph) and the entire piece (piece) in a subset of *he's* and *he'd* paradigms.

Figure 147 shows the mean durations of the phonetic events in a subset of *he's* and *he'd*. It can be observed that the main durational difference is in the duration of the alveolar friction. The shorter duration of all the phonetic events in *he'd* tokens could be due to a slightly higher degree of reduction in these tokens. On the one hand, if we consider the definition of reduction as 'articulatory undershoot' (Lindblom, 1963; Bauer, 2008), it could be assumed that when /d/ is realised with friction instead of a complete closure, the degree of reduction is higher. Therefore, the tokens chosen for this analysis (instances of *he'd* in which there is friction instead of a complete closure) exhibit more reduction than the rest of items in the *he'd* datasets (instances of *he'd* articulated with a complete closure). On the other hand, some of the *he'd* tokens in which /d/ is realised with friction occur in the first of the five repetitions. If repetition triggers reduction (see Section 6.3), the first mention of *he'd* should be the least reduced of the five. The fact that in some cases the first repetition of *he'd* exhibits friction instead of closure might indicate that this is a widespread feature that is not restricted to highly reduced speech. Findings by Buizza (2010) and Buizza and Plug (2012) indicate that /t/ frication in RP English is a common phenomenon in spontaneous speech.

The first four spectral moments of the alveolar friction in *he'd* and *he's* were analysed. The results of the four mixed-effects models reveal a main effect on *CoG* ($\chi^2(1) = 97.15$, $p < 0.001$) and *Skewness* ($\chi^2(1) = 79.66$, $p < 0.001$), but not *Amplitude* ($\chi^2(1) = 1.57$, $p = 0.21$), *Kurtosis* ($\chi^2(1) = 4.29$, $p = 0.038$) or *SD* ($\chi^2(1) = 2.58$, $p = 0.108$). Table 24 reports the mean values of the alveolar friction in *he's* and *he'd* subsets.

	CoG***	**SD**	**Skewness*****	**kurtosis**
mean *he's*	6983	2224	-0.70659	2.510948
mean *he'd*	2692	1903	1.59695	5.342809

Table 24. Mean spectral moments of the alveolar friction in the subsets of *he'd* and *he's*. In red the parameters that are significantly different (***).

The significant differences in spectral qualities for two of the four spectral moments indicate that the friction has different acoustic characteristics. First of all, the mean CoG is much higher in the alveolar friction in *he's* than in *he'd*. This suggests that the articulation of the friction in *he'd* is articulated with a different tongue shape creating a cavity of a different size or shape than that of *he's*. Moreover, the positive skewness of the alveolar friction in *he'd* indicates that the aperiodic energy is concentrated in the lower frequencies; while the negative skewness of the friction in *he's* indicates that the aperiodic energy is concentrated in the higher frequencies. The difference in kurtosis, although not significant, indicates that the alveolar friction in *he'd* is characterised by a more peaked distribution, while the friction in *he's* has a flatter distribution.

The mean values of the formant dynamics of the vocoid differ slightly in *he's* and *he'd* tokens, especially F2 (see Figure 148). F2 is lower in *he's* than *he'd*, suggesting that the place of articulation of the vocoid in *he's* might be further back than the place of articulation of the vocoid in *he'd*.

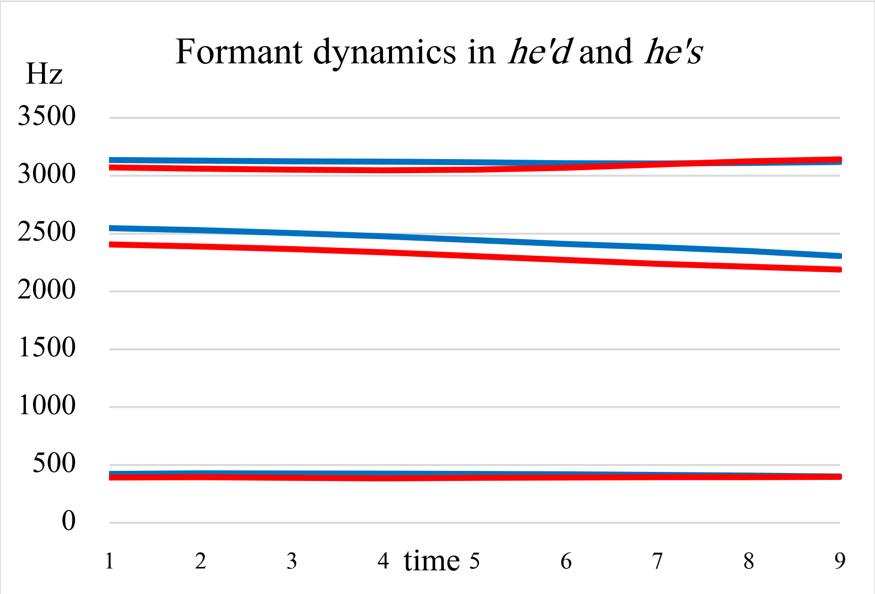


Figure 148. Formant dynamics of the vocoid in *he'd* (blue) and *he's* (red).

To summarise, the acoustic features of the alveolar friction in *he's* and *he'd* differ in duration and two out of four spectral moments. This would suggest that the two paradigms *he's* and *he'd* maintain their contrast even when *he'd* is produced with friction instead of a complete closure. The fact that /d/ is produced with alveolar friction even in the first repetition, might suggest that friction for closure might be a widespread phenomenon and not a feature of reduction. Simpson (2001) claims that fricated variants of alveolar stops in English might become habitual patterns for some speakers.

5.2.5. Contrast between pronouns

As reported in Chapter 3, English pronouns are characterised by a range of diverse features (see Table 8). However, *you* and *he* are both characterised by palatality. In addition to a palatal, or close front articulation, *you* is characterised also by labiality, but this feature can be weak. While *he* is characterised by glottal friction, but also this feature can be weak. This means that *you* and *he* share the feature palatality as their more prominent characteristic. In the data collected, several instances of *you* exhibit a period of voiceless friction with a palatal quality at the beginning of the piece (see Section 3.1.5). If these instances with onset friction do not exhibit the other feature of *you* – labiality – what are the acoustic features that maintain the contrast in pairs such as *he'd* and *you'd*? The next section reports on the comparison between the acoustic features of *he'd* and the instances of *you'd* with initial friction.

5.2.5.1. *He'd* versus *you'd*

In 70% of instances of *you*, there is a portion of friction at the beginning of the piece. This friction is due to the articulation in the oral cavity starting to move before the onset of vibration of the vocal folds, resulting in voiceless friction with a palatal quality and visible formants (mainly F2) in the spectrograms. Although in some cases the friction is too weak to be audible, when it is audible, it has a palatal quality. This feature raises the question of whether the contrast with *he* is maintained, as the glottal friction in onset of *he* has a similar close front quality. Moreover, during the acoustic analysis it was noticed that the formant dynamics of the two paradigms were rather similar. Figure 149 shows an example of *he'd* (left) and *you'd* (right) produced by the same speaker.

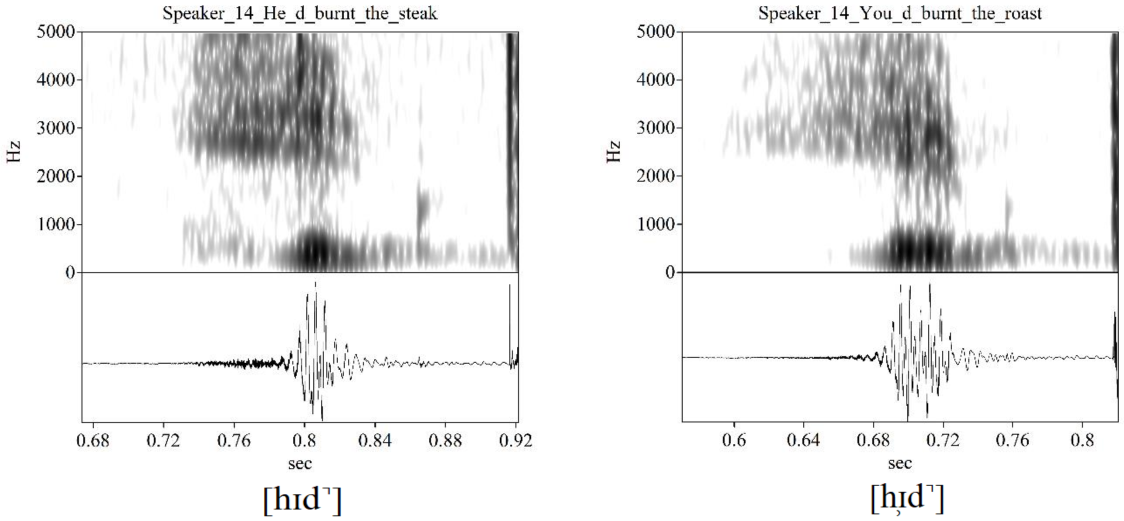


Figure 149. Spectrogram and waveform of an instance of *he'd* (left) and *you'd* (right) produced by the same speaker (S14).

Both tokens in Figure 149 are characterised by a short vocoid and a long portion of friction in onset. The main differences that can be observed in Figure 149, are the slightly stronger friction at the beginning of *he'd*, which is also characterised by clearer F2 and F3 compared to the weaker and less well-defined friction at the beginning of *you'd*. Figure 150 shows the formant dynamics of the vocoid in *he'd* and *you'd*.

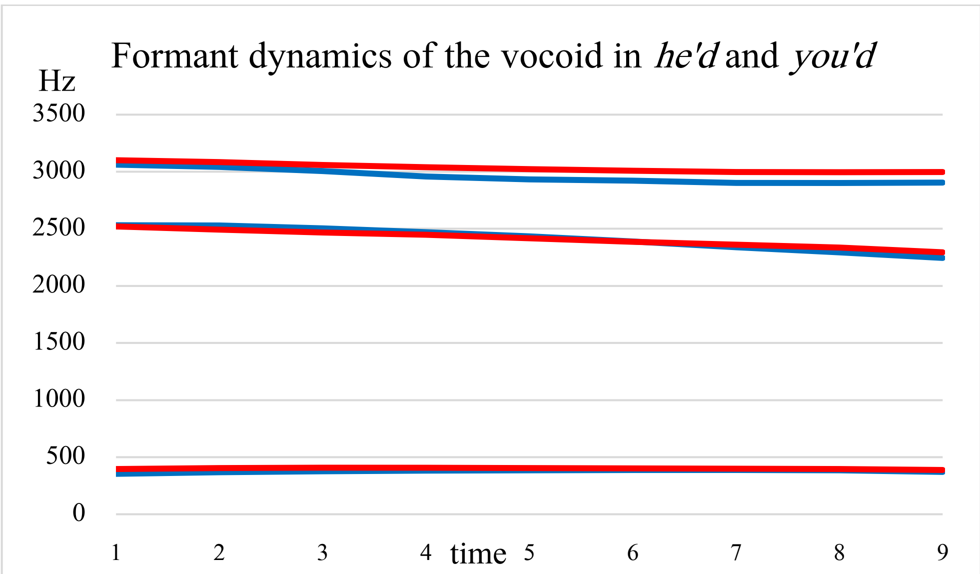


Figure 150. Formant dynamics of the vocoid in *he'd* (red) and *you'd* (blue).

The mean formant dynamics (calculated across speakers and repetitions) of the vocoid in *you'd* and *he'd* are rather similar. The most noticeable difference between them is the lower F3 in *you'd*, which might indicate that there is a trace of lip-rounding in *you*. However, F2 is very similar.

The mean durations of *you'd* and *he'd* were analysed. The results reveal a main effect of *Pronoun* on *Vocoid* ($\chi^2(1) = 43.59$, $p < 0.001$), *Voicing* ($\chi^2(1) = 32.46$, $p < 0.001$), and *Piece* ($\chi^2(1) = 79.23$, $p < 0.001$). *Pronoun* did not have a significant effect on the duration of the other intervals measured (all have p-value > 0.001). Figure 151 shows the mean durations in ms of the phonetic events measured in the two datasets.

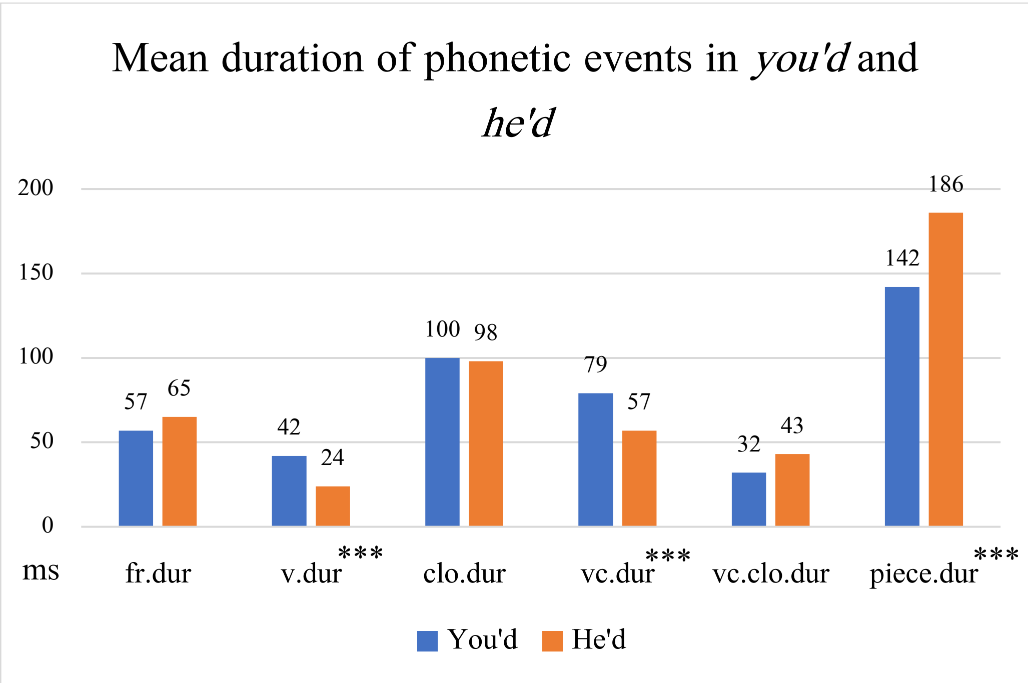


Figure 151. Mean duration of the friction in onset (fr), the vocoid (v), the closure (clo), voicing (vc), voicing during the closure (vc.clo) and the piece (piece) in *you'd* (blue) and *he'd* (orange).

The spectral qualities of the friction in onset position were compared. The results of the five mixed effects models reveal a significant effect of *Pronoun* on *CoG* ($\chi^2(1) = 13.07$,

$p < 0.001$), and *Skewness* ($\chi^2(1) = 12.2$, $p < 0.001$), but not on *Amplitude* ($\chi^2(1) = 0.56$, $p = 0.45$), *SD* ($\chi^2(1) = 3.9$, $p = 0.048$), and *Kurtosis* ($\chi^2(1) = 5.71$, $p = 0.017$). Figure 152 shows the CoG of the friction in *you'd* and *he'd*.

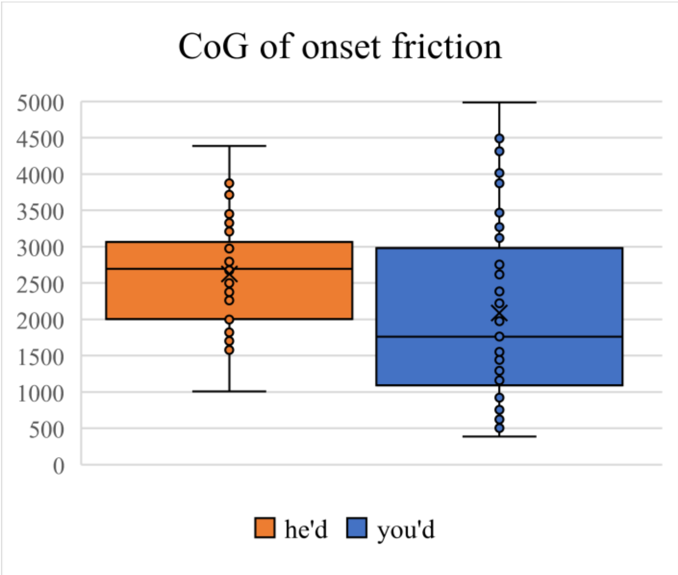


Figure 152. Distribution of the Centre of Gravity of the voiceless friction at the beginning of *he'd* and *you'd*.

The results of the analysis suggest that, despite the initial portion of friction possibly creating confusion between *you* and *he*, its spectral properties actually help maintain the contrast. When there is no friction at the beginning of *you*, the contrast is maintained by the presence of glottal friction in onset of *he* versus its absence. When the friction is present at the beginning of *you*, its acoustic properties maintain the contrast from the friction in onset of *he'd*. The perception experiment in chapter 7 investigates the intelligibility of these two paradigms.

5.3. Summary and discussion

Starting from the observations of the qualitative analysis reported in Chapters 3 and 4, the aim of this chapter was to identify the acoustic features that maintain the contrast between paradigms that convey grammatical information such as the tense of the verb. The high degree of reduction of some instances of pr + aux observed in the data, raised the question of how the contrast between pairs of paradigms that convey crucial grammatical information is maintained in reduced speech. This chapter reported on the acoustic analysis of several pairs of paradigms along several parameters such as duration of phonetic events, spectral moments of fricatives and formant dynamics. The comparison between the acoustic features of the paradigms in each pair indicates that they differ on two or more parameters, suggesting that the contrast is maintained even in reduced speech.

The first section reported the comparison of pronouns combined with the cliticised forms of *will* and *would*. The analysis of the paradigms *we'd* and *we'll*, and *she'd* and *she'll* showed that they differ in the duration of the closure and the resonances of the piece. The pr + *'d* are characterised by a longer closure (relative to that of pr + *'ll*) and clear resonances; while pr + *'ll* are characterised by a shorter closure (relative to that of pr + *'d*) and dark resonances. The paradigms *I'd* and *I'll* exhibit the same patterns for closure duration and resonances and, in addition, they differ also in the duration of the vocoid: *I'll* is characterised by a significantly longer vocoid duration than *I'd*.

The observation that the paradigms with the cliticised non-syllabic form of *will* are characterised by dark resonances, and the paradigms with the cliticised non-syllabic form of *would* are characterised by clear resonances, indicates that the non-syllabic forms exhibit the opposite polarity in the resonances compared to the syllabic forms of *will* and *would*. The syllabic form of *will* is characterised by clear resonances, while the syllabic form of *would* is characterised by dark resonances. The non-syllabic forms *'ll* and *'d* exhibit the opposite clear versus dark opposition. This feature confirms that the syllabic and non-syllabic forms belong to two separate systems, and that a polysystemic approach is suitable for their analysis.

To analyse the contrast between the cliticised forms of the present tense and the non-cliticised forms of the past tense of the auxiliary verb BE, the pairs of paradigms *she's* and *she was*, *you're* and *you were*, and *we're* and *we were* were compared. Although the past tense of the auxiliary BE does not have a cliticised form in the spelling, the analysis reported in Section 4.2 showed that reduced instances of pr + *was* and pr + *were* are monosyllabic and are thus comparable with the cliticised forms pr + *'s* and pr + *'re*. As mentioned in Section 1.2, this claim was made also by Ogden (1999). All three pairs compared in this chapter exhibit a significantly different vocoid duration. In addition, the paradigms *she's* and *she was* differ in their resonances: *she's* is characterised by clear resonances, while *she was* is characterised by dark resonances. However, a large number of instances of *she's* (32%) do not have an identifiable vocoid, and are characterised by palato-alveolar friction alone (and sometimes weak voicing). This can be a discriminating feature in itself, since no instance of *she was* is characterised by the absence of the vocoid. Surprisingly, the palato-alveolar friction in onset does not exhibit the same contrast in resonances when calculated across the entire datasets of *she's* and *she was*. However, the analysis of two smaller datasets of reduced instances of *she's* without a vocoid and *she was* with a shorter duration, revealed that the instances of *she's* in which there is no vocoid, the mean CoG of the palato-alveolar friction is much higher and significantly different from the mean CoG of the palato-alveolar friction of reduced instances of *she was*. This feature suggests that the contrasting resonances that are present in the vocoid in less reduced instances of *she's* and *she was*, are retained by the palato-alveolar friction in more reduced instances of *she's* and *she was.* From an articulatory point of view, a possible explanation for this feature is in the coarticulation and overlap of articulatory gestures. The absence of a time-delimited vocoid in reduced instances of *she's* does not necessarily imply that the gesture of the vocoid is not articulated. In fact, the qualitative analysis reported in Chapter 3 and 4 showed that even in highly reduced tokens, the gestures for the apparently missing sounds can be articulated. The difference between the mean CoG in the entire dataset of *she's* and the mean CoG of the subset of *she's* without a vocoid is 226 Hz. While the difference between the mean CoG in the entire dataset of *she was* and the mean CoG of the subset of reduced tokens of *she was* is 79 Hz. This suggests that the degree of coarticulation is higher in instances in which the vocoid is not identifiable as a separate segment. From a perceptual point of view, the absence of a vocoid with contrasting resonances in reduced instances of *she's* can have implications for word recognition. That is, if the contrast between *she's* and *she was* is not conveyed by the resonances in the vocoid, there must be other acoustic features that maintain the contrast. Since the palato-alveolar friction is the only apparent acoustic element remaining in reduced instances of *she's*, the friction must convey the contrast with *she was*, hence the increased difference in the CoG of the palato-alveolar friction in onset of *she's* and *she was*.

As for the contrast between *you're* and *you were*, and *we're* and *we were*, the present and past tense of each pair differ significantly in vocoid and, therefore, piece duration. Surprisingly, the formant dynamics of the vocoids in the present and past tense of each pair are very similar. This is surprising because from the qualitative analysis reported in Chapter 4, it emerged that *you're* is characterised by palatality and openness, while *you were* is characterised by palatality and labiality. However, this is not reflected in the formant dynamics. The acoustic correlate of labiality (or lip-rounding in *you were*) is a lower frequency of F3. This is not apparent in the formants of *you're* and *you were*. As for the pronoun *we*, auditorily, *we're* sounds less rounded and more open than *we were*. In the formant dynamics, there is a difference in F1 that indicates that *we're* has a more open articulation, but F3 is less distinct (see Figure 145). The main difference between the formant dynamics of the present and past tense is in the magnitude of the gesture as seen in the larger movement of F2 in pr + *were* compared to the F2 movement in pr + *'re*. The analysis and comparison between the reduced subsets of each paradigm display an even more striking similarity, due to the decreased magnitude of gestures, which correlates with a flatter F2 in pr + *were*. This raises the question of whether the difference in duration alone is distinctive enough to maintain the contrast between the present and past tense of BE in the pairs *you're* and *you were*, and *we're* and *we were*.

The analysis of the contrast between *has* and *had* reported in this chapter focussed on those instances of pr + *'d* in which the phonological plosive is realised with friction. The presence of friction at the same place of articulation of the friction in pr + *'s*, raised the question of whether and how the contrast between the two paradigms is maintained in reduced speech. The comparison between the acoustic features of *he's* and *he'd* revealed that the spectral qualities of the voiced alveolar frictions in coda position differ between the two paradigms.

Two pronouns were also analysed. The presence of voiceless friction with palatal quality at the beginning of *you*, and the similarity between the formant dynamics of the vocoid in the pronouns *he* and *you* in *he'd* and *you'd* tokens, raised the question of whether and how the contrast between *you* and *he* is maintained. The acoustic analysis of the spectral qualities of the friction in onset revealed that they are significantly different in duration and two spectral moments, including CoG.

Table 25 summarises the acoustic features that differ between the pairs of paradigms compared in this chapter.

Present ~ past contrast (except *)	Pairs of paradigms analysed	Durational differences		Resonance/spectral differences
will ~ would	*we'd ~ we'll*		closure	clear ~ dark
	I'd ~ I'll	vocoid	closure	clear ~ dark
	she'd ~ she'll		closure	clear ~ dark
is ~ was	*she's ~ she was*	vocoid		clear ~ dark
are ~ were	*you're ~ you were*	vocoid		
	we're ~ we were	vocoid		
has ~ had	*he's ~ he'd*		closure	spectral qualities of coda friction
he ~ you *	*he'd ~ you'd*	vocoid		spectral qualities of onset friction

Table 25. Summary of the contrasts analysed, the pairs of paradigms compared for each contrast, and the acoustic features that differ between the items in each pair, divided into two columns: one for the differences in duration, and one for the differences in the resonances or spectral qualities of the friction. All contrasts are between present and past tense, except for *he* versus *you* (marked with an asterisk *).

The analysis reported in this chapter has led to a few observations that add to our knowledge and understanding of reduction, and how the contrast between paradigms conveying grammatical information is maintained in reduced speech.

First of all, it emerged that the items in each pair of paradigms differ in the duration of one of their phonetic events – either the vocoid or the closure, or both in the case of *I'd* and *I'll* – and in their resonances. That is, two main parameters maintain the distinction between paradigms: duration and resonances.

The analysis of the formant dynamics in subsets of pr + aux datasets with instances with a shorter vocoid duration than the mean vocoid duration across the dataset confirmed that in reduced (shorter) tokens the magnitude of gesture is reduced. This feature is visible in the formant dynamics which show that the movements of the first three formants, especially F2, are flatter and more restricted in shorter instances than in the whole dataset of each pr + aux. In the literature, reduction has often been described as decreased magnitude of gesture and the analysis presented here confirms this description.

Interestingly, the decrease in the magnitude of gestures does not lead to a loss or weakening of the contrast between paradigms. In fact, if we consider the difference between the frequencies of F2 in the present and past tense of the paradigms as the acoustic correlate of the polarity of the resonances (clear versus dark), the contrast between the resonances is not only maintained in reduced instances but even enhanced. The evidence for this claim can be found in the comparison between the formant dynamics of reduced tokens and the formant dynamics measured in the whole dataset of the paradigms under observation. For instance, it was reported that the difference between the frequency of F2 at the beginning of *we'd* and *we'll*, calculated in the entire datasets (which include reduced and unreduced tokens) is 571 Hz. While the difference between the frequency of F2 at the beginning of *we'd* and *we'll*, calculated in the subsets of reduced instances is 718 Hz. This means that the distinction between F2 in the reduced subsets is greater than in the whole datasets. This suggests that the polarity between the resonances of the pieces, not only are maintained in reduced speech, but are even enhanced, and as a consequence, the contrast between the paradigms might be enhanced too.

Finally, the analysis reported in Section 5.2.3.1 showed that the resonances in *she's* and *she was* in the whole datasets differ only in the vocoid, while the palato-alveolar friction in onset does not exhibit a significant difference in CoG.[13] The presence of a high number of

[13] Although the other three spectral moments (SD, skewness, and kurtosis) measured in the palato-alveolar friction of *she's* and *she was* differ significantly, the CoG is treated here as the acoustic correlate of the resonances.

instances of *she's* realised without a vocoid raised the question of whether the contrast is maintained in reduced instances. The analysis of the spectral moments of the palato-alveolar friction in reduced instances of *she's* and *she was* revealed that the difference in CoG is greater and significant in reduced items. This again indicates that in reduced instances the contrast between paradigms is not lost. This feature can be explained as an increase in the degree of coarticulation, or temporal realignment, and can have implications for the correct identification of reduced word forms.

On the basis of the observations and the acoustic analysis reported in this chapter, Chapters 7 and 8 investigate the perception of reduced instances of pr + aux. In particular, Chapter 8 investigates whether the acoustic features identified in this chapter (duration and resonances), are available to perception and are used by listeners for the correct identification of contrasting paradigms. Chapter 6 describes three aspects of reduction observed during the analysis carried out in Chapters 3 to 5: utterance beginnings or silent articulations (Section 6.1), the variability of reduced speech (Section 6.2), and reduction in repetition (Section 6.3).

6. Aspects of variability in reduced speech

Despite the growing number of studies on reduced speech, our knowledge of the nature of reduction is still limited (Ernestus and Smith, 2018). The auditory and spectral analysis carried out on the data collected led to the observation of some aspects of the reduced pr + aux that can advance our understanding of the phenomenon of reduction. This chapter illustrates three aspects of reduction that emerged during the qualitative analysis described in Chapters 3 and 4, and that can help shed light on the phenomenon. Firstly, it was observed that in several utterances, the beginning of the vibration of the vocal folds is preceded by a portion of voiceless friction. As mentioned in previous chapters (see e.g. Sections 3.1.5 and 3.1.6, and Section 5.2.5.1), this is a widespread phenomenon in the data collected, and it occurs across speakers and paradigms. This feature is analysed in more detail in Section 6.1. A second aspect described in this chapter is the variability observed in reduced items. It was noticed that not all the elements of a piece are reduced in the same way. In particular, two patterns, or types of variability, emerged. The first type occurs when the sounds in a piece are articulated with different degrees of articulatory effort. For example, in the piece *it's*, the friction can be well-articulated, while the vocoid is reduced. In this thesis, this type of variability is termed 'vertical variability' and it is described in Section 6.2.2.1 The second type of variability occurs when the magnitude of gesture is reduced in a piece that does not exhibit temporal reduction, or vice-versa. For example, the piece *I'd* can exhibit a long vocoid duration while the formants are flat suggesting that there is no tongue movement during the production of the vocoid. In this thesis, this type of variability is termed 'horizontal variability' and it is described in Section 6.2.2.2. The third aspect analysed in this chapter is the relationship between reduction and repetition. Although the correlation between repetition and reduction has been investigated before (e.g.

Fowler, 1988; Fowler and Housum, 1987), in the data collected the expected correlation was not observed. This raised the question of what the role of repetition in the degree of reduction of repeated items is, and led to a deeper investigation of this issue. The relationship between repetition and reduction is analysed in Section 6.3.2.

6.1. Utterance beginnings or 'silent articulations'

During the auditory analysis and spectral observation of the data collected in the production study, it was noticed that several pieces which have a vocoid in utterance-initial position exhibit a portion of voiceless friction with clear formant structure before the onset of voicing. Figure 153 shows an instance of *we're* in which the vocoid is preceded by a period of weak friction with clear F2 and F4, and a faint F3.

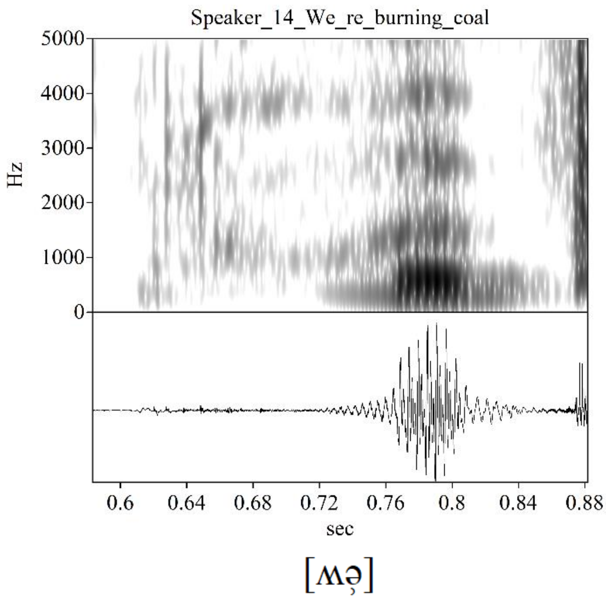


Figure 153. Spectrogram and waveform of an instance of *we're* with a period of weak friction before voicing starts.

This feature occurs frequently in the data collected, and although it is related to the beginning of speech, it can also shed light on the phenomenon of reduction. The following sections look at the literature on speech beginnings (Section 6.1.1), then describe the phonetic characteristics of the phenomenon observed in the data (Section 6.1.2), and conclude with a plausible explanation for the phenomenon observed (Section 6.1.3).

6.1.1. Background – beginning to speak

When we prepare to speak, a series of events occur in the vocal tract, including swallowing, breathing in, and the separation of the articulators. As a result of these events, the acoustic output can include noises such as percussives, clicks, and friction (Scobbie, Schaeffler and Mennen, 2011; Schaeffler, Scobbie and Schaeffler, 2014; Palo, Schaeffler and Scobbie, 2014, 2015). Any pre-speech noise is usually considered extra-linguistic and often ignored in speech analysis (Ogden, 2013). However, some pre-speech noises can reveal useful information about speech articulation and production. In the context of reduced speech, they can provide information about the mechanisms behind articulatory reduction and the perceptual salience of remaining (and audible) acoustic events.

Research on pre-speech noises has looked at the movements of the articulators before any audible speech noise is generated (Rasskazova, Mooshammer and Fuchs, 2019). Scobbie et al., (2011) looked at the noises produced in the vocal tract before the acoustic onset of speech. Using articulatory data, they found that these noises are typically generated by swallowing, the breathing activity before speech, and the movement of the articulators in preparation for the speech act. They looked in particular at the acoustic 'spikes' occurring before speech and found that they are all due to the opening of the vocal tract. When the articulators are pulled apart in preparation of speech, they can produce clicks or click-like noises at any place of articulation, from labial to lingual (Scobbie et al., 2011). They also measured the timing of the pre-speech noises and found that they occurred a quarter to half a second before the onset of the acoustic linguistic speech. Schaeffler et al. (2014) analysed the time between articulatory movements and the acoustic output using UTI and video imaging. They found that the articulators, such as the lips, the tongue, and the jaw, move into place well "before anything becomes audible" (Schaeffler et al., 2014: 379). For this reason, they called these movements "silent articulatory movements" (Schaeffler et al., 2014: 379). The specific characteristics of these pre-speech noises seem to be linked to the articulatory settings of the speaker and to be language-specific (Wilson, 2006; Mennen, Scobbie, de Leeuw, Schaeffler and Schaeffler, 2010).

The pre-speech noises such as clicks, percussives, and in-breaths described in the literature occur also in the data presented here. From auditory and visual observations, most of these noises can be attributed to the movements of the articulators in preparation for speech. In addition to these noises, a portion of friction was noticed in several instances in which the utterance started with a voiced sound. This portion of friction is much closer to the beginning of the utterance than the other articulatory noises observed in the data and described in the literature. Although Scobbie et al. (2011) mentioned the presence of "extended frication" in some of their data, they claimed that the origin of the frication was the in-breath. This is not the case in the data analysed here, in which the in-breath occurs much earlier than the voiceless friction produced just before voicing begins (see Figure 156 for an example). The following section describes the phonetic features of the friction observed before vocalisation starts at the beginning of speech.

6.1.2. Phonetic characteristics of utterance-initial friction

In the data collected, in several utterances characterised by an initial vocoid, before the vocal folds start vibrating, a period of voiceless friction can be observed in the spectrograms. One of the main features of this friction is to be characterised by clear formant dynamics that correspond to the gestures of the voiced sound in utterance-initial position. Figure 154 shows an example of friction before the beginning of *we'd*.

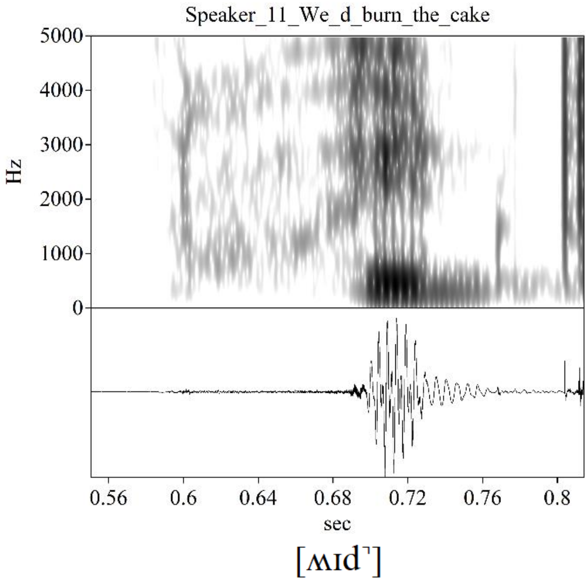


Figure 154. Spectrogram and waveform of an instance of *we'd* with weak friction at the beginning before the onset of voicing.

Figure 154 shows the spectrogram and waveform of *we'd* in the sentence *we'd burn the cake*. The vocoid is quite short (duration 24 ms). However, before the onset of voicing, F2 is already visible and moves from a low frequency (989 Hz) to a higher frequency (1948 Hz). This F2 movement is what we would expect to see when the tongue moves from a back position to a front position in the oral cavity. It is the movement we would expect to see in the spectrogram of a (voiced) labial-velar approximant. In the spectrogram in Figure 154, when voicing starts, the formant structure of the vocoid indicates that the tongue is already in a close front position. The waveform displays only a very weak friction. The resulting auditory impression is that of a weak voiceless labial-velar friction followed by a short near-close near-front vocoid [ʍɪ]. In this instance, the articulation before voicing starts can be heard only through careful listening in a quiet environment. In several instances in the dataset, the aperiodic energy produced before the vibration of the vocal folds is visible in the spectrogram, but it is not audible. For this reason, this phenomenon is referred to as 'silent articulation' in this thesis. The term is borrowed from Schaeffler et al. (2014), but it is used here in a narrower sense: it refers only to the articulation that is visible in the formants of the friction that precedes the onset of voicing. Although in the majority of cases the silent articulations are not audible, there are also instances in which they are audible. Figure 155 shows an example in which the friction before voicing at the beginning of *you've* is strong and audible.

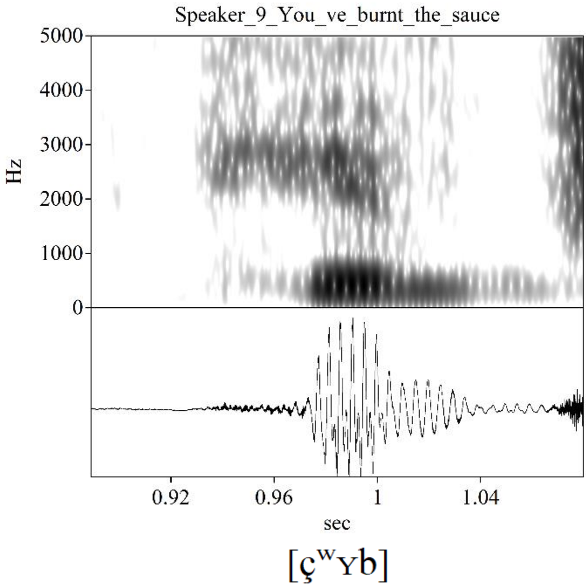


Figure 155. Spectrogram and waveform on an instance of *you've* with friction before the onset of voicing.

Figure 155 shows the spectrogram and waveform of an instance of *you've* in the sentence *you've burnt the sauce*. In this instance, the friction is audible and it is characterised by a high front (or palatal) quality with a hint of labiality. The position of the second formant also suggests that the tongue is in a palatal position during the friction. In this instance, both F2 and F3 are visible before the vocal folds start vibrating. At the beginning of the aperiodic energy, F2 is 2360 Hz and F3 is 2848 Hz. Both formants drop slightly but remain high for the duration of the friction, which is 40 ms. Both F2 and F3 decrease further during the vocoid and their values at the end of the vocoid are 2018 Hz and 2534 Hz respectively. As was described in Chapter 5 (Section 5.2.5.1), the phonetic characteristics of instances of *you* with palatal friction in initial position make them very similar to the pronoun *he*.

The utterance initial friction described here occurs at the beginning of the pronouns that begin with a vocoid, namely *I*, *it*, *we* and *you*. It does not occur at the beginning of the pronouns that have an obstruent sound in onset position, namely *he*, *she* and *they*. It occurs across speakers and repetitions, although some speakers tend to produce it more often than others. This period of aperiodic energy must not be confused with in- and out-breaths. In-breaths usually occur much earlier than the silent articulations and are louder, longer, and clearly identifiable in the spectrograms as in-breaths. Figure 156 shows an example of in-breath and silent articulation in the same token before the utterance *it'd burn the grass*.

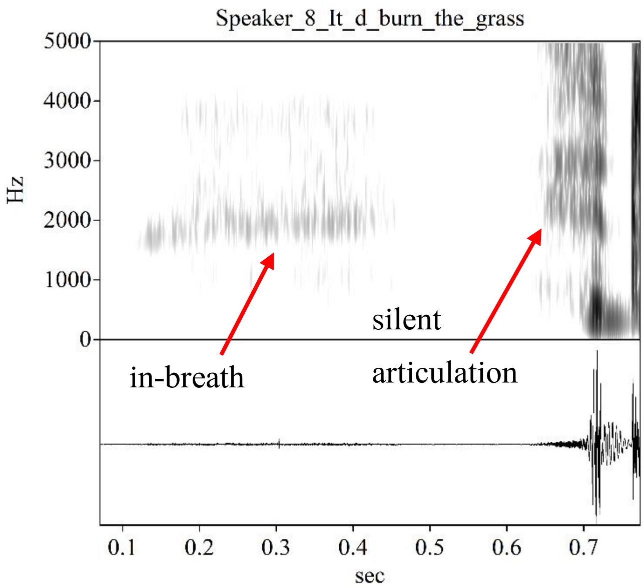


Figure 156. Spectrogram and waveform of an in-breath and a silent articulation.

Figure 156 shows an in-breath before *it'd* at the beginning of the sentence *it'd burn the grass*, produced by S8. The friction of the in-breath is weaker and only at a frequency range around 2000 Hz. The friction of the silent articulation is more distributed along the frequency scale, from around 2000 Hz to 13 kHz (only shown up to 5 kHz in Figure 156). The breath inspiration starts 583 ms before the beginning of voicing. Its duration is 322 ms. The duration of the silent articulation is 66 ms, after which voicing starts.

One of the questions raised by the presence of silent articulations, is where the origin of the friction is. There are two possibilities. First, the source of noise can be at the place of articulation, in the case of *you*, the place of articulation is the hard palate. Alternatively, the source of friction can be in the glottis. In this second case, glottal friction is generated in the larynx due to exhalation, and then shaped in the oral cavity by the articulators being in place for the first sound of the utterance. The shape of the oral cavity gives the glottal friction the palatal quality.

To test these hypotheses, the features of the silent articulations at the beginning of the pronoun *I* were analysed. The rationale for analysing the pronoun *I* is that it begins with an open articulation, which is unlikely to produce friction. The pronoun *I* begins with an open and back vocoid. Figure 157 shows an example of silent articulation at the beginning of *I'd*.

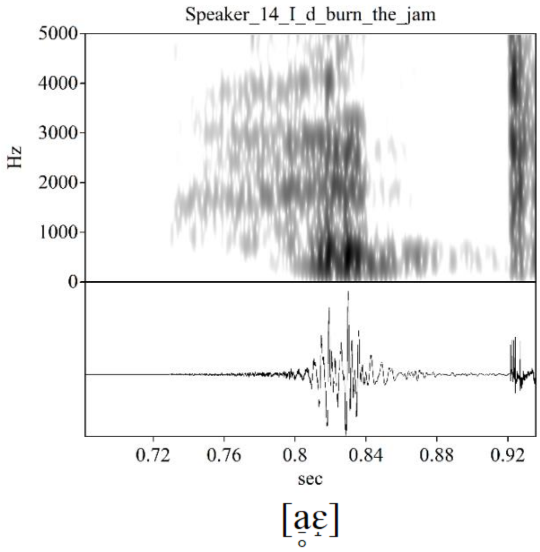


Figure 157. Spectrogram and waveform of an instance of *I'd* with initial friction before the onset of voicing.

Figure 157 shows an instance of *I'd* in the sentence *I'd burn the jam* uttered by S14. The long portion of friction (81 ms) at the beginning of the piece has an open front quality. Although the first formant is not usually visible in silent articulations, in this instance there are some traces of F1 in the spectrogram. Figure 158 shows a zoomed-in spectrogram of the same token with arrows pointing at the visible F1 points. The first three formants were measured at these points.

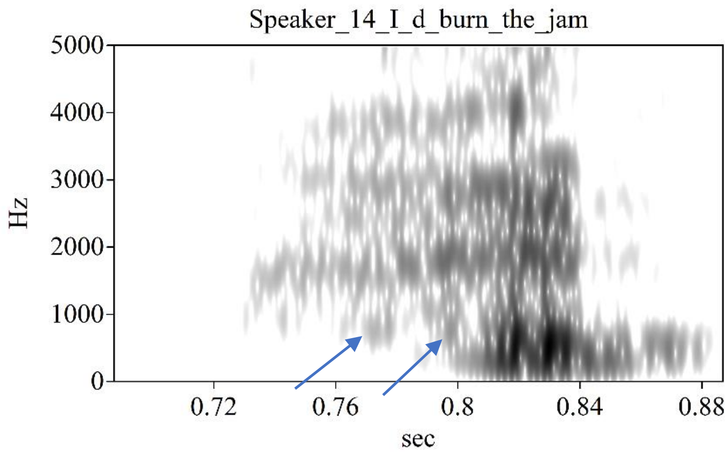


Figure 158. Spectrogram of the silent articulation at the beginning of *I'd*.

The first three formants were measured at the friction midpoint (left arrow) and friction end (right arrow). The values of the formants are reported in Table 26.

Hz	Friction midpoint	Friction end
F3	2785	2782
F2	1736	1819
F1	829	733

Table 26. Formant values of the first three formants measured at midpoint and at the end of the friction at the beginning of *I'd* in *I'd burn the jam* uttered by S14.

The high F1 at friction midpoint confirms that the tongue is in a low position at the silent articulation midpoint and that it moves to a slightly higher position before the onset of voicing. A low tongue position like the one articulated for the production of the first part of the vocoid in *I*, and shown in this token, cannot produce friction at the place of articulation in the oral cavity. The tongue is too distant from the passive articulators to create any turbulence. The occurrence of silent articulations at the beginning of the pronoun *I* suggests that friction is produced in the glottis and then shaped in the oral cavity.

Interestingly, the friction of the silent articulations can occur simultaneously to a glottal stop. The pronoun *it* is often realised with creaky voice and often starts with a glottal stop. Although from an articulatory point of view, a silent articulation and a glottal stop are not mutually exclusive, it might be assumed that the movement of the articulators during a glottal stop might not be audible and might not be visible in the spectrograms due to the glottal closure. That is, even in the event of moving articulators, any movement would not be audible or visible on the spectrogram during a glottal stop due to the complete closure in the glottis. However, there are a few instances in which both silent articulation and glottal stop are audible and visible in the spectrograms. Out of 240 instances of *it*+ aux, 22 have both silent articulation and glottal stop (N = 22/240, 9%). Figure 159 shows an example of initial friction followed by the release of a glottal stop and the beginning of creaky voice.

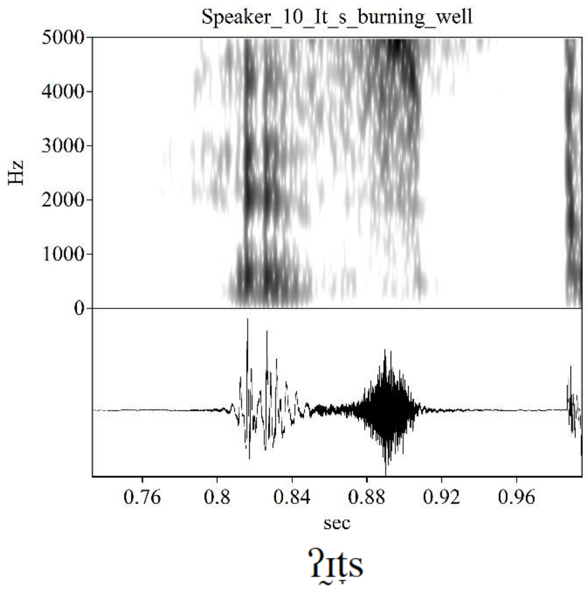


Figure 159. Spectrogram and waveform of an instance of *it's* with a short period of friction before the beginning of creaky voice.

Figure 159 shows an instance of *it's* in the sentence *it's burning well* uttered by S10. In this instance, the silent articulation (46 ms) is audible only at a careful listening.

A glottal burst can also be superimposed on the silent articulation, as in the instance in Figure 160.

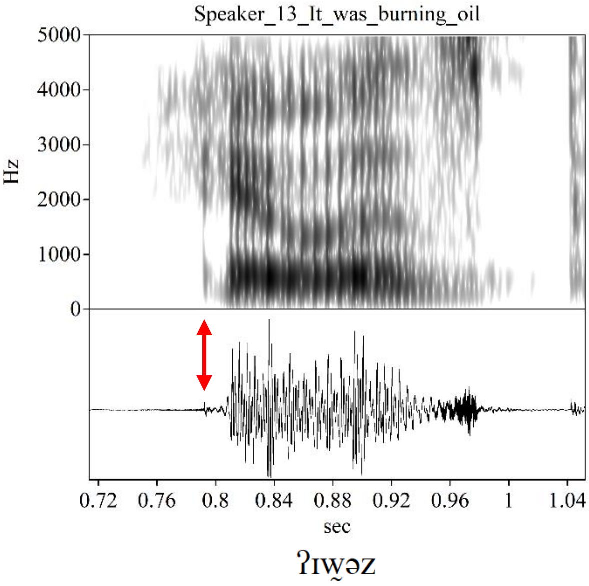


Figure 160. Spectrogram and waveform of an instance of *it was* with initial voiceless friction and a glottal stop (indicated by the red arrow).

Figure 160 shows an instance of *it was* in the sentence *it was burning oil* uttered by S13. In this piece, a portion of voiceless friction occurs simultaneously to a glottal stop. First the friction begins, then the glottal stop is released with a burst indicated by the red arrow just before 0.8 seconds.

Looking at the pronoun *I* subset, which also contains instances of initial glottal stop, all speakers produce silent articulations; they all produce initial glottal stops too; and all except one speaker produce simultaneous silent articulation and glottal stop. Table 27 shows a breakdown of the production of silent articulation, glottal stop and both, by speakers in the pronoun *I* subset.

	Silent Art. (1)	Silent Art. %	Glottal Stop (2)	Glottal Stop %	Both (3)	Both %	Total N sentences per speaker
Speaker 2	8	27	17	57	0	0	30
Speaker 3	16	55	9	31	1	3	29
Speaker 6	5	17	16	53	2	7	30
Speaker 7	16	53	4	13	2	7	30
Speaker 8	20	69	5	17	2	7	29
Speaker 9	11	39	1	4	2	7	28
Speaker 10	8	27	5	17	1	3	30
Speaker 11	14	54	2	8	10	38	26
Speaker 13	8	27	13	43	1	3	30
Speaker 14	22	79	3	11	3	11	28
Speaker 15	2	7	16	59	6	22	27
Total N/ mean %	130	41%	91	29%	30	9%	317

Table 27. Count and percentage of instances of silent articulation (1), glottal stop (2) and both (3) per speaker in the pronoun *I* subset.

The percentage of sentences per speaker in which a silent articulation is visible indicates that the feature is common across speakers.

6.1.3. Discussion

Although the silent articulations might not be audible, and therefore not relevant for speech communication, their observation is important for a better understanding of reduction and speech production in general. The most plausible explanation for the occurrence of silent articulations is the temporal reorganisation of two events – the vibration of the vocal folds and the articulatory gesture in the oral cavity. The observation of formant structure superimposed on weak voiceless friction before the beginning of the vibration of the vocal folds, suggests that the gestures for the articulation of the sounds at the beginning of the sentence start before the onset of phonation. In other words, the tongue and the lips start moving before the glottal folds start vibrating. This creates a period of time in which the sound gesture is articulated, but the voice source has not been activated yet. The temporal realignment of phonetic events seems to be one of the main causes of acoustic reduction in the data analysed. The fact that the acoustic output of the gestures is visible at a close inspection of the spectrograms, suggests that the gesture is articulated. However, the acoustic output of the gesture is not audible (or hardly audible) due to the lack of voicing from the glottis, which starts later. The resulting acoustic output is that of a short voiced sound, which is interpreted as an instance of reduction. For example, Figure 161 shows an instance of *we'll* in which the acoustic output is a short vocoid with flat formants suggesting little or no tongue movement. The interpretation of the short vocoid as an instance of reduction is justified by the duration of the vocoid (temporal reduction) and the decreased magnitude of gesture (articulatory reduction). However, the presence of the aperiodic energy before voicing starts, and the formant structure visible in the friction indicate that the gesture was articulated and it lasted longer than the resulting short sound.

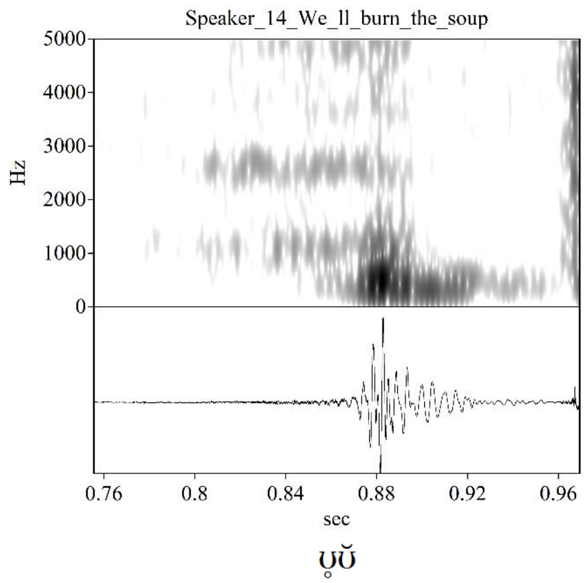


Figure 161. Spectrogram and waveform of an instance of *we'll* characterised by a short vocoid and a period of friction with visible formants before the onset of voicing.

The observation of silent articulations at the beginning of utterances raises the question of the implications for speech intelligibility. If the articulation of the first sound is not audible, the temporal realignment of articulatory gestures and voicing might result in the loss of information at the beginning of speech. However, if the articulation of the first sound superimposed on voiceless friction is audible, it might aid intelligibility by making the first sound available to perception for longer. For example, in the instance shown in Figure 161, the pr + aux piece is extremely reduced: while the vibration of the vocal folds lasts for 47 ms, the duration of the vocoid is only 15 ms. At the end of the vocoid, the lip closure for the hold phase of the bilabial stop begins. The quality of the vocoid is that of a close back rounded vocoid. The silent articulation is audible and it has the close back rounded quality of the vocoid. The duration of the silent articulation is 69 ms. This means that the presence of the silent articulation extends the duration of the sound from 15 ms to 84 ms so that it is available to perception for longer.

The temporal realignment of phonetic events and the presence of friction at the beginning of an utterance can have implications for talk-in-interaction. In face-to-face conversations, even if the silent articulations are not audible, the preparation to speak could be visible and affect turn-taking. It has to be remembered that if there is no superimposed friction, silent articulations are not only inaudible, but also not visible in the spectrograms. In these cases, only an articulatory investigation of the speaker's vocal tract during speech using equipments such as electropalatography or UTI could shed light on the nature of this phenomenon. Although the silent articulations seem to be a case of low-level phonetic effect, the observation of this phenomenon provides some insight into the nature of reduction.

6.2. Types of variability in reduced speech

Each pr + aux piece is a complex bundle of phonetic features and events. However, not all the phonetic features and events of each pr + aux are reduced in the same way or to the same extent. In fact, it was observed in the data collected that, even in highly reduced tokens, some features are not only unreduced, but fully articulated or even hyper-articulated. This section investigates this aspect by describing some instances found in the dataset. In particular, two types of variability were observed and are described in this section. The first type is the occurrence of both under-articulated and apparently over-articulated phonetic features and events in the same piece. In this thesis, this type of variability is referred to as 'vertical variability' and it is covered in Section 6.2.2.1. The second type of variability refers to the lack of positive correlation between reduction in duration and reduction in the magnitude of gestures in the same piece. In this thesis, this type of variability is referred to as 'horizontal variability' and it is described in Section 6.2.2.2.

6.2.1. Background – variability

Temporal reduction and articulatory reduction are often considered the main features of reduced speech (e.g. Gahl et al., 2012). However, the relation between these two aspects of reduction is still not clear. A few studies that have focussed on the relationship between articulation and speaking style, have shown that temporal and articulatory reduction are not necessarily correlated. Van Son and Pols (1992, 1993) showed that the spectral qualities of vowels uttered by the same speaker reading a text at normal and fast rate were not affected by the speaking rate, which means that fast tempo does not always imply vowel centralisation. Warner and Tucker (2011) showed that consonant reduction is pervasive also in careful speech, with plosives and flaps in intervocalic position being realised as approximant even in careful read speech. The lack of a correlation between duration and articulation has been noticed also in conversation analysis research. In a study on other-initiated repair utterances in American English conversations, Curl (2005) observed that the temporal compression of a repeated sentence did not always correspond to a less clear or compressed articulation, indicating that temporal and articulatory reduction are not always positively correlated.

The features described in this chapter under the term vertical variability have not been extensively covered in the literature on reduction. A possible reason for this gap is that most studies of reduction carry out quantitative analyses of counts of missing segments and syllables, rather than detailed qualitative analyses of the wide range of variations in the realisations of the target items. A welcome exception is a recent investigation by Ernestus and Smith (2018) on the realisation of the Dutch word *eigenlijk*. They carried out a qualitative and quantitative analysis of 159 tokens of *eigenlijk*. In their analysis, they took into account several features as measures of reduction. Some of these features included the number of syllables, the word duration, and the clarity of articulation. They found "surprisingly little clear correlation between the different indices of reduction" (Ernestus and Smith, 2018: 30). For example, they observed clear articulations in monosyllabic tokens, and reduced articulations in longer, disyllabic, and trisyllabic tokens.

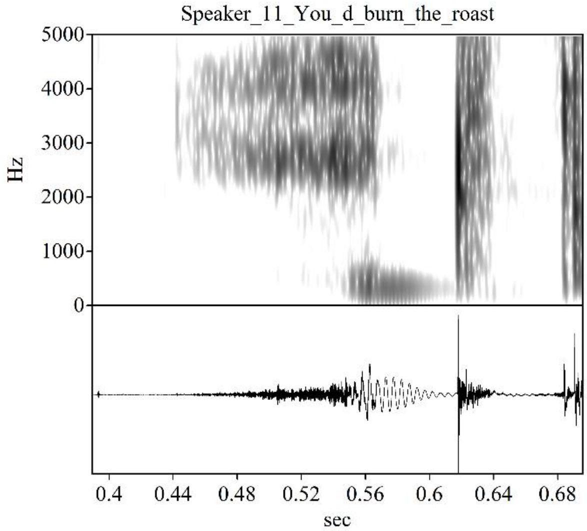
This section describes some of the variability observed in the data. In particular, two aspects are covered: the occurrence of both hypo- and hyper-articulated gestures in the same piece (vertical variability), and the correlation (or absence of) between articulatory and temporal reduction (horizontal variability).

6.2.2. Types of variability in the present data

The next two sections describe two aspects of variability found in the data that can shed light on the phenomenon of reduction. The first has been termed ‘vertical variability’, and is described in Section 6.2.2.1. The second has been termed ‘horizontal variability’, and is described in Section 6.2.2.2.

6.2.2.1. **Vertical variability**

In some instances of pr + aux, part of the piece is reduced while part of it is not. For example, in a sequence of a vocoid and a contoid, one of them can be highly reduced, and the other one can be not only unreduced, but fully articulated or even hyper-articulated. A pr + aux combination that well illustrates this point is pr + *’d*. As described in Section 4.1.1 (on aux *‘d*), the most frequent realisation of the contracted auxiliary *’d* in the dataset is unreleased. The unreleased realisation is due to the presence of a following plosive in onset of the verb *burn*. This is a common phenomenon in English connected speech: when two plosives are in succession, the first one is unreleased (Cruttenden, 2008). A realisation in which both adjacent plosives are released with a burst might be considered a case of careful articulation or hyper-articulation, as the release of the first plosive is unnecessary. In the dataset collected, there are a few instances of reduced pr + *’d* pieces in which the pronoun is highly reduced, while the contracted auxiliary *’d* is realised with a complete closure followed by a burst. Figure 162 shows an instance of *you’d*.



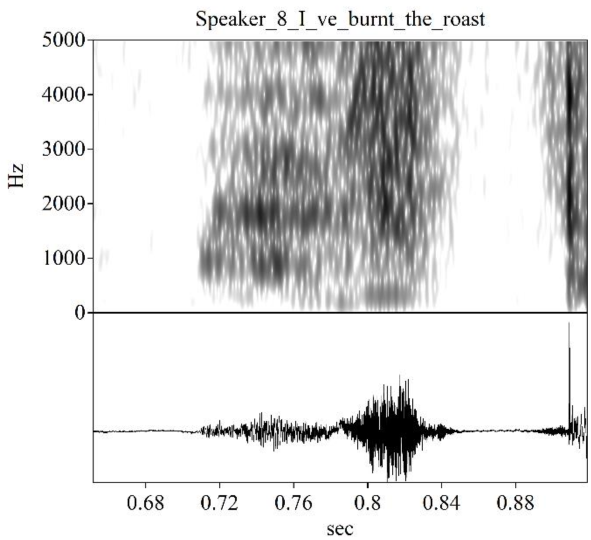
[i̥ĭd̥]

Figure 162. Spectrogram and waveform of an instance of *you'd* in which the vocoid is reduced, but the contoid is not.

Figure 162 shows an instance of *you'd* uttered by S11. There are several phonetic features that indicate that this piece is (at least partially) reduced. Firstly, the vocoid is very short. Its duration is only 15 ms, indicating temporal reduction. The vocoid is preceded by a portion of voiceless friction with palatal quality (duration 109 ms). Secondly, the formants are flat throughout the duration of the friction and the vocoid, indicating reduction in the magnitude of the gestures. Thirdly, the auditory impression suggests that there is no lip rounding throughout the piece. Therefore, the first part of the piece (the vocoid) is realised without tongue movement (indicated by the flat formants), without lip rounding (from auditory observation) and with little voicing. All these features indicate that the vocoid is reduced. On the other hand, the contoid is fully articulated with a complete closure at the place of articulation and a sharp burst at release. We would expect the first of two adjacent plosives to be unreleased and this is also the most common realisation of /d/ in pr + *'d* in the dataset. The articulation of the contoid with an abrupt release can be considered an instance of hyper-articulation in this context and it contrasts with the reduced articulation of the vocoid.

This means that two sounds that belong to the same piece are articulated with contrasting articulatory effort. Although it is not surprising to see phonetic features and events reduced in different degrees and ways in the same piece, it is to a certain extent unexpected to find under-articulated and over-articulated features occurring next to each other. That is, to find adjacent sounds belonging to the same piece produced with differing degrees of magnitude of gesture. Here, this feature is referred to as 'vertical' variability in that the contrasting degrees of articulatory effort occur in temporal succession and not simultaneously. Despite being intertwined, two adjacent sounds display two opposing magnitudes of gesture.

This phenomenon is not restricted to pr + *'d*. It occurs across pr + aux combinations and speakers. Figure 163 shows a similar example in an instance of *I've*.



[ḁ̆v]

Figure 163. Spectrogram and waveform of an instance of *I've* in which the vocoid is reduced, but the labio-dental friction is well-articulated.

Figure 163 shows an instance of *I've* in the sentence *I've burnt the roast* by S8. The vocoid at the beginning of the piece is voiceless although the formants can be observed in the friction. The formants are flat, indicating that there is little tongue movement. However, the labio-dental contoid is characterised by a strong friction. The amount of labio-dental friction produced in this instance is unusual for this auxiliary, which is normally characterised by a short portion of weak friction. The mean duration for the labio-dental friction in the pr + aux *I've* across speakers and repetitions is 30 ms. The duration of the labio-dental friction in the token shown in Figure 163 is 60 ms. This clearly suggests that there is a degree of hyper-articulation which contrasts with the reduction of the vocoid.

This type of articulatory variation occurs also the other way round: the vocoid can be fully articulated while the contoid is reduced. Figure 164 shows an instance of *I've* that is characterised not only by a long vocoid, but also by evident formant movement, which is unusual in most instances of *I*, as reported in Section 3.1.1 (on pronoun *I*). However, in this instance of *I've*, the labio-dental friction is apparently absent. In fact, the vocoid ends abruptly with the closure of the lips for the hold phase of the bilabial contoid that follows. In this case, while the vocoid is seemingly unreduced, the friction is highly reduced to apparent deletion.

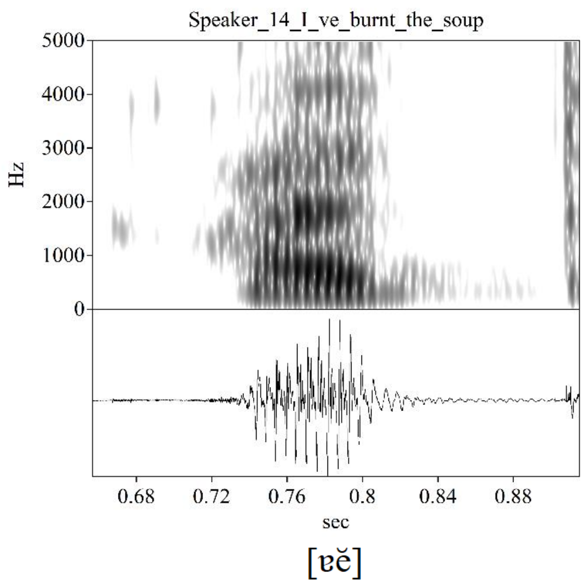


Figure 164. Spectrogram and waveform of an instance of *I've* with a well-articulated vocoid but no labio-dental friction.

Figure 164 shows an instance of *I've* in the sentence *I've burnt the soup* uttered by S14. The duration of the vocoid in this piece is 58 ms, while the mean duration of the vocoid in *I've* across speakers and repetitions is 37 ms. The absence of labio-dental friction in *I've* is not uncommon and it is found in 9 out of 55 tokens (N = 9/55, 16%) across speakers and repetitions.

The examples presented in this section highlight the wide range of variability in the degree of reduction of sounds in the same piece. They show the discrepancy in reduction and magnitude of gesture between adjacent sounds in the same piece. It has to be recalled that these sounds are not found across prosodic boundaries. On the contrary, they constitute a phonological and processing unit. This aspect raises the question of whether reduction is an automatic process or whether speakers are in control of the degree of reduction of their speech. This issue will be discussed in more detail at the end of the next section, which describes another type of variability in the way the pr + aux pieces are reduced.

6.2.2.2. **Horizontal variability**

The second aspect described in this section is the discrepancy observed in some tokens between the reduction in duration and the reduction in the magnitude of gestures. The duration of sounds and stretches of speech and the magnitude of gestures are the two main parameters analysed in reduced speech. These two parameters can be both reduced in any given sound or piece – in this case they are positively correlated: both duration and gestures are reduced. Alternatively, only one of them is reduced and the other is not – in this case they are not correlated: there is reduction in duration or articulatory gesture, but not both. In some cases, these two parameters can be negatively correlated: one of them is reduced while the other is not only unreduced, but 'enhanced' or hyper-articulated. This feature is termed ‘horizontal variability’ in this thesis because the two parameters (duration and magnitude of gesture) occur simultaneously in the same sound or piece. Figure 165 exemplifies the rationale behind the choice of naming the two features described in this section ‘vertical’ and ‘horizontal’ variability.

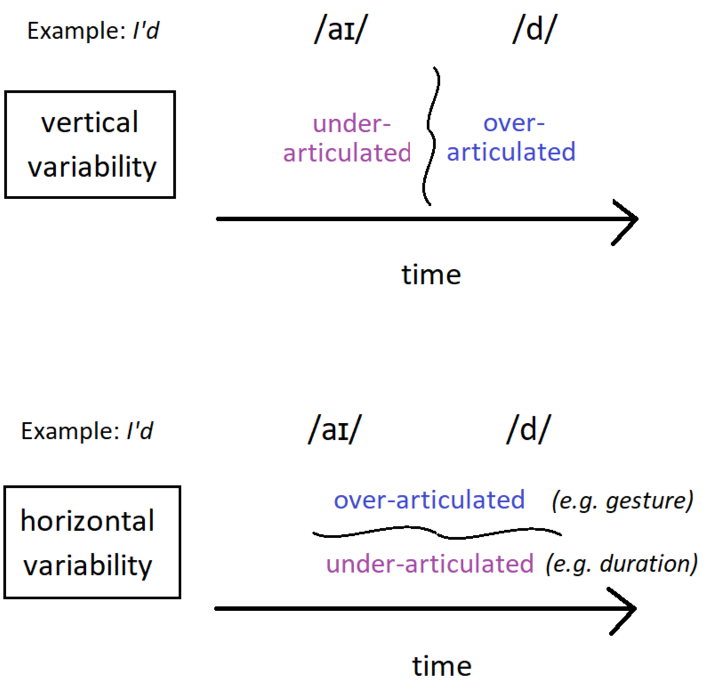


Figure 165. Representation of ‘vertical’ and ‘horizontal’ variability. The curved lines indicate that there are not always well-defined boundaries between sounds (vertical variability) and between parameters (horizontal variability).

Intuitively, we would expect a positive correlation between temporal and gestural compression. For example, in a shorter vocoid there is less time for the articulators to make a large movement. However, this is not always the case. For example, the instances in Figure 166 support the view that in a shorter vocoid the same gesture undergoes temporal compression and is articulated more quickly. Figure 166 shows two instances of *you’ve*. As described in Section 3.1.5 (pronoun *you*) the pronoun *you* can be realised as a vocoid with a single quality throughout its duration, as shown by flat formants in the spectrograms and through auditory observations. This feature can be explained as a decrease of magnitude of gesture. However, this feature is not necessarily accompanied by temporal reduction.

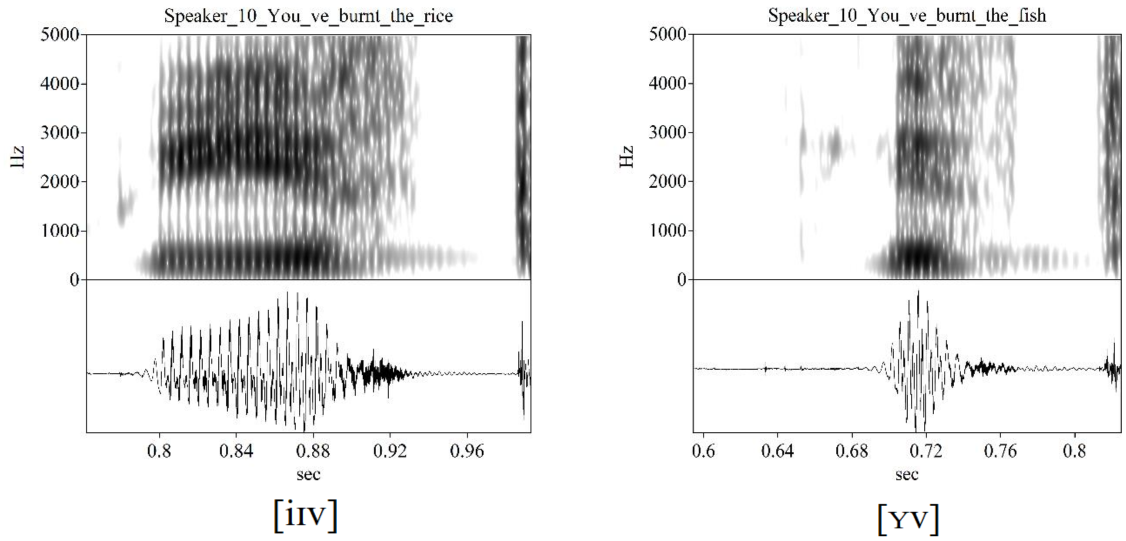
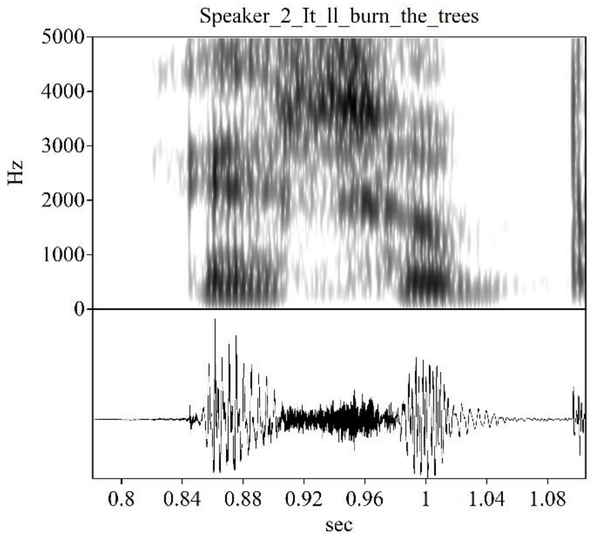


Figure 166. Spectrogram and waveform of two instances of *you've* produced by the same speaker (S10). The instance on the left is characterised by a long vocoid and relatively flat formants. The instance on the right is characterised by a short vocoid and visible formant movement. Both instances are displayed in the same time window (230 ms).

Figure 166 shows two instances of *you've* uttered by the same speaker (S10). The instance on the left is characterised by a long vocoid (duration 85 ms) with relatively flat formants considering its duration. The formant ranges are: 57 Hz for F1, and 579 Hz for F2. The instance on the right is characterised by a short vocoid (duration 23 ms) with some F2 and F3 movement. The formant ranges are: 28 Hz for F1, and 363 Hz for F2. This means that the token on the right shows 62% F2 frequency range in 27% of the duration of the token on the left. Moreover, from auditory observations, the instance on the right exhibits some lip-rounding, while the instance on the left does not. In other words, the temporal reduction does not correspond to the same amount of gestural reduction.

Figure 167 shows another example of discrepancy between temporal reduction and gestural reduction.



[ʔɪt̞ə]

Figure 167. Spectrogram and waveform of an instance of *it'll* characterised by friction instead of a complete closure for /t/.

Figure 167 shows an instance of *it'll* in the sentence *it'll burn the trees* by S2. Both vocoids are fully voiced. However, the phonological plosive /t/ is realised without a complete closure but with friction throughout its duration (81 ms). The tongue blade moves toward the target passive articulator but instead of blocking the air, a narrow constriction is created which produces a turbulent airstream. This is a case of reduction in articulatory effort, in which the target sound – in this case a stop – is undershot. However, there is no temporal reduction. In the *it'll* subset, 22 tokens are realised with a complete closure and an abrupt release in form of a burst (N = 22/41, 54%), 7 with a complete closure and gradual release of friction (N = 7/41, 17%), 7 are realised only with friction throughout their duration and no complete closure (N = 7/41, 17%), and 5 with creaky voice between the two vocoids (N = 5/41, 12%). This is a case of both vertical and horizontal variability. Vertical variability: the vocoids are not visibly reduced, while the contoid in intervocalic position is. Horizontal variability: the contoid exhibits articulatory reduction but not temporal reduction.

6.2.3. Discussion

The examples described in this section show that reduction is not a uniform phenomenon but displays a wide range of variation even in units of speech such as the pieces analysed here. As mentioned in Section 6.2.2.1, these pieces are not merely sequences of sounds, but they constitute phonological and processing units. Yet, within a piece, the degrees and types of reduction are highly variable.

These aspects raise the question of whether reduction is an automatic process or whether the speaker is in control of the degree of reduction. From the examples reported in Section 6.2.2.1, it seems that reduction cannot be viewed simply as an automatic process linked to a decrease in articulatory effort. A reduced articulatory effort would affect the entire piece in a more distributed way. The features discussed in this section point to view reduction as a rather controlled phenomenon in which the speaker has a choice of what to reduce and to what extent. This is in line with findings by Ernestus and Smith (2018) on the reduction of the Dutch word *eigenlijk*.

A possible explanation for the phenomena of vertical and horizontal variability is that the various phonetic features and events in a piece have different roles in perception. Some features might be more salient than others and need to be retained in the acoustic signal. However, the data analysed do not uniformly provide evidence to support this hypothesis. For example, with reference to Figure 162 in which /d/ is released with a burst, in 87% of pr + *'d* analysed across the dataset (N = 421/484), the stop in coda position is unreleased. This suggests that the release of the phonological plosive does not need to be perceived for the correct identification of the pr + *'d*.

To conclude this section, more research on these two aspects of variability in reduced speech is needed. Qualitative analyses of the type carried out in this research could provide valuable information on the motivations and mechanisms of reduction, and advance our understanding of the relation between the production and perception of reduced speech.

6.3. Reduction in repetition

As described in Chapter 2, to trigger a high degree of reduction, in the production experiment speakers were asked to utter each sentence a number of times. The rationale behind this choice was that repetition has been found to trigger articulatory reduction (see e.g. Fowler and Housum, 1987; Aylett and Turk, 2004; Pluymaekers et al., 2005b). However, the mechanisms behind the relationship between repetition and reduction are still not clear. On the basis of the quantitative analysis carried out on the data collected, the relationship between duration and repetition is further investigated in this section.

6.3.1. Background – reduction and repetition

Research by Fowler and colleagues (see e.g. Fowler, 1988; Fowler and Saltzman, 1993; Fowler, Levy and Brown, 1997) found that the relationship between repetition and reduction is linked to the informativeness of repeated items. While the first mention of a word or item constitutes 'new' information, any subsequent repetition constitutes 'old' information. As such, words carrying old information are shorter and less intelligible (Fowler and Housum, 1987). However, Fowler and colleagues suggest that intelligibility and reduction are more closely linked to redundancy and predictability than repetition. Bard, Sotillo, Anderson, Doherty-Sneddon and Newlands (1995) add that the redundancy of a word is not necessarily related to what is redundant for the listeners, but what is redundant for the speaker. They found that when repeating the same information, speakers produce less intelligible words even when the listeners had not heard the first mention of the same word. In contrast with the theory that redundancy and predictability are the most important factors in triggering reduction (Fowler et al., 1997), Hawkins and Warren (1994) argue that rather than the informativeness of the repetition, it is the lack of "informational focus" that leads repeated words or utterances to be less intelligible. According to Hawkins and Warren, sentence accent influences intelligibility and new information is normally accented. In other words, it is the presence versus absence of accent that affects intelligibility, rather than the "informational status" of a word (Hawkins and Warren, 1994: 494). In their experimental study, Hawkins and Warren (1994) found that repetition did not influence intelligibility, but that presence or absence of accent did. Contrasting results have been found by Aylett and Turk (2004). According to their research, the number of times a word has been previously mentioned influences the duration of the word. In other words, unaccented repetitions of the same word exhibit durational differences.

The current data can help shed light on some of the issues raised by previous research. Firstly, one of the issues (raised by Hawkins and Warren, 1994) is whether phrasal stress is the main factor in triggering reduction. In the present data, the items analysed never carry phrasal stress or accent; they are always unstressed. The current data is also preceded and followed by the same context, so the context cannot be a factor in influencing reduction. Secondly, Fowler (1988) found that shortening did not occur in word lists. She claims that shortening (as a measure of reduction) is linked to informational redundancy, and therefore can occur only in a communicative context. The data analysed here are more similar to a word list than to a conversational context. Finally, Aylett and Turk (2004) claimed that reduction is incremental with the number of repetitions. The present data provides a good opportunity to test the relationship between repetition and duration.

6.3.2. Reduction and repetition in the present data

In most studies of the relation between repetition and reduction, segmental duration and word duration are used as the main parameters to quantify the degree of reduction. For this reason, the duration of the vocoid and the duration of the piece in several paradigms were compared in the five repetitions of each piece. The paradigms analysed are the combinations of the pronoun *you* with the six auxiliaries analysed in this thesis.

Figure 168 shows the mean duration of the vocoid in the five repetitions of each of the six *you*+aux combinations. The mean duration for each repetition is calculated across speakers. For each repetition, 11 utterances were recorded and measured.

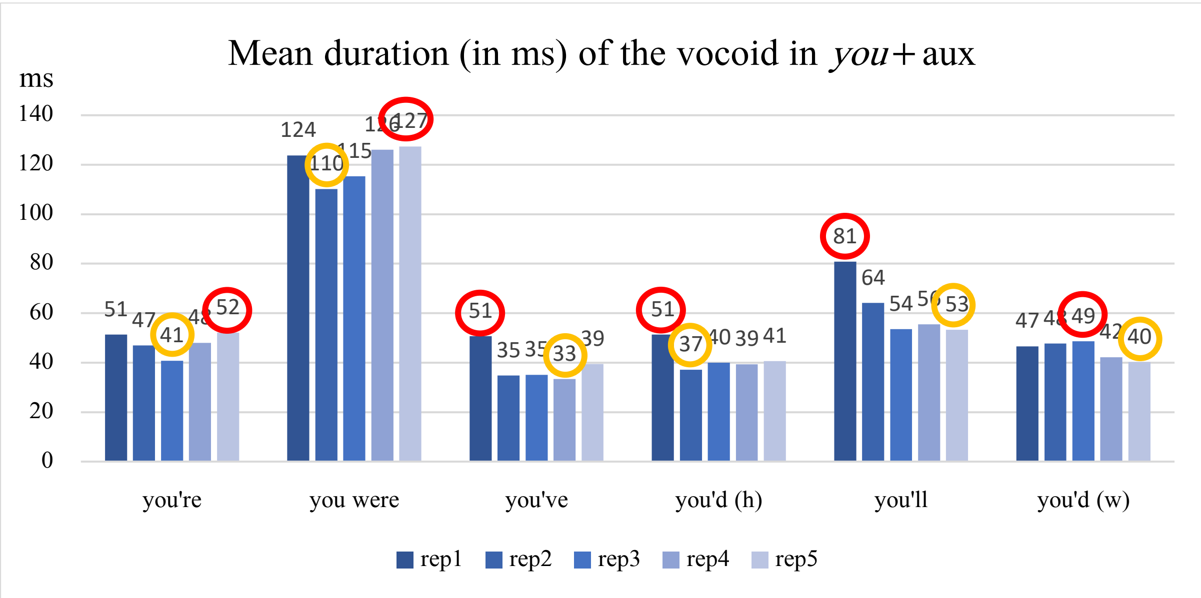


Figure 168. Mean duration across speakers of the vocoid in the six *you*+aux combinations, divided by repetition (in shades of blue), from repetition 1 (rep1) on the left to repetition 5 (rep5) on the right. In each *you*+aux, the red circle indicates the repetition with the longest vocoid duration, and the orange circle indicates the repetition with the shortest vocoid duration.

Figure 168 shows the mean duration of the vocoid in the five repetitions of each of the six *you*+aux paradigms. The red circles indicate the longest vocoid duration between the five repetitions and the orange circles indicate the shortest vocoid duration between the five repetitions of each paradigm. The first point to notice is that there is no clear pattern of decreasing vocoid duration from the first to the fifth repetition. In three out of six paradigms (50%), the first repetition is characterised by the longest vocoid duration. However, in two cases (*you're* and *you were*) the longest vocoid duration is in the fifth repetition. In two paradigms, the shortest vocoid duration is in the second repetition; in two paradigms it is in the fifth repetition; in one paradigm it is in the third repetition; and in one paradigm it is in the fourth repetition. The first repetition has never the shortest vocoid duration.

Six Spearman's Rank correlation tests – one for each of the paradigms *you're*, *you were*, *you'll*, *you'd* (*would*), *you've* and *you'd* (*had*) – were run to test whether there is a correlation between vocoid duration and number of repetitions in each paradigm. Five of the six tests revealed no correlation between vocoid duration and repetition (all $ps > 0.03$, $rhos < +0.11$). However, there was a significant negative correlation between vocoid duration and repetition for the paradigm *you'll* ($rho = -0.33$, $p = 0.03$).

The comparison between the five repetitions was carried out also by looking at the mean duration of the entire piece in the six *you*+aux paradigms. Figure 169 shows the mean duration of the pieces divided by repetition and paradigm.

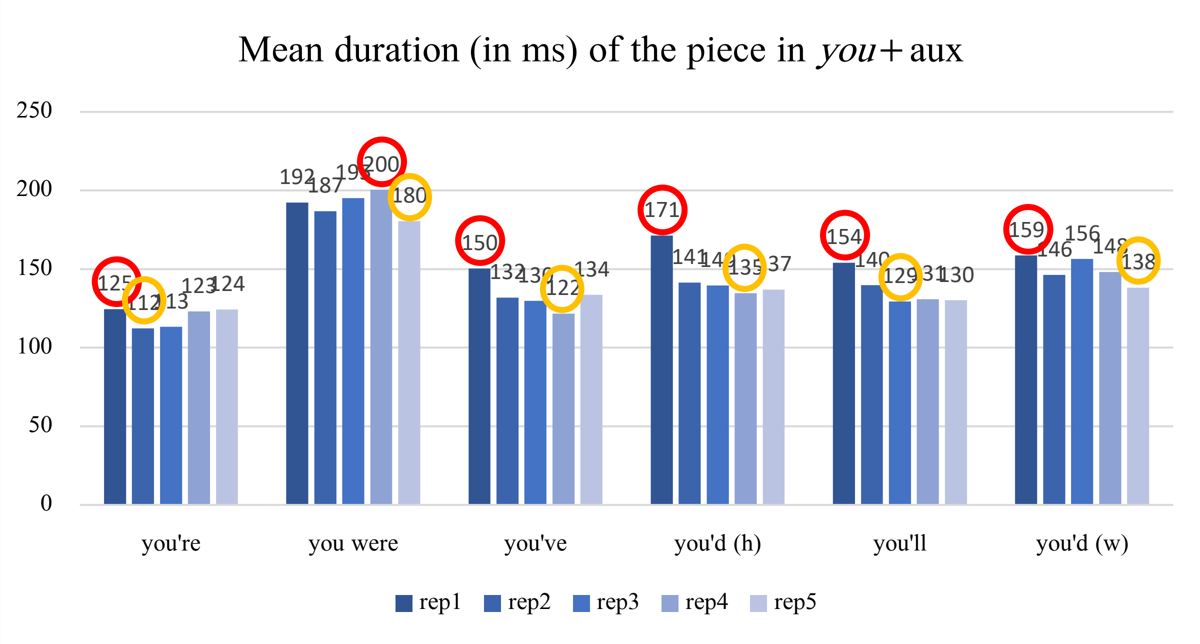


Figure 169. Mean duration across speakers of the piece in the six *you*+aux combinations, divided by repetition (in shades of blue), from repetition 1 (rep1) on the left to repetition 5 (rep5) on the right. In each *you*+aux, the red circle indicates the repetition with the longest piece duration, and the orange circle indicates the repetition with the shortest piece duration.

Figure 169 shows the mean duration across speakers of the five repetitions of each of the six *you*+aux recorded. In five out of six paradigms (83%) the first repetition is the one with the longest piece duration. However, the repetitions from 2 to 5 do not exhibit a clear pattern of decreased piece duration from the second to the fifth repetition. In two paradigms the last repetition is the shortest.

The same analysis and comparison was carried out for other pr+aux paradigms with similar results. While the comparison of the vocoid duration shows a less clear pattern, the comparison of the piece duration in several paradigms shows that the first utterance is the one with the longest duration in most paradigms (between 67% and 83% depending on pronoun). However, besides the first utterance, the subsequent repetitions do not exhibit a well-defined pattern even in the comparison between the piece durations. The fifth repetition is the shortest in only 20% of paradigms.

6.3.3. Discussion

The analysis presented in this section indicates that there is no clear correlation between repetition and duration of reduced word forms, except for the first utterance. The first mention of a piece exhibits the longest piece duration in the majority of cases. This is in line with the theory that new information (such as the first time a piece is uttered) is more fully articulated, while old information is indeed reduced. However, the subsequent repetitions (from 2 to 5) do not exhibit a decrease in duration for each repetition of the same item. In other words, repetition does not lead to incremental reduction in duration from the second to the fifth repetition of the same piece. Any one of the four repetitions from two to five can be the shortest one, and in several cases the last repetition is not the shortest of the five repetitions.[14]

Although the data analysed do not have a communicative function, in that the utterances were read from a screen and not part of an interaction, the results of the analysis on the correlation between repetition and duration support the theory that new information is more fully articulated, while old information is reduced. These results also support the theory that the informativeness of the utterance is not linked to the listeners, but to the speaker, as claimed by Bard et al. (1995). In the present data, the absence of a listener during the recording process suggests that the old information is old for the speaker alone. The relation between duration of repetition in the present data also suggests that reduction is linked to the informational focus of the sentence, rather than to the repetition itself. That is, the degree of reduction does not increase with the number of repetitions, as demonstrated by the

---

[14] Speakers did not know how many repetitions there were for each sentence, as some sentences were repeated three times, some four times, and some five times.

lack of a pattern between the duration of the repetitions from two to five. The results of the analysis presented in this section support only the theory that the first mention of a word or item is more fully articulated because it constitutes new information, while subsequent repetitions can be variably reduced as they do not provide any new information.

6.4. Summary and discussion

This chapter covered three aspects of reduction observed in the data that can advance our knowledge and understanding of the phenomenon of reduction. Firstly, Section 6.1 showed that there is more to speech beginnings than just the noise produced by in-breaths and the opening of the articulators in preparation of speech. The temporal reorganisation of voicing and articulatory gestures means that often the articulatory gestures occur before voicing starts. This phenomenon has been termed 'silent articulation' in this thesis because in most cases the initial friction is not audible. However, the presence of instances of audible initial friction raises the question of whether the presence of this friction can be problematic for speech intelligibility and spoken word recognition. For example, the discrimination between the pronouns *you* and *he* could be problematic in instances in which both of them are realised with voiceless close front (or palatal) friction followed by a close front vocoid. However, the intelligibility of speech beginnings could be facilitated by the availability of information in the acoustic signal for a longer period of time, provided that the silent articulations are audible (or visible in the case of face-to-face conversation). The issue of the intelligibility of the contrast between *he* and instances of *you* with audible friction is further investigated in the perception experiment reported in Chapter 7 (Section 7.3.2).

Section 6.2 reported on some of the variability found in reduced speech. In particular, two aspects were described. Firstly, the variability in the amount of articulatory effort found in adjacent sounds belonging to a single phonological and processing unit (vertical variability). Secondly, the lack of correlation between the two main aspects that traditionally characterise reduced speech: temporal reduction and articulatory reduction (horizontal variability). These findings have implications for the description of reduction as an automatic speech process or a speaker-controlled mechanism. The examples reported in this chapter suggest that at least some aspects of reduction can be controlled by the speaker and reduction cannot be considered a fully automatic process.

Finally, Section 6.3 shed some light on the correlation between reduction and repetition. While repetition has long been known to trigger reduction, the relationship between these two aspects is still unclear. Several other factors are at play in the relationship between reduction and repetition, such as sentence stress and phonological context, new and old information, redundancy and predictability. In the data analysed in this thesis some of these factors were controlled for. For example, none of the items carried sentence stress and they could not be predicted by the context. In the present data the first mention of an item is typically longer than the subsequent repetitions, suggesting that new information is more fully articulated than old information. However, the lack of incremental reduction in subsequent repetitions suggests that the informational focus has a more prominent role in the degree of reduction than the number of repetitions.

The next two chapters (Chapters 7 and 8) report on a perception experiment to test the intelligibility of reduced speech and the perceptual salience of selected acoustic parameters.

7. Perception of reduced pr + aux

The analyses reported in Chapters 3 to 5 showed that the identity of pr + aux in reduced instances is maintained by fine phonetic detail, and that the difference between paradigms is manifest in sometimes subtle phonetic detail (see e.g. Section 5.2.5.1). The presence of such fine-grained phonetic detail raises the question of whether this detail is available to perception, and whether listeners are sensitive to it, and can make use of it for the correct interpretation of reduced words. From the acoustic and statistical analyses reported in Chapter 5, it emerged that pairs of paradigms do differ from an acoustic point of view. However, the acoustic and statistical analyses do not provide information on the availability of the subtle phonetic differences between paradigms to perception, or whether listeners use them for the correct identification of reduced pr + aux. The following two chapters (Chapter 7 and 8) investigate the intelligibility of reduced pr + aux (Chapter 7), and the perceptual salience of the fine phonetic detail identified in the previous chapters (Chapter 8). In particular, they address the following questions:

- Are highly reduced instances of pr + aux intelligible when heard in a limited semantic context? Is the fine phonetic detail that distinguishes pairs of paradigms sufficient to maintain the contrast in perception? (Chapter 7).
- Are some acoustic parameters more salient than others in spoken word recognition? In other words, do listeners rely on one parameter (e.g. duration) more than on others for the correct interpretation of reduced word forms? (Chapter 8).

Since the experiment aimed to address two questions, the description of the methodology and the analysis of the results have been divided into two chapters. For ease of reference, the two parts of the experiment will be referred to as Experiment A and Experiment B. The aim of Experiment A was to test the intelligibility of highly reduced word forms. For this experiment, 37 natural non-manipulated stimuli were used. The aim of Experiment B was to test the perceptual salience of selected acoustic parameters. For this experiment, 54 manipulated stimuli were used. The stimuli for Experiment A and Experiment B were randomised together with 21 additional stimuli not discussed in this thesis for a total of 112 stimuli. All 112 stimuli were presented to the participants in a single experimental session. Experiment A is reported in this Chapter 7. Experiment B is reported in Chapter 8. This chapter is structured as follows: Section 7.1 looks at the literature on the perception of reduced speech. Section 7.2 describes the methodology of the perception experiment. Section 7.3 reports on the results of Experiment A. Section 7.4 closes the chapter with the discussion on the results.

7.1. Perception of reduced speech

A number of studies have shown that, despite the high degree of reduction found in everyday speech, communication remains mostly intelligible (see e.g. Ernestus et al., 2002; Manuel et al., 1992; Ernestus, Kouwenhoven and van Mulken, 2017; Mitterer, Yoneyama and Ernestus, 2008; Janse et al., 2007; Mitterer and Ernestus, 2006; van de Ven et al., 2012). Several studies have shown that listeners rely on multiple types of cues for the correct identification of reduced speech (Ernestus et al., 2002; van de Ven et al., 2012), including the semantic and syntactic context, the acoustic cues remaining in the signal, and the situational context. Most studies on the intelligibility of reduced speech have focussed either on the recognition of reduced word forms in relation to the context given (see e.g. Ernestus et al., 2002; Janse and Ernestus, 2011; van de Ven et al., 2012;) or the reduction of single segments and how it affects intelligibility (see e.g. Mitterer and Ernestus, 2006; Janse et al., 2007; Warner et al., 2009; Mitterer, 2011; Pitt et al., 2011).

7.1.1. Acoustic cues versus the context

The perception experiment described in this chapter aims to shed light on the perception of reduced word forms. The auditory analysis of the pr + aux collected in the production experiment highlighted the presence of highly reduced items. The question is whether these items can be correctly identified when they are heard in a limited semantic context.

A number of studies have focussed on the relationship between the recognition of reduced word forms and the context given. Research has found that reduced speech is unproblematic when heard in context (see e.g. Ernestus et al., 2002). Ernestus et al. (2002) investigated the role of the phonetic, semantic, and syntactic contexts in spoken word recognition in casual speech. They categorised reduced word forms according to three degrees of reduction (low, middle, and high) and presented them in three different types of context: isolation (no context at all), limited context (minimal phonetic context of surrounding vowels and intervening consonants) and full context. They found that all three categories of reduced words were recognised in the full context; that the words with a low and medium degree of reduction were recognised well also in the limited context; and that the highly reduced forms needed their full context to be correctly recognised. They concluded that the recognition of reduced word forms is correlated to the degree of reduction and the context given. The higher the degree of reduction, the more the phonetic, semantic and syntactic contexts are needed to identify the words.

Research by van de Ven et al. (2012) also showed that listeners can better identify reduced pronunciation variants if they are presented in their preceding and following contexts. They found that the acoustic cues available in the preceding and following context helped listeners identify omitted reduced words. In their research, van de Ven et al. (2012) investigated the role of the semantic and syntactic context, and the role of the acoustic cues present in the surrounding context of reduced pronunciation variants. In four experiments they tested whether participants could predict a missing (reduced) word from the semantic and syntactic contexts alone (without auditory stimuli) and from the semantic and syntactic context together with the acoustic context of surrounding words. Listeners could better predict the missing reduced word when the auditory stimuli was presented together with the orthographic stimuli than from the orthographic stimuli alone, despite the high speech rate. Importantly, when presented with contrasting semantic and acoustic context, listeners relied more on the acoustic cues in predicting the missing words, rather than relying on the semantic and syntactic context. This demonstrates that listeners can and do make use of the information available in the acoustic signal for the correct interpretation of reduced speech (van de Ven et al., 2012).

The importance of acoustic information in the correct interpretation of reduced speech has been reported also by Janse and Ernestus (2011). In their experiments, participants who heard the preceding and following contexts were able to identify reduced words better than participants who read orthographic versions of the preceding and following contexts. This is surprising only if we consider that the auditory stimuli was reduced, while the semantic and syntactic contexts were clearly spelt out. That is, participants were better at identifying reduced words when they were presented in their reduced acoustic context than when they were presented in a “very clear orthographic representation” of the semantic and syntactic context (Janse and Ernestus, 2011: 12). As a result, Janse and Ernestus claim that “the semantic/syntactic cues are by themselves of little help” (Janse and Ernestus, 2011: 12).

To summarise, previous research has shown that the context needed for the correct interpretation of reduced speech is correlated to the degree of reduction – the more a word is reduced, the more context is needed. However, it has also been shown that listeners are sensitive to small changes in the acoustic signal and that they use them to identify the presence of absence of a word in a semantic and syntactic context in which both options are equally likely.

In the perception experiment reported here (Experiment A), the reduced tokens are presented in a limited semantic context. Only the following phonetic and syntactic contexts are available. Listeners hear the entire sentence. From the syntactic structure of the sentence, listeners know that the reduced tokens they have to identify are a pronoun and an auxiliary, but they do not have any other contextual information. For example, there is no preceding context indicating who the subject of the sentence is, or in which tense the action occurs. For the correct identification of the pronoun and auxiliary, listeners can rely only on the acoustic cues available in the signal.

7.1.2. Research question

This chapter investigates the intelligibility of reduced combinations of a pronoun and an auxiliary in sentence-initial position. It reports on a perception experiment that addresses the following question:

> RQ: Are highly reduced instances of pr + aux intelligible when heard in a limited semantic context? Is the fine phonetic detail that distinguishes pairs of paradigms sufficient to maintain the contrast in perception?

The hypothesis is that a fair amount of reduction does not impede intelligibility as reduced word forms retain crucial acoustic information that are available to perception. However, extreme reduction might not be intelligible without a preceding semantic context.

7.2. Methodology

In order to test the intelligibility of reduced pr + aux a two-category forced-choice word identification task was created. The stimuli were selected from the corpus of data collected in the production experiment. This section describes the methodology of the experiment, starting with the selection of the speakers (Section 7.2.1) and the selection of the stimuli (Section 7.2.2).

7.2.1. Selection of speakers

The data collected during the production study was used in the perception experiment. Two issues affected the decision of which speakers' material to include in the stimuli. On the one hand, listeners can become habituated to speaker-specific production (Smith, 2015). On the other hand, speaker-specific characteristics affect every aspect of speech (Smith, 2015), including reduction (Hanique, Ernestus and Boves, 2015), and there is evidence that listeners are sensitive to the characteristics of different speakers (e.g. Goldinger, Pisoni and Logan, 1991). The use of sentences produced by only one speaker was considered but discarded, because selecting only one speaker would have meant to test only a limited range of degrees of reduction and reduction features. In perception experiments in which the phonetic detail is relevant to the correct identification of the words, ideally, speaker-specific variations should be avoided. However, selecting material from more than one speaker made it possible to test a wider range of reduction features. Precisely because speakers vary the type and degree of reduction and the way they reduce, for the present research it was thought more important to test the intelligibility of a range of different features and, therefore, speakers. Although this means that there is great variation in the stimuli, it also made it possible to test the intelligibility of different types of variation, and the selection of more and more varied sentences.

7.2.2. Stimuli selection

The stimuli were selected from the utterances recorded during the production experiment. Careful auditory and spectral observations were carried out to select the most appropriate tokens. Although the aim of the perception experiment was to test the intelligibility of the *most* reduced pr + aux, a few of the most reduced items were actually excluded. In some of the most reduced tokens, the amplitude of the pr + aux compared to the amplitude of the main verb was so low that it was hardly audible. Manipulating the amplitude of the entire sentence would have meant that the stimuli in the experiment had different levels of intensity. A pilot experiment indicated that most of the reduced stimuli were correctly interpreted by the majority of listeners. Therefore, to keep the perception experiment short to avoid participants' fatigue, 112 stimuli were selected. Only the stimuli that address the two research questions are discussed in this thesis: 37 stimuli in Chapter 7 (Experiment A), and 54 stimuli in Chapter 8 (Experiment B). Within the stimuli selected for Experiment A, two sets of stimuli were selected to test a specific type of variation: friction in pr + *'d* (N = 5) and initial audible friction (N = 15). The following sections describe the two groups of stimuli.

7.2.2.1. **Friction in pr + *'d***

As reported in Chapter 4, in several instances the cliticised form *'d* in pr + *'d* is realised with friction at the place of articulation, the alveolar ridge. The presence of homorganic friction in coda of pr + *'d* pieces raised the question of whether the contrast with pr + *'s* pieces with the same pronoun is maintained. According to the acoustic analysis reported in Chapter 5, the friction produced in pr + *'d* paradigms has different acoustic characteristics from the friction produced in pr + *'s* paradigms. The parameters in which the two pieces are significantly different are duration, and two of the four spectral moments of the friction (CoG and skewness). A small number (N = 4) stimuli of pr + *'d* that exhibit friction at the end of the piece were selected to test whether the friction in pr + *'d* pieces can lead to the mis-identification of the auxiliary. The control stimuli of pr + *'s* had all been correctly identified during the pilot study, therefore, it was decided to include only one in the experiment to keep the experiment short. Table 28 lists the four *he'd* stimuli used to test the perception of the friction in *pr + 'd* and the control *he's* stimulus included in the experiment.

Speaker_15_He_d_burnt_the_pie
Speaker_6_He_d_burnt_the_fish
Speaker_8_He_d_burnt_the_cake
Speaker_8_He_d_burnt_the_pie
Speaker_6_He_s_burnt_the_fish_CONTROL

Table 28. List of stimuli selected to test the identification of pr + *'d* paradigms in which the phonological plosive is realised with friction at the alveolar ridge.

In the forced-choice word identification task, listeners would hear the *he'd* stimulus with friction and had to choose between the two options – *he'd* and *he's* – presented on the screen. Figure 170 shows two of the stimuli included in the experiment. On the left, an instance of *he's* with the phonological fricative /s/; on the right, an instance of *he'd* with the phonological plosive /d/ realised with friction. Both sentences were uttered by the same speaker (S6).

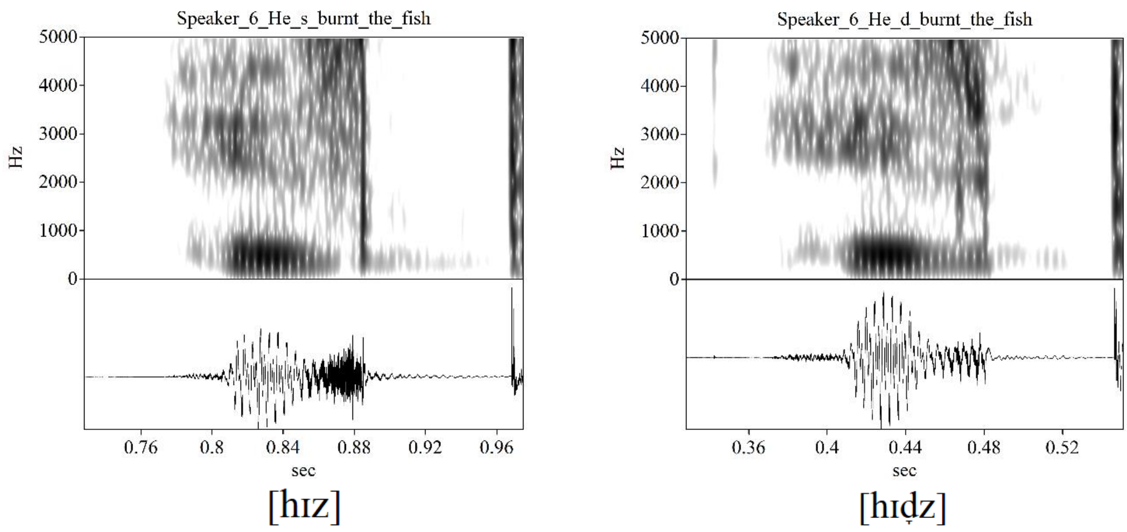


Figure 170. Spectrogram and waveform of *he's* on the left, and *he'd* on the right uttered by the same speaker (S6).

Figure 170 shows the pieces *he's* and *he'd* uttered by the same speaker. Despite their similarities, there are some acoustic cues in the signal that maintain their contrast from an acoustic point of view (e.g. duration of the alveolar friction and the amount of voicing in coda). The perception experiment aims at testing whether their acoustic cues are perceptually available to the listeners and help in the identification of the paradigms. The hypothesis is that the acoustic features of the friction in coda of pr + *'d* and pr + *'s* maintain the distinction between the two paradigms and that the cliticised form of the auxiliary in pr + *'d* can be correctly identified.

7.2.2.2. **Initial audible friction – *you'd* versus *he'd***

As reported in Chapter 6, several pr + aux displayed a portion of voiceless friction before the beginning of voicing. When this friction is audible at the beginning of instances of *you*, it is characterised by a palatal – or high front – quality. As described in Chapter 5 some reduced instances of the pronoun *you* preceded by audible voiceless (palatal) friction sound very similar to the pronoun *he*. In the pronoun *he* the friction is produced in the glottis, but the position of the tongue in preparation for the following high front vocoid makes the quality of the friction sound high front too. Moreover, in some reduced instances of *you*, the vocoid is temporally reduced resulting in a short vocoid with a single quality not dissimilar to the quality of the vocoid in *he*. The acoustic analysis reported in Section 5.2.5.1 showed that the formant dynamics in the two vocoids in *he* and *you* calculated across speakers and repetitions are strikingly similar. The only noticeable difference being the slightly lower F3 in the second half of the vocoid in *you*. Figure 171 shows an instance of *he'd* (left) and an instance of *you'd* with a portion of voiceless friction in onset (right), uttered by the same speaker. Both tokens were included as stimuli in the perception experiments.

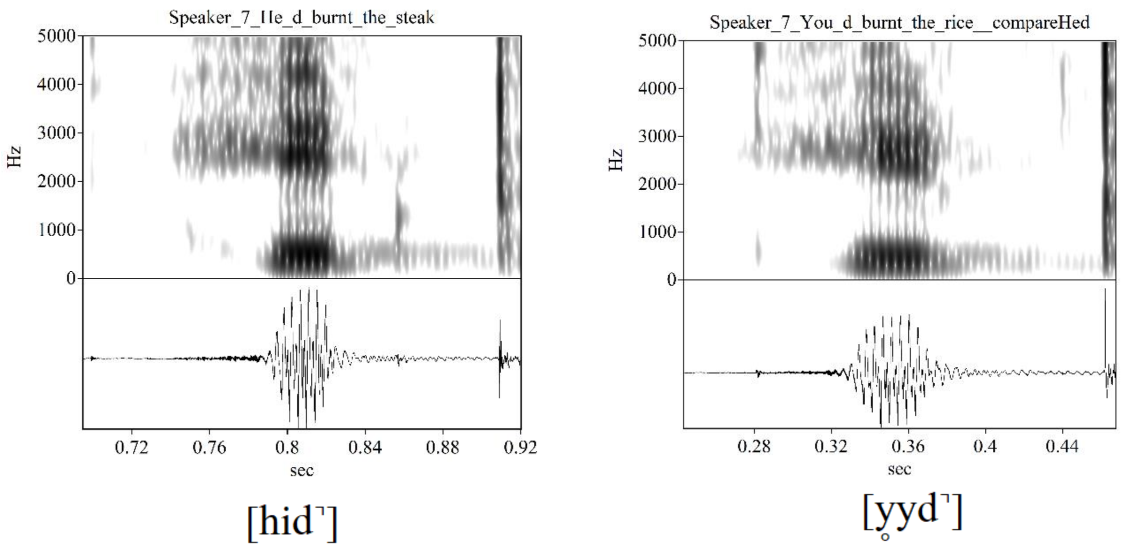


Figure 171. Spectrogram and waveform of two stimuli included in the perception experiment: an instance of *he'd* (left) and an instance of *you'd* (right) produced by the same speaker.

Twelve stimuli of *you'd* and three of *he'd* (total N = 15) were selected to test whether the friction at the beginning of *you* can lead to the mis-identification of the pronoun with *he*. In the forced-choice task, listeners would hear an instance of *you'd* with initial friction (or *he'd* in the case of the control stimuli) and had to choose between the two options written on the screen – *you'd* and *he'd*. Table 29 reports the stimuli included in the experiment to test the perception of the initial friction in *you*.

Speaker_10_You_d_burn_the_fish_compareHed
Speaker_10_You_d_burnt_the_rice_compareHed
Speaker_11_You_d_burn_the_roast_compareHed
Speaker_15_You_d_burn_the_pie_compareHed
Speaker_7_You_d_burn_the_fish_compareHed
Speaker_7_You_d_burn_the_jam_compareHed
Speaker_7_You_d_burnt_the_rice_compareHed
Speaker_7_You_d_burnt_the_roast_compareHed
Speaker_8_You_d_burnt_the_pie_compareHed
Speaker_8_You_d_burnt_the_rice_compareHed
Speaker_9_You_d_burn_the_fish_compareHed
Speaker_9_You_d_burn_the_pie_compareHed
Speaker_7_He_d_burnt_the_fish
Speaker_8_He_d_burn_the_roast
Speaker_8_He_d_burn_the_soup

Table 29. List of stimuli selected to test the perception of the initial friction in *you'd* including three control stimuli of *he'd*.

7.2.2.3. **Highly reduced pr + aux**

Sixteen stimuli were selected to test the intelligibility of reduced speech. These stimuli exhibit typical features of temporal and gestural reduction and temporal reorganisation of phonetic events that cannot be grouped under a single feature. They also did not display acoustic features typical of other pr + auxs, such as the ones described above. In other words, the acoustic features of their reduced forms did not seem to neutralise the contrast with any other pr + aux. For this reason, it was problematic to establish which 'plausible' alternative option to present to the participants on the screen besides the correct option. To avoid giving participants a biased choice based on the expectations of the researcher, a preliminary survey was carried out to investigate what the listeners might hear if they were unable to identify the correct pr + aux.

In the preliminary survey, twelve native speakers of British English took part. The participants were asked to listen to a number of short sentences (selected from the corpus collected for the production study) and to write down what they thought they heard. They could listen to each sentence multiple times for a maximum of five times. They had to write the entire sentence, not just the pr + aux. This survey was carried out using Qualtrics. Although the participants had been clearly informed that all the tokens contained at the beginning of each sentence a pronoun AND an auxiliary, several participants for several tokens wrote only a pronoun stating that they could not hear any auxiliary at all. This confirmed that the open answer paradigm was not suitable for the perception experiment and the forced-choice paradigm was used. As for the alternative option to present on the screen, for each token, the incorrect answer with the highest number of mentions was chosen. The alternative options are reported in Table 30 with the corresponding stimuli. Figure 172 shows two tokens that were selected as stimuli for the perception experiment.

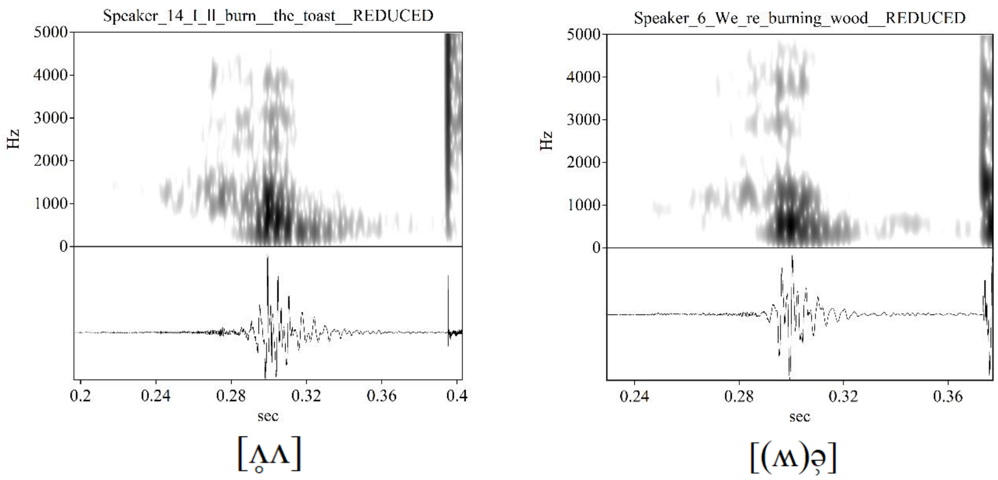


Figure 172. Spectrogram and waveform of two stimuli included in the perception experiment to test the intelligibility of highly reduced pr + aux. On the left, an instance of *I'll*, and on the right, an instance of *we're*.

The 16 stimuli selected to test the intelligibility of reduced pr + aux include a range of pr + aux combinations uttered by a range of speakers, as shown in Table 30.

Stimulus	Alternative option
Speaker_6_We_re_burning_wood	*we were*
Speaker_6_We_were_burning_wood	*we're*
Speaker_7_I_d_burn_the_pie	*I'll*
Speaker_7_It_was_burning_down	*it's*
Speaker_7_It_was_burning_oil	*it's*
Speaker_7_She_was_burning_leaves	*she's*
Speaker_7_You_were_burning_wood	*you're*
Speaker_8_He_d_burn_the_roast	*he'll*
Speaker_8_He_d_burn_the_soup	*he'll*
Speaker_8_I_ll_burn_the_chips	*I'd*
Speaker_9_She_ll_burn_the_cake	*she'd*
Speaker_11_You_were_burning_coal	*you're*
Speaker_14_I_ll_burn_the_toast	*I'd*
Speaker_14_You_d_burn_the_rice	*he'd*
Speaker_15_We_were_burning_straw	*we're*
Speaker_15_You_d_burn_the_pie	*he'd*

Table 30. List of stimuli selected to test the intelligibility of reduced pr + aux (on the left) and the alternative options presented on screen.

7.2.3. Experiment procedures

The stimuli of Experiment A and the stimuli of Experiment B were combined together so that only one experimental session was run. Therefore, the experiment task and procedures are the same for Experiment A and Experiment B (Experiment B is described in Chapter 8). Due to the pandemic, the experiment had to be carried out online instead of in the Recording Studio of the Department of Languages and Linguistic Science of the University of York as planned. The experiment was created in PsychoPy3 (Peirce, Gray, Simpson, MacAskill, Höchenberger, Sogo, Kastman and Lindeløv, 2019) and run on Pavlovia (https://pavlovia.org). This section describes the recruitment of the participants and the set-up of the experiment.

7.2.3.1. **Subjects**

The participants were recruited online and by email. Before the experiment started, participants were told that they could withdraw at any time and that their participation was voluntary. They were asked for their consent to use their data for research purposes.

Because of the online nature of the experiment, it was difficult to control for two factors: the participants' accent, and the environment in which they took the experiment. To compensate for this lack of control, the participants were asked to answer a few questions before starting the experiment. To assess the participants' accent, participants were asked where they were born and where they had spent the majority of their lives. Although these two questions do not automatically define the accent of a speaker, it was considered unhelpful to ask participants what accent they spoke, mainly because non-linguists might not be aware of their accent. Standard Southern British English (SSBE) speakers were the target participants because the stimuli used were spoken in SSBE and it has been found that intelligibility decreases with unfamiliar accents (Ernestus and Warner, 2011). However, speakers of other British English accents have some familiarity with SSBE due to exposure. The instructions at the beginning of the experiment stated that any native speaker of British English could participate in the experiment. The idea was to exclude at a later stage any participant that could not identify the 'control' stimuli correctly. This does not resolve the issue of the participants' accent, but at least gives the researcher some degree of control.

Due to the nature of the stimuli, it was crucial that the experiment occurred in a quiet environment. Ideally, the experiment would have taken place in a sound-proof room of the university, in which it would have been possible to control that the environment was as quiet as possible and that participants could wear appropriate headphones and listen to the sentences at an adequate volume. Since the experiment was online and it was not possible to control the participants' environment, a few questions regarding their environment were asked before the beginning of the experiment. Participants had to state whether they were in a quiet environment and whether they were wearing headphones. In addition, the experiment could not be taken on a mobile phone but had to be taken on a laptop or desktop computer. This again was a form of control of the environment. In addition, participants were asked if they had any hearing impairment. Finally, participants were asked to state their year of birth and gender.

Sixty-six participants took part, but two participants were discarded. One participant was discarded because she declared a minor hearing loss on the left ear. One participant was discarded because she mis-identified 50% of control tokens. The responses from 64 participants were thus analysed. Fifty-six participants were born in the UK (N = 60/64, 94%), and four abroad[15] (N = 4/64, 6%). Of the participants that were born in the UK, one was born in Northern Ireland, two in Scotland and two in Wales. At the question where they had spent most of their lives, 58 participants (N = 58/64, 91%) answered a location in England, two in Scotland, three in Wales and one abroad. The age of the participants ranged from 19 to 66 years of age (mean 30 years, median 23 years). Forty-five participants were female (70%), eighteen male (28%) and one non-binary (2%).

[15] The experiment instructions indicated that the experiment was only for native speakers of British English. Therefore, it was assumed that even those that were born abroad were native speakers of British English. Their responses have been manually checked and there was no reason to doubt that this was the case.

7.2.3.2. **Experiment setup**

After answering the questions outlined in the previous section, the experiment would start. Participants would see on their screen the instructions of the experiment and could move forward by pressing any key on their keyboard. The instructions informed the participants that they were required to choose between two possible 'beginnings' of the sentences they would hear. For each stimulus, two options of beginning would be displayed on the screen, one on the left and one on the right. Participants were instructed to press the right key (›) if they thought they heard the option written on the right and the left key (‹) if they thought they had heard the option written on the left. Pressing the right or left key would progress the experiment to a new stimulus and a new page on the screen. Participants could listen to each stimulus multiple times for a maximum of ten times by pressing the spacebar on their keyboard.

In a pilot experiment, 170 stimuli were included. A few participants were asked at what point they started to feel tired and distracted and the answers ranged from 70 to 90 stimuli. Therefore, the experiment was substantially shortened to a total of 112 stimuli (37 stimuli for Experiment A, 54 stimuli for Experiment B, and 21 stimuli to analyse a feature that is not discussed in this thesis). Additionally, a break was added. After the first 56 stimuli, a page would be displayed on the screen stating that participants could pause for a short time or go on with the experiment. The second set of stimuli consisted of 56 sentences too. The pages on the screen were numbered so that the participants were aware of their progress.

The stimuli were manually ordered to make sure that similar stimuli were not next to each other. Although it is possible to randomise stimuli automatically, too many factors had to be controlled for. Adjacent stimuli had to differ on several parameters including: speaker, pronoun, auxiliary, noun (object) of the sentence, and type of reduction.

In addition to recording the participants' responses, the number of repetitions that each participant played of each sentence was also recorded. The reaction time to each stimulus and repetition was recorded too. However, participants were not instructed to respond quickly or as quickly as possible. It was considered more important that participants listened carefully to each stimulus. The next section reports the analysis of the responses of the perception experiment.

7.2.3.3. Statistical analysis

For the analysis of the responses of the perception experiments (Chapters 7 and 8), exact two-sided binomial tests were run to determine whether group responses were significantly above a chance value of 50%. That is, when two options for response were presented, whether participants were responding in a consistent manner. Since the binomial tests do not take into account structured variability such as by speakers and sentences, the results need to be interpreted cautiously.

7.3. Data analysis and results

This section reports on the analysis of the responses to the 37 stimuli of Experiment A, that is, those stimuli selected for their high degree of reduction or for a specific reduction feature. The analysis of the responses has been divided according to the feature analysed, following the structure of Section 7.2.2 on stimuli selection.

7.3.1. Friction in pr + *'d* – results

Four *he'd* stimuli and one *he's* stimulus (N = 5) were included in the experiment to test whether the presence of friction in coda of pr + *'d* paradigms would neutralise the contrast between pr + *’s* and pr + *’d*. An exact two-sided binomial test reveals that the correct responses to the stimuli testing the intelligibility of *he’d* realised with friction in coda position were significantly above chance ($p < 0.001$; 95% CI [0.19, 0.3]). Figure 173 and Figure 174 show the participant responses by stimulus. Figure 173 shows the responses to *he’d* stimuli, and Figure 174 shows the responses to the stimulus *he’s*.

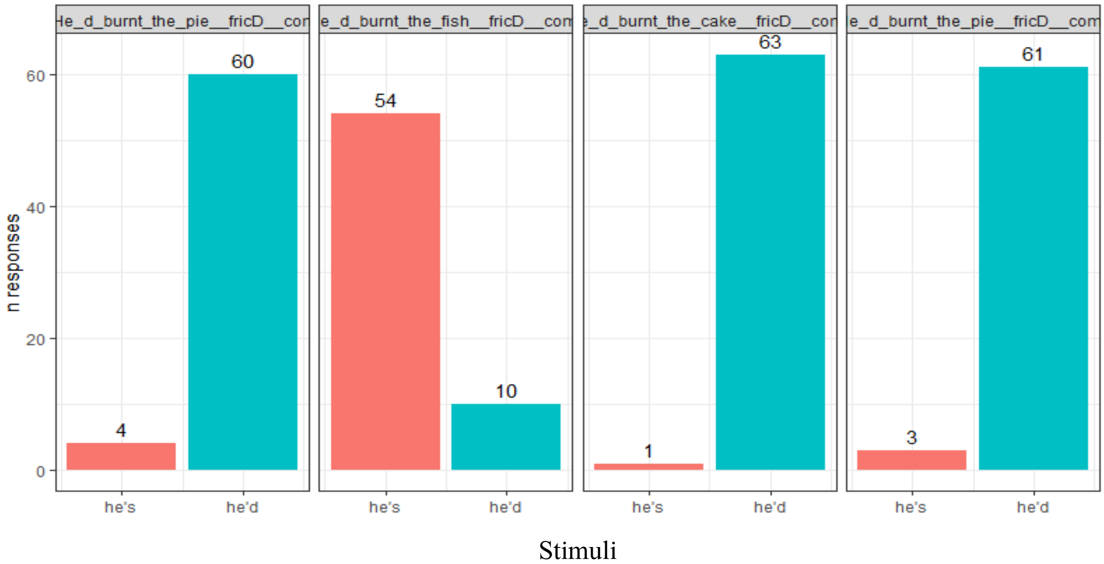


Figure 173. Counts of *he’d* (blue) and *he’s* (orange) responses for each of the four stimuli of *he’d* with friction in coda.

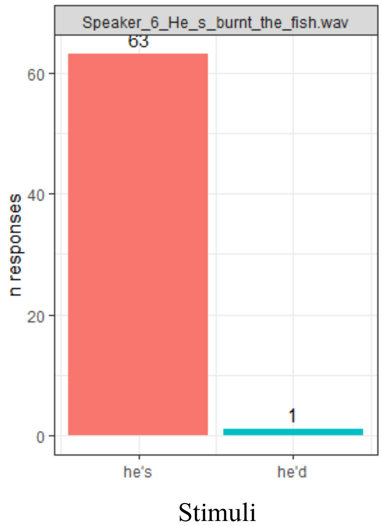


Figure 174. Counts of *he’d* (blue) and *he’s* (orange) responses for the control stimulus *he’s*.

The control stimulus *he's* was correctly identified by 63/64 participants (98%), as shown in Figure 174. Figure 173 shows that three out of four stimuli (75%) were correctly interpreted by the large majority of participants (N = 60/64, 94%; N = 63/64, 98%; N = 61/64, 95%). The high percentage of participants (N = 54/64, 84%) who identified *he'd burnt the fish* uttered by S6 as *he's* is worth an in-depth investigation. Only 10 listeners listened to the stimuli twice and only one listened to the stimuli three times. This indicates that most listeners were fairly confident as to which pr + aux they heard. Chapter 5 (Section 5.2.4.1) highlighted the acoustic differences between the friction in coda of pr + *'s* and pr + *'d* pieces. The parameters that were significantly different were the duration of the friction and some of the spectral qualities of the friction (CoG and skewness, but not kurtosis and SD). Looking at the acoustic characteristics of this specific token (*he'd burnt the fish* uttered by S6), the duration and kurtosis of the alveolar friction are more similar to the mean values of the alveolar friction in *he's* tokens than *he'd* tokens, while the CoG and skewness are more in line with the mean values of the spectral qualities of the alveolar friction in *he'd* tokens (Table 31).

Alveolar friction	duration (ms)	CoG (Hz)	skewness	kurtosis
Mean values in *he's* (N = 50)	49	6984	-0.7066	2.5109
Mean values in *he'd* (N = 24)	18	2692	1.5969	5.3428
Speaker_6_He_d_burnt_the_fish	*36*	*1584*	*1.8962*	*3.1146*

Table 31. Mean duration and spectral moments of the alveolar friction in coda of *he's* and *he'd* instances (calculated across speakers and repetitions) compared to the individual values of the token *Speaker_6_He_d_burnt_the_fish* which was perceived as *he's* by 84% of participants.

Although all the other parameters analysed during the acoustic and statistical analysis did not exhibit a significant difference, on a closer spectrographic inspection, it seems that the duration of friction and the amount of voicing in this specific token might make it more perceptually similar to *he's* paradigm. Except for this token, the results of the perception experiment of the pr + aux *he'd* show that the presence of friction at the place of articulation do not hinder the correct identification of pr + *'d*. This confirms the findings of the acoustic analysis which showed that the temporal and spectral properties of the friction in underlying /s/ and /d/ are distinct.

7.3.2. Initial audible friction – *you'd* versus *he'd* – results

This section reports on the results of the identification of stimuli that have audible voiceless friction at the beginning of *you* + aux. Twelve tokens with audible friction at the beginning of *you* uttered by seven different speakers and were used. In addition, three control stimuli with the pronoun *he* were included in the experiment, as shown in Table 29. An exact two-sided binomial test revealed that the correct responses to the stimuli testing the intelligibility of *you'd* realised with onset friction were significantly above chance ($p < 0.001$; 95% CI [0.28, 0.34]). Figure 175 and Figure 176 show the breakdown of the responses by stimulus. Figure 175 shows the responses to *you'd* stimuli, and Figure 176 shows the responses to *he'd* stimuli.

Figure 175. Counts of *you'd* (blue) and *he'd* (orange) responses for each of the 12 stimuli of *you'd* with initial friction.

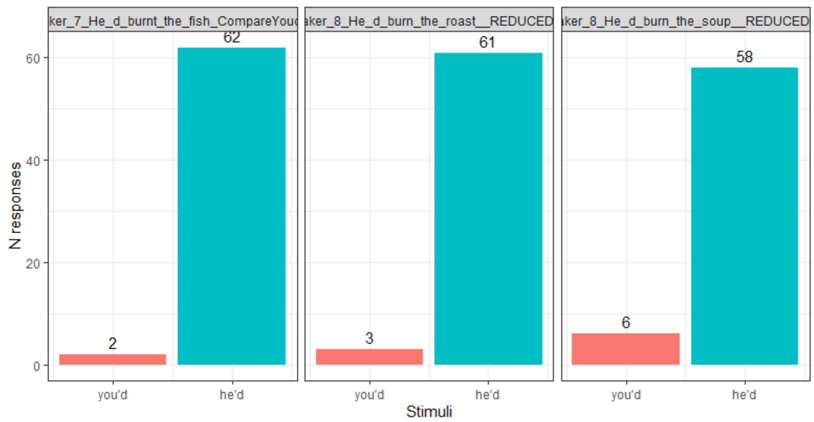


Figure 176. Counts of *you'd* (orange) and *he'd* (blue) responses for each of the 3 stimuli of *he'd* (control stimuli).

Figure 176 shows the number of *he'd* (blue) and *you'd* (orange) responses for each *he'd* stimulus. These were the control stimuli and they were all correctly identified by the majority of participants. Figure 175 shows the number of *you'd* (blue) and *he'd* (orange) responses for each stimulus. All the stimuli used were original *you'd* tokens collected during the production study, and were not manipulated. The majority of the stimuli (N = 8/12, 67%) were correctly identified as the pr + aux *you'd*. Five stimuli received more *he'd* responses than *you'd*. However, the number of participants that perceived these five stimuli as *he'd* instead of *you'd* is only minimally higher than the number of participants who perceived them as *you'd* in four cases out of five.

The sentence *Speaker_8_You_d_burnt_the_rice* is an exception in that it was perceived as *he'd* by 72% of participants (N = 46/64, 72%). Figure 177 shows two stimuli of *you'd* with initial friction produced by the same speaker (S8). The stimulus on the left was interpreted as *he'd* by 72% of participants (N = 46/64). The stimulus on the right was interpreted as *you'd* by 95% of participants (N = 61/64).

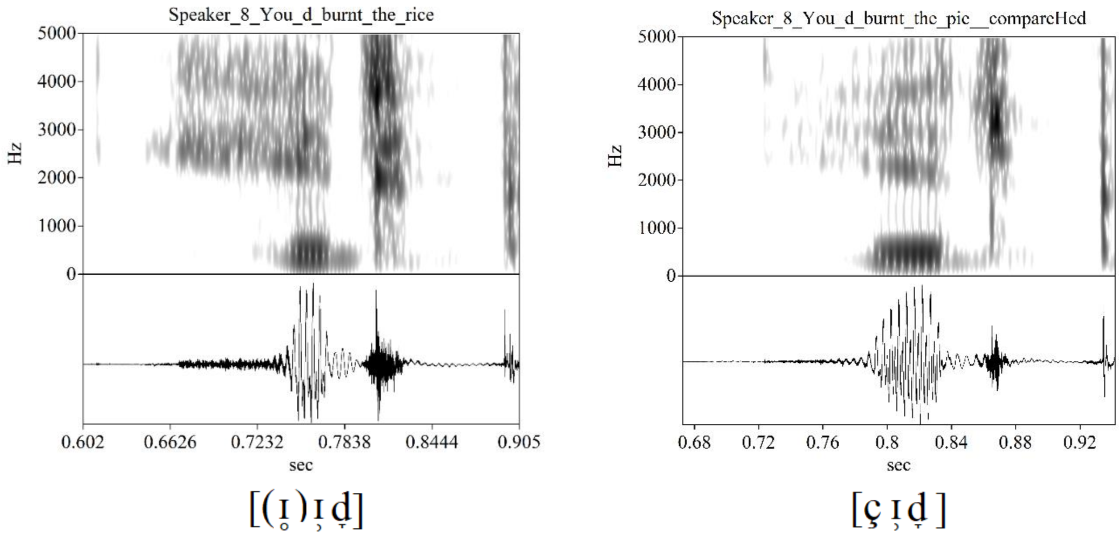


Figure 177. Spectrogram and waveform of two instances of *you'd*. The one on the left was identified as *he'd* by 72% of participants. The one on the right was identified as *you'd* by 95% of participants.

Figure 177 shows two stimuli that were perceived differently by the majority of listeners. The stimulus on the left was interpreted as *he'd*, while the stimulus on the right was interpreted as *you'd*. It can be noticed that the friction in onset of the token on the left is much stronger than the friction in onset of the instance on the right. Table 32 compares the duration and amplitude of the friction and the vocoid in *you'd* and *he'd* (across speakers and repetitions) with the individual values of the stimulus that most speakers interpreted as *he'd* (*Speaker_8_You'd_burnt_the_rice*).

	Friction duration (ms)	**Friction amplitude (dB)**	**Vocoid duration (ms)**	**Vocoid amplitude (dB)**
Mean values in *you'd* (N=41)	59	44	42	69
Mean values in *he'd* (N=53)	65	45	24	69
Speaker_8_You'd_burnt_the_rice	*83*	*46*	*21*	*70*

Table 32. Mean duration and amplitude of the friction and vocoid in *you'd* and *he'd* instances (calculated across speakers and repetitions) compared to the individual values of the token *Speaker_8_You_d_burnt_the_rice* which was perceived as *he'd* by 72% of participants.

Table 32 shows that the duration of the vocoid of the stimulus that was perceived as *he'd* (21 ms) is much more similar to the mean duration of the vocoid in *he'd* tokens (24 ms) than the mean duration of the vocoid in *you'd* tokens (42 ms). The duration of the friction in the same token (83 ms) is also much longer than the mean duration of the friction in either *he'd* (65 ms) and *you'd* (59 ms) datasets. The amplitude of the friction in the same token (46 dB) is also closer to the mean amplitude of the friction across the *he'd* dataset (45 dB), then the *you'd* dataset (44 dB). The combination of these acoustic features makes the stimulus *Speaker_8_You'd_burnt_the_rice* more similar to the instances of *he'd* analysed and are likely to have triggered its identification as *he'd*.

Some of the stimuli were identified at chance level. *Speaker_7_You_d_burnt_the_roast* was identified as *he'd* by 34 participants (53%), and as *you'd* by 30 participants (47%). Compared to the other *you'd* stimuli produced by S7 that were identified as *you'd*, this stimulus is characterised by less lip-rounding. The stimulus *Speaker_11_You_d_burn_the_roast* was identified as *he'd* by 35 participants (N = 35/64, 55%). This token is characterised by strong onset friction and a very short vocoid. The initial friction is 109 ms long, while the vocoid has a duration of only 17 ms. Although the onset friction is a feature of the pronoun *he*, in this token the friction is also characterised by a hint of lip-rounding. The combination of these two contrasting features – onset friction and lip-rounding – is likely to be the reason behind the mixed responses to the other two stimuli interpreted at chance level too (*Speaker_9_You_d_burn_the_fish* and *Speaker_10_You_d_burnt_the_rice*).

7.3.3. Highly reduced pr + aux – results

Sixteen stimuli were selected to test the intelligibility of highly reduced pr + aux, regardless of the type or degree of reduction they exhibited. The alternative options presented on screen were determined according to the results of the survey reported in Section 7.2.2.3, and were reported in Table 30 (page 346). Unfortunately, it was not possible to indicate in Figure 178 the two options given and, thus, they are indicated as 'incorr' and 'corr' response, where 'corr' stands for the option that corresponds to the stimulus played, and 'incorr' to the alternative option. An exact two-sided binomial test indicates that the responses to the highly reduced stimuli were significantly above chance ($p < 0.001$; 95% CI [0.14, 0.18]). Figure 178 shows the responses by stimulus with the count of the correct responses in blue, and the incorrect responses in orange.

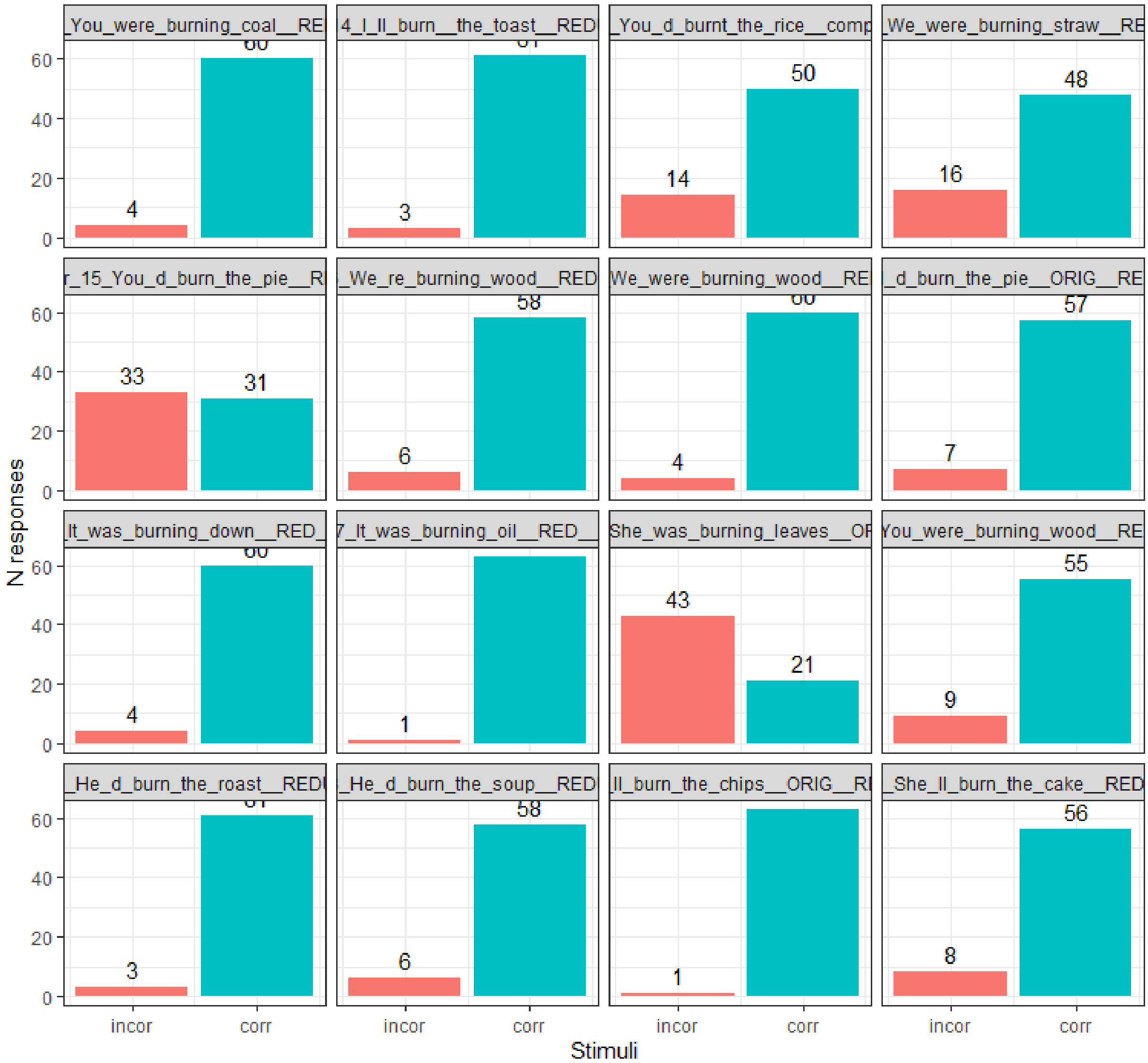


Figure 178. Count of incorrect (orange) and correct (blue) responses for the 16 stimuli testing the intelligibility of reduced pr + aux.

The majority of stimuli (N = 14/16, 87%) was correctly identified by a large number of participants. Only two stimuli were not correctly identified. The majority of participants (N = 43/64, 67%) mis-identified the stimulus *Speaker_7_ She_was_burning_leaves* as *she's*. This token of *she was* undergoes a fair amount of temporal reduction: it exhibits the shortest vocoid duration of all the items in the *she was* dataset. Its vocoid duration is 33 ms, while the mean vocoid duration of the instances of *she was* across speakers and repetitions is 86 ms, and the mean vocoid duration of the instances of *she's* is 36 ms. Figure 179 shows the spectrogram and waveform of the stimulus *Speaker_7_ She_was_burning_leaves.*

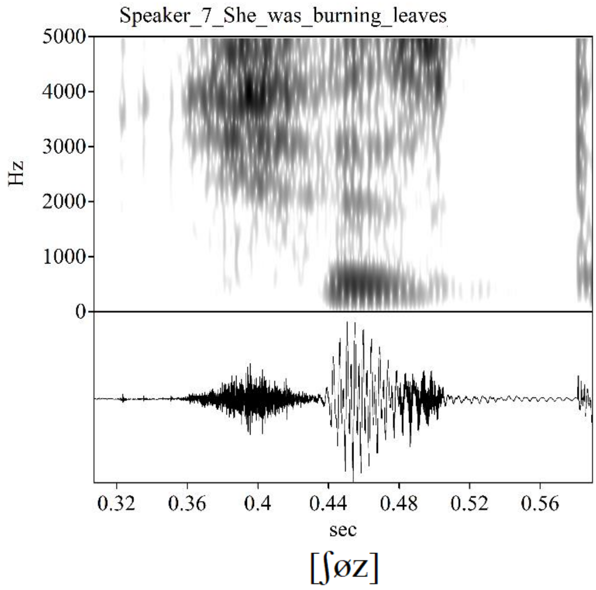


Figure 179. Spectrogram and waveform of the token *Speaker_7_ She_was_burning_leaves* that was incorrectly interpreted by the majority of participants.

Although the duration of the vocoid is more similar to the duration of the vocoid in tokens of *she's*, the piece sounds slightly labialised. However, 67% of participants chose the option *she's* when they heard this stimulus.

The stimulus *Speaker_15_ You_d_burn_the_pie* was identified as *he'd* by 33 participants (N = 33/64, 52%). The vocoid in this piece is realised with creaky voice throughout. Its quality is palatal but with no lip-rounding. The duration of the vocoid in this token is 34 ms as shown in Figure 180.

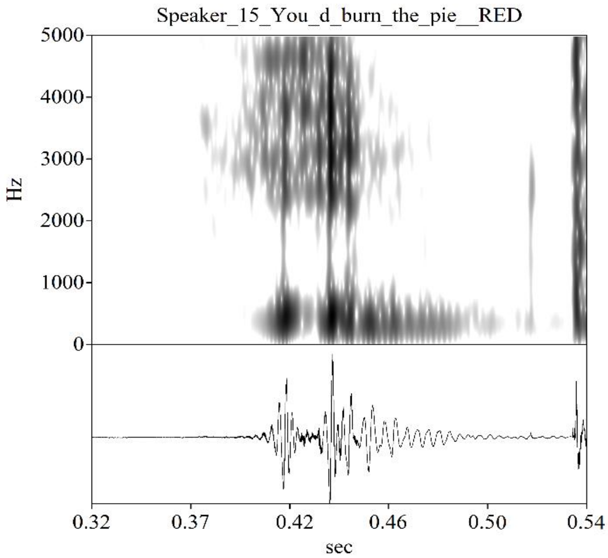


Figure 180. Spectrogram and waveform of the token *Speaker_15_ You_d_burn_the_pie* which was incorrectly interpreted by the majority of participants.

The pronouns *you* and *he* are characterised by a similar quality of the vocoid, which is palatal or close front. The essential phonetic features that distinguish *you* and *he* (in this data) are the lip-rounding in *you* and the glottal friction in onset of *he*. Neither of these two features is present in this pr + aux and the vocoid is characterised by a close front or palatal quality, which is the feature that the pronouns *you* and *he* share. This is likely to be the explanation for the identification at chance level of this stimulus.

7.4. Summary and discussion

This chapter described the aims, methodology and results of Experiment A on the identification of reduced word forms. The aim of the experiment was to investigate whether highly reduced instances of pr + aux can be correctly identified when a limited context is given. That is, whether the fine phonetic detail that maintains the contrast between paradigms is available to perception in reduced speech. The forced-choice word identification task reported in this chapter included stimuli that can be grouped under three types of reduction. Firstly, pr+ *'d* instances in which the phonological plosive /d/ is realised with friction produced at the alveolar ridge. The rationale for investigating this feature was to determine whether the perceptual contrast with the homorganic fricative /z/ in pr+ *'s* is maintained. The quantitative analysis reported in Chapter 5 indicated that the two frictions differ along several spectral parameters. The results of the perception experiment indicate that the majority of instances of pr+ *'d* (N=3/4, 75%) is correctly identified by the large majority of participants (between 94% and 98% of participants).

The second group of stimuli was selected to determine whether the voiceless friction at the beginning of reduced instances of *you'd* could lead listeners to identify the piece as *he'd* instead of *you'd*. Although the majority of the stimuli (67%) were correctly identified by the large majority of listeners (between 61% and 97%), four stimuli were incorrectly identified by a small majority of listeners (between 55% and 72%). The statistical analysis indicated that the correct identification of the reduced stimuli is significantly above chance. The perception experiment showed that the stimuli that were misidentified are acoustically more similar to the instances of *he'd* than the instances of *you'd* analysed.

The third group of stimuli was selected to determine the intelligibility of a range of highly reduced items that did not share a specific type or degree of reduction. Also for this group, the large majority of stimuli (N=14/16, 87%) was correctly identified by the large majority of participants (between 75% and 98% of participants), and only two stimuli were misidentified.

The results of the perception experiment reported in this chapter indicate that even highly reduced pr+aux presented in a limited semantic context can be correctly identified on the basis of the acoustic signal. By 'limited' context it is meant that the context does not provide any useful information for the discrimination between the two options and the choice of one or the other of the options provided on the screen. Either option is grammatically correct and equally likely. In absence of a preceding semantic context, the identification of the pr + aux relies only on the acoustic material. The main limitation of the experiment is that listeners were given two options to choose from, which means that they had some information (besides the acoustic cues) about the items to be identified. In an open answer paradigm, the identification of the stimuli would be more challenging, as observed in a pilot experiment carried out with the same stimuli in which listeners could write what they thought they had heard.

The results reported in this chapter confirm the findings by previous research on the role of acoustic cues in the identification of reduced speech (e.g. Kohler and Niebuhr, 2011; van de Ven et al., 2012). In particular, the results of the first two groups of stimuli (friction in pr + *'d* versus pr + *'s*, and initial friction in *you'd* versus *he'd*) suggest that the identification of target items can rely on subtle acoustic features such as the spectral properties of the friction. The importance of subtle acoustic cues was highlighted by the analysis of two of the stimuli that were misidentified. In both *S6_He_d_burnt_the_fish* (which was identified as *he's* by the majority of participants) and *S8_You_d_burn_the_rice* (which was identified as *he'd* by the majority of participants), the incorrect identification could be explained by looking at the acoustic features of the two stimuli. In both cases, their acoustic features (such as duration and spectral moments of the friction) were more similar to the acoustic features of the tokens of the 'incorrect' option: *he's* for *S6_He_d_burnt_the_fish*, and *he'd* for *S8_You_d_burn_the_rice*. These results indicate that listeners are sensitive to fine-grained phonetic detail that remains in the acoustic signal and rely on it for the correct identification of reduced word forms.

The results of the experiment reported in this chapter have implications for models of speech perception and for phonological theories of the pr + aux system. They confirm that the crucial information for word identification in reduced speech is conveyed by distributed, non-segmental phonetic features, i.e. fine phonetic detail, which is available to perception in highly reduced speech. The importance of the fine phonetic detail in spoken word recognition challenges the idea that the identity of function words, and the contrast between function words, is conveyed by segmental units. Yet, our knowledge and understanding of how fine phonetic detail is used in reduced speech and in perception is still limited. The next chapter aims to advance our knowledge of the perceptual salience of fine phonetic detail by investigating the roles of such detail in spoken word identification.

8. Perceptual salience of selected acoustic parameters

The acoustic analysis reported in Chapter 5 highlighted the parameters that discriminate between highly reduced pairs of pr + aux from an acoustic point of view. It emerged that pairs of pr + aux differ along two main parameters: duration and resonances. This raised the question of whether the acoustic differences between these paradigms are available to perception, and whether listeners use them to identify the identity of reduced function words. The perception experiment reported in Chapter 7 revealed that, given two options, listeners can correctly identify highly reduced pr + aux on the basis of their acoustic features. The question that remains to be answered is which phonetic feature (if any) maintains the contrast between paradigms in perception. Are listeners more sensitive to one of the two parameters identified by the acoustic analysis? In other words, are both parameters equally salient for the correct interpretation of reduced function words? Or is one parameter more perceptually salient than the other? This chapter investigates these issues. It reports on the analysis of a subset of data of the perception experiment described in the previous chapter. In particular, it investigates the perceptual salience of duration and resonances. By manipulating the duration in a subset of stimuli, this experiment aims to establish whether listeners rely more on the duration or on other properties of the stimuli, such as the resonances.

8.1. Background

The seminal works by Hawkins and Smith (2001), Hawkins (2003) and Local (2003) highlighted the importance of fine phonetic detail in speech comprehension and communication. They stressed the role of phonetic detail in carrying crucial information, not only about the lexical identity of words, but also about grammatical, linguistic and conversational information. Most studies focus on the recognition of words carrying semantic meaning, but listeners have been found to be sensitive to the acoustic cues that signal linguistic information too. For example, Kemps et al. (2005) showed that listeners are sensitive to phonetic detail carrying morphological information about inflectional boundaries. Davis et al. (2002) found that listeners make use of phonetic detail in word segmentation. Baker (2008) and Clayards et al. (2021) showed that listeners can discriminate between words that contain similar sound sequences that differ only in their morphological status. Despite the role of fine phonetic detail in speech communication, and the fact that reduction is pervasive in everyday speech, only a few studies have focused on the perception of fine phonetic detail in reduced speech. The following sections look at studies that focus on the perception of the two parameters that are investigated in this chapter: resonances and duration.

8.1.1. Perception of resonances

Resonances are related to the secondary articulations of sounds (Kelly and Local, 1989). West (1999a) investigated the perceptual salience of the resonances associated with the English consonants /l/ and /ɹ/. She focussed on the long-domain effects of resonances on neighbouring sounds and syllables. In her experiments, she tested whether the long-domain resonances are available to perception even when the sounds responsible for them are masked by noise so that listeners cannot rely on them in the identification of the words. She also investigated how much acoustic information has to be available to perception for listeners to discriminate between minimal pairs without the aid of any semantic context. The results of her experiment showed that listeners can and do use the acoustic cues of long-domain resonances that remain in the signal to correctly identify minimal pairs. Although the speech used in her experiments was not articulatorily reduced, the sounds that triggered the resonances and some neighbouring sounds were masked by noise and the semantic context was absent altogether. Her findings demonstrate that the acoustic cues associated with the resonances are used in word recognition (West, 1999a).

Kohler and Niebuhr (Niebuhr and Kohler, 2011; Kohler and Niebuhr, 2011) focussed on the role of "articulatory prosodies" in word identification. The original concept of 'prosodies' comes from Firthian linguistics (e.g. Firth, 1948; see Section 1.3.1). In Kohler's view, articulatory prosodies are traces of reduced segments and words that "persist as non-linear, suprasegmental features of syllables, reflecting e.g. nasality or labiality that is no longer tied to specific segmental units" (Kohler, 1999: 89). In a series of experiments, Kohler and Niebuhr tested the role of articulatory prosodies in spoken word recognition. In one of their experiments (reported in Kohler and Niebuhr, 2011), the degrees of palatality and nasality associated with the word *Ihnen* in the phrase *ich kann Ihnen das ja mal sagen* (which is grammatical also without *Ihnen*) were manipulated to generate perceptually ambiguous stimuli. That is, in the absence of a segmental realisation of *Ihnen*, different degrees of palatality and duration of the nasal contoid in the manipulated stimuli, led listeners to interpret the phrase as being uttered either with or without the word *Ihnen* (Kohler and Niebuhr, 2011). They found that the degree of palatality significantly affected the interpretation of the utterance as including the word *Ihnen*, while the role of the variations in the duration of the nasal contoid was less prominent.

8.1.2. Perception of duration

Duration is one of the main parameters used to quantify reduction. However, duration is affected by several factors, of which speech rate is an obvious one. Lehiste (1970, 1976) includes duration in the suprasegmental features of speech, and points out that segmental duration can be affected by articulatory and phonological factors. For example, the duration of vowels seems to be correlated with tongue height so that high vowels are shorter than low vowels. While preceding consonants might not exert a considerable influence on a vowel, it has been found that following consonants do affect the duration of a vowel extensively (Lehiste, 1976). From a perceptual point of view, she claims that "small, short-term variations in the timing of speech intervals have perceptual value" (Lehiste, 1976: 226). Several studies have investigated the perceptual salience of duration trying to establish the "just noticeable difference" in duration that listeners can perceive. Huggins (1972a, 1972b) found that listeners can perceive a durational difference of 15 ms and that they are more sensitive to durational changes in vowels than consonants. An experiment by Klatt and Cooper (1975) showed that listeners are sensitive to a change in segmental duration of at least 25 ms. However, Lehiste points out that the durational differences might be relative to the duration of a given sound. For example, a duration difference of 10 ms might be perceptually relevant if the sound has the duration of 100 ms, but not if it has a duration of 200 ms. She concludes that "in the range of the durations of speech sounds – usually from 30 to about 300 msec – the just-noticeable differences in duration are between 10 and 40 msec" (Lehiste, 1970: 13). Bochner, Snell and MacKenzie (1988) found that the minimal durational changes perceived by listeners are in the range between 10% and 15% of the segmental duration of a given sound. However, they also found that listeners are more sensitive to durational differences of sounds than to durational differences of silences. This has implications for the durational changes in plosive duration that are available to perception.

8.1.3. Research question

This chapter investigates the perceptual salience of selected acoustic parameters in instances of pr + aux that have been manipulated. It reports on a perception experiment in which the duration of selected phonetic events (such as vocoids) has been manipulated. It addresses the following research question:

> RQ: Are some acoustic parameters more salient than others in spoken word recognition? In other words, do listeners rely on one parameter (e.g. duration) more than on others (e.g. resonances) for the correct interpretation of reduced word forms?

The hypothesis is that if listeners prioritise duration over other acoustic cues, they will respond differently to the same base stimulus when it has different manipulated durations. If the differences in duration have little or no effect on the responses of the listeners, the assumption is that the resonances play an important role in the correct identification of reduced words. Although it is not excluded that a cluster of acoustic cues (other than duration) is used by listeners for the correct interpretation of speech, the resonances are here considered the main factor in light of the auditory and acoustic analyses reported in Chapters 3 to 5.

8.2. Methodology

To test the perceptual salience of duration and resonances, a two-alternative forced choice word identification task was used. The methodology of this experiment (Experiment B) is largely the same as the methodology of Experiment A reported in Chapter 7 (Section 7.2).

This section describes only the parts of the perception experiment that differ from what was reported there.

The acoustic and statistical analysis of contrastive pairs of paradigms reported in Chapter 5 showed that pairs of pr + aux differed along two or three parameters depending on the pr + aux combination. The aim of this perception experiment is to investigate whether listeners rely on one parameter more than others or if all the acoustic cues together are needed for the correct identification of words. To do this, pairs of contrasting paradigms have been manipulated.

8.2.1. Stimuli manipulation

In a pair such as *she's* and *she was*, the duration of the vocoid is significantly different: in *she's* the vocoid is SHORT relative to the vocoid in *she was*, which is LONG. As for the resonances, *she's* is characterised by CLEAR resonances, while *she was* is characterised by DARK resonances.[16] To test whether listeners rely more on the parameter 'duration' or the parameter 'resonances' to correctly identify a piece, some of the stimuli were acoustically manipulated. The rationale was to create a set of stimuli carrying mismatching acoustic information and test which acoustic cues the listeners would be more reliant on in deciding which word they heard. The following table shows the features that characterise the original *she's* and *she was* and the new stimuli that have been created through the manipulation of the original sound files. There are two variables in each of the two parameters, which means that four combinations are possible: two matching (naturally-occurring) combinations and two non-matching (manipulated) combinations. The two matching combinations are: SHORT

[16] The small capitals for the labels SHORT and LONG, CLEAR and DARK are used to indicate that these features are relative qualities rather than absolute values.

and CLEAR (*she's*), and LONG and DARK (*she was*). The two non-matching combinations are: *SHORT and DARK, and *LONG and CLEAR. Table 33 exemplifies the four combinations.

Pr + aux	Resonances	Duration
She's	CLEAR	SHORT
She was	DARK	LONG
Manipulated	CLEAR	LONG
Manipulated	DARK	SHORT

Table 33. The four combinations of features. Only the first two combinations exist in natural speech; the third and the fourth combinations need to be created.

The assumption is that listeners can perceive and make use of both parameters – duration and resonances – to correctly interpret what they hear. However, one of the aims of the experiment is to find out if listeners rely on one of the two parameters more than the other if they are forced to choose between them. To test whether listeners rely more on one parameter than another, the stimuli have been manipulated to contain contrasting information. In the experiment, the listener is forced to choose between two visual options displayed on the screen. One of the options matches the duration of the auditory stimulus; the other option matches the resonances of the auditory stimulus. Figure 181 illustrates what the participants of the experiment hear and the phonetic features of the two options they see on screen.

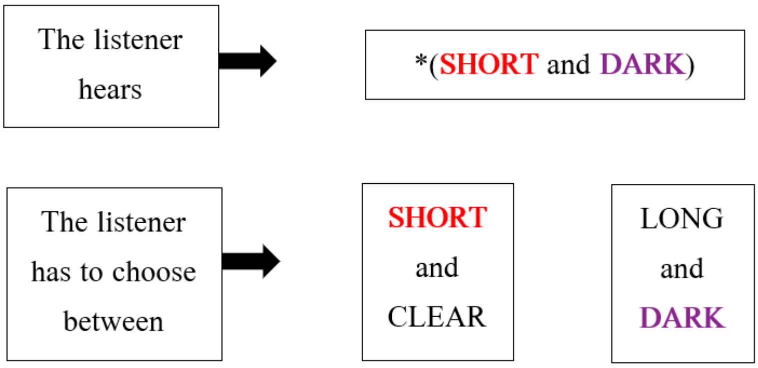


Figure 181. At the top, the mismatching acoustic cues in the signal, such as *(SHORT and DARK); at the bottom the phonetic features of the two options given to the listeners which contain matching information with only one parameter (the one highlighted).

For the non-matching stimuli, the original sound files recorded during the production study were used. Because manipulating the duration usually results in more natural sounding stimuli than manipulating the resonances, only the duration was manipulated and everything else was kept as in the original tokens, including the resonances.

The resonances can be modified by manipulating the formants of sonorant sounds. However, it seemed that the output of manipulating the duration alone instead of the resonances was appropriate for the aims of the experiment. The resulting non-matching stimuli have a manipulated duration that contrasts with the resonances of the paradigm. Table 34 illustrates the matching and non-matching stimuli (in blue the manipulated parameters).

Pr + aux	Resonance	Duration
She's	CLEAR	SHORT
She was	DARK	LONG
She's + lengthened vocoid	CLEAR	*LONG
She was + shortened vocoid	DARK	*SHORT

Table 34. The four combinations of features in the four stimuli with the addition of the two non-matching stimuli in the third and fourth rows. In blue the manipulated parameter (duration).

The selection of stimuli was carried out through careful auditory and spectrographic inspection. On the one hand, the resonances of each stimuli had to be unmistakably CLEAR or unmistakably DARK. Pr + aux in which the resonances sounded ambiguous were excluded. On the other hand, the formants of the vocoid had to be quite stable so that removing a portion of the vocoid would not incur in loss of spectral information. Luckily, most of the data collected in the production study are characterised by reduction in gesture magnitude which leads to flat steady formants. Fifty-four stimuli were included in Experiment B.

To increase the duration of a vocoid, splicing was used. The rationale for choosing this method to increase the vocoid duration, instead of time-domain manipulation, is that it allows more control over the output. During the selection of the stimuli, it was possible to test a range of degrees of increased duration, as well as the naturalness of the output, through manual manipulation. This method was therefore deemed suitable for the manipulation of duration. To increase the duration of a vocoid, suitable glottal cycles of vocal fold vibration were duplicated and inserted next to the original ones they were duplicated from. A glottal cycle was considered suitable for duplication if its formants were steady and aligned with the preceding and following glottal cycles, and its amplitude was the same or similar to the amplitude of the preceding and following glottal cycles, as visible from spectral observation. In other words, if a glottal cycle did not exhibit any considerable variation from the preceding and following glottal cycles, it was considered suitable for duplication. The first two glottal cycles and the last two glottal cycles of each vocoid were usually discarded as they carry spectral information about the preceding and following contoid. Figure 182 and Figure 183 illustrate an example of glottal cycle duplication and insertion.

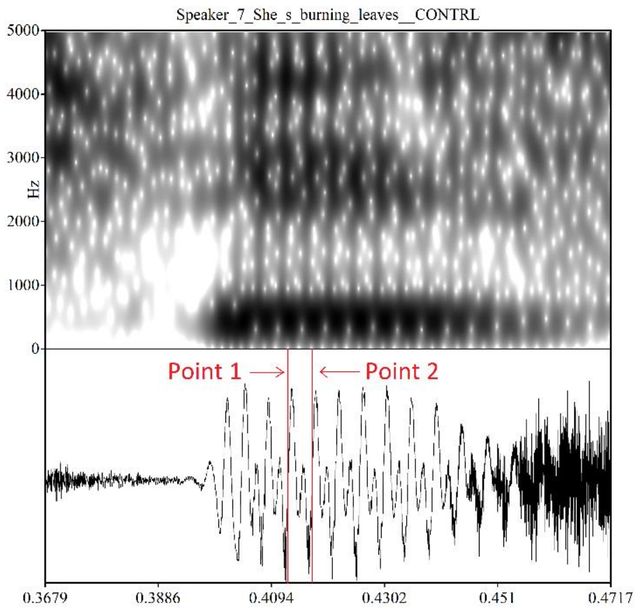


Figure 182. Selection of a cycle of vocal fold vibration for duplication.

Figure 182 shows the selection of a glottal cycle from a zero-crossing (Point 1) to a zero-crossing (Point 2). In this instance, the vocoid's duration is 35 ms. The duration of the glottal cycle selected from Point 1 to Point 2 is 4.3 ms. The glottal cycle selected would then be duplicated and inserted at Point 2. Figure 183 shows where the duplicated glottal cycle is inserted.

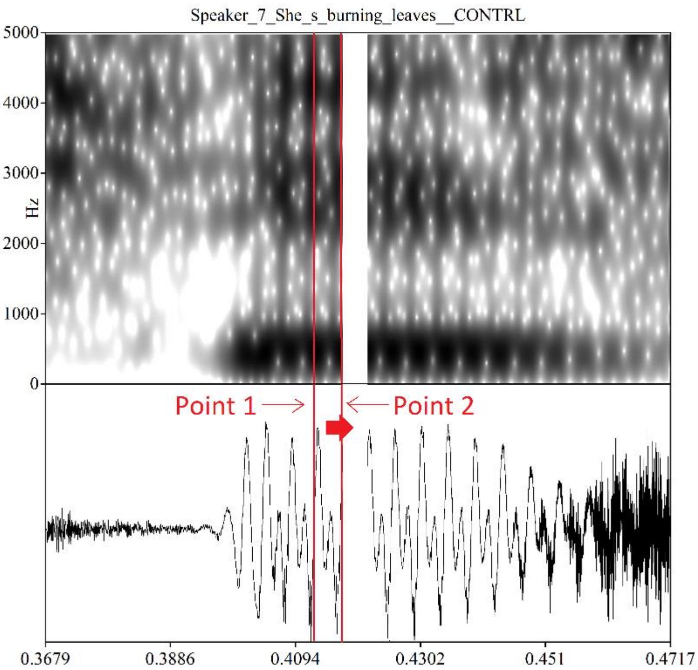


Figure 183. Insertion of a cycle of vocal fold vibration for duplication.

The next suitable glottal cycle would then be duplicated and inserted at the end of the glottal cycle itself. This process would be repeated until the chosen vocoid duration was reached or until no more duplication was possible because the resulting vocoid would sound unnatural. This was the case in instances in which the original vocoid was particularly short. It was of prime importance to maintain the stimuli as perceptually natural as possible. Increasing the duration too much would result in a distorted sound.

To decrease the duration of a vocoid, suitable glottal cycles were selected and removed. A glottal cycle was considered suitable for removal if its formants were steady, and its amplitude was similar to the amplitude of the preceding and following glottal cycles, as visible from spectral observation.

The acoustic analysis reported in Chapter 5 showed that in pairs of paradigms in which one of the contrasting pairs is pr + *'d*, such as *I'd* and *I'll*, the duration of the closure is significantly different too. Therefore, also the duration of the closure had to be increased or decreased in the stimuli selected for the perception experiment. To increase the duration of the closure, a portion of silence of the original closure was duplicated and inserted next to the original portion of silence. The next section describes the stimuli selected for the perception experiment.

8.2.2. Stimuli selection

To test the perceptual salience of duration and resonances, contrasting pairs of paradigms were selected and manipulated as explained in the previous section. The contrasting pairs tested in the perception experiment are: *will* versus *would*, *is* versus *was*, and *are* versus *were*. Due to the small number of pr + aux sentences recorded by each speaker (five repetitions for each pr + aux per speaker), it was not possible to have all the stimuli produced by only one speaker. The selection of the stimuli had to take into account several factors, including the duration and quality of the vocoid in each pr + aux. However, in most cases, it was possible to use the sentences produced by one speaker for all the combinations of each contrasting pair (but different speakers for different pairs). In order to keep the perception experiment short to avoid listeners fatigue, only a few paradigms were tested. For example, *you're* versus *you were* and *we're* versus *we were* were included in the experiment, but not *they're* versus *they were*. The following sections describe the tokens selected and manipulated for each contrasting pair of paradigms.

8.2.2.1. Stimuli for the contrast between *will* and *would*

In the pair *will* and *would*, two parameters are involved in maintaining the contrast: the closure duration and the vocoid resonances. Table 35 shows the possible combinations of duration and resonances in *will* and *would*, including the manipulated combinations. The asterisk (*) indicates the parameter that was manipulated.

Closure Duration Category	Vocoid Resonances Category	Actual pr + aux
SHORT	DARK	pr + *will*
LONG*	DARK	
LONG	CLEAR	pr + *would*
SHORT*	CLEAR	

Table 35. The four combinations of features used to test the contrast between pr + *'ll* and pr + *'d*. The asterisk (*) indicates that the duration had to be manipulated.

As shown in Table 35, the resonances of the vocoid are DARK in pr + *'ll* and CLEAR in pr + *'d*. The duration of the closure is LONG in pr + *'d* , and SHORT in pr + *'ll*. Therefore, the combinations of the natural-occurring paradigms are: DARK resonances with SHORT closure duration (pr + *'ll*), and CLEAR resonances with LONG closure duration (pr + *'d*). Two non-matching combinations of resonances and duration were created. One combination combines a SHORT closure duration with CLEAR vocoid resonances. The other combination combines a LONG closure duration with DARK vocoid resonances. While the natural resonances of the pieces were maintained, the duration of the closures were manipulated.

In addition, the paradigms *I'll* and *I'd* differ significantly also in the duration of the vocoid. In this case, three parameters had to be tested and eight combinations were needed. Section 8.2.2.1.3 describes the stimuli selection of *I'll* and *I'd*.

The contrast between the cliticised forms of *will* and *would* was tested in combination with four pronouns: *I, she, you*. The following sections describe the sets of stimuli included in the perception experiment: *she'll* versus *she'd*, *you'll* and *you'd*, and *I'll* versus *I'd*.

8.2.2.1.1. *She'll* versus *she'd*

The pieces *she'll* and *she'd* differ in duration of the closure and resonances of the vocoid. In *she'd*, the resonances are CLEAR. In *she'll*, the resonances are DARK. The duration of the closure in *she'd* is LONG. The duration of the closure in *she'll* is SHORT. Therefore, there are four possible combinations of paradigms. The two matching combinations are: LONG and CLEAR (original paradigm *she'd*), and SHORT and DARK (original paradigm *she'll*). The two non-matching combinations are: LONG and DARK (duration manipulated from the base token *she'll*), and SHORT and CLEAR (duration manipulated from the base token *she'd*).

The mean duration of the closure in *she'll* is 72 ms. The mean duration of the closure in *she'd* is 95 ms. To increase and decrease the duration of the closure, the method explained in Section 8.2.1 was followed. Eight stimuli produced by the same speaker (S8) were selected and are reported in Table 36 with the actual closure duration (in ms).

Stimuli name	**Vocoid Resonances Category**	**Closure Duration Category**	**Actual Closure Duration**
Speaker_8_She_d_burn_the_cake_ShortVDur-27ms_LongHphDur-106ms_ CONTROL	CLEAR	LONG	106 ms
Speaker_8_She_d_burn_the_cake_LongVDur-36ms_LongHphDur-104ms_ CONTROL	CLEAR	LONG	104 ms
Speaker_8_She_d_burn_the_cake_ShortVDur-27ms_ShortHphDur-68ms	CLEAR	SHORT*	68* ms
Speaker_8_She_d_burn_the_cake_LongVDur-36ms_ShortHphDur-68ms	CLEAR	SHORT*	71* ms
Speaker_8_She_ll_burn_the_roast_LongVDur-37ms_ShortHphDur-69ms_ CONTROL	DARK	SHORT	69 ms
Speaker_8_She_ll_burn_the_rice_ShortVDur-14ms_ShortHphDur-71ms_ CONTROL	DARK	SHORT	71 ms
Speaker_8_She_ll_burn_the_rice_ShortVDur-14ms_LongHphDur-108ms	DARK	LONG *	108* ms

Speaker_8_She_ll_burn_the_roast_LongVDur-37ms_LongHphDur-112ms	DARK	LONG *	112* ms

Table 36. The eight stimuli used to test the resonances and duration in the contrasting pair *she'd* and *she'll*. Duration category and values followed by an asterisk (*) indicate that the duration was manipulated.

8.2.2.1.2. *You'll* versus *you'd*

Similarly to *she'll* and *she'd*, *you'll* and *you'd* differ in duration of the closure and resonances of the vocoid. In *you'd*, the resonances are CLEAR. In *you'll*, the resonances are DARK. The duration of the closure in *you'd* is LONG. The duration of the closure in *you'll* is SHORT. Therefore, there are four possible combinations. Six stimuli were selected and are reported in Table 37.

Stimuli name	**Vocoid Resonances Category**	**Closure Duration Category**	**Actual Closure Duration**
Speaker_7_You_d_burn_the_pie_LongVDur-45ms_LongHphDur-102ms_CONTROL	CLEAR	LONG	106 ms
Speaker_7_You_d_burn_the_pie_LongVDur-45ms_ShortHphDUR-62ms	CLEAR	SHORT*	62* ms
Speaker_7_You_ll_burn_the_roast_LongVDur-47ms_ShortHphDur-71ms_ CONTROL	DARK	SHORT	71 ms
Speaker_7_You_ll_burn_the_pie_RED_ShortVDur-32ms_ShortHphDur-78ms_ CONTROL	DARK	SHORT	78 ms
Speaker_7_You_ll_burn_the_pie_ShortVDur-32ms_LongHphDur-102ms	DARK	LONG*	102* ms
Speaker_7_You_ll_burn_the_roast_LongVDur-47ms_LongHphDur-105ms	DARK	LONG*	105* ms

Table 37. The six stimuli used to test the resonances and duration in the contrasting pair *you'd* and *you'll*. Duration category and values followed by an asterisk (*) indicate that the duration was manipulated.

8.2.2.1.3. *I'll* versus *I'd*

In the pair *I'll* and *I'd*, three parameters are involved in maintaining the contrast. In addition to the vocoid resonances and the duration of the closure, the acoustic analysis (see Section 5.2.2.2) revealed that also the duration of the vocoid is significantly different between *I'll* and *I'd*. The paradigm *I'll* is characterised by DARK resonance, LONG vocoid duration, and SHORT closure duration. In contrast, the paradigm *I'd* is characterised by CLEAR resonance, SHORT vocoid duration, and LONG closure duration. To test the perceptual salience of the three parameters, eight combinations had to be created. Table 38 exemplifies the eight combinations included in the experiment.

Vocoid Duration Category	Closure Duration Category	Vocoid Resonances Category	Original pr + aux
LONG	SHORT	DARK	pr + *will*
SHORT*	SHORT	DARK	
LONG	LONG*	DARK	
SHORT*	LONG*	DARK	
SHORT	LONG	CLEAR	pr + *would*
LONG*	LONG	CLEAR	
SHORT	SHORT*	CLEAR	
LONG *	SHORT*	CLEAR	

Table 38. The eight combinations of features used to test the contrast between pr + *'ll* and pr + *'d*. The asterisk (*) indicates that the duration had to be manipulated.

The mean duration of the vocoid in *I'll* is 53 ms, while the mean duration of the closure is 76 ms. The mean duration for the vocoid in *I'd* is 42 ms, while the mean duration of the closure is 97 ms. It was not possible to find suitable instances of all natural (matching) combinations uttered by a single speaker. So the stimuli from five speakers (S8, S9, S10, S11, S14) were selected. Table 39 reports the 14 stimuli included in the experiment with the categories for each parameter and the manipulated durations in ms.

Stimuli name	Vocoid Resonances Category	Vocoid Duration Category	Closure Duration Category and actual duration
Speaker_11_I_d_burn_the_fish_CONTROL	CLEAR	SHORT	LONG
Speaker_11_I_d_burn_the_fish_ShortHphDur-74ms	CLEAR	SHORT	SHORT* 74 ms
Speaker_14_I_d_burn_the_cake_ShortHphDur-68ms	CLEAR	SHORT	SHORT* 68 ms
Speaker_9_I_d_burn_the_pie_ShortHphDur-70ms	CLEAR	SHORT	SHORT* 70 ms
Speaker_9_I_d_burn_the_pie_LongVDur-56ms_LongHphDur96ms	CLEAR	LONG* 56 ms	LONG
Speaker_10_I_d_burn_the_pie_LongVDur-56ms_LongHphDur-101ms	CLEAR	LONG* 32 ms	LONG
Speaker_10_I_d_burn_the_pie_LongVDur-56ms_ShortHphDur-85ms	CLEAR	LONG* 56 ms	SHORT* 85 ms
Speaker_14_I_ll_burn_the_roast_CONTROL	DARK	LONG	SHORT
Speaker_8_I_ll_burn_the_roast_LongHphDur-102ms	DARK	LONG	LONG* 102 ms
Speaker_8_I_ll_burn_the_chips_RED_LongHphDur-103ms	DARK	LONG	LONG* 103 ms
Speaker_14_I_ll_burn_the_steak_ShortVDur-32ms_ShortHphDur-78ms	DARK	SHORT* 32 ms	SHORT
Speaker_14_I_ll_burn_the_roast_ShortVdur-43ms	DARK	SHORT* 43 ms	SHORT
Speaker_14_I_ll_burn_the_steak_ShortVDur-32ms_LongHphDur-102ms	DARK	SHORT* 32 ms	LONG* 102 ms
Speaker_14_I_ll_burn_the_roast_ShortVdur-43ms_LongHphDur-108ms	DARK	SHORT* 43 ms	LONG* 108 ms

Table 39. The 14 stimuli used to test the durations in the contrasting pair *I'd* and *I'll*. Duration category and values followed by an asterisk (*) indicate that the duration was manipulated.

8.2.2.2. Present versus past tense of BE

In Chapter 5 it was reported that the contrast between the present and past tense of the auxiliary verb BE is conveyed by the duration of the vocoid and, in some cases, by the resonances of the vocoid. Three pairs of paradigms were selected and manipulated: *she's ~she was*, *we're ~ we were*, and *you're ~ you were*. All three pairs of paradigms differ in the duration of the vocoid. In addition, *she's* and *she was* differ in the vocoid resonances. Although the pairs *we're* and *we were*, and *you're* and *you were* exhibit similar formant dynamics, auditorily their resonances are CLEAR in the present tense and DARK in the past tense. The next sections describe the matching and non-matching stimuli for each pair of paradigms.

8.2.2.2.1. *She's* versus *she was*

She's is characterised by SHORT vocoid duration and CLEAR resonances. *She was* is characterised by LONG vocoid duration and DARK resonances. There are four possible combinations of paradigms to test: SHORT and CLEAR (original paradigm *she's*), LONG and CLEAR (duration manipulated from *she's* base), LONG and DARK (original paradigm *she was*), and SHORT and DARK (duration manipulated from *she was* base).

The mean duration of the vocoid in *she's* is 36 ms (N = 34)[17]. The mean duration for the vocoid in *she was* is 85 ms (N = 43). The duration of the vocoid in *she's* was increased following the process of glottal cycle reduplication explained above. Unfortunately, it was not possible to increase the duration of the chosen instance of *she's* to a level that would

[17] The low number of tokens in which the vocoid duration could be measured is due to the high degree of reduction. Several instances of *she's* were produced without a vocoid or without voicing.

make it comparable to the duration of *she was*. The threshold after which the chosen stimuli *she's burning leaves* uttered by S7 would start sounding unnatural was 57 ms. The duration of the vocoid of the original piece was 40 ms. In hindsight, a different token should have been chosen to create a stimulus with a longer duration closer to the mean duration of 85 ms. The duration of the vocoid in *she was* was decreased following the process of removal explained above. The stimulus *she was burning straw* uttered by the same speaker (S7) was chosen. The four stimuli used are shown in Table 40.

Stimuli name	Vocoid Resonance Category	Vocoid Duration Category	Actual Vocoid Duration
Speaker_7_She_s_burning_leaves_CONTROL	CLEAR	SHORT	40 ms
Speaker_7_She_s_burning_leaves	CLEAR	LONG*	57* ms
Speaker_7_She_was_burning_straw_CONTROL	DARK	LONG	78 ms
Speaker_7_She_was_burning_straw	DARK	SHORT*	36* ms

Table 40. The four stimuli used to test the resonances and duration in the contrasting pair *she's* and *she was*. Duration category and values followed by an asterisk (*) indicate that the duration was manipulated.

8.2.2.2.2. *We're* versus *we were*

The pieces *we're* and *we were* differ in duration of the vocoid, which is SHORT in *we're* and LONG in *we were.* The mean duration of the vocoid in *we're* is 46 ms (N = 46). The mean duration of the vocoid in *we were* is 136 ms (N = 49). Although the acoustic analysis reported in Chapter 5 did not show a clear-cut distinction between the formant dynamics of *we're* and *we were*, the auditory analysis suggested that the quality of the two paradigms differ, albeit slightly. In *we're*, the resonances are CLEAR. In *we were*, the resonances are DARK. Therefore, the four possible combinations of paradigms are: SHORT and CLEAR (original paradigm *we're*), LONG and CLEAR (duration manipulated from *we're* base), LONG and DARK (original paradigm *we were*), and SHORT and DARK (duration manipulated from *we were* base). Twelve stimuli were used and are shown in Table 41.

Stimuli name	**Vocoid Resonance Category**	**Vocoid Duration Category**	**Actual Vocoid Duration**
Speaker_6_We_re_burning_logs_CONTROL	CLEAR	SHORT	48 ms
Speaker_6_We_re_burning_wood_CONTROL	CLEAR	SHORT	26 ms
Speaker_10_We_re_burning_wood_CONTROL	CLEAR	SHORT	40 ms
Speaker_6_We_re_burning_logs_LongVDur-104ms	CLEAR	LONG*	104* ms
Speaker_10_We_re_burning_logs_LongVDur-103ms	CLEAR	LONG*	103* ms
Speaker_15_We_re_burning_leaves_LongVDur-120ms	CLEAR	LONG*	120* ms
Speaker_6_We_were_burning_leaves_CONTROL	DARK	LONG	144 ms
Speaker_6_We_were_burning_straw_CONTROL	DARK	LONG	138 ms
Speaker_10_We_were_burning_leaves_CONTROL	DARK	LONG	114 ms
Speaker_6_We_were_burning_leaves_ShortVDur-58ms	DARK	SHORT*	58* ms
Speaker_10_We_were_burning_leaves_ShortVDur-52ms	DARK	SHORT*	52* ms
Speaker_10_We_were_burning_leaves_ShortVDur-73ms	DARK	SHORT*	73* ms

Table 41. The 12 stimuli used to test the resonances and duration in the contrasting pair *we're* and *we were*. Duration category and values followed by an asterisk (*) indicate that the duration was manipulated.

8.2.2.2.3. *You're* versus *you were*

Similarly to *we're* and *we were*, the pieces *you're* and *you were* differ in duration of the vocoid, which is SHORT in *you're* and LONG in *you were*. Although the acoustic analysis of the resonances was inconclusive, the auditory impression is that the resonances in *you're* are CLEAR, and the resonances *you were* are DARK. Four possible combinations of paradigms are thus possible: SHORT and CLEAR (original paradigm *you're*), LONG and CLEAR (duration manipulated from *you're* base), LONG and DARK (original paradigm *you were*), and SHORT and DARK (duration manipulated from *you were* base).

The mean duration of the vocoid in *you're* is 50 ms (N = 50). The mean duration of the vocoid in *you were* is 114 ms (N = 49). Nine stimuli were selected and are shown in Table 42.

Stimuli name	Vocoid Resonance Category	Vocoid Duration Category	Actual Vocoid Duration
Speaker_13_You_re_burning_straw_ CONTROL	CLEAR	SHORT	51 ms
Speaker_13_You_re_burning_straw_LongVDur-86ms	CLEAR	LONG*	86* ms
Speaker_13_You_re_burning_straw_LongVDur-105ms	CLEAR	LONG*	105* ms
Speaker_13_You_were_burning_straw_ CONTROL	DARK	LONG	104 ms
Speaker_13_You_were_burning_logs_ShortVDur-57ms	DARK	SHORT*	57* ms
Speaker_15_You_were_burning_straw_ CONTROL	DARK	LONG	81 ms
Speaker_15_You_were_burning_straw_ShortVDur-58ms	DARK	SHORT*	58* ms
Speaker_7_You_were_burning_logs_ CONTROL	DARK	LONG	93 ms
Speaker_7_You_were_burning_logs_ShortenedVDur-58ms	DARK	SHORT*	58* ms

Table 42. The nine stimuli used to test the duration in the contrasting pair *you're* and *you were*. Duration category and values followed by an asterisk (*) indicate that the duration was manipulated.

8.2.3. Participants and procedure

The participants and procedure of the perception experiment were explained in Chapter 7, Section 7.2.3.

8.3. Data analysis and results

This section reports the results of Experiment B. The perceptual salience of the duration of the vocoid and/or closure, and by extension the perceptual salience of the resonances of the vocoids in contrasting pairs of paradigms were tested. Both matching and non-matching (manipulated) stimuli were included in the experiment. The matching stimuli acted as control stimuli. In total, the stimuli in Experiment B were 54. The total number of responses for Experiment B was 3,456. This section is divided according to the paradigms tested and follows the structure of the stimuli preparation reported in Section 8.2.2.

8.3.1. *Will* versus *would*

The auxiliaries *will* and *would* in their contracted forms *'ll* and *'d* were tested in combination with the pronouns *I*, *she*, and *you*. The following sections report on the results of the perception experiment divided by pr + aux. First, the results for the paradigms *she'll~she'd*, and *you'll~you'd* are reported, then *I'll~I'd*.

8.3.1.1. *She'll* versus *she'd*

The paradigms *she'll* and *she'd* have significantly different closure duration. In addition, they contrast also in their resonances: *she'll* is characterised by SHORT vocoid duration and DARK resonances, while *she'd* is characterised by LONG closure duration and CLEAR resonances.

Eight stimuli were included in the experiment including four control (matching) stimuli. Figure 184 shows the responses to the *she'd* stimuli: on the left, the two control (matching CLEAR and LONG) stimuli, and on the right the two non-matching (CLEAR and SHORT) stimuli manipulated from *she'd* base tokens.

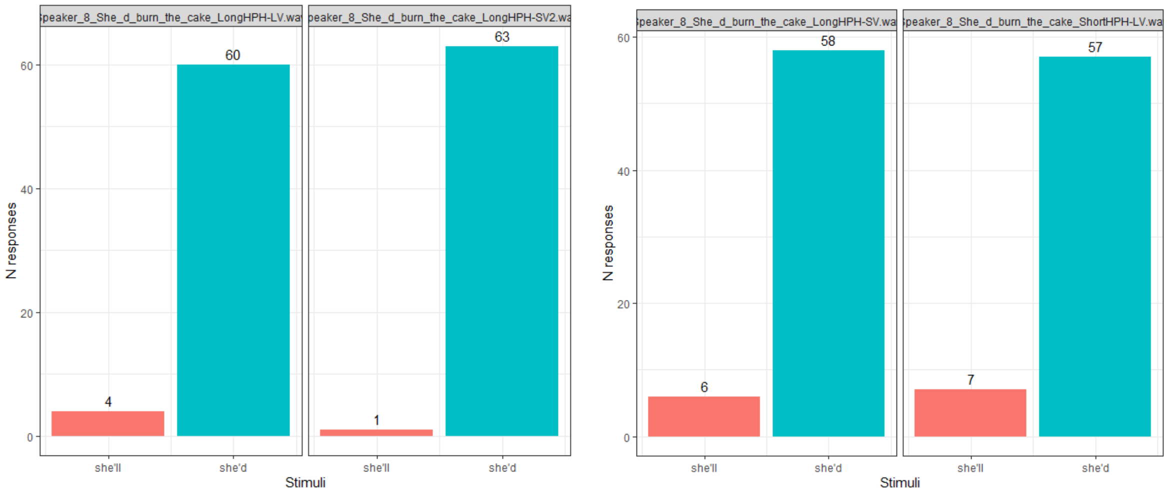


Figure 184. Responses counts for the four *she'd* stimuli including the two control stimuli (left), and the two manipulated non-matching stimuli (right).

Figure 184 shows that for both matching and non-matching *she'd* stimuli, the large majority of participants chose the option *she'd* on screen and not *she'll*. That is, even the *she'd* stimuli with a shortened closure duration to match the closure duration of *she'll*, were identified as *she'd*. This suggests that the duration of the closure did not play a major role in the identification of the pr + aux, and that listeners attended to other phonetic features such as the resonances to identify the pr + aux.

Figure 185 shows the responses to the *she'll* stimuli: on the left, the two control (matching DARK and SHORT) stimuli, and on the right the two non-matching (DARK and LONG) stimuli manipulated from *she'll* base tokens.

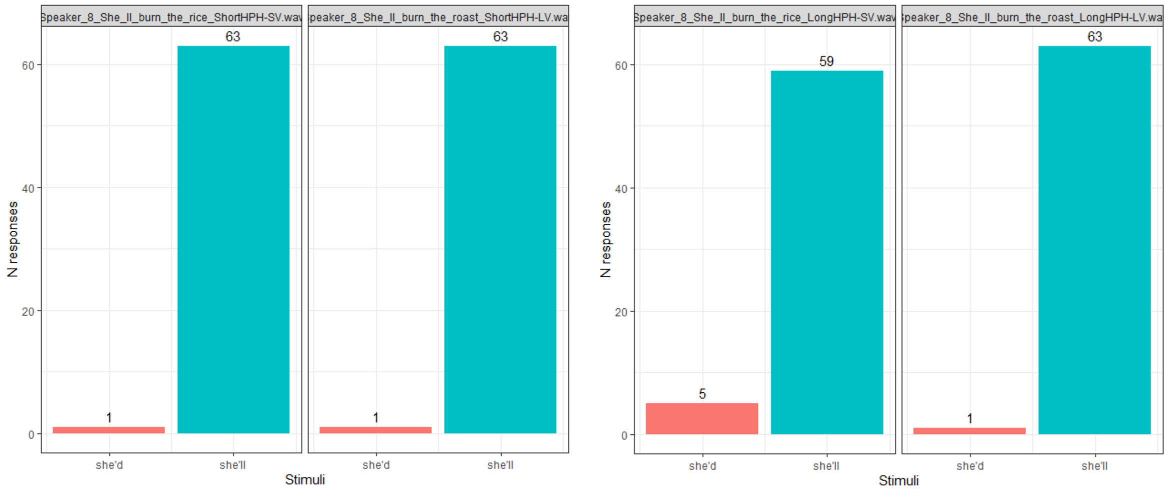


Figure 185. Responses counts for the four *she'll* stimuli including the two control stimuli (left), and the two manipulated non-matching stimuli (right).

Figure 185 shows that for both matching and non-matching *she'll* stimuli, the large majority of participants chose the option *she'll* on screen and not *she'd*. That is, even *she'll* stimuli with a manipulated LONG closure duration to match the closure duration of *she'd*, were identified as *she'll*. Similarly to the responses for the *she'd* stimuli, also for the *she'll* stimuli most listeners did not seem to identify the stimuli on the basis of the duration of the closure.

As shown in Figure 184 and Figure 185, in all cases (control and non-matching), listeners chose the option on the screen that represented the 'base' stimuli, which means that the change in duration did not affect the identification of the pr + aux. That is, the large majority of participants chose the option that matched the resonances of the non-matching stimuli regardless of the duration of the closure. An exact binomial test indicates that the responses for the *she'd* and *she'll* stimuli (N = 512) were significantly above chance ($p < 0.001$; 95% CI [0.06, 0.11]).

8.3.1.2. *You'll* versus *you'd*

The paradigms *you'll* and *you'd* have significantly different closure duration and differ in the vocoid resonances. While *you'll* is characterised by SHORT closure duration and DARK resonances, *you'd* is characterised by LONG closure duration and CLEAR resonances.

Six stimuli were used including three control (matching) stimuli. Figure 186 shows the responses to the *you'll* non-matching stimuli in which the duration of the closure of *you'll* base stimuli was manipulated to match the duration of the closure in *you'd*. These are the responses for the DARK and LONG non-matching *you'll* stimuli. Figure 187 reports the responses for the control matching *you'll* stimuli. These are the responses for the DARK and SHORT stimuli.

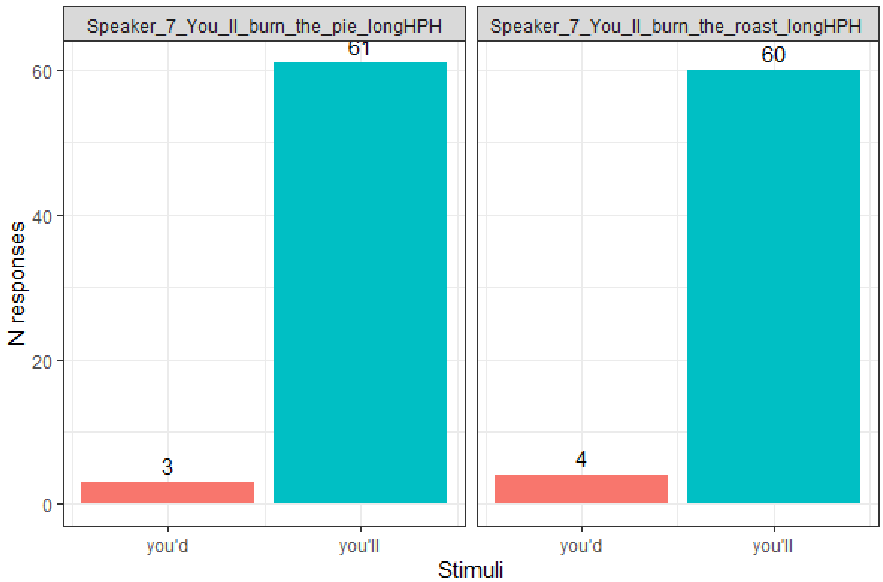


Figure 186. Responses counts for the two non-matching *you'll* stimuli in which the *you'll* base was manipulated to have a LONG closure duration. These are the responses for the non-matching DARK and LONG stimuli.

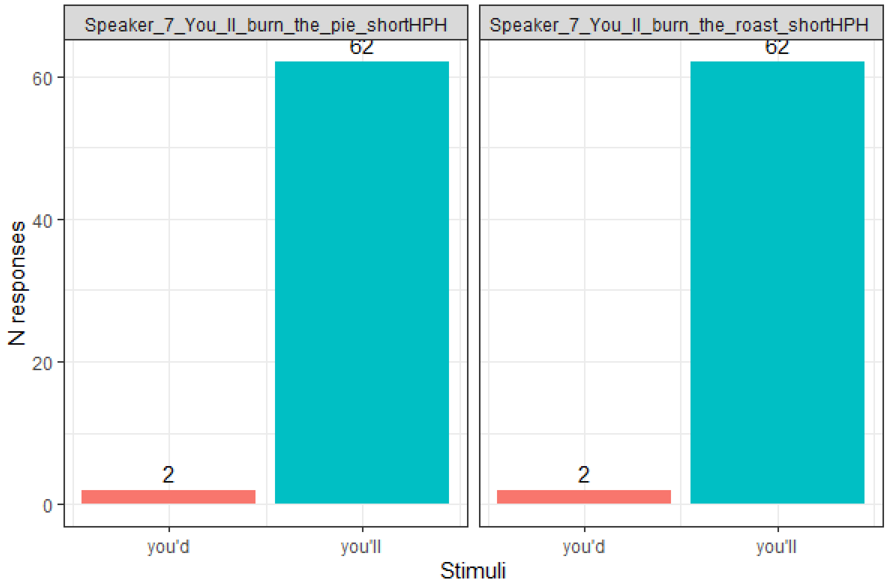


Figure 187. Responses counts for the control *you'll* stimuli. These are the responses to the natural DARK and SHORT *you'll* stimuli.

Figure 186 and Figure 187 show that both for the control and the non-matching manipulated stimuli, the large majority of listeners chose the option on the screen that represented the base stimuli *you'll*, regardless of the duration of the closure. This means that the change in duration did not affect the identification of the pr + aux.

Figure 188 reports the count of the responses to the *you'd* stimuli: on the left, the responses to the control (matching CLEAR and LONG) *you'd* stimulus, and on the right the responses to the non-matching (CLEAR and SHORT) stimulus manipulated from the same *you'd* base token.

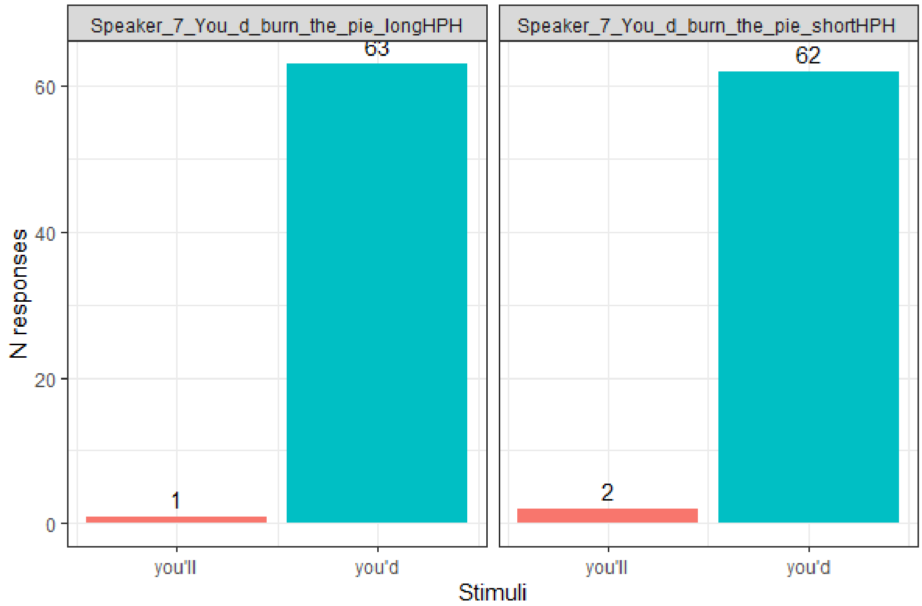


Figure 188. Responses counts for the control (left) and manipulated (right) *you'd* stimuli. On the right the responses to the non-matching CLEAR and SHORT stimulus.

Similarly to the responses for *you'll* stimuli, also in the identification of the non-matching *you'd* stimulus, listeners seem to rely on features other than the duration of the closure to correctly identify the pr + aux. An exact binomial test indicates that the responses to the *you'll* and *you'd* stimuli (N = 384) were significantly above chance ($p < 0.001$; 95% CI [0.02, 0.06]).

8.3.1.3. *I'll* versus *I'd*

The paradigms *I'll* and *I'd* have contrasting resonances and significantly different vocoid duration, and closure duration: *I'll* is characterised by DARK resonances, LONG vocoid duration and SHORT closure duration; while *I'd* is characterised by CLEAR resonances, SHORT vocoid duration and LONG closure duration.

Fourteen stimuli were used including two control stimuli. The following figures show, first, the counts of responses to the *I'll* stimuli (N = 7) with a manipulated SHORT vocoid duration

(Figure 189) and then the counts of responses to the *I'll* stimuli with a LONG vocoid duration (Figure 190).

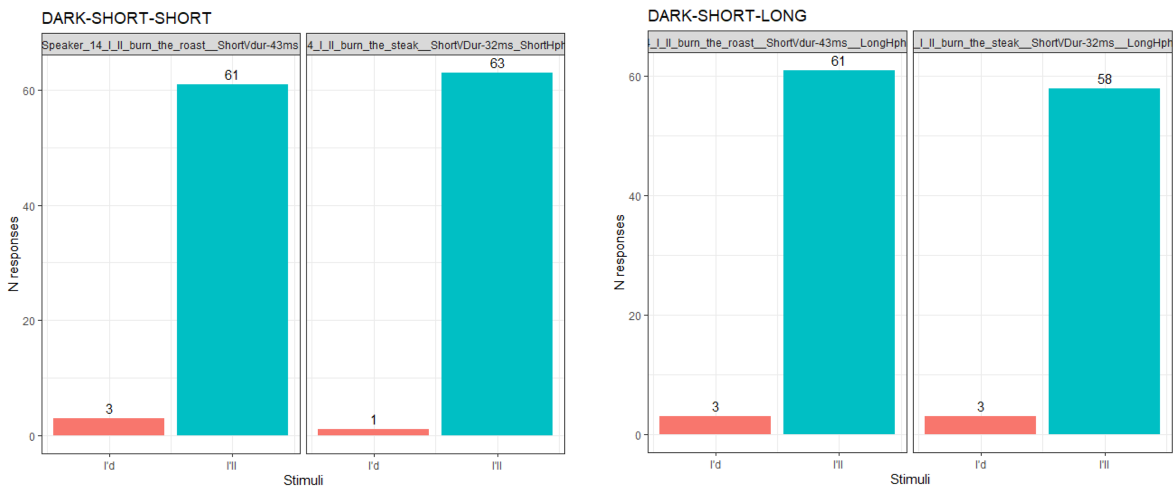


Figure 189. Responses counts for all the *I'll* stimuli with a DARK and SHORT (manipulated) vocoid duration. On the left, responses for the *I'll* stimuli with SHORT (manipulated) vocoid duration and SHORT (natural) closure duration (DARK-SHORT-SHORT). On the right, responses to the stimuli with SHORT (manipulated) vocoid duration and LONG (manipulated) closure duration (DARK-SHORT-LONG).

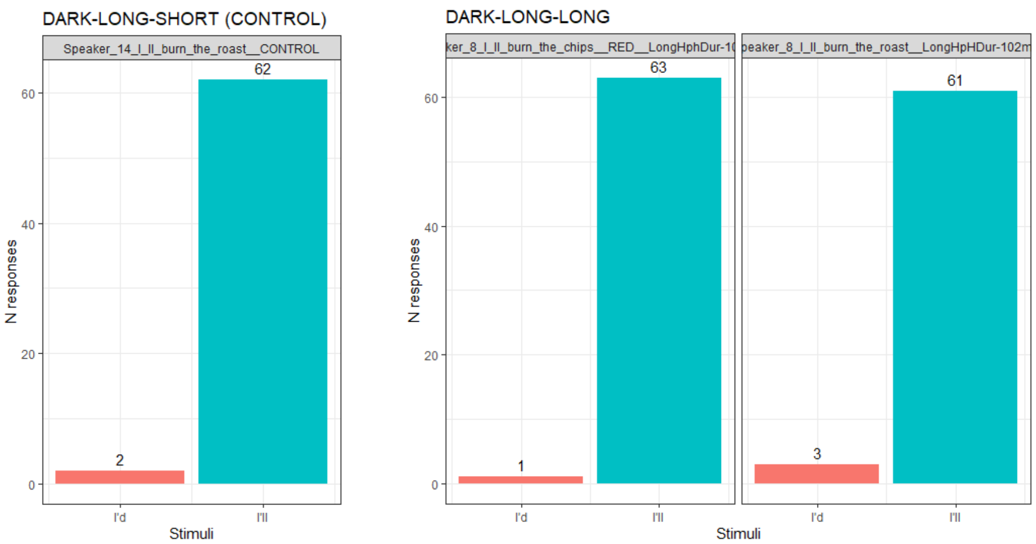


Figure 190. Responses counts for all the *I'll* stimuli with a DARK and LONG (natural) vocoid duration. On the left, responses for the *I'll* control stimuli with LONG vocoid duration and SHORT closure duration (DARK-LONG-SHORT). On the right, responses to the stimuli with LONG vocoid duration and LONG (manipulated) closure duration (DARK-LONG-LONG).

Figure 189 and Figure 190 show the responses to the stimuli with *I'll* base. All four combinations of SHORT and LONG vocoid and closure duration were tested with the DARK resonances of *I'll*. The responses indicate that participants are not affected by the durations of the vocoid and/or the closure to identify the stimuli. The great majority of participants identified the stimuli on the basis of the resonances and other phonetic features that were not manipulated, but not on the basis of the manipulated durations, as shown by the very similar results for all the four combinations tested. The participants that identified *I'll* as *I'd* in any of the combinations ranged only from 1 to 3 (1-5%).

The following figures show, first, the counts of responses to the *I'd* stimuli (N = 7) with a manipulated LONG vocoid duration (Figure 191) and then the counts of responses to the *I'd* stimuli with a SHORT vocoid duration (Figure 192).

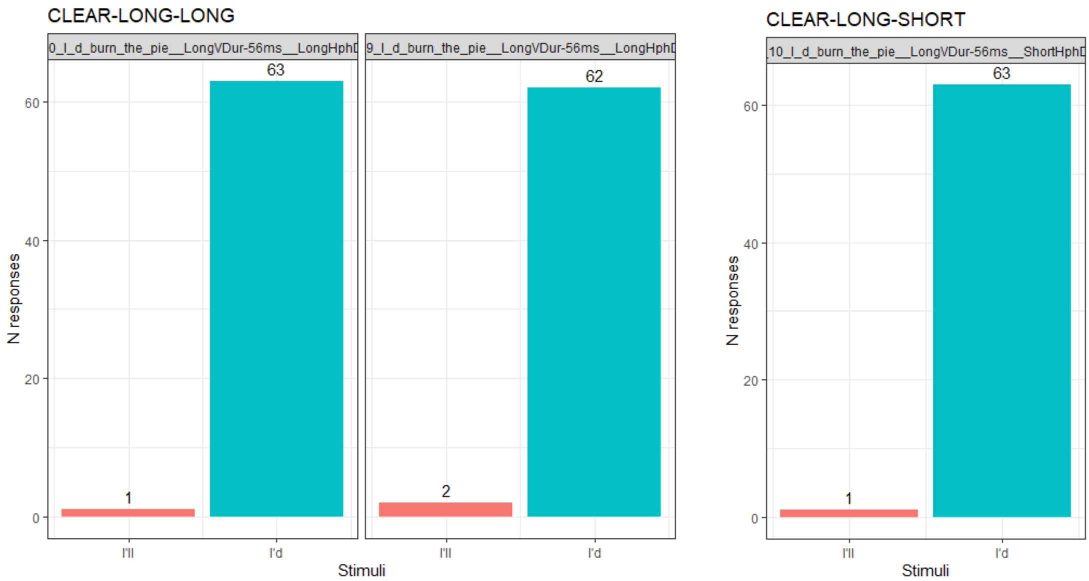


Figure 191. Responses counts for all the *I'd* stimuli with a CLEAR and LONG (manipulated) vocoid duration. On the left, responses for the *I'd* stimuli with LONG (manipulated) vocoid duration and LONG (natural) closure duration (CLEAR-LONG-LONG). On the right, responses to the stimuli with LONG (manipulated) vocoid duration and SHORT (manipulated) closure duration (CLEAR-LONG-SHORT).

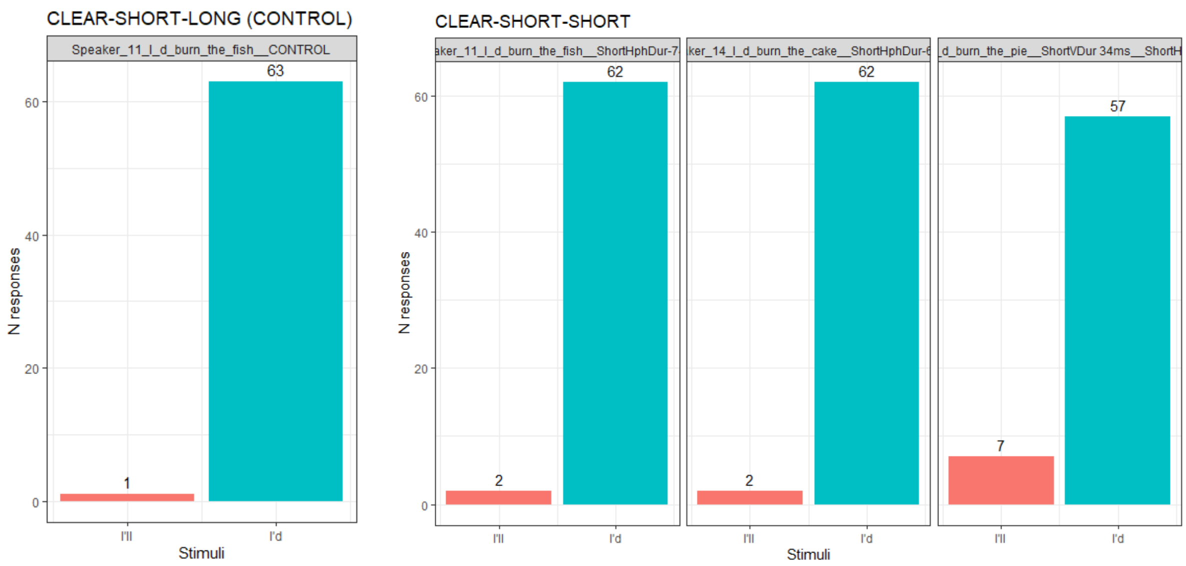


Figure 192. Responses counts for all the *I'd* stimuli with a CLEAR and SHORT vocoid duration. On the left, responses for the *I'd* control stimuli (CLEAR-SHORT-LONG). On the right, responses to the stimuli with SHORT (natural) vocoid duration and SHORT (manipulated) closure duration (CLEAR-SHORT-SHORT).

Similarly to the results for the *I'll* stimuli, the responses to the *I'd* stimuli presented in Figure 191 and Figure 192 show that the manipulation of the duration did not affect the responses of the participants and that listeners did not rely on the durations of the vocoid and/or the closure to identify the pr + aux they heard, but rather on other aspects of the stimuli such as the resonances. An exact binomial test indicates that the responses to the *I'll* and *I'd* stimuli (N = 896) were significantly above chance ($p < 0.001$; 95% CI [0.02, 0.05]).

8.3.2. Present versus past tense of BE

Three pairs of paradigms were selected and manipulated to test the perceptual salience of the duration and resonances in present versus past tense paradigms of BE: *she's ~ she was*, *we're ~ we were*, and *you're ~ you were*. The next sections report the analysis of the participant responses.

8.3.2.1. *She's* versus *she was*

This section reports the analysis of the participant responses to the control and manipulated stimuli for *she's* and *she was* (N = 4). *She's* is characterised by CLEAR resonances and SHORT vocoid duration, while *she was* is characterised by DARK resonances and LONG vocoid duration. Two *she's* stimuli and two *she was* stimuli were used. Figure 193 shows the number of responses for the two *she's* stimuli: on the left, the control stimulus *she's* (CLEAR and SHORT), and on the right the non-matching *she's* stimulus with a manipulated LONG vocoid duration (CLEAR and LONG). Figure 194 shows the count of responses for the *she was* stimuli: on the left, the control stimulus *she was* (DARK and LONG), and on the right the non-matching *she was* stimulus with a manipulated SHORT vocoid duration (DARK and SHORT).

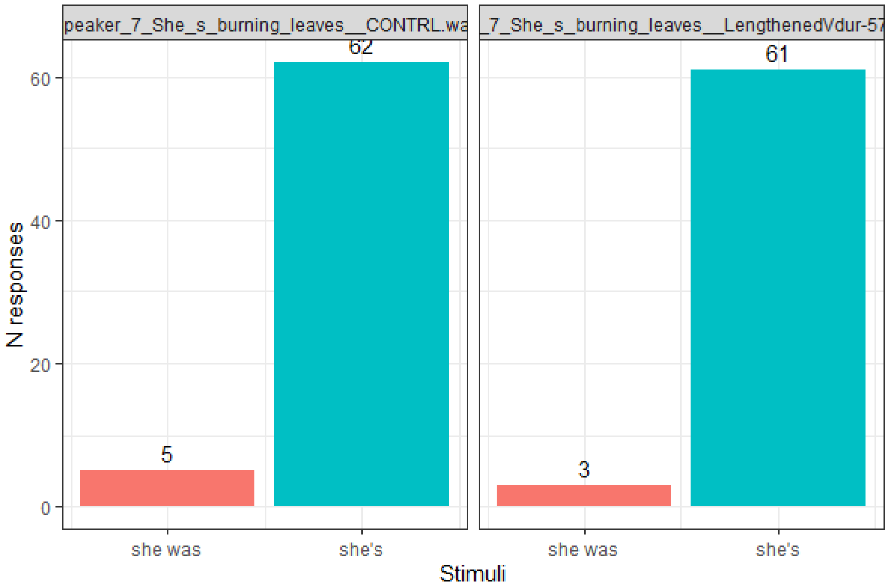


Figure 193. Responses counts for the control (left) and manipulated (right) *she's* stimuli. On the right the responses to the non-matching CLEAR and LONG stimulus with a manipulated vocoid duration.

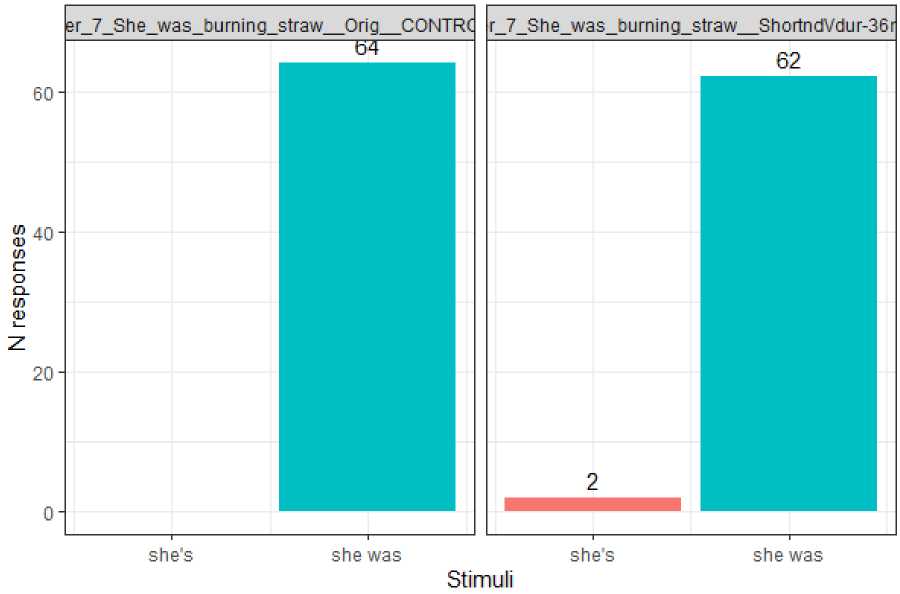


Figure 194. Responses counts for the control (left) and manipulated (right) *she was* stimuli. On the right, the responses to the non-matching DARK and SHORT stimulus with a manipulated vocoid duration.

The responses to *she's* and *she was* stimuli reported in Figure 193 and Figure 194 show that the variation in duration of the vocoid has little or no effect on participant responses, with a difference of only 3% between the number of responses to the control and manipulated stimuli. Unfortunately, the duration of the vocoid in *she's burning leaves* could not be increased more than 20 ms. This makes the results of this part of the test unreliable. An exact two-sided binomial test revealed that the responses to the *she's* and *she was* stimuli (N = 256) were significantly above chance ($p < .001$; 95% CI [0.02, 0.07]).

8.3.2.2. *We're* versus *we were*

This section reports the analysis of the responses of the participants to the control and manipulated stimuli for *we're* and *we were* (N = 12).

We're is characterised by CLEAR resonances and SHORT vocoid duration, while *we were* is characterised by DARK resonances and LONG vocoid duration. Twelve stimuli were included in the experiment: six *we're* stimuli (three control and three manipulated stimuli), and six *we were* stimuli (three control and three manipulated stimuli).

Figure 195 shows the number of responses for the three *we're* manipulated stimuli. The vocoid duration of original *we're* stimulus was manipulated to increase the vocoid duration. These are the CLEAR and LONG non-matching stimuli. Figure 196 shows the number of responses for the control *we're* stimuli. These are the original *we're* stimuli with matching resonances and vocoid duration (CLEAR and SHORT).

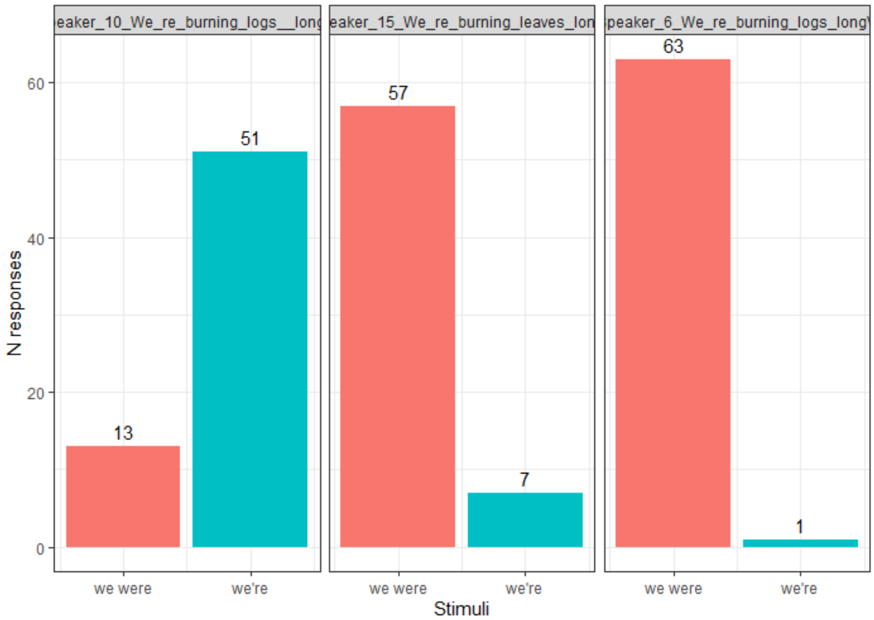


Figure 195. Responses counts for the manipulated *we're* stimuli. These are the responses for the non-matching CLEAR and LONG stimuli in which the vocoid duration was manipulated from *we're* base tokens.

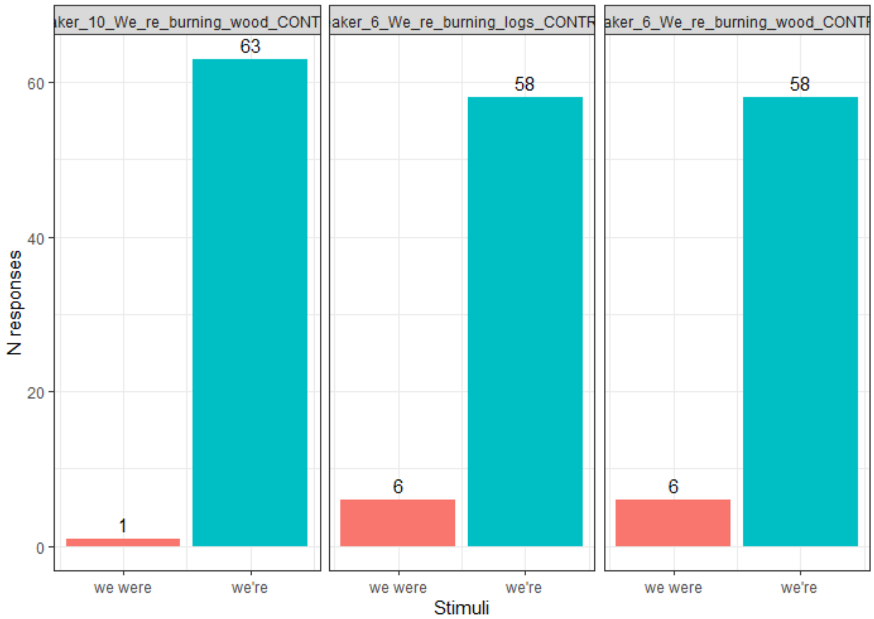


Figure 196. Responses counts for the control *we're* stimuli. These are the responses for the original matching CLEAR and SHORT stimuli.

Figure 195 shows that for two of the CLEAR and LONG non-matching stimuli, the large majority of participants (N = 57/64, 89% for the stimulus in the middle; N = 63/64, 98% for the stimulus on the right) chose the option *we were* rather than *we're*. This result seems to indicate that the participant responses were affected by the manipulated duration, as for two stimuli the participants chose the option that corresponded to the pr + aux with a LONG duration. However, one stimulus (*Speaker_10_We_re_burning_logs_LongVDur-103ms*) was identified as *we're* by the majority of participants (N = 51/64, 80%). On auditory inspection, the vocoid in this stimulus is characterised by a more open and less rounded quality, compared to the other two stimuli. Figure 197 shows the formant dynamics of two of the three CLEAR and LONG stimuli (*Speaker_6_We_re_burning_logs_LongVDur-104ms* and *Speaker_10_We_re_burning_logs_LongVDur-103ms*).

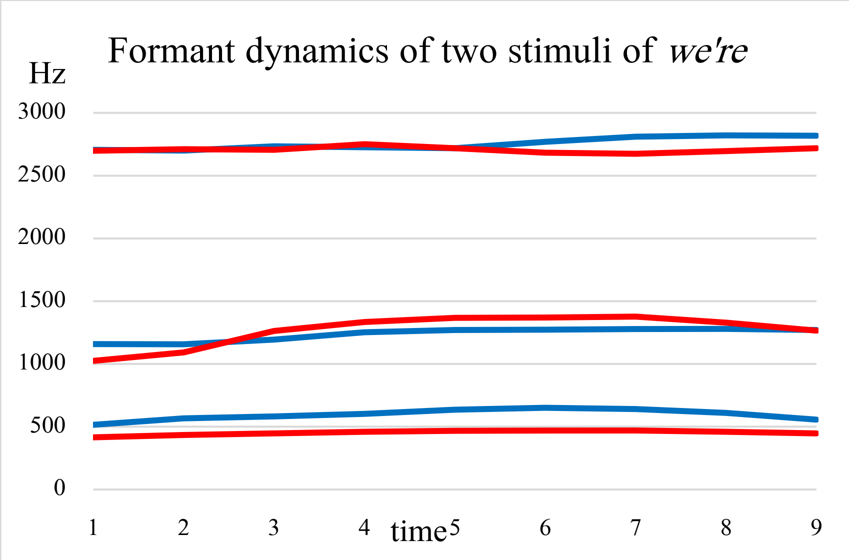


Figure 197. Formant dynamics of two of the three CLEAR and LONG non-matching stimuli. The instance in blue was identified as *we're* by the majority of listeners. The instance in red was identified as *we were* by the majority of listeners.

Figure 197 shows the formant dynamics of two instances of CLEAR and LONG non-matching stimuli of *we're*. In blue are the formant dynamics of the LONG instance of *we're* that was identified as *we're* by the majority of participants. In red are the formant dynamics of the LONG instance of *we're* that was identified as *we were* by the majority of participants. As expected, F1 is higher in the stimulus perceived as *we're*, which is characterised by a more open vocoid, while the stimulus that was interpreted as *we were* is characterised by a higher and moving F2 indicating that the quality of the vocoid changes in time.

Figure 198 shows the number of responses for the three *we were* manipulated stimuli. These are the DARK and SHORT non-matching stimuli. Figure 199 shows the number of responses for the control *we were* stimuli. These are the original *we were* stimuli with matching resonances and vocoid duration (DARK and LONG).

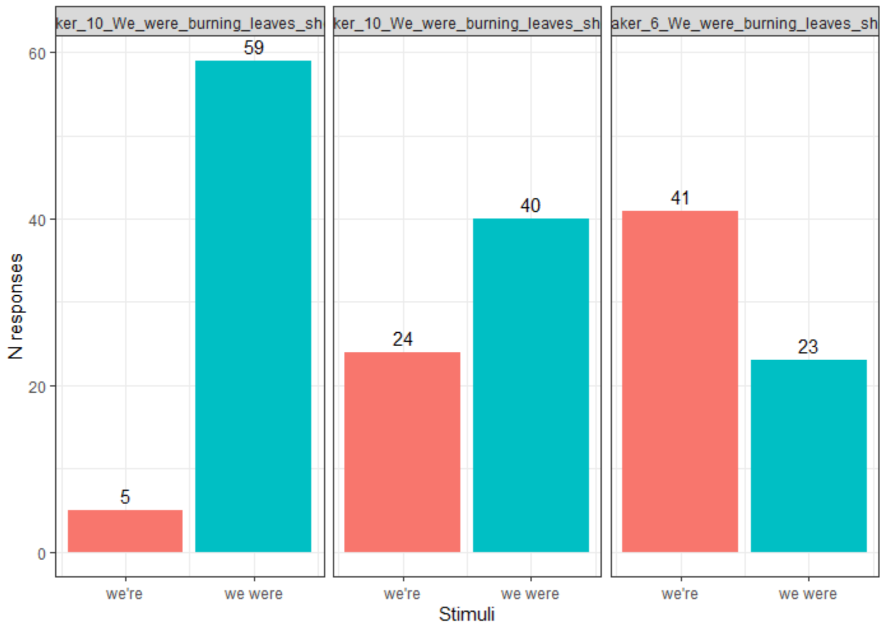


Figure 198. Responses counts for the manipulated *we were* stimuli. These are the responses to the non-matching DARK and SHORT stimuli manipulated from *we were* base tokens.

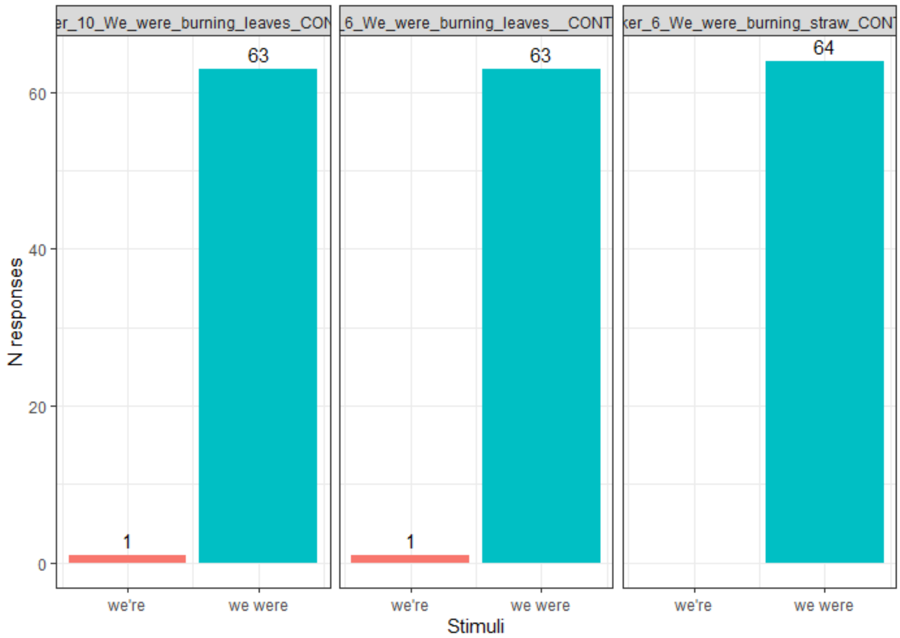


Figure 199. Responses counts for the control *we were* stimuli. These are the responses for the matching DARK and LONG stimuli.

Figure 198 shows that for two of the DARK and SHORT non-matching stimuli, the majority of participants (N = 40/64, 63%; N = 59/64, 92%) chose the option *we were*, but for one of the stimuli, *Speaker_6_We_were_burning_leaves_ShortVDur-58ms*, the majority of participants (N = 41/64, 64%) chose the option *we're* against *we were*. Similarly to the responses to the manipulated *we're* stimuli (Figure 195), the responses to the manipulated *we were* stimuli (Figure 198) suggest that the duration manipulation had some effect on the identification of the stimuli.

The responses to the manipulated *we're* and *we were* stimuli do not show a clear pattern. In four out of six stimuli (67%), participants chose the option *we were* – for two stimuli with DARK and SHORT manipulated stimuli and two CLEAR and LONG manipulated stimuli. Despite the absence of a clear pattern, an exact two-sided binomial test revealed that the responses to the *we're* and *we were* stimuli (N = 768) were significantly above chance ($p < .001$; 95% CI [0.25, 0.32]).

8.3.2.3. *You're* versus *you were*

This section reports the responses of the participants to the control and manipulated stimuli for *you're* and *you were* (N = 9).

You're is characterised by CLEAR resonances and SHORT vocoid duration, while *you were* is characterised by DARK resonances and LONG vocoid duration. Nine stimuli were included in the experiment: three *you're* stimuli (including one control stimulus), and six *you were* stimuli (three control and three manipulated stimuli).

Figure 200 shows the number of responses for the two *you're* manipulated stimuli. These are the CLEAR and LONG non-matching stimuli. Figure 201 shows the number of responses for the control *you're* stimulus. This is the original *you're* stimulus with matching resonances and vocoid duration (CLEAR and SHORT).

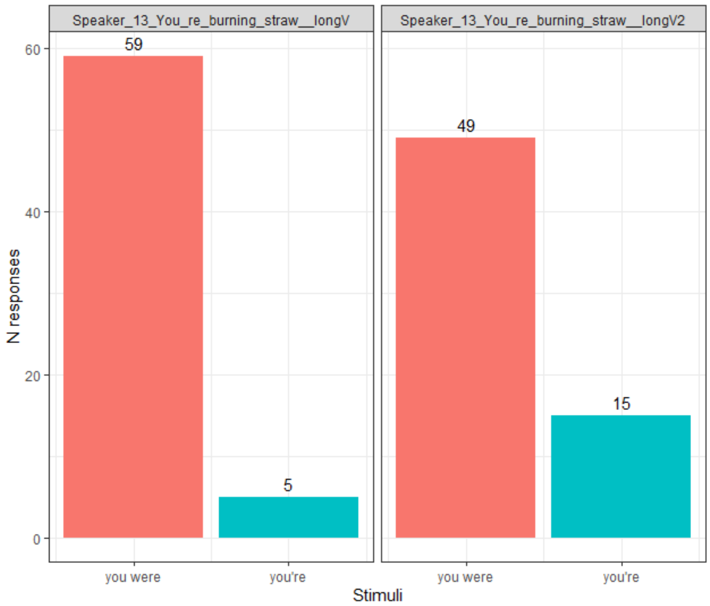


Figure 200. Responses counts for the manipulated *you're* stimuli. These are the responses for the non-matching CLEAR and LONG stimuli in which the vocoid duration was manipulated from an original *you're* token.

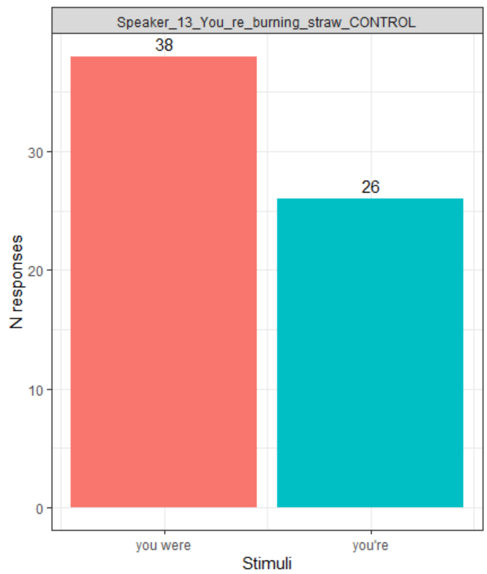


Figure 201. Responses count for the control *you're* stimuli. These are the responses to the original CLEAR and SHORT stimulus.

Figure 201 shows the participant responses to the control stimulus *you're*. Unfortunately, this stimulus was identified as *you were* by the majority of participants (N = 38/64, 59%), which means that it was not a suitable stimulus to be used in the experiment. However, the count of responses reported in Figure 200 shows that incresing the duration of the vocoid, increased the number of responses for *you were*. The vocoid duration of the base (control) stimulus was 51 ms. The vocoid duration of the stimulus on the right in Figure 200 was increased by 35 ms to 86 ms. The responses for *you were* increased from 59% (N = 38) to 77% (N = 49/64, + 11). The vocoid duration of the stimulus on the left was icreased by 54 ms to 105 ms. The responses for *you were* increased from 59% to 92% (N = 59/64, + 21). That is, the change in duration of the vocoid had an incremental effect on the participant responses.

Figure 202 shows the responses count to the three manipulated *you were* stimuli with DARK resonances and a non-matching SHORT vocoid duration.

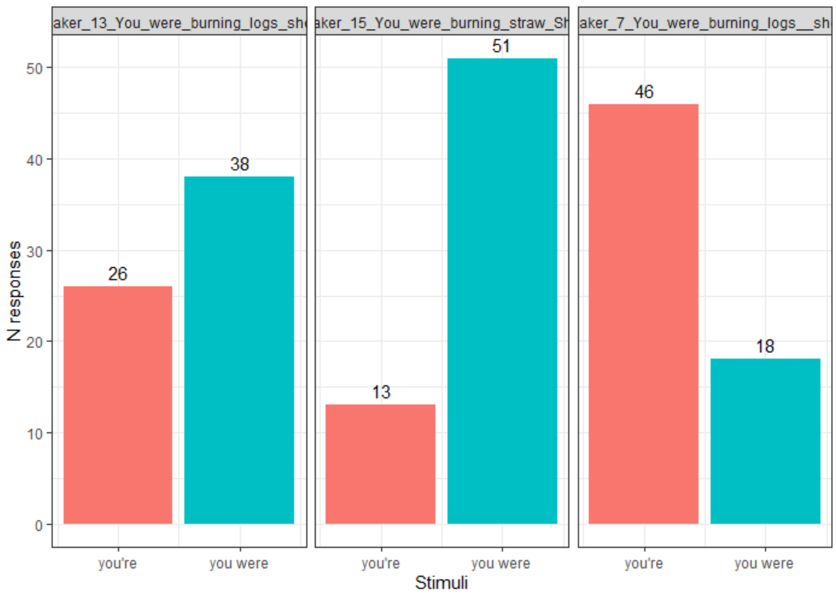


Figure 202. Responses counts for the manipulated *you were* stimuli. These are the responses to the non-matching DARK and SHORT stimuli manipulated from *you were* base tokens.

Figure 202 shows the responses to the three DARK and SHORT manipulated *you were* stimuli. Two stimuli were identified as *you were* by the majority of participants (N = 38/64, 59%, and N = 51/64, 80%), and one was identified as *you're* by the majority of participants (N = 46/64, 72%).

Figure 203 reports the counts of the responses for the control *you were* stimuli. These stimuli were not manipulated and are the original DARK and LONG stimuli.

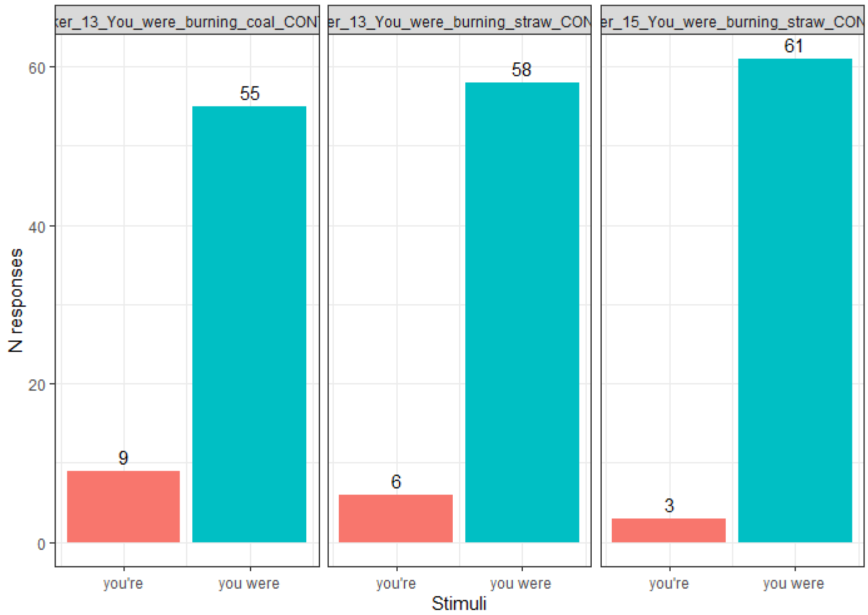


Figure 203. Responses counts for the control *you were* stimuli. These are the responses to the original matching DARK and LONG *you were* stimuli.

Figure 203 shows that the large majority of participants correctly identified the control *you were* stimuli. An exact two-sided binomial test revealed that the responses to the *we're* and *we were* stimuli (N = 576) were significantly above chance (p = .003; 95% CI [0.4, 0.48]).

To summarise the results of the responses to the stimuli that investigate the roles of duration and resonances in the identification of paradigms of the present and past tense of BE, the responses to the pairs *she's* and *she was* are unequivocal, while the responses to *we're* and *we were*, and *you're* and *you were*, are more ambiguous. In the identification of the paradigms *she's* and *she was* listeners seemed little affected by the changes in duration of the vocoids in *she's* and *she was*. This result is in line with the analysis of the responses to the contrast between *will* and *would*: participants chose the option that corresponded to the resonances of the stimuli rather than the duration.

However, the responses to the stimuli pr + *'re* and pr + *were* did not exhibit such unequivocal results and no clear pattern emerged. In combination with the pronoun *we*, the auxiliary *were* received more responses than the auxiliary *are*, both when the base token *we're* was manipulated resulting in a LONG vocoid duration, and when the base token *we were* was manipulated resulting in a SHORT vocoid duration. A similar result is observed for the manipulated *you were* and *you're* stimuli. Although the base token used for the manipulation of *you're* cannot be considered reliable as it was identified as *you were* by the majority of participants, it showed that increasing the vocoid duration increased the responses for *you were* incrementally.

8.4. Summary and discussion

This chapter investigated the perceptual salience of duration and, by extension, resonances in word identification. From the acoustic analysis reported in Chapter 5, it emerged that most pairs of contrasting paradigms are characterised by contrasting resonances and duration of the vocoid and/or the closure. Chapter 5 also highlighted the similarity in the trajectories of the formant dynamics within the pairs *we're* and *we were*, and *you're* and *you were*. The similarity between their formant dynamics, raised the question of whether the present and past tense *are* and *were* contrast only in duration.

The experiment reported in this chapter tested the perceptual salience of the duration (and by extension the resonances) in a range of pr + aux. Although only the duration was manipulated, the fact that the auditory and acoustic analyses reported in Chapters 3 to 5 highlighted the differences in the resonances of contrasting paradigms, led to the assumption that the resonances would play a major role in the identification of pr + aux. That is, when the duration does not affect the responses of the participants, it is assumed that participants rely on other phonetic features to identify the paradigms, of which the resonances are one of the most prominent.

To investigate the role of these two parameters in word identification, a forced-choice word identification task was created. For this task, the duration was manipulated to create non-matching stimuli with a combination of resonance and duration that does not occur in natural speech. For example, in the case of a pair such as *she's* and *she was*, the non-matching stimulus with a SHORT and DARK vocoid was created. Listeners had to indicate which one of two choices they heard: one choice had matching resonances with the stimulus played, the other choice had matching duration. In this way, the option chosen by the listeners should indicate which one of the two parameters is more perceptually prominent or salient in word identification.

The results of the experiment indicate that for the pairs of paradigms with the cliticised forms of *will* and *would*, the manipulation of the duration did not have any effect on the participant responses. Three pairs of paradigms were included in the experiment: *she'll* and *she'd*, *you'll* and *you'd* and *I'll* and *I'd*. To determine whether differences in duration of 20 ms and 40 ms would change the participant responses, a range of durations was included in the experiment. None of the manipulated stimuli was interpreted on the basis of the duration. Although the results of this subset of stimuli are unambiguous, one issue must be borne in mind: the differences in duration of periods of silence are less perceptually salient than the differences in duration of periods of sound (Bochner et al, 1988). However, in the pair *I'll* and *I'd*, the duration of the vocoid was manipulated too. The duration of the vocoids differs by 24 ms, which is slightly longer than the "just noticeable difference" of 15 ms found by Huggins (1972a, 1972b), and close to the change in duration of 25 ms that Klatt and Cooper (1975) claim listeners are sensitive to. Moreover, as claimed by Lehiste

(1970), the durational differences that are perceptually relevant are relative to the duration of a given sound and are in the range between 10-40 ms for sounds that are in the range of 30-300 ms, that is 13-33%. The stimuli *I'll* and *I'd* had a vocoid duration of 32 ms and 56 ms respectively, which means that a difference of 24 ms is between 43-75%. However, none of these durational differences affected the participant responses. Since only the parameter duration was controlled, it is difficult to establish exactly which properties of the stimuli the listeners relied on in the identification task. As reported in this thesis (see e.g. Section 4.1.3 on the acoustic differences between pr+ *'ve* and pr+ *'d*) in reduced speech the contrast between pr+aux can be conveyed by subtle phonetic features. However, the identification of the resonances as essential phonetic features in some of the pronouns and auxiliaries, including the velarity in pr+ *'ll*, might be sufficient evidence to indicate that, in the case of pr+ *'ll* and pr+ *'d*, the resonances have a prominent role in the identification of the pr+aux.

The responses to the stimuli testing the perceptual salience of the duration in the present and past tense of the auxiliary BE show two distinct patterns. In the case of *she's* and *she was*, the listeners' responses reflect the results obtained for *will* and *would*: listeners' responses were not affected by the durational differences between stimuli. However, the responses to the paradigms pr+ *'re* and pr+ *were* were not so categorical. The paradigms tested were *we're* and *we were*, and *you're* and *you were*. The acoustic analysis in Chapter 5 revealed that the trajectories and values of the formant dynamics in each present-past tense pair were fairly similar. For this reason, the resonances of the vocoid cannot be considered a contrasting feature. That is, the items in each pair differ only in vocoid duration, but not in their resonances. The question is whether the duration is a sufficient parameter to maintain the contrast between paradigms in reduced speech. The overall responses to the stimuli *we're* and *we were* do not provide a conclusive answer. When the non-matching stimuli were CLEAR and LONG, the majority of the participants (69%) chose the options that matched the duration of the stimuli. When the non-matching stimuli were DARK and SHORT, the majority of participants (64%) chose the option that matched the resonances of the stimuli. The analysis of the breakdown of the counts of responses for each stimulus confirms the ambivalent results: in 3 out of 6 stimuli (50%) the participants chose the option that matched the stimulus' resonances and in the other 3 out of 6 (50%) they chose the option that matched the stimulus' duration.

A very similar pattern emerged from the responses to the *you're* and *you were* non-matching stimuli. While the overall response count exhibits the same pattern, the breakdown of the counts for each stimulus reveals that for the two CLEAR and LONG non-matching stimuli, the participants chose the option that matched the duration, while for the two DARK and SHORT non-matching stimuli, the participants chose the option that matched the resonances. In other words, participants chose the option *you were* in all four cases. This is the case also in 4 out of 6 stimuli of the pair *we're* and *we were*: the majority of participants chose the option *we were* in 4 out of 6 stimuli. A possible explanation can be found in the perceptual adaptation to the style of speaking listeners are exposed to. Participants can become habituated to the degree of reduction of the stimuli they hear and expect a high degree of reduction throughout the experiment. In other words, if listeners become habituated to the degree of reduction of the stimuli, they might expect a stimulus to be more reduced than it actually is, and therefore, choose the most complex option. More research is needed to untangle the role of the duration in the identification of these paradigms. For now, it can be said that the analysis of the responses to the finite forms *are* and *were* indicate that in absence of clearly distinct resonances between the two paradigms

– as observed in the acoustic analysis – listeners do not necessarily rely on the duration to make their judgement about the identity of the pr + aux.

The results of this experiment are in line with findings by Kohler and Niebuhr (2011) on the relevance of resonances and duration in word identification. They tested the role of palatality and the duration of nasality in the recognition of the word *Ihnen* in German. They found that the degree of palatality did affect the responses, but the effect of the duration of nasality was less strong than the effect of the resonances. They report that "[w]hen palatality is strong […], nasal duration has very little influence on *Ihnen* judgements; when palatality is weak […] or absent […], duration can only weakly compensate for it" (Kohler and Niebuhr, 2011: 25). This seems to be the case also in the results of Experiment B: when there is a clear polarity between the resonances of contrasting pairs, the duration does not affect the word identification, and listeners base their decision on the resonances alone. When the resonances are not an essential element of the contrast, the influence of duration on listeners' responses is unclear.

The results reported in this chapter show that listeners are sensitive to the fine phonetic detail that maintains the contrast between paradigms conveying grammatical information. This phonetic detail can be considered the essential elements that constitute the identity of the piece and maintain the contrast in production and perception.

9. Discussion and conclusion

This chapter summarises the findings reported in the thesis, addresses the research questions introduced in Chapter 1, and concludes with some final remarks on the nature of reduction and the role of fine phonetic detail.

9.1. Introduction

The main aim of this research was to investigate the production and perception of reduced speech, and how highly reduced speech remains intelligible. Two aspects and their relation were particularly important: reduction and fine phonetic detail. In the description of the theoretical approach (Section 1.3) it was highlighted how a meaningful observation of reduced speech can only occur through a detailed qualitative analysis of the wide range of variation that characterises reduced speech, paying attention to the fine phonetic detail that remains in the acoustic signal and conveys linguistic meaning. Bearing this in mind, the thesis sought to address the following questions:

(1) Are function words characterised by fine phonetic detail that remains in the signal in reduced speech?

- What are the main phonetic features of function words that remain in the signal in highly reduced speech?

(2) Is the paradigmatic contrast between function words maintained by fine phonetic detail?

(3) Is the fine phonetic detail that maintains the contrast in reduced speech available to perception and sufficient for the correct identification of words in reduced speech?

In addition, this research aimed to document the range and types of variation in reduced speech, to shed light on the role of the context and the acoustic cues in spoken word recognition, and to advance our knowledge and understanding of speech reduction.

To this end, the present research analysed the combinations of pronouns and contracted auxiliaries in British English. The rationale behind the choice of pr + aux for an investigation of reduction is that pr + aux exhibit a wide range of variation and high degree of reduction. Chapter 1 (Section 1.1.2) described the factors that influence the degree of reduction of words and stretches of speech. Words exhibit variable degrees of reduction depending on the word class they belong to, their frequency and predictability given the context, the paradigmatic system they belong to, and the rhythmic and prosodic structure of their sentential context. Pronouns and auxiliaries perfectly reflect all these factors: they are function words, they occur frequently, they are predictable, they belong to a small paradigmatic system of contrasts, they undergo grammaticalisation, and they can occur in weak, unstressed prosodic positions.

Although in most traditional phonological accounts, the weak forms of function words that have a strong and one or more weak forms are treated as derivational through processes of deletion and modification, this thesis adopted a polysystemic approach. This means that the various forms of the English auxiliaries are treated as belonging to different systems and having different phonological characteristics and phonetic exponents. Chapter 5 provided evidence that a polysystemic approach is more appropriate for the analysis of English auxiliary weak forms. It showed that a phonological account of the polarity of labiality (rounding versus non-rounding) in *will* and *would* syllabic forms, cannot be applied to the non-syllabic forms of *will* and *would* which exhibit the exact opposite qualities as far as labiality is concerned.

An important aspect of pr + aux that is relevant to the work presented in this thesis is their belonging to a small paradigmatic system of contrast. The system of pr + aux combinations contains only a few contrasting items (relative to other systems such as that of content words). Moreover, this system exploits only a limited number of sounds and sound combinations. Because of these two features, the items in this system can be highly reduced. However, at the same time, each item in the system conveys grammatical information that must be retained in the signal and be available to perception for speech to be understood. For these reasons, pr + aux combinations are the ideal object for an investigation of the production and perception of reduced speech.

In order to carry out a detailed acoustic analysis of the pr + aux and to be able to compare the subtle phonetic differences between pairs of paradigms, it was crucial to collect high quality material. As explained in the methodology of the production experiment described in Chapter 2, the design of a carefully controlled experiment meant that it was possible to record all the paradigms in the system of pr + aux uttered in the same phonological and prosodic environment. It also made it possible to trigger a high degree of reduction by placing the target items in the prosodic position of anacrusis, and by repeating each sentence multiple times introducing new information in the phrasal stress position. As a result, the data collected exhibit a wide range of variation including high degrees of reduction, and made it possible to address the research questions stated at the beginning of the research. The next sections address the research questions with reference to the results of the analyses reported in Chapters 3 to 8.

9.2. Research questions

This thesis set out to address three main research questions. The next sections attempt to answer them with reference to the findings reported in this thesis.

9.2.1. Phonetic features of function words

The first research question was:

(1) Are function words characterised by fine phonetic detail that remains in the signal in reduced speech?

- What are the main phonetic features of function words that remain in the signal in highly reduced speech?

In order to address the first question, a thorough auditory and acoustic analysis of the pr + aux collected in the production study was carried out. To our knowledge, this is the first systematic and detailed phonetic analysis of pronouns and contracted auxiliaries in British English. The analysis reported in Chapters 3 and 4 highlighted the wide range of variation found in the realisation of pr + aux. It showed some of the most common realisations as well as the unusual features observed in the data. It highlighted the various degrees of reduction of the pr + aux by showing a range of variants from the most unreduced to the most reduced ones. The production of all the speakers were used to illustrate the data. Despite the great deal of variation and reduction observed, the analysis led to the identification of phonetic features that characterise each pronoun and each auxiliary and, in turn, each pr + aux. 'Phonetic features' here refers to the fine phonetic detail that remains in the signal in reduced speech and that constitutes the identity of each piece. As explained in Section 1.3.3, following Local (2003), and Hawkins (2003, 2010) among others, the term fine phonetic detail is used to indicate distributed and systematic phonetic features that convey linguistic meaning. Since the main function of pr + aux is to convey grammatical information, the phonetic features identified express the grammatical meaning of each pr + aux. Table 43 summarises the phonetic features that characterise the pronouns. Table 44 summarises the phonetic features that characterise the cliticised forms of the auxiliaries.

Pronouns	Phonetic features
I	Open front articulation
She	Palato-alveolar friction
He	Close front articulation Glottal friction (which can be very weak)
It	Glottality (but not always) If glottality is not present, alveolarity
You	Palatality and labiality
We	Velarity and labiality
They	Dentality and a mid front articulation

Table 43. Summary of the phonetic features that characterise the pronouns.

Clitics	Phonetic features
'd	Closure and/or alveolarity
's	Alveolar friction
've	Labiodentality and/or voicing
'm	Nasal contoid
'll	Velarity
're	Openness
was	Labiovelarity and friction
were	Labiovelarity

Table 44. Summary of the phonetic features that characterise the cliticised forms of the auxiliaries.

Table 43 and Table 44 summarise the phonetic features that characterise the pronouns and the cliticised forms of the auxiliaries. When the pr + aux are combined, both features are present. For example, *they're* is characterised by both dentality and openness. When the pr + aux are reduced, the phonetic features can be reduced too, but they are always there.

Most of the phonetic features that were identified during the auditory and acoustic analysis are distributed long-domain features that are not tied to segmental units, and can be described as resonances of the piece. For example, the glottality of *it* is usually distributed throughout the piece in form of creaky voice. The velarity of *'ll* can be heard throughout the piece, regardless of the primary articulation of the lateral.

As mentioned in Section 1.3.2, Niebuhr and Kohler (2011) and Kohler and Niebuhr (2011) proposed the concept of essential components of function words that constitute their "phonetic essence" (Niebuhr and Kohler, 2011: 320). They state that the essential features of function words can be present in form of segmental units in unreduced word forms, or articulatory prosodies in reduced word forms. They claim that even in extreme cases of reduction, function words retain the articulatory prosodies that constitute their identity. According to Kohler and Niebuhr, this is the reason reduced speech remains intelligible: the articulatory prosodies are available in the signal when little else is, and listeners attend to them for the correct interpretation of speech. The concept of the phonetic essence of words has recently been evaluated also by Ernestus and Smith (2018) in their analysis of the word *eigenlijk* in Dutch spontaneous speech. They found that *eigenlijk* is characterised by articulatory prosodies that are present in the signal in highly reduced instances of the word, although they state that "even the essential parts allow for variation" (Ernestus and Smith, 2018: 30).

The analysis of pr + aux reported in this thesis provides some support for the theory of the essential components that constitute the phonetic essence of function words. In particular, features such as the glottality in *it*, the palatality in *you*, and the labiovelarity in *we* are long-domain resonances that characterise the whole piece, rather than time-delimited local features. Importantly, these resonances are always retained in reduced speech. However, it was pointed out in Section 4.3 that due to the great deal of variation of some pr + aux it was problematic to define what is ‘essential’.

This is the case, for example, of the cliticised form *’d*. It might seem straightforward to suggest that *’d* is characterised by closure and alveolarity. However, it was observed that just one of these two features is needed, which means that neither of them is essential. An even more complex case is that of the cliticised form *’ve*. In the great majority of instances, pr + *’ve* is characterised by either friction or labiodentality or both. However, in a few instances of pr + *’ve*, neither of these two features is present. The acoustic analysis and comparison with comparable instances of pr + *’d* revealed that the two paradigms differ along a number of subtle acoustic cues, including the duration of the closure, the duration of the vocoid, the frequencies of F2 in the final portion of the vocoid, and the abrupt versus gradual end of the vocoid. For this reason, it was suggested that (at least some) pr + aux are better described in terms of a set of phonetic features that distinguish them from the paradigms they contrast with. In other words, the ‘essential’ phonetic features are those that maintain the contrast between pairs of paradigms that convey grammatical meaning. That is, the phonetic essence of function words is in relation to their system of contrasts and the essential features are those that maintain the contrast between paradigms.

To answer the first research question: **function words are characterised by fine phonetic detail that remains in the signal in reduced speech**. The phonetic features of pr + aux can be described as long-domain resonances that characterise the whole piece. These features retain the identity of the pr + aux in reduced speech. To establish whether they are ‘essential’ more exploratory work is needed, possibly using several types of speech material, such as conversational speech.

9.2.2. Contrast in reduced speech

The second research question was:

> (2) Is the paradigmatic contrast between function words maintained by fine phonetic detail?

One of the aims of this research was to investigate how the contrast between function words such as pr + aux combinations is maintained in reduced speech. Pr + aux belong to a small system of paradigmatic contrast and for this reason they can be highly reduced. At the same time, pairs of paradigms that share similar features (e.g. the pronoun but not the auxiliary) convey crucial grammatical information that must be retained in the signal for the message to be unambiguous. For example, if in reduced instances of *we'd* and *we'll* the contoid in coda position is not realised as a segmental unit, what are the phonetic features that distinguish the two paradigms?

Following the observations of the qualitative analysis reported in Chapters 3 and 4, the quantitative analysis reported in Chapter 5 compared the acoustic features of the pairs of paradigms that convey grammatical information.

The main focus of Chapter 5 was the analysis of the contrast between the present and past tense of the auxiliaries, such as *she'll* versus *she'd*, and *she's* versus *she was*. The pairs of paradigms selected were analysed along several acoustic parameters, including duration, spectral moments, amplitude, and formant dynamics of every phonetic event in each piece. The results revealed that the main acoustic parameters that distinguish the present and past tense in each pair of paradigms are the duration and the resonances. Table 45 summarises the present versus past tense contrasts analysed in Chapter 5, the pairs selected for the analysis, and the acoustic features that differ between the items in each pair.

Present ~ past contrast	Pairs of paradigms analysed	Durational differences		Resonance/spectral differences
will ~ would	*we'd ~ we'll*		closure	clear ~ dark
	I'd ~ I'll	vocoid	closure	clear ~ dark
	she'd ~ she'll		closure	clear ~ dark
is ~ was	*she's ~ she was*	vocoid		clear ~ dark
are ~ were	*you're ~ you were*	vocoid		
	we're ~ we were	vocoid		
has ~ had	*he's ~ he'd*		closure	spectral qualities of coda friction

Table 45. Summary of the present versus past tense contrasts analysed, the pairs of paradigms compared for each contrast, and the acoustic features that differ between the items in each pair, divided into two columns: one for the differences in duration, and one for the differences in the resonances or spectral qualities.

All the pairs analysed differ in the duration either of the vocoid or the closure or both (in the case of *I'll* versus *I'd*). By extension, in most cases, the duration of the piece is a distinguishing factor too, as it is positively correlated to the duration of the vocoid. The statistical analysis confirmed that the duration was a significant factor in all the pairs analysed.

The other acoustic parameter that distinguishes the items in each pair that contrast in the tense of the auxiliary verb is the resonances. The resonances of each piece were identified through auditory inspection and the acoustic analysis of the formant dynamics. The formants, and in particular F2, are the acoustic correlates of the resonances. In this thesis, the difference between the resonances was broadly defined as a polarity between clear and dark resonances. However, the clear or dark quality of a piece can be the output of different articulatory gestures. For example, while the dark resonance that characterises pr + *'ll* is due to the secondary articulation of /l/, the dark resonance of pr + *was* is due to the labiovelarity of the initial /w/. This means that the dark resonance of pr + *'ll* does not have the same quality of the dark resonance of pr + *was*. Similarly, while the clear resonance of *I'd* (relative to *I'll*) is due to the openness of the vocoid, the clear resonance of *you'd* (relative to *you'll* or *we'd*) is due to the palatality of the vocoid. Nevertheless, regardless of the actual articulatory gesture that produces the clear and dark resonances, the items in each pair of paradigms exhibit opposite resonance qualities.

Surprisingly, the analysis of the pairs *you're* versus *you were*, and *we're* versus *we were* revealed that they contrast only in duration, but not in the resonances of the piece. This is surprising because the past tense *were* is characterised by labiovelarity, which is not a feature of the present tense *are*, which is characterised by openness. The labial-velar gesture for /w/ at the beginning of *were* should result in a low F2 and therefore in dark resonances. However, in the pairs compared, the opposition between clear and dark resonances in pr + *'re* and pr + *were* was not observed. However, the items in the pairs *you're* and *you were*, and *we're* and *we were* differ greatly in their duration. The difference in duration is 'almost categorical' except for a few outliers. In the paradigms that exhibit a strong polarity of their resonances, such as *I'd* and *I'll*, the durations of the two paradigms overlap greatly (see e.g. Figure 130). As suggested in Chapter 5, a possible explanation for the lack of polarity in the resonances of the two items in each pair (pr + *'re* versus pr + *were*) is that the pronouns must retain their resonances to maintain their identity. The identity of the pronouns *you* and *we* is solely in their resonances: *you* is characterised by palatality, and *we* by labiovelarity. The large difference in duration might be sufficient for the correct identification of the paradigms that contrast in the present versus past tense of BE.

In addition to the contrast between present and past tense, the contrast between the pronouns *he* and *you* was investigated too. The rationale for the comparison between *he* and *you* came from the observation of voiceless palatal friction at the beginning of several instances of *you*. This raised the question of whether the presence of the initial friction would make *you* sound similar to *he*.

To answer the second research question: **the comparison between the acoustic features of contrasting pairs of paradigms showed that the contrast is maintained in reduced speech by two parameters: duration and resonances**. This raised the question of whether the acoustic differences in duration and resonances are available to perception and whether listeners use them for the correct identification of pr + aux in reduced speech.

9.2.3. Speech intelligibility and perceptual salience in reduced speech

The third research question was:

(3) Is the fine phonetic detail that maintains the contrast in reduced speech available to perception and sufficient for the correct identification of words in reduced speech?

Chapter 7 and 8 reported on a perception experiment investigating the intelligibility of highly reduced word forms, and the role of fine phonetic detail in the correct interpretation of reduced function words.

Experiment A aimed to answer the first part of the question – whether the fine phonetic detail identified in Chapters 3 and 4, and quantitatively analysed in Chapter 5, is available to perception. The results of Experiment A showed unambiguously that reduced paradigms remain intelligible even when highly reduced, and that listeners do not have any difficulties in correctly identifying reduced pr + aux solely on the basis of their acoustic features. Crucially, in the perception experiments reported in this thesis, the semantic and syntactic contexts did not provide any cues for the correct identification of the pr + aux – both options presented to the participants were grammatical, and no preceding contextual factors were given. In other words, either paradigm was equally possible in the sentence the participants heard, which means that their decision was based only on the acoustic cues they heard from the signal.

After identifying the two parameters that distinguish paradigms acoustically, Experiment B aimed to investigate the perceptual salience of these two parameters and to test whether one of them was more salient for the correct identification of pr + aux. The results of Experiment B revealed that listeners rely on the resonances for the correct identification of pr + aux. From the results of the experiment, the role of the duration is not clear. When the resonances of pairs of paradigms show a clear polarity, the duration does not seem to have a role in word identification. When the resonances are not an important element of the pair of paradigms, such as in pr + *'re* versus pr + *was/were*, the experiment did not show a clear role of duration in word identification.

Besides shedding light on the role of fine phonetic detail in word recognition, these experiments add new evidence to the long-debated issue of the role of the context and the role of the acoustic cues in speech understanding. Much of the literature on word identification in reduced speech has focussed on the role of the context and the acoustic features present in the signal. On the one hand, Ernestus et al. (2002), claimed that the intelligibility of reduced speech depends on two aspects that are negatively correlated: the degree of reduction and the context given. That is, higher degrees of reduction require more contextual information to be understood. On the other hand, the experiments by Kohler and Niebuhr (2011) show that listeners can correctly identify the presence or absence of a function word based solely on the acoustic cues. The results of Experiment B are in line with the findings by Kohler and Niebuhr. The two options given to the listeners were both grammatical and appropriate for the context given. Therefore, the listeners could rely only on the acoustic features of the stimuli to make their decision.

However, the main drawback of the present experiments is that two options were given to the participants and they were visually presented on a screen. It was observed in a pilot experiment that if participants were asked to write down what they heard at the beginning of the sentence in the stimuli, the identification rate was much lower, and in several cases participants could not hear an auxiliary at all. This means that the role of the acoustic cues and the context can be evaluated only in this restricted setting and cannot be extended to word recognition in, for example, spontaneous speech. Nevertheless, the experiments presented here provide evidence that listeners are sensitive to fine phonetic detail and that they use this detail for the correct identification of reduced function words in a limited context.

To answer the third research question: **the fine phonetic detail that maintains the contrast between paradigms is available to perception and sufficient for the correct identification of function words**. The perception experiment reported in this thesis shows that listeners rely on the resonances of the piece to make a decision on which paradigm they have heard.

9.3. Reduction

In addition to answering the research questions introduced in Chapter 1, this thesis has contributed to our knowledge of speech reduction. This section summarises the main findings of this thesis.

9.3.1. Temporal and gestural reduction

Reduction is a broad term used to refer to patterns of variation of speech sounds that involve modifications such as decrease in duration and/or articulatory effort. The auditory and acoustic analysis carried out in this research confirms that reduced pr + aux exhibit either a decrease in time, or a decrease in the magnitude of gestures, or both.

The evidence that reduction is characterised by a decrease in duration is provided by two sources in the thesis. Firstly, the analysis of reduction in repeated items (Section 6.3) showed that the first item has a longer duration than the subsequent repetitions. Although the analysis of repetitions did not show a clear correlation between duration and repetition, the point here is that the first time the target item was uttered, it exhibits a longer duration than the subsequent ones. That is, the first time a target item is uttered, it is more fully articulated, while in the subsequent repetitions the duration is shorter than in the first mention.

Secondly, the qualitative analysis reported in Chapter 3 showed that in the great majority of instances, what is perceived as reduction, is given by the temporal reorganisation of the phonetic events in the vocal tract. The presence of voiceless friction at the beginning of the utterances with a vocoid at the beginning of the utterance, led to the observation of the formant structure during the friction. The formant structure clearly indicates that the gesture for the first sound in the utterance is often articulated before the onset of voicing. The phasing of the articulatory gesture in the supralaryngeal vocal tract and the vibration of the vocal folds in the larynx results in a short vocoid being audible. In other words, the gesture of the sound is articulated, and voicing is produced, but they occur simultaneously only for a short portion of time. Figure 204 illustrates the temporal reorganisation of the phonetic events that occur in the supralaryngeal vocal tract (light blue) and in the larynx (dark blue), and the reduced pr + aux *we're* in the acoustic output (red).

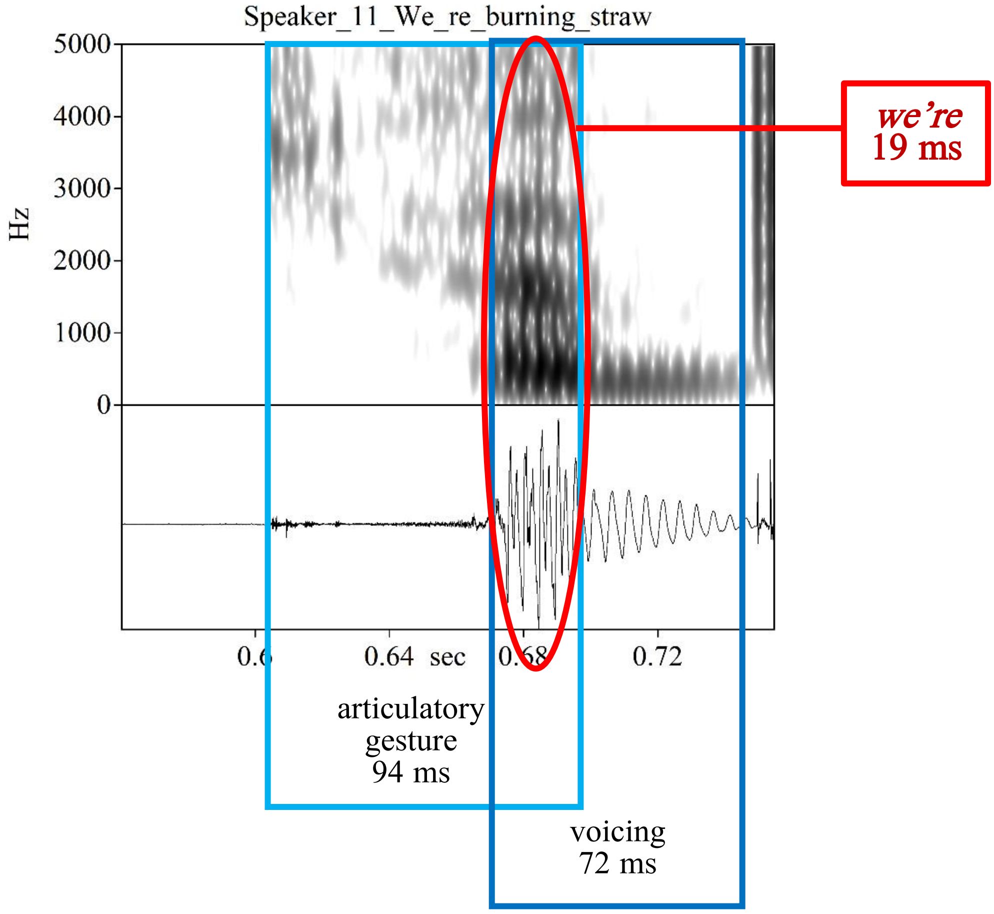


Figure 204. Temporal reorganisation of phonetic events occurring in the supralaryngeal vocal tract (light blue) and the larynx (dark blue), and the acoustic output (red).

The fact that the voiceless friction in onset is often inaudible, means that the acoustic output is a short vocoid. If the cause of much of the acoustic reduction observed in the data is the temporal realignment of the phonetic events, it can be claimed that reduction is indeed characterised by a decrease in duration.

The analyses reported in this thesis confirm that reduced items exhibit a decrease in the magnitude of gestures. For several pr + aux, a subset of items with shorter durations were selected and analysed. Their acoustic features were compared to the acoustic features of the whole dataset. For example, the analysis of the formant dynamics of several subsets of pr + aux revealed that in the reduced instances the formants, especially F2, are flatter and exhibit less movement than the formant dynamics measured in the whole datasets. The formant dynamics are the acoustic correlates of the shape, size and movement of the oral cavity and the articulators and, as such, provide information on the articulation of speech sounds. Flatter formants indicate that there is less articulatory movement in the production of the sound. This feature confirms that reduction reflects a decrease in the magnitude of gestures. Moreover, the reduction in the magnitude of gesture was observed also in the contoids. The alveolar plosives /t/ and /d/ are particularly prone to reduction. Several instances are realised with friction instead of a complete closure. The lateral /l/ is realised without the primary articulation in the majority of cases. That is, the tip of the tongue does not touch the alveolar ridge. As for the fricatives, such as /v/ in pr + *'ve*, and /z/ in pr + *'s*, they are realised with weak aperiodic energy to the extent that sometimes it is difficult to identify and hear the portion of friction, especially /v/. All these aspects indicate that reduced speech is characterised by a decrease in the magnitude of gestures.

9.3.2. Reduction is not loss of information

Crucially, the analysis reported in Chapter 5 suggests that the decrease in the magnitude of gestures does not lead to the loss of acoustic information, but can actually lead to their enhancement. The comparison between the acoustic features of pairs of contrasting paradigms such as *we'd* and *we'll*, highlighted the difference in the formant dynamics of the vocoids in the two paradigms. From the comparison between the two subsets of reduced (shorter) instances of the paradigms it emerged that the distance between F2 of *we'd* and F2 of *we'll* is larger in the reduced subsets than in the whole datasets. Since one of the two features that discriminate between the two paradigms is their resonances, the larger difference in F2 in the reduced subset suggests that the distinction between the two paradigms is enhanced rather than weakened in reduced speech. Similarly, the analysis of the spectral properties of the palato-alveolar friction in *she's* and *she was* showed that the difference in the mean CoG calculated across the whole datasets was 87 Hz, while the difference in the mean CoG between the two reduced subsets of *she's* and *she was* was 392 Hz. That is, the acoustic feature that discriminates between the two paradigms exhibits a larger difference in reduced instances of the two paradigms than in the whole datasets. This suggests that the contrast between the two paradigms might be enhanced rather than weakened in reduced speech.

These findings support the theory that reduced speech does not imply loss of information. The concept that reduction leads to informational loss is linked to the segmental view of speech, according to which if a segment is deleted in the reduced word form compared to the canonical word form found in dictionaries, then some information must be missing (Cangemi and Niebuhr, 2018). This is not the case in our data. Firstly, it has been demonstrated that the identity of the pr + aux is maintained by fine phonetic detail, such as the resonances, throughout the dataset. Secondly, in the most reduced variants of pr + aux, the resonances seem to be enhanced rather than weakened. This feature provides more evidence that fine phonetic detail conveys essential information for speech recognition.

9.3.3. Variability in reduced speech

As reported in Chapter 6 (Section 6.2), temporal reduction and articulatory reduction do not necessarily occur together in the same piece, and a piece that exhibits temporal reduction might not exhibit articulatory reduction, and vice-versa (horizontal variability). It was also reported that a piece can be produced with different degrees of magnitude of gesture. That is, a sound or phonetic feature in a piece can be hypo-articulated while another sound or feature in the same piece can be hyper-articulated (vertical variability). The lack of correlation between temporal and articulatory reduction has been observed in studies on speaking style and rate (e.g. van Son and Pols, 1992, 1993; Warner and Tucker, 2011). However, to our knowledge, the co-occurrence of hypo- and hyper-articulation in the same piece (vertical variability) has not been described in the literature on reduction before. A possible reason for this gap is that most research on reduction focusses on single segments, rather than stretches of sounds, and/or applies quantitative methodologies, missing to notice variations such as these which were observed during the qualitative analysis. An exception is Ernestus and Smith (2018: 34) who carried out a qualitative analysis of the Dutch word *eigenlijk* and found that a word form "may be reduced in one aspect, but not in another".

The variability observed suggests that reduction is not an automatic process that consistently occurs to all the elements of a piece in the same way. The presence of hypo- and hyper-articulation in the same phonological and processing unit suggests that the degree of articulation might be under the control of the speaker. In light of the features observed, it seems that reduction cannot be viewed simply as an automatic process due to a decrease in articulatory effort. The variability observed in the data suggests that reduction is a controlled phenomenon on which the speaker has a choice of what to reduce and to what extent. This is in line with findings by Ernestus and Smith (2018) on the reduction of the Dutch word *eigenlijk*.

A possible explanation for the phenomenon of vertical variability is that the various phonetic features and events in a piece have different roles in perception. Some features might be more salient than others and need to be retained in the acoustic signal. However, the data do not provide evidence to support this hypothesis. For example, the fact that /d/ is unreleased in 87% of pr + *'d* suggests that the stop release itself is not a perceptually salient feature for spoken word recognition. Therefore, the few instances in which the piece pr + *'d* is reduced, but the plosive is released with a burst (as shown in Figure 162) cannot be explained with reference to speech intelligibility. More research into the subtle variations observed in reduced speech is needed to shed light on this phenomenon. Despite leaving unanswered questions, the phenomena observed confirm that speech is characterised by a great deal of variability and that some of it has still not been described. However, these observations have advanced our knowledge of reduction and variation in speech.

9.3.4. Reduction in repetitions

The analysis of the relationship between reduction and repetition did not reveal a clear correlation between them. In the elicitation task, each pr + aux was repeated five times in sequence. The mean duration of the vocoid and the mean duration of the piece in several pr + aux were analysed grouped by repetition. The results indicate that the first utterance exhibits the longest piece duration in the majority of cases. However, the subsequent repetitions do not exhibit an incremental decrease in duration for each repetition of the same item. In the literature, several hypotheses were formulated about the correlation between reduction and duration. The data analysed in this thesis do not support the hypothesis that the duration is positively correlated to the number of repetitions put forward by Aylett and Turk (2004). The data partially support the theory that new information is more fully articulated and that old information is shorter (Fowler and Housum, 1987). They support this claim only 'partially' because the lack of a communicative context in the data analysed means that there is no recipient for the 'information' uttered, whether it is new or old information. However, the fact that the first mention of the pr + aux is the longest, does suggest that new information has a longer duration. The new information might be new only for the speaker, which supports Bard et al.'s (1995) theory that the informativeness of the utterance is not linked to the listeners, but to the speaker. A more plausible explanation for the lack of a pattern observed in the data is that the repetition of the same articulatory gesture leads to faster movements as a consequence of repeated physical behaviour. That is, repeated motor patterns become quicker with practice and can lead to temporal or articulatory reduction.

9.3.5. Silent articulations

Chapter 6 reported that throughout the dataset, a portion of weak voiceless friction was observed at the beginning of the utterances. This feature was observed in all speakers at the beginning of sentences that started with a vocoid. Although in most instances analysed the friction is too weak to be heard, in a few instances the friction is audible. The quality of the friction reflects the articulatory gestures of the first sound in the utterance. The friction is also characterised by (usually clear) formant structure corresponding to the first sound of the utterance. Following Schaeffler et al. (2014), this phenomenon has been referred to as 'silent articulations' in this thesis, even though it is not always 'silent'. To our knowledge, this phenomenon was not observed before. There are several reasons for this gap. Firstly, the research on speech beginnings (e.g. Mennen et al., 2010; Scobbie et al., 2011; Schaeffler et al., 2014; Rasskazova et al., 2019) does not look at reduced speech. Secondly, it focusses on articulatory movements occurring well before any speech sound is articulated (＜250 ms). Thirdly, some experiments analysed the articulatory movements occurring before the beginning of words that had a contoid in onset (e.g. Scobbie et al., 2011) while the phenomenon described here was noticed when the utterance had a vocoid at the beginning.

The presence of the initial friction with visible formant structure highlighted the role of the temporal realignment of the phonetic events occurring in the vocal tract in leading to acoustic reduction. That is, the visible formants during the initial voiceless friction suggest that the gesture for the first sound is articulated before the onset of voicing. This results in a very short vocoid in the acoustic output. However, what is a case of reduction in the acoustic output, is not necessarily a case of articulatory reduction. The visible formants indicate that the gesture of the first sound is articulated, but the temporal realignment of voicing makes it inaudible.

This phenomenon can have implications for speech understanding. Two hypotheses can be formulated. On the one hand, the inaudible friction could impede word recognition in those instances in which the first sound is inaudible. On the other hand, when the friction is audible, it could aid word recognition by prolonging the availability of information in the acoustic signal. This feature could also have implications for talk-in-interaction. For example, in face-to-face conversations the visible movements of the silent articulations could affect turn-taking. Further research is needed to determine whether the silent articulations are indeed inaudible or if they are audible, and in the latter case, what their role in speech intelligibility and communication is.

9.4. Final remarks

This thesis has provided new knowledge and understanding of two fundamental aspects of speech and their relation: reduction and fine phonetic detail. It has shown that reduction is systematic and meaningful, and not an automatic phenomenon that affects every sound in the same way. Reduced speech is characterised by fine phonetic detail that is informationally rich and relates to linguistic function and structure. In the system of pr + aux, fine phonetic detail conveys crucial linguistic information. It reflects, and provides information about, the linguistic structure – e.g. the paradigmatic and syntagmatic systems – and function – e.g. grammatical meaning.

The fine phonetic detail of function words is distributed rather than segmental. It can be thought of as an essential element that is always articulated and constitutes the identity of function words. For this reason, it is always present in production and it is available to perception.

The importance of fine phonetic detail is demonstrated by its role in spoken word recognition. Fine phonetic detail is sufficient for the correct identification of reduced function words in a limited context. This provides evidence that reduced speech does not imply loss of information. In fact, the linguistic contrast conveyed by fine phonetic detail in the system of pr + aux is even enhanced in reduced speech.

This thesis has also shown that a polysystemic approach is more appropriate in the analysis of speech. In the system of function words, interconnected sub-systems – e.g. syllabic and non-syllabic forms of auxiliaries – are characterised by different phonologies and phonetic forms. Bearing this in mind, future research should look into the systematic variation of different systems of function words in different communicative contexts.

The importance of carrying out a qualitative analysis, and the benefits of adopting a non-segmental approach to investigating reduced speech have been demonstrated throughout this research. Moreover, this thesis has highlighted the importance of focussing on what remains in the acoustic signal in reduced speech, rather than on what is missing compared to a reference form that rarely occurs in everyday speech.

The findings reported in this thesis have implications for models of speech production and perception. Such models should integrate fine phonetic detail, information about its variability, and its relation to linguistic structure. These findings can also have an impact on related fields, in particular psycholinguistics, sociolinguistics, forensic speech science, speech synthesis, and foreign language teaching.

To conclude, the systematic variation in the phonetic detail characterising function words in reduced speech suggests that fine phonetic detail is more meaningful than it is accounted for. Many aspects of the wide range of variation observed in speech are still unexplained. This thesis has contributed to our knowledge and understanding of this variation by providing new insights into the role of fine phonetic detail in reduced speech.

Appendices

Appendix A – Accompanying ethics documentation

A.1 Information sheet, Production experiment

THE UNIVERSITY *of York*

DEPARTMENT OF LANGUAGE AND LINGUISTIC SCIENCE
Heslington, York, YO10 5DD, UK

INFORMATION SHEET

PLEASE KEEP THIS INFORMATION SHEET AND A SIGNED COPY OF THE CONSENT FORM FOR YOUR RECORDS

You are invited to take part in a research study. Before you decide whether to participate it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information carefully. If there is anything you do not understand, or if you want more information, please ask the researcher.

Title of study:

Phonetic features of speech sounds

Researcher:

Emanuela Buizza

What is the research about?

This research investigates the fine phonetic details of speech sounds. To avoid affecting the results of the study, I am afraid I cannot say more than this.

Who is carrying out the research?

Emanuela Buizza, PhD student in Linguistics in the Department of Language and Linguistic Science, University of York.

Who can participate?

Any native speaker of English with a Standard Southern British accent can participate.

What does the study involve?

You will be asked to watch a PowerPoint presentation and to read out all the sentences written on it. On each slide of the presentation there will be a sentence accompanied by pictures. You will be asked to read out the sentences in a natural way, but with a specific stress pattern. The sentences are short and simple. Your voice will be recorded during the entire duration of the PowerPoint presentation. Both the presentation and the recording can be stopped at any time.

The recording session will take place in the Psycholinguistic Laboratory in the Department of Language and Linguistic Science. The session will last approximately 45 minutes.

Do I have to take part?

You do not have to take part in the study. If you do decide to take part you will be given this information sheet to keep and will be asked to sign two copies of the consent form (one copy is for you to keep). If you decide to take part you will still be free to withdraw without giving a reason, even during the session itself. If you withdraw from the study, we will destroy your data and will not use it in any way.

What are the possible risks of taking part?

Every step has been taken to ensure that the risks of participating in this study are as close to zero as possible. The slides you will have to look at do not contain any sensitive material. In any case, you are free to withdraw from it at any time should you feel uncomfortable in any way.

Are there any benefits to participating?

You will contribute to the ongoing research on speech production and perception and to our knowledge of how human communication works.

What will happen to the data I provide?

The data you provide will be used alongside the data of other participants to analyse subtle phonetic details used in speech. Your data will be stored securely in the Department of Language and Linguistic Science, University of York.

What about confidentiality?

Your identity will be kept strictly confidential. No real names will be used in any presentations or publications.

Will I know the results?

I am happy to send you the results once the study is completed.

*This study has been reviewed and approved by the Departmental Ethics Committee of the Department of Language and Linguistic Science at the University of York. If you have any questions regarding this, you can contact the chair of the L&LS Ethics Committee, Márton Sóskuthy (**email:** marton.soskuthy@york.ac.uk; **Tel:** (01904) 324171).*

If you have further questions regarding this study, please feel free to contact:

Researcher:
Emanuela Buizza
Department of Language and Linguistic Science
University of York, Heslington, York, YO10 5DD
email: emanuela.buizza@york.ac.uk

Supervisor:
Dr Richard Ogden
Department of Language and Linguistic Science
University of York, Heslington, York, YO10 5DD
email: richard.ogden@york.ac.uk

A.2 Consent form, Production experiment

Phonetic features of speech sounds

Lead researcher: Emanuela Buizza

Consent form

This form is for you to state whether or not you agree to take part in the study. Please read and answer every question. If there is anything you do not understand, or if you want more information, please ask the researcher.

Have you read and understood the information leaflet about the study?	Yes ❒ No ❒
Have you had an opportunity to ask questions about the study and have these been answered satisfactorily?	Yes ❒ No ❒
Do you understand that the information you provide will be held in confidence by the research team, and your name or identifying information about you will not be mentioned in any publication?	Yes ❒ No ❒
Do you understand that you may withdraw from the study at any time before the end of the data collection session without giving any reason, and that in such a case all your data will be destroyed?	Yes ❒ No ❒
Do you understand that the information you provide may be kept after the duration of the current project, to be used in future research on language?	Yes ❒ No ❒
Do you agree to take part in the study?	Yes ❒ No ❒
If yes, do you agree to your interview being recorded on audio?	Yes ❒ No ❒
Do you agree to excerpts from your audio recordings to be used in presentations or in teaching by the researcher, without disclosing your real name? *(You may take part in the study without agreeing to this).*	Yes ❒ No ❒
Do you agree to the researcher's keeping your contact details after the end of the current project, in order that s/he may contact you in the future about possible participation in other studies? *(You may take part in the study without agreeing to this).*	Yes ❒ No ❒

Your name (in BLOCK letters): ______________________________

Your signature: ______________________________

Researcher's name: ______Emanuela Buizza______________________

Date: ______________________________

One copy to be retained by the researcher, one copy to be kept by the participant.

A.3 Information sheet and consent form (online Google document), Perception experiment

Perception of reduced word forms in English

Researcher: Emanuela Buizza

Department of Language and Linguistic Science, University of York

Information for participants

This research investigates the perception of reduced word forms in English. To take part in this experiment you must be over 18 years old and a native speaker of British English.

If you decide to participate, you will be asked to listen to some short phrases and to indicate which one of two options presented on the screen you have heard. You will be able to play each phrase multiple times. If possible, you should carry out the experiment in a quiet environment and using headphones. The experiment takes approximately 5 minutes. This experiment must be carried out on a laptop or desktop computer as some features do not work on mobile devices.

Participation in this study is voluntary. You do not have to take part. If you decide to take part you can withdraw at any time without giving a reason, even during the session. If you withdraw from the study, your data will be destroyed and not used in any way.

The data you provide will be processed under the General Data Protection Regulation (GDPR). The University of York is committed to the principle of data protection by design and default and will collect the minimum amount of data necessary for the project. In addition, all data will be anonymised. No personal data will be retained. There are no risks associated with this study.

If you have any questions or concerns, please contact the researcher: Emanuela Buizza, Department of Language and Linguistic Science, University of York, Heslington, York, YO10 5DD; email: emanuela.buizza@york.ac.uk.

By clicking the link below, you acknowledge that your participation is voluntary, you meet the inclusion criteria outlined above, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

Before the experiment begins, you will be asked to answer a few questions. The questions marked with an asterisk (*) are required.

Start Experiment

PS The experiment seems to have some issues with the browser 'Safari'. If you are using a Mac, please open this link in Chrome:

https://run.pavlovia.org/EmanuelaB/perception-experiment-5/html/

Appendix B – List of sentences and list of stimuli

B.1 List of the sentences for production experiment

Full list of the sentences elicited in the production experiment in the order they were presented to the speakers (Chapters 2-6).

1. I'm burning wood
2. I'm burning leaves
3. I'm burning straw
4. I'm burning coal
5. I'm burning logs
6. He'll pack his books
7. He'll pack his shoes
8. He'll pack his clothes
9. We're pouring tea
10. We're pouring milk
11. We're pouring drinks
12. She'd burn the cake
13. She'd burn the toast
14. She'd burn the jam
15. She'd burn the roast
16. She'd burn the pie
17. We've bought a map
18. We've bought a tent
19. We've bought a torch
20. He's been abroad
21. He's been unwell
22. He's been unkind
23. We'd burnt the soup
24. We'd burnt the toasts
25. We'd burnt the rice
26. We'd burnt the cake
27. We'd burnt the fish
28. He could park the van
29. He could park the car
30. He could park the bike
31. We're teaching maths
32. We're teaching French
33. We're teaching Scots
34. You'll burn the toast
35. You'll burn the jam
36. You'll burn the cake
37. You'll burn the roast
38. You'll burn the pie
39. He's pecked at seeds
40. He's pecked at crumbs
41. He's pecked at worms

42. We’d boost the sales
43. We’d boost the trade
44. We’d boost the funds
45. I could burn the sticks
46. I could burn the logs
47. I could burn the straw
48. I could burn the wood
49. I could burn the leaves
50. He’d book a cab
51. He’d book a seat
52. He’d book a room
53. You’ve passed the church
54. You’ve passed the pub
55. You’ve passed the shop
56. We’ll burn the cake
57. We’ll burn the roast
58. We’ll burn the rice
59. We’ll burn the soup
60. We’ll burn the pie
61. He’d pack the bowls
62. He’d pack the mugs
63. He’d pack the plates
64. You can pause to breathe
65. You can pause to rest
66. You can pause to drink
67. It’s burnt the house
68. It’s burnt the tree
69. It’s burnt the chair
70. It’s burnt the park
71. It’s burnt the car
72. He’s potted plants
73. He’s potted bulbs
74. He’s potted herbs

BREAK 1

75. She’d bought a hat
76. She’d bought a dress
77. She’d bought a bag
78. I’d burn the cake
79. I’d burn the toast
80. I’d burn the jam
81. I’d burn the fish
82. I’d burn the pie
83. He can park the car
84. He can park the van
85. He can park the bus
86. We were pouring drinks
87. We were pouring tea
88. We were pouring juice
89. He could burn the wood
90. He could burn the leaves
91. He could burn the straw

92. He could burn the logs
93. He could burn the sticks
94. You'll pass the park
95. You'll pass the gate
96. You'll pass the house
97. He can beat his dad
98. He can beat his mum
99. He can beat his friend
100. We were burning sticks
101. We were burning wood
102. We were burning logs
103. We were burning straw
104. We were burning leaves
105. He could pick a date
106. He could pick a time
107. He could pick a seat
108. You've been abroad
109. You've been unwell
110. You've been unkind
111. He can pause the film
112. He can pause the tape
113. He can pause the game
114. They could burn the wood
115. They could burn the straw
116. They could burn the leaves
117. They could burn the sticks
118. They could burn the logs
119. You'll teach him maths
120. You'll teach him Welsh
121. You'll teach him Dutch
122. He's boosted funds
123. He's boosted trade
124. He's boosted sales
125. It'll burn the town
126. It'll burn the park
127. It'll burn the farm
128. It'll burn the trees
129. It'll burn the bridge
130. He's pushed a rock
131. He's pushed a car
132. He's pushed a door
133. We can park the car
134. We can park the van
135. We can park the bus
136. You've burnt the rice
137. You've burnt the sauce
138. You've burnt the steak
139. You've burnt the chips
140. You've burnt the fish
141. He'll beat his boss
142. He'll beat his aunt

143. He'll beat his friend

BREAK 2

144. She'd teach them French
145. She'd teach them Dutch
146. She'd teach them maths
147. He can burn the wood
148. He can burn the logs
149. He can burn the sticks
150. He can burn the leaves
151. He can burn the straw
152. We've passed the shop
153. We've passed the pub
154. We've passed the church
155. He could book a car
156. He could book a guide
157. He could book a room
158. You've paused to eat
159. You've paused to drink
160. You've paused to rest
161. I'd burnt the rice
162. I'd burnt the soup
163. I'd burnt the meat
164. I'd burnt the cream
165. I'd burnt the bread
166. He could peck at seeds
167. He could peck at worms
168. He could peck at crumbs
169. You're pushing rocks
170. You're pushing walls
171. You're pushing sacks
172. He'd boost the sales
173. He'd boost the funds
174. He'd boost the trade
175. She was burning logs
176. She was burning wood
177. She was burning straw
178. She was burning sticks
179. She was burning leaves
180. He'll punch the cards
181. He'll punch the bag
182. He'll punch the ball
183. You're teaching Scots
184. You're teaching French
185. You're teaching Dutch
186. It's burning down
187. It's burning fast
188. It's burning oil
189. It's burning well
190. He'd pot the herbs
191. He'd pot the bulbs
192. He'd pot the plants

193. You could pause the film
194. You could pause the tape
195. You could pause the game
196. We’d burn the roast
197. We’d burn the cake
198. We’d burn the jam
199. We’d burn the pie
200. We’d burn the toast
201. He was pouring tea
202. He was pouring juice
203. He was pouring drinks
204. They can burn the straw
205. They can burn the wood
206. They can burn the leaves
207. They can burn the logs
208. They can burn the sticks
209. You were parking too
210. You were parking there
211. You were parking fast
212. He’s packed the books
213. He’s packed the clothes
214. He’s packed the shoes

BREAK 3

215. She’d burnt the toast
216. She’d burnt the cake
217. She’d burnt the jam
218. She’d burnt the roast
219. She’d burnt the pie
220. He was passing fast
221. He was passing there
222. He was passing too
223. We can beat their team
224. We can beat their group
225. We can beat their friends
226. You’d bought a bag
227. You’d bought a cap
228. You’d bought a pen
229. I was burning straw
230. I was burning logs
231. I was burning coal
232. I was burning leaves
233. I was burning sticks
234. You can teach him Dutch
235. You can teach him Scots
236. You can teach him French
237. He’d pick a date
238. He’d pick a time
239. He’d pick a seat
240. They’ve burnt the jam
241. They’ve burnt the pie
242. They’ve burnt the toast

243. They’ve burnt the cake
244. They’ve burnt the soup
245. He was pecking seeds
246. He was pecking crumbs
247. He was pecking worms
248. We’re parking soon
249. We’re parking now
250. We’re parking fast
251. You could burn the leaves
252. You could burn the sticks
253. You could burn the logs
254. You could burn the wood
255. You could burn the straw
256. He’s punched the card
257. He’s punched the bag
258. He’s punched the ball
259. She’ll burn the roast
260. She’ll burn the cake
261. She’ll burn the rice
262. She’ll burn the bread
263. She’ll burn the fish
264. He can boost the funds
265. He can boost the sales
266. He can boost the trade
267. You’d pass the park
268. You’d pass the farm
269. You’d pass the bridge
270. He could beat his dad
271. He could beat his mum
272. He could beat his boss
273. It’d burn the house
274. It’d burn the grass
275. It’d burn the town
276. It’d burn the trees
277. It’d burn the park
278. He was pushing sacks
279. He was pushing rocks
280. He was pushing doors
281. You could teach him maths
282. You could teach him Welsh
283. You could teach him Scots
284. He’s burnt the soup
285. He’s burnt the rice
286. He’s burnt the sauce
287. He’s burnt the fish
288. He’s burnt the steak
289. We were parking too
290. We were parking there
291. We were parking fast

BREAK 4

292. He can pot the plants

293. He can pot the herbs
294. He can pot the bulbs
295. They're burning wood
296. They're burning leaves
297. They're burning straw
298. They're burning logs
299. They're burning sticks
300. He'll boost the sales
301. He'll boost the funds
302. He'll boost the trade
303. We can pause to rest
304. We can pause to drink
305. We can pause to eat
306. I'll burn the roast
307. I'll burn the pie
308. I'll burn the chips
309. I'll burn the toast
310. I'll burn the steak
311. He'd pass the park
312. He'd pass the church
313. He'd pass the shop
314. We could beat their friends
315. We could beat their group
316. We could beat their team
317. He was packing too
318. He was packing well
319. He was packing fast
320. He'd burnt the cake
321. He'd burnt the steak
322. He'd burnt the sauce
323. He'd burnt the pie
324. He'd burnt the fish
325. We've been abroad
326. We've been unkind
327. We've been unwell
328. He could punch the card
329. He could punch the ball
330. He could punch the bag
331. You're burning leaves
332. You're burning straw
333. You're burning sticks
334. You're burning logs
335. You're burning grass
336. He'd peck at worms
337. He'd peck at crumbs
338. He'd peck at seeds
339. You can park the van
340. You can park the car
341. You can park the bus
342. They'll burn the roast
343. They'll burn the cake

344. They'll burn the jam
345. They'll burn the toast
346. They'll burn the pie
347. He could pour the tea
348. He could pour the wine
349. He could pour the juice
350. We were teaching Dutch
351. We were teaching Welsh
352. We were teaching French
353. He'll pass the farm
354. He'll pass the pub
355. He'll pass the church
356. She's burning leaves
357. She's burning logs
358. She's burning coal
359. She's burning wood
360. She's burning straw
361. He can bend his knees
362. He can bend his back
363. He can bend his arms

BREAK 5

364. You're pouring tea
365. You're pouring wine
366. You're pouring drinks
367. He'll burn the cake
368. He'll burn the toast
369. He'll burn the pie
370. He'll burn the jam
371. He'll burn the roast
372. He was picking pears
373. He was picking fruit
374. He was picking nuts
375. We'll pause to eat
376. We'll pause to rest
377. We'll pause to drink
378. I can burn the logs
379. I can burn the leaves
380. I can burn the straw
381. I can burn the grass
382. I can burn the sticks
383. We'd beat their team
384. We'd beat their friends
385. We'd beat their group
386. He's bought a book
387. He's bought a cap
388. He's bought a bike
389. We've burnt the cake
390. We've burnt the steak
391. We've burnt the jam
392. We've burnt the rice
393. We've burnt the pie

394. She’d park the car
395. She’d park the van
396. She’d park the bus
397. He could pot the plants
398. He could pot the bulbs
399. He could pot the herbs
400. It can burn the trees
401. It can burn the house
402. It can burn the cars
403. It can burn the park
404. It can burn the town
405. He’ll book a cab
406. He’ll book a flight
407. He’ll book a guide
408. You can pause to breathe
409. You can pause to rest
410. You can pause to drink
411. He’d punched the card
412. He’d punched the bag
413. He’d punched the ball
414. They were burning leaves
415. They were burning coal
416. They were burning sticks
417. They were burning straw
418. They were burning logs
419. He can pack the books
420. He can pack the clothes
421. He can pack the shoes
422. You’d burnt the toast
423. You’d burnt the rice
424. You’d burnt the pie
425. You’d burnt the fish
426. You’d burnt the roast

BREAK 6

427. He’d pour the tea
428. He’d pour the drinks
429. He’d pour the wine
430. We could pass the farm
431. We could pass the church
432. We could pass the park
433. He was boosting sales
434. He was boosting funds
435. He was boosting trade
436. She can burn the leaves
437. She can burn the wood
438. She can burn the logs
439. She can burn the straw
440. She can burn the sticks
441. He’ll peck at crumbs
442. He’ll peck at seeds
443. He’ll peck at worms

444. You’d teach him French
445. You’d teach him Dutch
446. You’d teach him Scots
447. I’ve burnt the soup
448. I’ve burnt the toast
449. I’ve burnt the pie
450. I’ve burnt the roast
451. I’ve burnt the rice
452. He’s picked a seat
453. He’s picked a date
454. He’s picked a gift
455. You’ll pause to rest
456. You’ll pause to drink
457. You’ll pause to sleep
458. He’s burning coal
459. He’s burning straw
460. He’s burning leaves
461. He’s burning wood
462. He’s burning logs
463. You could park the bike
464. You could park the van
465. You could park the bus
466. He can pick a date
467. He can pick a time
468. He can pick a seat
469. We could burn the sticks
470. We could burn the logs
471. We could burn the straw
472. We could burn the leaves
473. We could burn the wood
474. He was potting plants
475. He was potting bulbs
476. He was potting herbs
477. We’d pass the bridge
478. We’d pass the park
479. We’d pass the farm
480. It could burn the park
481. It could burn the house
482. It could burn the trees
483. It could burn the town
484. It could burn the cars
485. He’ll pause to breathe
486. He’ll pause to rest
487. He’ll pause to drink

BREAK 7

488. You were pouring tea
489. You were pouring milk
490. You were pouring wine
491. He’d burn the pie
492. He’d burn the soup
493. He’d burn the roast

494. He’d burn the rice
495. He’d burn the fish
496. We’ll beat their friends
497. We’ll beat their team
498. We’ll beat their group
499. He could boost the trade
500. He could boost the funds
501. He could boost the sales
502. We’re burning leaves
503. We’re burning coal
504. We’re burning straw
505. We’re burning wood
506. We’re burning logs
507. He can book a seat
508. He can book a date
509. He can book a guide
510. You were teaching French
511. You were teaching Scots
512. You were teaching Welsh
513. It was burning well
514. It was burning fast
515. It was burning down
516. It was burning oil
517. He’s passed the shop
518. He’s passed the pub
519. He’s passed the farm
520. She could burn the logs
521. She could burn the sticks
522. She could burn the leaves
523. She could burn the wood
524. She could burn the straw
525. He’ll pot the plants
526. He’ll pot the bulbs
527. He’ll pot the herbs
528. We’re pouring tea
529. We’re pouring milk
530. We’re pouring drinks
531. They’d burnt the chips
532. They’d burnt the steak
533. They’d burnt the pie
534. They’d burnt the sauce
535. They’d burnt the bread
536. He could pack the clothes
537. He could pack the books
538. He could pack the shoes
539. She’d boost the sales
540. She’d boost the trade
541. She’d boost the funds
542. We can burn the straw
543. We can burn the wood
544. We can burn the logs

545. We can burn the sticks
546. We can burn the leaves
547. He was beating Nick
282. He was beating Pam
283. He was beating Jane
548. We'll park the bus
549. We'll park the car
550. We'll park the van

BREAK 8

551. You'd burn the roast
552. You'd burn the pie
553. You'd burn the jam
554. You'd burn the toast
555. You'd burn the fish
556. He was punching cards
557. He was punching holes
558. He was punching balls
559. We could pour the tea
560. We could pour the wine
561. We could pour the drinks
562. She'd bought a hat
563. She'd bought a dress
564. She'd bought a bag
565. You were burning logs
566. You were burning coal
567. You were burning wood
568. You were burning leaves
569. You were burning straw
570. She'd pass the gate
571. She'd pass the bench
572. She'd pass the pond
573. He'll pick a pear
574. He'll pick a fig
575. He'll pick a fruit
576. She's burnt the soup
577. She's burnt the rice
578. She's burnt the fish
579. She's burnt the pie
580. She's burnt the toast
581. We'd pause to eat
582. We'd pause to drink
583. We'd pause to rest
584. He was burning straw
585. He was burning logs
586. He was burning grass
587. He was burning leaves
588. He was burning sticks
589. She'll bend her knees
590. She'll bend her head
591. She'll bend her back
592. You can burn the logs

593. You can burn the leaves
594. You can burn the sticks
595. You can burn the grass
596. You can burn the straw
597. He’d beat his dad
598. He’d beat his mum
599. He’d beat his friend
600. It’d burnt the park
601. It’d burnt the car
602. It’d burnt the house
603. It’d burnt the town
604. It’d burnt the tree
605. He can punch the card
606. He can punch the bag
607. He can punch the ball
608. You’re passing now
609. You’re passing there
610. You’re passing fast
611. They’d burn the steak
612. They’d burn the chips
613. They’d burn the pie
614. They’d burn the bread
615. They’d burn the sauce

B.2 List of stimuli for perception experiment

Full list of stimuli (with the original name of the sound files) included in the Perception Experiment in the order they were presented to the listeners (Chapters 7-8).

S11_I_d_burn_the_fish_CONTROL
S13_You_were_burning_coal_VDur-85ms
S15_She_ll_burn_the_cake_REDUCED_LengthendHphDur-103ms
S7_You_d_burnt_the_pie_VDur-45ms_HphDur-102ms_burn-fish
S10_We_re_burning_wood_Original
S8_He_d_burn_the_soup_REDUCED
S7_It_was_burning_well_shortnedV-to-38ms
S14_I_d_burn_the_cake_Shortened-HPHdur-to-68ms
S6_We_were_burning_wood_REDUCED
S7_He_ll_burn_the_jam_REDUCED_LengthndHphDur-109ms
S9_You_d_burn_the_pie_compareHed
S7_She_s_burning_leaves_FRIC-IS_VW-WS-sticks_shortnVDUR-46ms
S14_I_ll_burn_the_steak_ShortenedVDur32ms_AverHphDur78ms
S6_We_were_burning_straw_dark
S7_You_d_burn_the_jam_compareHed
S11_It_was_burning_down_shortnedV-to-47ms
S8_She_d_burn_the_cake_LengthndVDur-36ms_ShortndHphDur-68ms_roast
S7_He_ll_burn_the_roast_REDUCED_LengthndHphDur-107ms
S13_You_re_burning_straw_LengthenedVDur-86ms_clear
S15_She_s_burnt_the_toast_REDUCED_OriginHphDur-79ms
S7_You_d_burn_the_fish_compareHed
S6_We_were_burning_leaves_CONTROL
S8_She_ll_burn_the_bread_REDUCED_LengthHphDur-106ms
S10_I_d_burn_the_pie_ORIG_LongVDur56ms_AverageHphDur101ms
S7_She_was_burning_straw_Orig_CONTROL
S9_You_d_burn_the_fish_compareHed
S8_She_d_burn_the_cake_LengthndVDur-36ms_LengthndHphDur-104ms_roast
S6_We_re_burning_logs_LengthndVDur-104ms_clear
S7_He_ll_burn_the_jam_REDUCED_held4
S15_She_s_burnt_the_fish_REDUCED_OriginalHphDur-76ms
S7_You_were_burning_logs_ORIG_CONTRL_VDur-93ms
S10_I_d_burn_the_pie_ORIG_LongVDur56ms_ShortHphDur85ms
S7_He_ll_burn_the_roast_REDUCED_held5
S15_She_s_burnt_the_fish_REDUCED_LengthendHphDur-101ms
S7_You_d_burnt_the_rice_compareHed

S6_We_re_burning_wood_REDUCED
S9_I_d_burn_the_pie_ORIG_LongVDur56ms_AverageHphDur96ms
S8_She_ll_burn_the_rice_VDur-14ms_LengthndHphDur-108ms
S7_You_were_burning_logs_ShortenedVDur-58ms_dark
S14_I_ll_burn_the_roast_Shortened-Vdur-to-43ms
S9_It_was_burning_fast_ORIG
S6_He_d_burnt_the_fish_fricD_compareHes
S7_She_was_burning_leaves_ORG-RED
S15_You_d_burn_the_pie_compareHed
S8_I_ll_burn_the_roast_CANONICAL_Lengthened-HPHdur-to-102ms
S7_You_were_burning_wood_REDUCED
S15_He_d_burnt_the_pie_fricD_compareHes
S14_I_ll_burn__the_toast_REDUCED
S8_She_d_burnt_the_toast_REDUCED_FricatedD_compareSHES
S9_It_was_burning_fast_ShortndVDur-23ms_dark
S14_You_d_burnt_the_rice_compareHed
S8_I_ll_burn_the_chips_ORIG_REDUCED
S15_You_were_burning_straw_OrigVDur-81ms
S7_She_s_burning_leaves_LengthenedVdur-57ms
S8_He_d_burnt_the_pie_fricD_compareHes
S14_I_ll_burn_the_steak_ShortenedVDur32ms_LengthHphDur102ms
S6_It_s_burnt_the_chair_Control

BREAK

S7_She_s_burning_leaves_CONTRL
S6_We_re_burning_logs_Original_clear
S7_He_d_burnt_the_fish_CompareYoud
S8_She_ll_burn_the_bread_REDUCED_OrignHphDur-74ms
S10_You_d_burn_the_fish_compareHed
S7_She_was_burning_straw_FRIC-WS_VW-IS-leaves_origVDUR-35ms
S14_I_ll_burn_the_roast_Shortnd-Vdur-to-43ms_Lengthnd-HPHdur-108ms
S10_We_were_burning_leaves_ShortndVDur-73ms_dark
S8_She_d_burn_the_cake_ShortndVDur-27ms_ShortndHphDur-68ms_roast
S15_You_were_burning_straw_ShortndVDur-58ms
S8_I_ll_burn_the_chips_REDUCED_Lengthened-HPHdur-to-103ms
S7_It_was_burning_well_CONTRL_ORIG
S11_You_d_burn_the_roast_compareHed
S9_She_ll_burn_the_bread_REDUCED_LengthendHphDur-104ms
S10_We_were_burning_leaves_original
S7_You_ll_burn_the_roast_ORIG_VDur-47ms_HphDur-71ms
S9_I_d_burn_the_pie_ShortndVDur34msBegKept_ShortHphDur-70ms

S13_You_were_burning_logs_ShortndVDur-57ms_Short_dark
S7_It_was_burning_oil_RED_ORIG
S8_She_d_burn_the_cake_ShortndVDur-27ms_LengthndHphDur-106ms_roast
S7_You_ll_burn_the_roast_VDur-47ms_LengthndHphDur-105ms
S11_It_was_burning_down_ORIG
S14_I_ll_burn_the_roast_ORIG_CONTROL
S7_She_was_burning_straw_ShortndVdur-36ms
S15_You_d_burn_the_pie_RED
S8_She_ll_burn_the_roast_LengthndVDur-37ms_HphDur-69ms
S10_We_re_burning_logs_LengthndVDur-103ms_clear
S8_You_d_burnt_the_pie_compareHed
S15_She_ll_burn_the_cake_REDUCED_OriginHphDur
S13_You_re_burning_straw_Short_clear
S7_I_d_burn_the_pie_ORIG_REDUCED
S10_We_were_burning_leaves_ShortenedVDUR_52ms
S7_You_ll_burn_the_pie_RED_ORIG_VDur-32ms_HphDur-78ms
S8_He_d_burn_the_roast_REDUCED
S7_She_s_burning_leaves_FRIC-IS_VW-WS-sticks_shortnVDUR-26ms
S10_You_d_burnt_the_rice_compareHed
S11_I_d_burn_the_fish_Shortened-HPHdur-to-74ms
S7_She_was_burning_straw_FRIC-WS_VW-IS-leaves_LengthndVDUR-78ms
S13_You_re_burning_straw_LengthenedVDur-105ms_clear
S8_He_d_burnt_the_cake_fricD_compareHes
S6_We_were_burning_leaves_ShortndVDur-58ms_dark
S7_You_d_burnt_the_roast_compareHed
S15_She_s_burnt_the_toast_REDUCED_LengthendHphDur-114ms
S7_You_ll_burn_the_pie_ VDur-32ms_LengthndHphDur-102ms
S15_We_were_burning_straw_REDUCED
S8_You_d_burnt_the_rice_compareHed
S15_We_re_burning_leaves_LengthndVDur-120ms_clear
S7_It_was_burning_down_RED_ORIG
S8_She_ll_burn_the_rice_VDur-14ms_HphDur-71ms
S13_You_were_burning_straw_VDur-104ms_Long_dark
S9_She_ll_burn_the_cake_REDUCED
S7_You_d_burnt_the_pie_VDur-45ms_ShortnHphDUR-62ms_burn-fish
S8_She_ll_burn_the_roast_LengthndVDur-37ms_LengthndHphDur-112ms
S11_You_were_burning_coal_REDUCED
S7_She_s_burning_leaves_FRIC-IS_VW-WS-sticks_OrigVDUR-69ms

Appendix C – R model syntax and outputs

C1. Linear mixed-effects models for the analysis of duration (Chapter 5)

Model syntax:

Six mixed effects models were run to test the difference in duration between the phonetics events in *we'll* and *we'd* (Section 5.2.1.1), *I'll* and *I'd* (Section 5.2.1.2), *she'll* and *she'd* (Section 5.2.1.3), *she's* and *she was* (Section 5.2.2.1), *you're* and *you were* (Section 5.2.2.2.1), *we're* and *we were* (Section 5.2.2.2.2), *he's* and *he'd* (Section 5.2.3.1.1), and *you'd* and *he'd* (Section 5.2.4.1). The durations analysed are: duration of silent articulation (`sa.dur`) or duration of the onset friction (`fr1.dur`), vocoid (`v.dur`), voicing (`vc.dur`), closure (`closure.dur`), voicing in closure (`vc.clo.dur`), and piece (`piece.dur`). *Auxiliary* (`Auxiliary`) was included in each model as a fixed effect with two levels and *Speaker* (`Speaker`) as a random intercept. The two levels of *Auxiliary* are *will~would*, *is~was*, *are~were*, *has~had*. The two levels of *Pronoun* are *you~he*.

The significant results that are reported in Chapter 5 are highlighted in yellow here for ease of reference.

```
Model.duration <- lmer (duration ~ Auxiliary + (1|Speaker), data = data, REML = F)
Model.duration.null <- lmer (duration ~ (1|Speaker), data = data, REML = F)
anova(model.duration, model.duration.null)
```

C1.A Outputs

Output of the likelihood ratio test for the pair *we'd* and *we'll* (Section 5.2.2.1)

```
> model.sa <- lmer (sa.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.sa.null <- lmer(sa.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.sa, model.sa.null)
Data: data
Models:
model.sa.null: sa.dur ~ (1 | Speaker)
model.sa: sa.dur ~ Auxiliary + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Model.sa.null  3 474.72 480.74 -234.36   468.72
Model.sa       5 478.70 488.73 -234.35   468.70 0.0263      2     0.9869
> model.v <- lmer (v.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
Model.v.null: v.dur ~ (1 | Speaker)
Model.v: v.dur ~ Auxiliary + (1 | Speaker)
```

```
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Model.v.null  3 1127.2 1135.8 -560.61   1121.2
Model.v       5 1126.3 1140.7 -558.15   1116.3 4.9373      2     0.0847 .
> model.vc <- lmer (vc.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Model.vc.null  3 702.37 709.32 -348.19   696.37
Model.vc       4 703.99 713.26 -347.99   695.99 0.3828      1     0.5361
> model.clo <- lmer (closure.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.clo, model.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
model.clo: closure.dur ~ Auxiliary + (1 | Speaker)
            Df  AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 1092 1100.6 -543.02     1086
model.clo       5 1028 1042.2 -509.01     1018 68.019      2  1.697e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.vc.clo <- lmer (vc.clo.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.clo.null <- lmer(vc.clo.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc.clo, model.vc.clo.null)
Data: data
Models:
model.vc.clo.null: vc.clo.dur ~ (1 | Speaker)
model.vc.clo: vc.clo.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.clo.null  3 579.83 586.49 -286.92   573.83
model.vc.clo       4 578.86 587.74 -285.43   570.86 2.9722      1    0.08471
> model.piece <- lmer (piece.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.piece.null  3 1201.7 1210.2 -597.85   1195.7
model.piece       5 1163.7 1178.0 -576.86   1153.7 41.976      2  7.674e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output of the likelihood ratio test for the contrasting pair *I'd* and *I'll* (Section 5.2.2.2)

```
> model.sa <- lmer (sa.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.sa.null <- lmer(sa.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.sa, model.sa.null)
Data: data
Models:
model.sa.null: sa.dur ~ (1 | Speaker)
model.sa: sa.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
```

```
model.sa.null  3 322.90 327.73 -158.45   316.90
model.sa       4 324.87 331.31 -158.44   316.87 0.0289      1      0.865
> model.v <- lmer (v.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
model.v.null: v.dur ~ (1 | Speaker)
model.v: v.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.v.null  3 652.57 659.49 -323.29   646.57
model.v       4 645.75 654.97 -318.88   637.75 8.8198      1    0.00298 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> model.vc <- lmer (vc.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.null  3 699.99 707.02 -346.99   693.99
model.vc       4 693.51 702.88 -342.75   685.51 8.4837      1   0.003583 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> model.clo <- lmer (closure.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.clo, model.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
model.clo: closure.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 612.20 619.03 -303.10   606.20
model.clo       4 571.56 580.66 -281.78   563.56 42.642      1  6.573e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> model.vc.clo <- lmer (vc.clo.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.clo.null <- lmer(vc.clo.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc.clo, model.vc.clo.null)
Data: data
Models:
model.vc.clo.null: vc.clo.dur ~ (1 | Speaker)
model.vc.clo: vc.clo.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.clo.null  3 564.22 571.22 -279.11   558.22
model.vc.clo       4 565.64 574.96 -278.82   557.64 0.5865      1     0.4438
> model.piece <- lmer (piece.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.piece.null  3 645.39 652.17 -319.69   639.39
model.piece       4 641.75 650.80 -316.87   633.75 5.6363      1    0.01759 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output of the likelihood ratio test for the contrasting pair *she'd* and *she'll* (Section 5.2.2.3)

```
> model.fr1 <- lmer (fr.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.null <- lmer(fr.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1, model.fr1.null)
Data: data
Models:
model.fr1.null: fr.dur ~ (1 | Speaker)
model.fr1: fr.dur ~ Auxiliary + (1 | Speaker)
            Df   AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.null  3 765.6 773.23 -379.8    759.6
model.fr1       4 764.6 774.77 -378.3    756.6 3.0032      1     0.0831 .
> model.v <- lmer (v.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
model.v.null: v.dur ~ (1 | Speaker)
model.v: v.dur ~ Auxiliary + (1 | Speaker)
            Df   AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
model.v.null  3 749.2 756.80 -371.60    743.2
model.v       4 747.7 757.83 -369.85    739.7 3.499      1     0.0614 .
> model.vc <- lmer (vc.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.null  3 817.67 825.33 -405.84   811.67
model.vc       4 818.74 828.96 -405.37   810.74 0.9297      1     0.3349
> model.clo <- lmer (closure.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.clo, model.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
model.clo: closure.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 771.04 778.57 -382.52   765.04
model.clo       4 711.41 721.45 -351.70   703.41 61.634      1  4.137e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.vc.clo <- lmer (vc.clo.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.clo.null <- lmer(vc.clo.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc.clo, model.vc.clo.null)
Data: data
Models:
model.vc.clo.null: vc.clo.dur ~ (1 | Speaker)
model.vc.clo: vc.clo.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.clo.null  3 676.46 684.06 -335.23   670.46
model.vc.clo       4 678.35 688.48 -335.18   670.35 0.1119      1      0.738
> model.piece <- lmer (piece.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
```

```
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
model.piece.null  3 877.29 884.86 -435.65   871.29
model.piece       4 874.08 884.17 -433.04   866.08 5.208      1    0.02248 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output of the likelihood ratio test for the pair *she's* and *she was* (Section 5.2.3.1)

```
> model.fr1 <- lmer (fr1.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.null <- lmer(fr1.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1, model.fr1.null)
Data: data
Models:
Model.fr1.null: fr1.dur ~ (1 | Speaker)
model.fr1: fr1.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Model.fr1.null  3 644.99 652.10 -319.49   638.99
Model.fr1       4 639.03 648.51 -315.51   631.03 7.9584      1   0.004786 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.v <- lmer (v.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
model.v.null: v.dur ~ (1 | Speaker)
model.v: v.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.v.null  3 797.01 804.19 -395.50   791.01
model.v       4 715.79 725.36 -353.89   707.79 83.225      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr2 <- lmer (fr2.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr2.null <- lmer(fr2.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.fr2, model.fr2.null)
Data: data
Models:
model.fr2.null: fr2.dur ~ (1 | Speaker)
model.fr2: fr2.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
model.fr2.null  3 608.84 615.95 -301.42   602.84
model.fr2       4 607.46 616.93 -299.73   599.46 3.382      1    0.06591 .
> model.vc <- lmer (vc.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.null  3 948.17 955.77 -471.09   942.17
model.vc       4 866.20 876.34 -429.10   858.20 83.968      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.clo <- lmer (closure.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.clo, model.fr.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
model.clo: closure.dur ~ Auxiliary + (1 | Speaker)
```

```
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 687.02 694.49 -340.51   681.02
model.clo       4 686.04 695.99 -339.02   678.04 2.9872      1    0.08392 .
> model.piece <- lmer (piece.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
model.piece.null  3 857.29 864.82 -425.64   851.29
model.piece       4 838.75 848.79 -415.37   830.75 20.54      1   5.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

<u>Output of the likelihood ratio test for the pair *you're* and *you were* (Section 5.2.3.2.1)</u>

```
> model.sa <- lmer (sa.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.sa.null <- lmer(sa.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.sa, model.sa.null)
Data: data
Models:
model.sa.null: sa.dur ~ (1 | Speaker)
model.sa: sa.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.sa.null  3 670.78 677.92 -332.39   664.78
model.sa       4 663.67 673.20 -327.83   655.67 9.1076      1   0.002545 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.v <- lmer (v.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
model.v.null: v.dur ~ (1 | Speaker)
model.v: v.dur ~ Auxiliary + (1 | Speaker)
            Df     AIC     BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.v.null  3 1084.52 1092.48 -539.26  1078.52
model.v       4  950.16  960.78 -471.08   942.16 136.36      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.vc <- lmer (vc.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Auxiliary + (1 | Speaker)
            Df     AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.null  3 1079.70 1087.6 -536.85  1073.70
model.vc       4  932.73  943.3 -462.36   924.73 148.98      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.clo <- lmer (closure.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.clo, model.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
model.clo: closure.dur ~ Auxiliary + (1 | Speaker)
```

```
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 788.76 796.73 -391.38   782.76
model.clo       4 788.58 799.19 -390.29   780.58 2.1895      1      0.139
> model.piece <- lmer (piece.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Auxiliary + (1 | Speaker)
             Df     AIC     BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.piece.null  3 1089.28 1097.21 -541.64  1083.28
model.piece       4  965.26  975.84 -478.63   957.26 126.02      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output of the likelihood ratio test for the pair *we're* and *we were* (Section 5.2.3.2.2)

```
> model.sa <- lmer (sa.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.sa.null <- lmer(sa.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.sa, model.sa.null)
Data: data
Models:
model.sa.null: sa.dur ~ (1 | Speaker)
model.sa: sa.dur ~ Auxiliary + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.sa.null  3 459.05 464.90 -226.53   453.05
model.sa       4 459.75 467.55 -225.87   451.75 1.3057      1     0.2532
> model.v <- lmer (v.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
model.v.null: v.dur ~ (1 | Speaker)
model.v: v.dur ~ Auxiliary + (1 | Speaker)
             Df     AIC     BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.v.null  3 1027.81 1035.48 -510.91  1021.81
model.v       4  893.44  903.66 -442.72   885.44 136.37      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> model.vc <- lmer (vc.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Auxiliary + (1 | Speaker)
             Df    AIC     BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.null  3 1005.6 1013.18 -499.81   999.62
model.vc       4  881.9  891.99 -436.95   873.90 125.71      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> model.clo <- lmer (closure.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.clo, model.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
```

```
model.clo: closure.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 672.22 679.62 -333.11   666.22
model.clo       4 670.84 680.70 -331.42   662.84 3.3777      1    0.06608
> model.vc.clo <- lmer (vc.clo.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.clo.null <- lmer(vc.clo.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc.clo, model.vc.clo.null)
Data: data
Models:
model.vc.clo.null: vc.clo.dur ~ (1 | Speaker)
model.vc.clo: vc.clo.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.clo.null  3 669.53 677.13 -331.77   663.53
model.vc.clo       4 671.30 681.43 -331.65   663.30 0.2311      1     0.6307
> model.piece <- lmer (piece.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.piece.null  3 969.01 976.48 -481.51   963.01
model.piece       4 857.93 867.88 -424.96   849.93 113.08      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output of the likelihood ratio test for the contrasting pair *he's* and *he'd* (Section 5.2.4.1.1)

```
> model.fr1 <- lmer (fr1.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.null <- lmer(fr1.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1, model.fr1.null)
Data: data
Models:
model.fr1.null: fr1.dur ~ (1 | Speaker)
model.fr1: fr1.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.null  3 515.13 521.75 -254.57   509.13
model.fr1       4 516.90 525.72 -254.45   508.90 0.2297      1     0.6318
> model.v <- lmer (v.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
model.v.null: v.dur ~ (1 | Speaker)
model.v: v.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.v.null  3 453.27 459.65 -223.63   447.27
model.v       4 454.30 462.81 -223.15   446.30 0.9697      1     0.3248
> model.vc <- lmer (vc.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
model.vc.null  3 575.68 581.96 -284.84   569.68
model.vc       4 577.68 586.06 -284.84   569.68 3e-04      1      0.986
> model.clo <- lmer (closure.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.clo, model.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
model.clo: closure.dur ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 526.81 533.33 -260.40   520.81
model.clo       4 528.72 537.42 -260.36   520.72 0.0876      1     0.7672
> model.fr2 <- lmer (fr2.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
```

```
> model.fr2.null <- lmer(fr2.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.fr2, model.fr2.null)
Data: data
Models:
model.fr2.null: fr2.dur ~ (1 | Speaker)
model.fr2: fr2.dur ~ Auxiliary + (1 | Speaker)
               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr2.null  3 452.32 458.34 -223.16   446.32
model.fr2       4 410.16 418.19 -201.08   402.16 44.163      1  3.021e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.piece <- lmer (piece.dur ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Auxiliary + (1 | Speaker)
               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.piece.null  3 590.88 597.31 -292.44   584.88
model.piece       4 587.37 595.95 -289.69   579.37 5.5028      1    0.01899 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output of the likelihood ratio test for the contrasting pair *he'd* and *you'd* (Section 5.2.5.1)

```
> model.fr1 <- lmer (fr1.dur ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.fr1.null <- lmer(fr1.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1, model.fr1.null)
Data: data
Models:
model.fr1.null: fr1.dur ~ (1 | Speaker)
model.fr1: fr1.dur ~ Pronoun + (1 | Speaker)
               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.null  3 788.80 796.46 -391.40   782.80
model.fr1       4 786.03 796.24 -389.01   778.03 4.7689      1    0.02898 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.v <- lmer (v.dur ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.v.null <- lmer(v.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.v, model.v.null)
Data: data
Models:
model.v.null: v.dur ~ (1 | Speaker)
model.v: v.dur ~ Pronoun + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.v.null  3 863.20 871.07 -428.60   857.20
model.v       4 821.61 832.11 -406.81   813.61 43.586      1  4.058e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.vc <- lmer (vc.dur ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.vc.null <- lmer(vc.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.vc, model.vc.null)
Data: data
Models:
model.vc.null: vc.dur ~ (1 | Speaker)
model.vc: vc.dur ~ Pronoun + (1 | Speaker)
              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.vc.null  3 944.87 952.72 -469.44   938.87
model.vc       4 914.41 924.87 -453.20   906.41 32.461      1  1.216e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.clo <- lmer (closure.dur ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.clo.null <- lmer(closure.dur ~ (1|Speaker), data = data, REML = F)
```

```
> anova(model.clo, model.clo.null)
Data: data
Models:
model.clo.null: closure.dur ~ (1 | Speaker)
model.clo: closure.dur ~ Pronoun + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.clo.null  3 782.08 789.84 -388.04   776.08
model.clo       4 780.26 790.60 -386.13   772.26 3.8233      1    0.05054
> model.piece <- lmer (piece.dur ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.piece.null <- lmer(piece.dur ~ (1|Speaker), data = data, REML = F)
> anova(model.piece, model.piece.null)
Data: data
Models:
model.piece.null: piece.dur ~ (1 | Speaker)
model.piece: piece.dur ~ Pronoun + (1 | Speaker)
            Df    AIC     BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.piece.null  3 993.71 1001.58 -493.85   987.71
model.piece       4 916.47  926.97 -454.24   908.47 79.234      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

C2. Linear mixed-effects models for the analysis of spectral moments (Chapter 5)

Model syntax:

Five mixed effects models were run to test the difference in spectral properties between the palato-alveolar friction in onset of *she'll* and *she'd* (Section 5.2.2.3) and *she's* and *she was* (Section 5.2.3.1); the initial friction in *he'd* and *you'd* (Section 5.2.5.1); and the alveolar friction in coda of *he's* and *he'd* (Section 5.2.4.1.1). The spectral moments analysed are: centre of gravity (`fr.cog`), Standard Deviation (`fr.SD`), skewness (`fr.skewness`) and kurtosis (`fr.kurtosis`). In addition, the amplitude of the friction was analysed too (`fr.dB`). *Auxiliary* (`Auxiliary`) was included in each model as a fixed effect with two levels and *Speaker* (`Speaker`) as a random intercept. The two levels of *Auxiliary* are *will~would*, *is~was*, *has~had*. For the contrast between *he'd* and *you'd*, *Pronoun* was included as a fixed effect with two levels instead of *Auxiliary*. The two levels of *Pronoun* are *you~he*.

The significant results that are reported in Chapter 5 are highlighted in yellow here for ease of reference.

```
Model.fr.cog <- lmer (fr.cog ~ Auxiliary + (1|Speaker), data = data, REML = F)
Model.fr.cog.null <- lmer (fr.cog ~ (1|Speaker), data = data, REML = F)
anova(model.fr.cog, model.fr.cog.null)
```

C2.A Outputs

Output of the likelihood ratio test for the spectral properties of the palato-alveolar friction in onset of *she'd* and *she'll* (Section 5.2.2.3)

```
> model.fr.dB <- lmer (fr.dB ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr.dB.null <- lmer(fr.dB ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.dB, model.fr.dB.null)
Data: data
Models:
model.fr.dB.null: fr.dB ~ (1 | Speaker)
model.fr.dB: fr.dB ~ Auxiliary + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.dB.null  3 506.64 514.31 -250.32   500.64
model.fr.dB       4 508.39 518.61 -250.19   500.39 0.2552      1     0.6135
> model.fr.cog <- lmer (fr.CoG ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr.cog.null <- lmer(fr.CoG ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.cog, model.fr.cog.null)
Data: data
Models:
model.fr.cog.null: fr.CoG ~ (1 | Speaker)
model.fr.cog: fr.CoG ~ Auxiliary + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.cog.null  3 1404.3 1412.0 -699.16   1398.3
model.fr.cog       4 1388.5 1398.7 -690.26   1380.5 17.806      1  2.446e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr.SD <- lmer (fr.SD ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr.SD.null <- lmer(fr.SD ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.SD, model.fr.SD.null)
Data: data
Models:
model.fr.SD.null: fr.SD ~ (1 | Speaker)
model.fr.SD: fr.SD ~ Auxiliary + (1 | Speaker)
```

```
                Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.SD.null  3 1296.7 1304.3 -645.34   1290.7
model.fr.SD       4 1292.6 1302.8 -642.32   1284.6 6.0354      1    0.01402 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr.skewness <- lmer (fr.skewness ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr.skewness.null <- lmer(fr.skewness ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.skewness, model.fr.skewness.null)
Data: data
Models:
model.fr.skewness.null: fr.skewness ~ (1 | Speaker)
model.fr.skewness: fr.skewness ~ Auxiliary + (1 | Speaker)
                Df    AIC    BIC   logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.skewness.null  3 212.61 220.27 -103.303   206.61
model.fr.skewness       4 207.55 217.77  -99.777   199.55 7.0529    1   0.007914 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr.kurtosis <- lmer (fr.kurtosis ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr.kurtosis.null <- lmer(fr.kurtosis ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.kurtosis, model.fr.kurtosis.null)
Data: data
Models:
model.fr.kurtosis.null: fr.kurtosis ~ (1 | Speaker)
model.fr.kurtosis: fr.kurtosis ~ Auxiliary + (1 | Speaker)
                Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.kurtosis.null  3 592.01 599.67 -293.00   586.01
model.fr.kurtosis       4 587.09 597.31 -289.55   579.09 6.9166    1   0.00854 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output of the likelihood ratio test for the spectral properties of the palato-alveolar friction in onset of *she's* and *she was* (Section 5.2.3.1)

```
> model.fr1.dB <- lmer (fr1.dB ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.dB.null <- lmer(fr1.dB ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1.dB, model.fr1.dB.null)
Data: data
Models:
model.fr1.dB.null: fr1.dB ~ (1 | Speaker)
model.fr1.dB: fr1.dB ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.dB.null  3 460.20 467.60 -227.10   454.20
model.fr1.dB       4 456.64 466.51 -224.32   448.64 5.5551      1    0.01843 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr1.cog <- lmer (fr1.CoG ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.cog.null <- lmer(fr1.CoG ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1.cog, model.fr1.cog.null)
Data: data
Models:
model.fr1.cog.null: fr1.CoG ~ (1 | Speaker)
model.fr1.cog: fr1.CoG ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.cog.null  3 1424.9 1432.6 -709.44   1418.9
model.fr1.cog       4 1426.3 1436.6 -709.15   1418.3 0.5947      1     0.4406
> model.fr1.SD <- lmer (fr1.SD ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.SD.null <- lmer(fr1.SD ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1.SD, model.fr1.SD.null)
Data: data
Models:
model.fr1.SD.null: fr1.SD ~ (1 | Speaker)
model.fr1.SD: fr1.SD ~ Auxiliary + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.SD.null  3 1367.3 1375.0 -680.64   1361.3
model.fr1.SD       4 1359.8 1370.1 -675.90   1351.8 9.4787      1   0.002079 **
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr1.skewness <- lmer (fr1.skewness ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.skewness.null <- lmer(fr1.skewness ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1.skewness, model.fr1.skewness.null)
Data: data
Models:
model.fr1.skewness.null: fr1.skewness ~ (1 | Speaker)
model.fr1.skewness: fr1.skewness ~ Auxiliary + (1 | Speaker)
             Df    AIC    BIC   logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.skewness.null  3 224.10 231.82 -109.051   218.10
model.fr1.skewness       4 206.32 216.62  -99.158   198.32 19.785      1  8.666e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr1.kurtosis <- lmer (fr1.kurtosis ~ Auxiliary + (1|Speaker), data = data, REML = F)
> model.fr1.kurtosis.null <- lmer(fr1.kurtosis ~ (1|Speaker), data = data, REML = F)
> anova(model.fr1.kurtosis, model.fr1.kurtosis.null)
Data: data
Models:
model.fr1.kurtosis.null: fr1.kurtosis ~ (1 | Speaker)
model.fr1.kurtosis: fr1.kurtosis ~ Auxiliary + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr1.kurtosis.null  3 591.70 599.42 -292.85   585.70
model.fr1.kurtosis       4 580.15 590.45 -286.08   572.15 13.543      1  0.0002332 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output of the likelihood ratio test for the spectral properties of the initial friction in *he'd* and *you'd* (Section 5.2.4.1)

```
> model.fr.dB <- lmer (fr.dB ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.fr.dB.null <- lmer(fr.dB ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.dB, model.fr.dB.null)
Data: data
Models:
model.fr.dB.null: fr.dB ~ (1 | Speaker)
model.fr.dB: fr.dB ~ Pronoun + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.dB.null  3 560.97 568.63 -277.49   554.97
model.fr.dB       4 562.41 572.63 -277.21   554.41 0.5583      1     0.4549
> model.fr.cog <- lmer (fr.cog ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.fr.cog.null <- lmer(fr.cog ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.cog, model.fr.cog.null)
Data: data
Models:
model.fr.cog.null: fr.cog ~ (1 | Speaker)
model.fr.cog: fr.cog ~ Pronoun + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.cog.null  3 1558.5 1566.1 -776.23   1552.5
model.fr.cog       4 1547.4 1557.6 -769.70   1539.4 13.066      1  0.0003007 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr.SD <- lmer (fr.SD ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.fr.SD.null <- lmer(fr.SD ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.SD, model.fr.SD.null)
Data: data
Models:
```

```
model.fr.SD.null: fr.SD ~ (1 | Speaker)
model.fr.SD: fr.SD ~ Pronoun + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.SD.null  3 1495.8 1503.4 -744.89   1489.8
model.fr.SD       4 1493.8 1504.1 -742.92   1485.8 3.9251      1    0.04757 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr.skewness <- lmer (fr.skewness ~ Pronoun + (1|Speaker), data = datashe, REML = F)
> model.fr.skewness.null <- lmer(fr.skewness ~ (1|Speaker), data = datashe, REML = F)
> anova(model.fr.skewness, model.fr.skewness.null)
Data: datashe
Models:
model.fr.skewness.null: fr.skewness ~ (1 | Speaker)
model.fr.skewness: fr.skewness ~ Pronoun + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.skewness.null  3 319.46 327.12 -156.73   313.46
model.fr.skewness       4 309.25 319.47 -150.62   301.25 12.209      1  0.0004756 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr.kurtosis <- lmer (fr.kurtosis ~ Pronoun + (1|Speaker), data = data, REML = F)
> model.fr.kurtosis.null <- lmer(fr.kurtosis ~ (1|Speaker), data = data, REML = F)
> anova(model.fr.kurtosis, model.fr.kurtosis.null)
Data: data
Models:
model.fr.kurtosis.null: fr.kurtosis ~ (1 | Speaker)
model.fr.kurtosis: fr.kurtosis ~ Pronoun + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr.kurtosis.null  3 761.18 768.84 -377.59   755.18
model.fr.kurtosis       4 757.47 767.69 -374.74   749.47 5.7096      1    0.01687 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output of the likelihood ratio test for the spectral properties of the alveolar friction in *he's* and *he'd* (Section 5.2.4.1.1)

```
> model.fr2.dB <- lmer (fr2.dB ~ Phon + (1|Speaker), data = data, REML = F)
> model.fr2.dB.null <- lmer(fr2.dB ~ (1|Speaker), data = data, REML = F)
> anova(model.fr2.dB, model.fr2.dB.null)
Data: data
Models:
model.fr2.dB.null: fr2.dB ~ (1 | Speaker)
model.fr2.dB: fr2.dB ~ Phon + (1 | Speaker)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr2.dB.null  3 399.56 406.26 -196.78   393.56
model.fr2.dB       4 399.99 408.93 -196.00   391.99 1.5718      1       0.21
> model.fr2.cog <- lmer (fr2.cog ~ Phon + (1|Speaker), data = data, REML = F)
> model.fr2.cog.null <- lmer(fr2.cog ~ (1|Speaker), data = data, REML = F)
> anova(model.fr2.cog, model.fr2.cog.null)
Data: data
Models:
model.fr2.cog.null: fr2.cog ~ (1 | Speaker)
model.fr2.cog: fr2.cog ~ Phon + (1 | Speaker)
            Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
model.fr2.cog.null  3 1278.3 1285.0 -636.14   1272.3
model.fr2.cog       4 1183.1 1192.1 -587.57   1175.1 97.15      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> model.fr2.SD <- lmer (fr2.SD ~ Phon + (1|Speaker), data = data, REML = F)
> model.fr2.SD.null <- lmer(fr2.SD ~ (1|Speaker), data = data, REML = F)
> anova(model.fr2.SD, model.fr2.SD.null)
Data: data
Models:
model.fr2.SD.null: fr2.SD ~ (1 | Speaker)
model.fr2.SD: fr2.SD ~ Phon + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr2.SD.null  3 1096.3 1103.0 -545.15   1090.3
model.fr2.SD       4 1095.7 1104.7 -543.86   1087.7 2.5847      1     0.1079
> model.fr2.skewness <- lmer (fr2.skewness ~ Phon + (1|Speaker), data = data, REML = F)
> model.fr2.skewness.null <- lmer(fr2.skewness ~ (1|Speaker), data = data, REML = F)
> anova(model.fr2.skewness, model.fr2.skewness.null)
Data: data
Models:
model.fr2.skewness.null: fr2.skewness ~ (1 | Speaker)
model.fr2.skewness: fr2.skewness ~ Phon + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr2.skewness.null  3 235.54 242.25 -114.77   229.54
model.fr2.skewness       4 157.88 166.82  -74.94   149.88 79.665      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.fr2.kurtosis <- lmer (fr2.kurtosis ~ Phon + (1|Speaker), data = data, REML = F)
> model.fr2.kurtosis.null <- lmer(fr2.kurtosis ~ (1|Speaker), data = data, REML = F)
> anova(model.fr2.kurtosis, model.fr2.kurtosis.null)
Data: data
Models:
model.fr2.kurtosis.null: fr2.kurtosis ~ (1 | Speaker)
model.fr2.kurtosis: fr2.kurtosis ~ Phon + (1 | Speaker)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.fr2.kurtosis.null  3 376.73 383.43 -185.36   370.73
model.fr2.kurtosis       4 374.43 383.37 -183.22   366.43 4.2957      1    0.03821 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C3. Exact two-sided binomial test reported in Chapter 7

<u>Model syntax:</u>

A series of exact two-sided binomial tests were run to determine whether group responses were significantly above a chance value of 50%. The features analysed in the perception experiment reported in Chapter 7 (Experiment A) were: friction of *'d* in the contrast *he'd* and *he's* (Section 7.3.1), initial audible friction in *you'd* and *he'd* (Section 7.3.2), highly reduced pr + aux (Section 7.3.2).

The significant results reported in Chapter 7 are highlighted in yellow here for ease of reference.

```
data %>% filter(contrast == "will-would" & pron == "she") %>%
  group_by(corr_resp) %>% tally()

binom.test(c(0s, 1s), p = 0.5, alternative = "two.sided")
```

C3.A Outputs

<u>Output of the binomial test for the intelligibility of *he's* and *he'd* with friction in coda.</u>

```
> data %>% filter(contrast == "fricD" & pron == "he",
+                          param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0   125
2         1   195
> binom.test(c(125, 195), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(125, 195)
number of successes = 125, number of trials = 320, p-value = 0.0001082
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.3368393 0.4464554
sample estimates:
probability of success
              0.390625
```

<u>Output of the binomial test for the intelligibility of *you'd* and *he'd* with friction in onset.</u>

```
> data %>% filter(contrast == "you'd-he'd",
```

```
+                  param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0   258
2         1   574
> binom.test(c(258, 574), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(258, 574)
number of successes = 258, number of trials = 832, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2787860 0.3427584
sample estimates:
probability of success
             0.3100962
```

Output of the binomial test for the intelligibility of highly reduced pr + aux.

```
> data %>% filter(contrast == "reduction",
+                  param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0   162
2         1   862
> binom.test(c(162, 862), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(162, 862)
number of successes = 162, number of trials = 1024, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1363709 0.1820200
sample estimates:
probability of success
             0.1582031
```

C4. Exact two-sided binomial test reported in Chapter 8

Model syntax:

A series of exact two-sided binomial tests were run to determine whether group responses were significantly above a chance value of 50%. The features analysed in the perception experiment reported in Chapter 8 (Experiment B) were: the contrast *will~would* in combination with the pronouns *she* (Section 8.3.1.1), *you* (Section 8.3.1.2), and *I* (Section 8.3.1.3); the contrast *is~was* in combination with the pronoun *she* (Section 8.3.2.1); and the contrast *are~were* in combination with the pronouns *we* (Section 8.3.2.2) and *you* (Section 8.3.2.3).

The significant results reported in Chapter 8 are highlighted in yellow here for ease of reference.

```
data %>% filter(contrast == "will-would" & pron == "she") %>%
  group_by(corr_resp) %>% tally()

binom.test(c(0s, 1s), p = 0.5, alternative = "two.sided")
```

C4.A Outputs

Output of the binomial test for the contrast between *will* and *would* in combination with the pronoun *she*.

```
> data %>% filter(contrast == "will-would" & pron == "she",
+                       param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0   112
2         1   720
> binom.test(c(112, 720), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(112, 720)
number of successes = 112, number of trials = 832, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1121494 0.1597050
sample estimates:
probability of success
             0.1346154
```

Output of the binomial test for the contrast between *will* and *would* in combination with the pronoun *you*.

```
> data %>% filter(contrast == "will-would" & pron == "you",
+                        param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0    14
2         1   370
> binom.test(c(14, 370), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(14, 370)
number of successes = 14, number of trials = 384, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.02007286 0.06041481
sample estimates:
probability of success
            0.03645833
```

Output of the binomial test for the contrast between *will* and *would* in combination with the pronoun *I*.

```
> data %>% filter(contrast == "will-would" & pron == "I",
+                        param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0    32
2         1   861
> binom.test(c(32, 861), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(32, 861)
number of successes = 32, number of trials = 893, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.02463728 0.05021276
sample estimates:
probability of success
            0.03583427
```

Output of the binomial test for the contrast between *is* and *was* in combination with the pronoun *she*.

```
> data %>% filter(contrast == "is-was" & pron == "she" & matching == "ch8",
+                       param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0    10
2         1   249
> binom.test(c(10, 249), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(10, 249)
number of successes = 10, number of trials = 259, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.01866767 0.06985723
sample estimates:
probability of success
            0.03861004
```

Output of the binomial test for the contrast between *are* and *were* in combination with the pronoun *you*.

```
> data %>% filter(contrast == "are-were" & pron == "you",
+                       param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
1         0   261
2         1   379
> binom.test(c(261, 379), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(261, 379)
number of successes = 261, number of trials = 640, p-value = 3.531e-06
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.3694594 0.4470199
sample estimates:
probability of success
             0.4078125
```

Output of the binomial test for the contrast between *are* and *were* in combination with the pronoun *we*.

```
> data %>% filter(contrast == "are-were" & pron == "we",
+                       param_resp != "space" ) %>%
+   group_by(corr_resp) %>% tally()
# A tibble: 2 x 2
  corr_resp     n
      <dbl> <int>
```

```
1          0   212
2          1   492
> binom.test(c(212, 492), p = 0.5, alternative = "two.sided")

        Exact binomial test

data:  c(212, 492)
number of successes = 212, number of trials = 704, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2674322 0.3365172
sample estimates:
probability of success
             0.3011364
```

Appendix D – Praat scripts

D1 – Praat script that reads the time of each marked point in a point tier.

```
Create Strings as file list... list
'sound_directory$'*'sound_file_extension$'
numberOfFiles = Get number of strings
fileappend "'resultfile$'" 'newline$' 'file$'
for ifile to numberOfFiles
	filename$ = Get string... ifile
	Read from file... 'sound_directory$''filename$'
	soundname$ = selected$ ("Sound", 1)
	gridfile$ = "'textGrid_directory$' 'soundname$'
'textGrid_file_extension$'"
	if fileReadable (gridfile$)
		Read from file... 'gridfile$'
		call GetTier 'tier$' tier
		fileappend "'resultfile$'" 'newline$' 'file$'
		numberOfPoints = Get number of points... tier
		for point from 1 to numberOfPoints
			label$ = Get label of point... tier point
			if label$ <> ""
				point_t = Get time of point... 1 point
				resultline$ = "'soundname$' 'label$' 'point_t'  "
				fileappend "'resultfile$'" 'resultline$'
				select TextGrid 'soundname$'
			endif
		endfor
		select TextGrid 'soundname$'
		Remove
	endif
	select Sound 'soundname$'
	Remove
	select Strings list
endfor
Remove
print Done.
```

D2 – Praat script that calculates the formant values of the first three formants at 9 equidistant points in time in each labelled interval in an interval tier.

```
Create Strings as file list... list
'sound_directory$'*'sound_file_extension$'
numberOfFiles = Get number of strings
fileappend "'resultfile$'" 'newline$' 'file$'
for ifile to numberOfFiles
      filename$ = Get string... ifile
      Read from file... 'sound_directory$''filename$'
      soundname$ = selected$ ("Sound", 1)
      To Formant (burg)... time_step maximum_number_of_formants
            maximum_formant window_length preemphasis_from
      gridfile$ =
            "'textGrid_directory$''soundname$''textGrid_file_extension$'"
      if fileReadable (gridfile$)
            Read from file... 'gridfile$'
            call GetTier 'tier$' tier
            fileappend "'resultfile$'" 'newline$' 'file$'
            numberOfIntervals = Get number of intervals... tier
            for interval from 1 to numberOfIntervals
                  label$ = Get label of interval... tier interval
                  if label$ <> ""
                        start = Get starting point... tier interval
                        end = Get end point... tier interval
                        quartpoint = start + ((end - start) / 4)
                        midpoint = start + ((end - start) / 2)
                        threequartpoint = start + (3 * ((end - start) / 4))
                        oneoctpoint = start + (1 * ((end - start) / 8))
                        threeoctpoint = start + (3 * ((end - start) / 8))
                        fiveoctpoint = start + (5 * ((end - start) / 8))
                        sevenoctpoint = start + (7 * ((end - start) / 8))
                        select Formant 'soundname$'
                        f1_a = Get value at time... 1 start Hertz Linear
                        f2_a = Get value at time... 2 start Hertz Linear
                        f3_a = Get value at time... 3 start Hertz Linear
                        f1_b = Get value at time... 1 oneoctpoint Hertz
                            Linear
                        f2_b = Get value at time... 2 oneoctpoint Hertz
                            Linear
                        f3_b = Get value at time... 3 oneoctpoint Hertz
                            Linear
                        f1_c = Get value at time... 1 quartpoint Hertz
                            Linear
                        f2_c = Get value at time... 2 quartpoint Hertz
                            Linear
                        f3_c = Get value at time... 3 quartpoint Hertz
                            Linear
                        f1_d = Get value at time... 1 threeoctpoint Hertz
                            Linear
                        f2_d = Get value at time... 2 threeoctpoint Hertz
                            Linear
                        f3_d = Get value at time... 3 threeoctpoint Hertz
                            Linear
```

```
                        f1_e = Get value at time... 1 midpoint Hertz Linear
                        f2_e = Get value at time... 2 midpoint Hertz Linear
                        f3_e = Get value at time... 3 midpoint Hertz Linear
                        f1_f = Get value at time... 1 fiveoctpoint Hertz
                            Linear
                        f2_f = Get value at time... 2 fiveoctpoint Hertz
                            Linear
                        f3_f = Get value at time... 3 fiveoctpoint Hertz
                            Linear
                        f1_g = Get value at time... 1 threequartpoint Hertz
                            Linear
                        f2_g = Get value at time... 2 threequartpoint Hertz
                            Linear
                        f3_g = Get value at time... 3 threequartpoint Hertz
                            Linear
                        f1_h = Get value at time... 1 sevenoctpoint Hertz
                            Linear
                        f2_h = Get value at time... 2 sevenoctpoint Hertz
                            Linear
                        f3_h = Get value at time... 3 sevenoctpoint Hertz
                            Linear
                        f1_i = Get value at time... 1 end Hertz Linear
                        f2_i = Get value at time... 2 end Hertz Linear
                        f3_i = Get value at time... 3 end Hertz Linear
                        resultline$ = "'soundname$'   'label$'    'f1_a'
      'f1_b'      'f1_c'      'f1_d'      'f1_e'      'f1_f'      'f1_g'
      'f1_h'      'f1_i'      'f2_a'      'f2_b'      'f2_c'      'f2_d'
      'f2_e'      'f2_f'      'f2_g'      'f2_h'      'f2_i'      'f3_a'
      'f3_b'      'f3_c'      'f3_d'      'f3_e'      'f3_f'      'f3_g'
      'f3_h'      'f3_i'      "
                        fileappend "'resultfile$'" 'resultline$'
                        select TextGrid 'soundname$'
                  endif
            endfor
            select TextGrid 'soundname$'
            Remove
      endif
      select Sound 'soundname$'
      plus Formant 'soundname$'
      Remove
      select Strings list
endfor
Remove
print Done.
```

D3 – Praat script that measures the amplitude and the first four spectral moments in a time window of 50% the duration of a labelled interval centred at mid-point.

```
Read Strings from raw text file... 'inputdir$'\'listfile$'.txt
numberOfFiles = Get number of strings
for fileteller to numberOfFiles
  select Strings 'listfile$'
  file$ = Get string... 'fileteller'
  Read from file... 'inputdir$'\'file$'.TextGrid
  Read from file... 'inputdir$'\'file$'.wav
  select TextGrid 'file$'
fileappend "'textfile$'" 'newline$' 'file$'
numberOfIntervals = Get number of intervals... 2
      for interval from 1 to numberOfIntervals
            select TextGrid 'file$'
            label$ = Get label of interval... 2 interval
            if label$ <> ""
                  start = Get starting point... 2 interval
                  end = Get end point... 2 interval
                  mid_point_first = start + end
                  mid = mid_point_first / 2
                  onequart = (start + mid) / 2
                  threequart = (mid + end) / 2
                  select Sound 'file$'
                  To Intensity... 100 0 yes
                  intensityID = selected("Intensity")
                  intensity_at_mid_point = Get mean... 'onequart'
                       'threequart' dB
                  Read from file... 'inputdir$'\'file$'.wav
                  select Sound 'file$'
                    plus TextGrid 'file$'
                  Edit
                    editor TextGrid 'file$'
                  Select... 'onequart' 'threequart'
                  Extract sound selection (time from 0)
                  Close
                  endeditor
                    select Sound untitled
                  To Spectrum... yes
                    select Spectrum untitled
                  Filter (pass Hann band)... 200 12000 100
                  Edit
                  cog = Get centre of gravity... 2
                  sdev = Get standard deviation... 2
                  skew = Get skewness... 2
                  kurt = Get kurtosis... 2
                  Remove
                  select Sound untitled
                  Remove
                  resultline$ = " 'label$' 'amplitude' =
                        'intensity_at_mid_point' 'CoG' = 'cog'     'SD' =
                        'sdev' 'skewness' = 'skew' 'kurtosis' = 'kurt'   "
                  fileappend "'textfile$'" 'resultline$'
            endif
      endfor
endfor
```

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Aguilar, L. (1999). Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication*, 28 (1), 57–74.

Aguilar, L. and Machuca, M. (1995). Intentionality in the speech act and reduction phenomena. In *Proceedings of the 13th International Congress of Phonetic Sciences*, *ICPhS*, 13–19.

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47 (1), 31–56.

Baker, R. (2008). The production and perception of morphologically and grammatically conditioned phonetic detail. PhD dissertation, University of Cambridge.

Baker, R., Smith, R. and Hawkins, S. (2007). Phonetic differences between mis- and dis- in English prefixed and pseudo-prefixed words. In *Proceedings of the 16th International Congress of Phonetic Sciences, ICPhS*, 553–556.

Ball, M. J., Howard, S. J. and Miller, K. (2018). Revisions to the extIPA chart. *Journal of the International Phonetic Association*, 48 (2), 155-164.

Bard, E. G., Sotillo, C., Anderson, A., Doherty-Sneddon, G. and Newlands, A. (1995). The control of intelligibility in running speech. In *Proceedings of the 13th International Congress of Phonetic Sciences*, *ICPhS*, 188-191.

Barry, W. and Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, 31 (1), 51–66.

Bates, D. M., Maechler, M. and Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.9999999-0.

Bauer, L. (2008). Lenition revisited. *Journal of Linguistics*, 44 (3), 605–624.

Bell, A., Brenier, J. M., Gregory, M., Girand, C. and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60 (1), 92–111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M. and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, 113 (2), 1001–1024.

Boersma, P. and Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37. Retrieved February 20, 2018.

Bochner, J. H., Snell, K. B. and MacKenzie, D. J. (1988). Duration discrimination of speech and tonal complex stimuli by normally hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 84 (2), 493–500.

Bouzon, C. and Hirst, D. J. (2004). Isochrony and prosodic structure in British English. In Bel, B. and Marlien, I. (eds.) *Proceedings of the 2nd International Conference on Speech Prosody*, 223–226.

Browman, C. P. and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6 (2), 201–251.

Browman, C. P. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J. and Beckman, M. E. (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 341–376.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49 (3-4), 155–180.

Buizza, E. (2010). Plosive lenition: Frication and affrication of /t/ in RP English spontaneous speech. MA dissertation, University of Leeds.

Buizza, E. and Plug, L. (2012). Lenition, fortition and the status of plosive affrication: The case of spontaneous RP English /t/. *Phonology*, 29 (1), 1–38.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14 (3), 261–290.

Bybee, J. and Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics*, 37 (4), 575–596.

Bybee, J., File-Muriel, R. J. and De Souza, R. N. (2016). Special reduction: A usage-based approach. *Language and Cognition*, 8 (3), 421-446.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24 (2), 209–244.

Cangemi, F. and Niebuhr, O. (2018). Rethinking reduction and canonical forms. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B. and Zellers, M. (eds.) *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. Berlin: De Gruyter Mouton, 277–302.

Carter, P. G. (2002). *Structured variation in British English liquids: The role of resonance*. PhD dissertation, University of York.

Carter, P. and Local, J. (2007). F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association*, 37 (2), 183–199.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Catford, J. C. (1988/2001). *A practical introduction to phonetics*. Oxford: Oxford University Press.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass*, 10 (3), 120–141.

Cho, T., Jongman, A., Wang, Y. and Sereno, J. A. (2020). Multi-modal cross-linguistic perception of fricatives in clear speech. *Journal of the Acoustical Society of America*, 147 (4), 2609–2624.

Cho, T., McQueen, J. M. and Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35 (2), 210–243.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clayards, M., Gaskell, M. G. and Hawkins, S. (2021). Phonetic detail is used to predict a word's morphological composition. *Journal of Phonetics*, 87, 101055.

Clopper, C. G. and Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B. and Zellers, M. (eds.) *Rethinking reduction*. Berlin: De Gruyter Mouton, 25–72.

Cole, J. and Shattuck-Hufnagel, S. (2018). Quantifying phonetic variation: Landmark labelling of imitated utterances. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B. and Zellers, M. (eds.) *Rethinking reduction*. Berlin: De Gruyter Mouton, 164–204.

Crowe, A. and Jack, M. A. (1987). Globally optimising formant tracker using generalised centroids. *Electronics Letters*, 23 (19), 1019–1020.

Cruttenden, A. (2008). *Gimson's pronunciation of English*. 8th ed. London: Hodder Education.

Cruz-Ferreira, M. (1998). Intonation in European Portuguese. In Hirst, D. and Di Cristo, A. (eds.) *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press, 167–178.

Davis, M. H., Marslen-Wilson, W. D. and Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal* *of Experimental Psychology: Human Perception and Performance*, 28 (1), 218–244.

Dilley, L. and Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America*, 122 (4), 2340–2353.

Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24 (4), 423–444.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, 142, 27–41.

Ernestus, M. and Smith, R. (2018). Qualitative and quantitative aspects of phonetic variation in Dutch *eigenlijk*. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B. and Zellers, M. (eds.) *Rethinking reduction*. Berlin: De Gruyter Mouton, 129–163.

Ernestus, M. and Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39, 253–260.

Ernestus, M., Baayen, H. and Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81 (1-3), 162–173.

Ernestus, M., Hanique, I. and Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics*, 48, 60–75.

Ernestus, M., Kouwenhoven H. and van Mulken, M. (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics*, 62, 50–64.

Fant, G. (1960). *Acoustics theory of speech production*. The Hague: Mouton.

Fant, G. (1992). Vocal tract area functions of Swedish vowels and a new three-parameter model. In *Proceedings of the 2nd International Conference on Spoken Language Processing*, 807–810.

Farnetani, E. and Recasens, D. (1999). Coarticulation models in recent speech production theories. In Hardcastle, W. J. and Hewlett, N. (eds.) *Coarticulation: Theory, data, technique*. Cambridge: Cambridge University Press, 31–65.

Firth, J. R. (1948). Sounds and prosodies. *Transactions of the Philological Society*. Reprinted in Palmer (1970), 1–26.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29, 109–135.

Fougeron, C. and Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101 (6), 3728–3740.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, 24 (1), 127–139.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, 31 (4), 307–319.

Fowler, C. A. and Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26 (5), 489–504.

Fowler, C. A. and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36 (2-3), 171–195.

Fowler, C. A., Levy, E. T. and Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, 37 (1), 24–40.

Gafos, A. and Dye, A. (2011). Vowel harmony: Transparent and opaque vowels. In van Oostendorp, M. and Rice, K. (eds.) *The Blackwell companion to phonology*. Malden: Wiley-Blackwell, 2164–2189.

Gahl, S., Yao, Y. and Johnson, K. (2012). Why reduce? Phonological neighbourhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66 (4), 789–806.

Gimson, A. C. (1989). *An introduction to the pronunciation of English*. 4th rev. ed. by Susan Ramsaran. London: Arnold.

Goldinger, S. D., Pisoni, D. B. and Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17 (1), 152–162.

Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E. and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, 35, 151–166.

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, 3 (1), 1–22.

Hanique, I., Ernestus, M. and Boves, L. (2015). Choice and pronunciation of words: Individual differences within a homogeneous group of speakers. *Corpus Linguistics and Linguistic Theory*, 11 (1), 161–185.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31 (3-4), 373–405.

Hawkins, S. (2010). Phonetic variation as communicative system: Perception of the particular and the abstract. In Fougeron, C., Kuehnert, B., D'Imperio, M. and Valee, N. (eds.) *Laboratory Phonology 10*. Berlin: De Gruyter Mouton, 479–510.

Hawkins, S. and Nguyen, N. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics*, 32 (2), 199–231.

Hawkins, S. and Slater, A. (1994). Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. In *Proceedings of the 1994 International Conference on Spoken Language Processing*, 57–60.

Hawkins, S. and Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics*, 13 (1), 99–188.

Hawkins, S. and Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, 22 (4), 493–511.

Heid, S. and Hawkins, S. (2000). An acoustical study of long domain /r/ and /l/ coarticulation. *Proceedings of the fifth seminar on Speech Production: Models and Data, ISCA*, 77–80.

Heine, B. (2003). Grammaticalisation. In Joseph, B. D. and Janda, R. D. (eds.) *Handbook of historical linguistics*. Oxford: Blackwell, 575–601.

Hertrich, I. and Ackermann, H. (1995). Coarticulation in slow speech: Durational and spectral analysis. *Language and Speech*, 38 (2), 159–187.

Hill, D. R., Jassem, W. and Witten, I. H. (1979). A statistical approach to the problem of isochrony in spoken British English. In Hollien, H. and Hollien, P. (eds.) *Current Issues in the Phonetic Science: Proceedings of the IPS 77 Congress*. Amsterdam: John Benjamins, 285–294.

Hughes, A., Trudgill, P. and Watt, D. (2005). *English accents and dialects: An introduction to social and regional varieties of English in the British Isles*. 4th ed. London: Arnold.

Huggins, A. W. F. (1972a). Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America*, 51 (4B), 1270–1278.

Huggins, A. W. F. (1972b). On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America*, 51 (4B), 1279–1290.

IPA (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Janse, E. and Ernestus, M. (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics*, 39 (3), 330–343.

Janse, E., Nooteboom, S. G. and Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes*, 22 (2), 161–200.

Jassem, W. (1952). *Intonation of conversational English (educated Southern British).* Wroclaw: Wroclawskie Towarzystwo Naukow.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *OSU Working Papers in Linguistics*, 50, 101–113.

Johnson, K. (2003). *Acoustic and auditory phonetics.* 2nd ed. Oxford: Blackwell.

Johnson, K. (2004). Massive reduction in conversational American English. In Yoneyama, K. and Maekawa, K. (eds.) *Proceedings of the first session of the 10th International Symposium on Spontaneous Speech: Data and analysis*, 29–54.

Johnson, W. and Britain, D. (2007). L-vocalisation as a natural phenomenon: Explorations in sociophonology. *Language Sciences*, 29 (2-3), 294–315.

Jones, D. (1960). *An outline of English phonetics.* 9th ed. Cambridge: W. Heffer & Sons.

Jones, M. J. and McDougall, K. (2009). The acoustic character of fricated /t/ in Australian English: A comparison with /s/ and /ʃ/. *Journal of the International Phonetic Association*, 39 (3), 265–289.

Jongman, A., Wayland, R. and Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108 (3), 1252–1263.

Jurafsky, D., Bell, A., Gregory, M. and Raymond, W. (2001a). The effect of language model probability on pronunciation reduction. *Proceedings of ICASSP-01 II*, 801–804.

Jurafsky, D., Bell, A., Gregory, M. and Raymond, W. (2001b). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P. J. (eds.) *Frequency and emergence of linguistic structure*. Amsterdam: John Benjamins, 229–254.

Kaisse, E. M. (1983). The syntax of auxiliary reduction in English. *Language*, 59 (1), 93–122.

Kaisse, E. M. (1985). *Connected speech: The interaction of syntax and phonology*. New York: Academic Press.

Kelly, J. and Local, J. (1989). *Doing phonology: Observing, recording, interpreting*. Manchester: Manchester University Press.

Kemps, R. J. J. K., Wurm, L. H., Ernestus, M., Schreuder, R. and Baayen, H. (2005). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, 20 (1-2), 43–73.

Klatt, D. H. and Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In Cohen, A. and Nooteboom, S. (eds.) *Structure and process in speech perception*. Heidelberg: Springer Verlag.

Kohler, K. J. (1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica*, 48 (2-4), 180–192.

Kohler, K. J. (1998). The disappearance of words in connected speech. *ZAS Papers in Linguistics*, 11, 21–33.

Kohler, K. J. (1999). Articulatory prosodies in German reduced speech. In *Proceedings of the 14th International Congress of Phonetic Sciences*, *ICPhS*, 89–92.

Kohler, K. J. (2011). Communicative functions integrate segments in prosodies and prosodies in segments. *Phonetica*, 68 (1-2), 26–56.

Kohler, K. J. and Niebuhr, O. (2011). On the role of articulatory prosodies in German message decoding. *Phonetica*, 68 (1-2), 57–87.

Kühnert, B. and Nolan, F. (1999). The origin of coarticulation. In Hardcastle, W. J. and Hewlett, N. (eds.) *Coarticulation: Theory, data, technique*. Cambridge: Cambridge University Press, 7–30.

Ladefoged, P. and Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lakoff, G. (1970). Global rules. *Linguistic Society of America*, 46 (3), 627–639.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lavoie, L. M. (2001). *Consonant strength: Phonological patterns and phonetic manifestations*. New York, London: Garland.

Lavoie, L. M. (2002). Some influences on the realisation of *for* and *four* in American English. *Journal of the International Phonetic Association*, 32 (2), 175–202.

Leech, G., Rayson, P. and Wilson, A. (2001). Word frequencies in written and spoken English: based on the British National Corpus. Word frequencies retrieved at: http://ucrel.lancs.ac.uk/bncfreq/

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1976). Suprasegmental features of speech. In Lass, N. J. (ed.) *Contemporary issues in experimental phonetics*. New York: Academic Press, 225–239.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6 (3), 172–187.

van Lieshout, P. H. H. M., Starkweather, C. W., Hulstijn, W. and Peters, H. F. M. (1995). Effects of linguistic correlates of stuttering on EMG activity in nonstuttering speakers. *Journal of Speech and Hearing Research*, 38 (2), 360–372.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35 (11), 1773–1781.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. J. and Marchal, A. (eds.) *Speech production and speech modelling*. Dordrecht: Kluwer, 403–439.

Lindblom, B. and Sundberg, J. (1971). Acoustic consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, 50 (4B), 1166–1179.

Local, J. (2003). Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics*, 31 (3-4), 321–339.

Local, J., Kelly, J. and Wells, W. H. G. (1986). Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics*, 22 (2), 411–437.

Luce, P. A., Pisoni, D. B. and Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In Altmann, G. T. M. (ed.) *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. London: The MIT Press, 122–147.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 25 (2), 187–205.

Manuel, S. Y. (1995). Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *Journal of Phonetics*, 25 (4), 453–476.

Manuel, S. Y., Shattuck-Hufnagel, S., Huffman, M., Stevens, K. N., Carlson, R. and Hunnicutt, S. (1992). Studies of vowel and consonant reduction. In *Proceedings of the 2nd International Conference on Spoken Language Processing (ICLSP 92)*, 943–946.

McDougall, K. (2004). Speaker-specific formant dynamics: An experiment on Australian English /aI/. *The International Journal of Speech, Language and the Law*, 11 (1), 103–130.

Mennen, I., Scobbie, J. M., de Leeuw, E., Schaeffler, S. and Schaeffler, F. (2010). Measuring language-specific phonetic settings. *Second Language Research*, 26 (1), 13–41.

Mitterer, H. (2011). Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics*, 39 (3), 298–303.

Mitterer, H. and Ernestus, M. (2006). Listeners recover/t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34 (1), 73–103.

Mitterer, H., Yoneyama, K. and Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of Memory and Language*, 59 (1), 133–152.

Niebuhr, O. (2016). Rich reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale. In *ExLing 2016: Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics*, 11–24.

Niebuhr, O. and Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, 39 (3), 319–329.

Noel, P. (2010). What constitutes anacrusis? Clause openings, Kuhn's Law, and the changing shape of Germanic syntax. In Seláf, L., Noel, P. and van Driel, J. (eds.) *Formes strophiques simples = Simple strophic patterns*. Budapest: Akadémiai Kiadó, 2010, 35-57.

Nooteboom, S. G. (1981). Lexical retrieval from spoken word fragments: Beginnings vs endings. *Journal of Phonetics*, 9 (4), 407–424.

Ogden, R. (1998). Dealing with destructive processes declaratively. Final report for ESRC award R000221880.

Ogden, R. (1999). A declarative account of strong and weak auxiliaries in English. *Phonology*, 16 (1), 55–92.

Ogden, R. (2009). *An introduction to English phonetics*. Edinburgh: Edinburgh University Press.

Ogden, R. (2012). Prosodies in conversation. In Niebuhr, O. (ed.) *Understanding prosodies*. Berlin: De Gruyter, 201–217.

Ogden, R. (2013). Clicks and percussives in English conversation. *Journal of the International Phonetic Association*, 43 (3), 299–320.

Ogden, R. and Local, J. (1994). Disentangling autosegments from prosodies: A note on the misrepresentation of a research tradition in phonology. *Journal of Linguistics*, 30 (2), 477–498.

Ogden, R., Hawkins, S., House, J., Huckvale, M., Local, J., Carter, P., Dankovicová, J. and Heid, S. (2000). ProSynth: An integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Computer Speech and Language*, 14 (3), 177–210.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39 (1), 151–168.

Palmer, F. R. (1965). *A linguistic study of the English verb*. London: Longman.

Palmer, F. R. (ed.) (1968). *Selected papers of J. R. Firth, 1952-59*. London/Harlow: Longman.

Palmer, F. R. (ed.) (1970). *Prosodic analysis*. London: Oxford University Press.

Palo, P., Schaeffler, S. and Scobbie, J. M. (2014). Pre-speech tongue movements recorded with ultrasound. In *Proceedings of the 10th International Seminar on Speech Production (ISSP)*, 300–303.

Palo, P., Schaeffler, S. and Scobbie, J. M. (2015). Effect of phonetic onset on acoustic and articulatory speech reaction times studied with tongue ultrasound. In *Proceedings of the 17th International Congress of Phonetic Sciences, ICPhS*, [paper no: 840].

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E. and Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. Behavior Research Methods, 51 (1), 195–203.

Pierrehumbert, J. and Talkin, D. (1992). Lenition of /h/ and glottal stop. In Docherty, G. and Ladd, D. R. (eds.) *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 90–117.

Pike, K. L. (1943). *Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds*. Ann Arbor: University of Michigan Press.

Pitt, M. A., Dilley, L. and Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, 39 (3), 304–311.

Pluymaekers, M., Ernestus, M. and Baayen, R. H. (2005a). Lexical frequency and acoustical reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118 (4), 2561–2569.

Pluymaekers, M., Ernestus, M. and Baayen, R. H. (2005b). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62 (2-4), 146–159.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org.

Rasskazova, O., Mooshammer, C. and Fuchs, S. (2019). Temporal coordination of articulatory and respiratory events prior to speech initiation. In *INTERSPEECH 2019*, 884–888.

Raymond, W. D., Dautrincourt, R. and Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18 (1), 55–97.

Recasens, D. (1989). Long range coarticulation effects for tongue dorsum contact in VCVCV sequences. *Speech Communication*, 8 (4), 293–307.

Roach, P. (2000). *English phonetics and phonology*. 3rd ed. Cambridge: Cambridge University Press.

Roach, P., Hartman, J. and Setter, J. (eds.) (2006). *Cambridge English Pronouncing Dictionary*. 17th ed. Cambridge: Cambridge University Press.

Schaeffler, S., Scobbie, J. M. and Schaeffler, F. (2014). Measuring reaction times: Vocalisation vs articulation. In *Proceedings of the 10th International Seminar on Speech Production (ISSP)*, 379–382.

Scheibman, J. (2000). *I dunno*: A usage-based account of the phonological reduction of *don't* in American English conversation. *Journal of Pragmatics*, 32 (1), 105–124.

Schuppler, B., Ernestus, M., Scharenborg, O. and Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, 39 (1), 96–109.

Scobbie, J. M., Schaeffler, S. and Mennen, I. (2011). Audible aspects of speech preparation. In *Proceedings of the 17th International Congress of Phonetic Sciences, ICPhS*, 1782–1785.

Selkirk, E. O. (1996). The prosodic structure of function words. In Demuth, K. and Morgan, J. L. (eds.) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, 187–213.

Sharoff, S. (2006). Creating general-purpose corpora using automated search engine queries. In Baroni, M. and Bernardini, S. (eds.) *Wacky! Working Papers on the Web as Corpus*. Bologna: GEDIT, 63–98. Collocations retrieved at http://corpus.leeds.ac.uk/protected/

Shockey, L. (2003). *Sound patterns of spoken English.* Oxford: Blackwell.

Simpson, A. (1992). The phonologies of the English auxiliary system. In Tracy, R. (ed.) *Who climbs the grammar-tree*. Tübingen: Niemeyer, 209–219.

Simpson, A. P. (2001). Does articulatory reduction miss more patterns than it accounts for? *Journal of the International Phonetic Association*, 31 (1), 29–39.

Simpson, A. P. and Pätzold, M. (eds.) (1996). *Sound patterns of connected speech: Description, models and explanation: Proceedings of the symposium held at Kiel University on 14-15 June 1996*. Kiel: Institut für Phonetik und Digitale Sprachverarbeitung, Universität Kiel.

Smith, R. (2015). Perception of speaker-specific phonetic detail. In Fuchs, S., Pape, D., Petrone, C. and Perrier, P. (eds.) *Individual differences in speech production and perception*. Frankfurt am Main: Peter Lang, 11–38.

Smith, R., Baker, R. and Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics*, 40 (5), 689–705.

van Son, R. J. J. H. and Pols, L. C. W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 88 (4), 1683–1693.

van Son, R. J. J. H. and Pols, L. C. W. (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 92 (1), 121–127.

van Son, R. J. J. H. and Pols, L. C. W. (1999). An acoustic description of consonant reduction. *Speech communication*, 28 (2), 125–140.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT.

Sussman, H. M., McCaffrey, H. A. and Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, 90 (3), 1309–1325.

Sweet, H. (1890). *A primer of phonetics*. Oxford: Clarendon Press.

Tench, P. (1992). From prosodic analysis to systemic phonology. In Tench, P. (ed.) *Studies in systemic phonology*, 1–17.

Tunley, A. (1999). *Coarticulatory influences of liquids on vowels in English*. PhD dissertation, University of Cambridge.

Van de Ven, M. and Ernestus, M. (2018). The role of segmental and durational cues in the processing of reduced words. *Language and Speech*, 61 (3), 358–383.

Van de Ven, M., Ernestus, M. and Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology*, 3 (2), 455–481.

Warner, N. and Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America*, 130 (3), 1606–1617.

Warner, N., Fountain, A. and Tucker, B. V. (2009). Cues to perception of reduced flaps. *Journal of the Acoustical Society of America*, 125 (5), 3317–3327.

Wells, J. C. (1982). *Accents of English 1: An introduction*. Cambridge: Cambridge University Press.

Wesener, T. (2001) Some non-sequential phenomena in German function words. *Journal of the International Phonetic Association*, 31 (1), 17–27.

West, P. (1999a). Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics*, 27 (4), 405–426.

West, P. (1999b). The extent of coarticulation of English liquids: An acoustic and articulatory study. In *Proceedings of the 14th International Congress of Phonetic Sciences*, *ICPhS*, 1901–1904.

Wilson, I. L. (2006). *Articulatory settings of French and English monolingual and bilingual speakers*. Ph.D. dissertation, University of British Columbia.

Wrench, A. A. (1995). Analysis of fricatives using multiple centres of gravity. In *Proceedings of the 13th International Congress of Phonetic Sciences*, *ICPhS,* Vol. 4, 460–463.

Zwicky, A. M. (1970). Auxiliary reduction in English. *Linguistic Inquiry*, 1 (3), 323–336.

Zwicky, A. M. (1977). *On clitics*. Bloomington, Indiana: IULC.