Patterns of Variation and Change in English Schwa

Sarah Tasker

PhD

University of York

Language and Linguistic Science

September 2020

Abstract

Schwa has been somewhat neglected in studies of spontaneous speech. This thesis addresses this gap by providing a detailed investigation of acoustic variation in schwa according to a number of variables. Large amounts of spontaneous speech are analysed from two different varieties of English; Derby, UK (Milroy et al, 1996), and New Zealand English (Gordon et al, 2007).

The work in this thesis contributes to the understanding of unstressed vowels in English in a number of ways. It is shown that schwa is distinct from /ɪ/ in unstressed syllables, although I also show that the way in which these vowel qualities are distributed differs amongst speakers. In addition, an extensive methodological analysis is presented, which explores the way that automated measurements of unstressed vowels can be filtered in order to make them suitable for analysis.

The thesis also contributes to debates about whether schwa has a phonetic target (cf. Browman and Goldstein, 1992; Flemming, 2009). A phonetic distinction between schwas that occur before a pause and before a consonant is found. Schwas before a pause are lower vowels and are also less variable in backness. Conversely, many pre-consonantal schwas are quite high in realisation, and vary widely in backness. However, when variation within schwa is considered, the evidence clearly points towards it having a phonetic target. Variation in schwa is explored according to its formant trajectory, its length, and also according to speaker year of birth. Clear evidence of schwa moving towards a phonetic target as it gets longer is found. Longer schwas are overall lower vowels, and also less variable in backness. Evidence of schwa undergoing change over time is also provided, with schwa having undergone substantial lowering in New Zealand English. Overall, the findings in this thesis provide clear evidence that schwa has a phonetic target.

Contents

6 CHANGE IN SCHWA IN NEW ZEALAND ENGLISH 177

List of Tables

List of Figures

Acknowledgements

I owe a great deal of gratitude to my supervisors, Márton Sóskuthy and Paul Foulkes. Their help and support throughout the PhD has been invaluable. I am immensely grateful or all the time they have spent reading my drafts, providing feedback and helping me work through my ideas, despite the difficulties with the distance over the past couple of years. Over the course of the PhD I have learnt a lot from both of them. I would also like to thank Márton for all the time spent helping me prior to the PhD with my initial application and applying for funding, without which I wouldn't have been doing the PhD at all.

I would also like to thank all the other linguists who have given various helpful suggestions and comments on my work over the past few years. In particular, I would like to thank my TAP member Vincent Hughes for all his comments on earlier drafts of these chapters. I was also lucky enough to have the opportunity to visit NZILBB, thanks to a generous grant by the AHRC, which was a really inspiring experience. I received so many useful suggestions whilst there, especially in relation to the methodological chapter in this thesis, which I am grateful for. I also thank NZILBB for kindly allowing me access to the ONZE data, and James Brand in particular for sharing his extracted formant data with me, and also the constant R help when I was there.

My fellow PhD colleagues also have my thanks for helping make this an enjoyable experience; especially Holly Dann, Jon Stevenson, Hielke Vriesendorp, Johanna Grace, and Nat Dziura for their friendship, and for all of the much needed laughs.

My family and friends outside academia have also helped me by supporting me and providing much needed distractions where necessary. In particular, I thank my parents for always supporting and believing in me, and also Abs and Joelle (including the last minute assistance!). I also thank Clare and Sam; without their generosity and kindness this last year would have been a lot more stressful. Finally, I thank Colin for making sure I didn't have to do anything other than finish this thesis over the past few weeks, for helping me stay motivated, and for all the support throughout.

This work was supported by the Arts & Humanities Research Council (grant number AH/L503848/1) through the White Rose College of the Arts & Humanities.

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.

Parts of the material from Chapter 3 have been presented at the following conferences; where presentations are co-authored I was responsible for the work presented in this thesis:

Tasker, S. (2017). Is levelling to schwa a change in progress? Evidence according to spelling.*11th Language Variation and Change conference*. Cardiff University, 29-31st August, 2017.

Tasker, S. (2018). Unstressed vowels in Derby: an investigation of the distinction between/ɪ/ and /ə/. *8th Northern Englishes Workshop.* Newcastle University, on 27-8th March 2018.

Tasker, S. (2018). Unstressed /ɪ/ and /ə/ in Derby: A categorical or gradient difference? *2018 Colloquium of the British Association of Academic Phoneticians*. University of Kent, 12-14th April 2018.

Tasker, S.; Sóskuthy, M. and Foulkes, P. (2018). How many reduced vowels are there in English: A Bayesian process-based approach. *16th Conference on Laboratory Phonology* .University of Lisbon, 19-22nd June 2018.

Tasker, S.; Sóskuthy, M. and Foulkes, P. (2018). Is variation between /ɪ/ and / ə/. Categorical? A Bayesian process-based approach. *4th Variation and Language Processing Conference*. Macquarie University on 23-25th January 2019.

Tasker, S (2019). Not just /ɪ/ and / ə/: Spelling predicts vowel quality in English. *19th International Congress of Phonetic Sciences*. Melbourne, Australia, 5-9th August 2019.

Elements of Chapter 3 have also previously been published:

Tasker, S. (2019). Not just /ɪ/ and / ə/: Spelling predicts vowel quality in English. Calhoun, P. Escudero, M. Tabain and P. Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (Pp 1879-1882). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

1 Introduction

This thesis addresses phonetic variation and change in the English schwa vowel. Unstressed vowels have been widely neglected in studies of spontaneous speech, and are rarely taken into account in research on language variation and change (e.g. Labov, 1994). Much of the work on schwa has tended to focus on speech in non-natural settings, such as read speech or schwas produced in experiments (e.g. Bates, 1995; Browman and Goldstein, 1992; Kondo, 1994; Gick, 2002; Flemming, 2009). There has also been considerable interest in schwa within phonological theory (e.g. Halle and Mohanan, 1985; van Oostendorp, 1998; Toft, 2002; Heselwood, 2007; Polgardi, 2015), due to its status as a 'neutral vowel'. However, there is still limited understanding of the way in which it is phonetically realised in spontaneous speech.

The mid central symbol [ə] is generally used to transcribe both word final unstressed vowels such as in *comm<u>a</u>* and *lett<u>er</u>*, and also non-final unstressed vowels in words such as *<u>a</u>bout* and *b<u>a</u>lloon*. However, there have been many studies using read speech which have found that the quality of schwa can be higher and much more variable in backness (e.g. Bates, 1995, Flemming and Johnson, 2007). This has led some to question whether schwa has its own vowel quality target or whether it is in fact a targetless vowel (e.g. Browman and Goldstein, 1992, Kondo, 1994; Bates, 1995; Flemming, 2009). If it is targetless, this would mean that speakers do not aim for a specific vowel quality target when producing schwa, and that its realisation is instead completely dependent on phonetic context. In this view, the articulators only move the minimum amount needed to produce a vowel between adjacent sounds. There is also debate as to whether word final schwa and non-final schwa are phonologically different. Despite them both generally being transcribed with [ə], some have found their realisation and behaviour to be quite different, and propose that they are fundamentally distinct (e.g. Flemming and Johnson, 2007). In light of the apparent variability of schwa and claims about its targetless nature, this also raises questions about how schwa is distinct, if at all, from other vowel qualities. This is particularly the case in respect of /ɪ/, which also appears in unstressed syllables.

This thesis examines these issues across two different varieties of English. I utilise large amounts of spontaneous speech data in corpora from Derby in England (Milroy et al, 1997), and from New Zealand using the Origins of New Zealand English (henceforth ONZE) corpus (Gordon et al, 2007), in order to explore variation in the schwa vowel. There are some particular issues with analysing measurements from schwa, insofar as it is generally very short compared to other vowels (e.g. Bates, 1995; Flemming, 2009), and does not always appear where transcriptions suggest that it should. For this reason, I also offer an analysis of how automated measurements can be used for reliable analysis of unstressed vowels. The rest of this introduction introduces the key concepts which are fundamental to this thesis, and explains how key terms will be defined. I also present the central research questions of the thesis, and provide a description of how the thesis progresses.

There is a particular inconsistency in the way in which the label 'schwa' is used. It can be used in a phonetic sense to describe the mid central vowel realisation transcribed with [ə], whether or not that particular vowel quality is stressed (Wells, 1982). Schwa is also used phonologically, to describe the vowel which is only found in unstressed syllables (e.g. in *comma, about*, *today*).This mixed usage creates a degree of confusion, as schwa in the phonological sense can of course have a variety of phonetic realisations, as any vowel can. As will be seen in this thesis, /ə/ is not always produced as mid central [ə]. In this thesis the label schwa will be used in this second phonological sense, referring to the vowels of unstressed syllables. I am thus not *a priori* prescribing any notion of what schwa phonetically is or is not. Rather, I look at the vowels that occur in unstressed syllables and explore their phonetic realisation.

It is important to note here that in descriptions of English, schwa is often described as one of three vowel qualities which can occur in unstressed syllables, the others being /ɪ/ and /ʊ/ (e.g. Gimson, 1962; Roach, 1983). In this thesis, however, I do not make any initial assumptions about which lexical items would have which vowel qualities. Rather, the topic of differences within unstressed vowels is explored in Chapter 3, and this analysis informs which vowels are used within the analyses in the rest of the thesis.

Only lexical schwas are included in the analyses, where the vowel 'aimed for' is a schwa. This means that epenthetic vowels which are a variable phonetic by-product between adjacent sounds will not be included. Although such vowels may often be phonetically similar to schwa (Silverman, 2011) studies have shown that they can be produced differently from phonological schwa (Gick and Wilson, 1999; Davidson, 2006).

Lexical items where schwa is in variation with the use of a full vowel are also not included in the analyses. Therefore words such as *advice* are not included as here the pronunciations [ədvaɪs] and [advaɪs] are both possible. The analysis instead only includes words such as *about,* where the only vowel possible (for native speakers) is a reduced one. It is important here to make a distinction between phonetic and phonological reduction (Fourakis, 1991; Van Bergem, 1993; referred to as acoustic and lexical reduction by Van Bergem). Phonetic reduction is the process whereby vowels move further away from their phonetic target according to factors such as duration, context, effort and stress, and is a gradient process which affects all vowels. Phonological vowel reduction, on the other hand, is where the vowel target is a reduced vowel quality which only occurs in unstressed syllables i.e. schwa. These two processes are of course connected, and it is thought that phonologically reduced vowels owe their origins to the neutralisation of other vowel quality contrasts due to phonetic reduction (e.g. Ladefoged and Johnson, 2014). However, in this thesis the aim is to look at the realisation of phonologically reduced vowels. In a word such as *November*, even if the vowel produced is perceived as a schwa there is no way of knowing whether the production of schwa was the aim of the speaker, or whether they simply produced a phonetically reduced /əʊ/. Therefore in order to have more certainty that a phonologically reduced vowel is the aim of the speaker, only lexical items where a phonologically reduced vowel is mandatory are investigated. The method for selecting which individual lexical items to include in the analyses is explained in Chapter 3.

It has long been considered that schwa is in some sense not like other vowels. It can only occur in unstressed syllables, its vowel quality can be similar to a phonetically reduced vowel, and it is often seen as occupying a central position in the vowel space. Schwa is as such often described as a neutral vowel (e.g. Chomsky and Halle, 1968; van Bergem, 1989; Browman and Goldstein, 1992). The meaning of neutrality can, however, take different forms, and depends whether it is phonological or phonetic neutrality that is referred to.

In regards to phonological neutrality, schwa has variously been described as an “empty slot” (Halle and Mohanan, 1985), as a “default vowel” (Toft, 2002, Polgardi, 2015), and as a vowel without features (Van Oostendorp, 1995). Similarly Heselwood (2007: 48) writes that schwa “is not in itself distinctive in the vowel system”, and seems to suggest that its only function is to separate consonants e.g. to separate consonants those that cannot occur together as in *to̲day*, or *de̲tain*.

In terms of phonetic neutrality, there are also various conceptions of how neutrality could manifest itself. Phonetic neutrality in schwa can refer to the idea that schwa occupies the centre of the vowel space. This is of course the position that is implied when schwa is transcribed with the symbol [ə]. Browman and Goldstein’s (1992) position is a variant on this idea, as they argue that “schwa seems to involve a warping of the trajectory toward an overall average or neutral tongue position” (p. 208). It has been noted by many scholars, however, that the vowel quality of schwa can be very variable and that it is often not produced with a mid central quality (e.g. Lass, 1986; Kondo, 1994; Bates, 1995; Flemming, 2009). This has led some to argue that schwa may have no phonetic target of its own at all (e.g. Bates, 1995). Targetlessness is, in a sense, a particular type of phonetic neutrality. However, rather than schwa itself having a particular neutral vowel target, its articulation is determined by the neutral position of the articulators within a given context.

Phonetic targetlessness in a vowel is generally used to mean a vowel which shows no evidence of movement to a phonetic target of its own. This is most commonly conceived of as complete predictability according to phonetic context (e.g. Kondo, 1994; Bates, 1995). The implication of this is that a targetless vowel would not involve any independent articulation, and that it would merely serve as an ‘in between’ movement between two adjacent segments. In terms of acoustics, this has been said to imply that a targetless vowel should show interpolation between adjacent contexts (Kondo, 1994; Flemming, 2009; Geng et al, 2010), meaning that it would be expected to show linear formant trajectories. For example, if schwa is surrounded by two segments with a high F2, it should also have a high F2 and should not show any movement away from this high F2. A slightly weaker version of this argument is made by Flemming (2009). Flemming argues that schwa should generally show complete interpolation between its adjacent contexts but only insofar as this allows it to be perceived as a vowel. For example, if schwa is surrounded by adjacent consonants which would predict an extremely low F1, schwa would still show a minimal movement to a higher F1 in order to be perceived as a vowel. Another variant of the idea of targetless is made by Bates (1995). In general, Bates concurs with the idea that a targetless schwa should move seamlessly between adjacent consonants. However, Bates also suggests that when there is more of a gap between the articulators, or when schwa is followed by silence that schwa my show movement to a neutral rest position.

In terms of the neutrality of schwa, this thesis only addresses the question of phonetic targetlessness. This is important as phonetic targetlessness and phonological neutrality do not necessarily equate to each other. It is possible to conceive of a situation where schwa could show clear evidence of movement towards a phonetic target, but could still be analysed as phonologically neutral. In this thesis, targetlessness is taken to mean a vowel that shows no evidence of movement towards an independent phonetic target in between contexts, above the minimum which is required to produce a vowel. Unlike many other studies of targetlessness, which have used more controlled speech, this thesis uses spontaneous speech. Therefore, due to the natural variation in speech, complete predictability from phonetic context is not expected. However, a targetless vowel is expected to show a higher level of assimilation to phonetic context than other vowels. The idea of schwa as a targetless vowel also implies that its vowel quality cannot be important in terms of contrast with other vowels, since great variability would be a poor vehicle for producing perceivable contrasts. It also suggests that it would not play the same role in vowel systems as other vowels. This means that a targetless vowel would not be expected to interact with other vowels in sound changes.

The word schwa typically refers to both unstressed vowels in word final position such as *comma* and non-final schwas such as *canal*. However, some have suggested that these two vowels are actually different phonologically (e.g. Flemming and Johnson, 2007). This thesis does not make any initial assumptions in this regard, but investigates both of these classes of words in order to see how they pattern. A division is made between word final schwa and non-final schwa. A division is also made between schwas that occur before a pause (pre-pausal schwa), and schwas that occur before a consonant (pre-consonantal schwa). Whilst non-final schwa is always pre-consonantal, word final schwa can either occur before a pause or a word beginning with a consonant. There are therefore three different groups of schwa that are examined in the thesis: non-final schwa, pre-consonantal word final schwa, and pre-pausal schwa.

The thesis addresses the following main research questions:

1) Is there evidence of a phonetic target in schwa?

This is the question at the heart of the thesis and is a theme that runs through most of the chapters. Although they are separate questions, questions 2 and 3 also feed into answering this question. This question is approached in several different ways. This includes examining whether schwa shows evidence of movement towards a phonetic target in the way it varies according to its formant trajectories and duration. The behaviour of schwa over time is also examined in order to see if it interacts with other vowels in sound changes. Throughout, there is a focus on comparing schwa to other vowels, in order to ensure that any claims of targetlessness are not based on behaviour which is also found in other vowels.

2) Are word final and non-final schwa phonologically distinct?

This thesis compares analyses across non-final schwa, pre-pausal schwa and pre-consonantal word final schwa. It is therefore possible to see to what degree the behaviour of schwa is determined by its word position, or by whether it precedes a pause or a consonant.

3) Is schwa distinct from other vowels in unstressed syllables?

As has already been mentioned, schwa is considered by some to be one of three vowel qualities which can occur in unstressed syllables. This thesis thus also examines how differences between unstressed vowel qualities manifest phonetically. This is an important issue in the description of schwa and how it contrasts from other vowels. In addition, addressing this issue is important in deciding what tokens to include in the schwa analyses in the thesis.

These are the main research questions addressed by the thesis as a whole. Within each of the analysis chapters there are more specific questions which are addressed, in order to contribute to answering these main questions. These sub-questions are set out within the individual chapters, where their contribution to the main thesis questions is also explained.

In the remainder of this chapter I provide an outline of the thesis.

Chapter 2 lays out the main issues which are addressed in the thesis, and hence provides a justification for the main research questions asked. I first discuss the more traditional impressionistic literature regarding schwa and its relationship to other unstressed vowel qualities. I also provide a detailed discussion of the main pieces of research which have previously investigated the idea of targetlessness in schwa. This chapter also provides a discussion of the various methods which will be used to investigate targetlessness in this thesis. Finally I review the current evidence on the realisation of schwa in different word positons.

Chapters 3 and 4 are both analyses of unstressed vowels within speakers from Derby. These chapters both use a dataset of manually segmented, and manually edited formant measurements. Chapter 3 focusses on the relationship between schwa and other unstressed vowel qualities. In doing so it addresses question 3. It also informs the criteria that are used to select schwa tokens in the rest of the thesis. In Chapter 4 the attention moves to the question of whether schwa has a phonetic target, thus addressing question 1. The way that schwa varies according to its formant trajectory and its duration are explored, as is its overall variability. Schwa is also compared to a range of stressed vowels in order to review how suitable the methods used are for identifying the existence of a phonetic target. This chapter also compares the realisation of schwa depending on both its phonetic context and its word positon, and in doing so also contributes to answering question 3.

Chapter 5 is a methodological chapter. The focus is on examining to what degree fully automated measurements are a suitable methodology to use for the analysis of unstressed vowels. The Derby corpus used for the analysis in Chapters 3 and 4 was small enough for every token to be manually segmented, and for the formant measurements to be manually checked. However, this was not the case for the data from the ONZE corpus, which is much larger. The purpose of this chapter is therefore to examine to what degree fully automated measurements are suitable for use in the analysis of unstressed vowels. The dataset that was created from the Derby corpus (used in Chapters 3 and 4) is compared with a fully automated set of measurements from the same corpus. The consistency across these two sets of measurements is compared, and there is a focus on how the automated data can be edited in order to make it more similar to the manual data and hence more suitable for analysis. The findings from this chapter are thus used to inform the data preparation steps used in Chapter 6 for the data taken from the ONZE corpus.

Chapter 6 presents an analysis and comparison of the changes in schwa and KIT over time in New Zealand English (NZE). Through assessing whether schwa shows any change over time, this chapter provides another angle of assessing whether it has a phonetic target, and thus answering question 1. In addition, this chapter also contributes to answering question 1 through investigating the comparative effect of duration on schwa and KIT. Findings from both pre-pausal and non-final schwa are explored in this chapter, and as such question 2 is also addressed.

Chapter 7 brings the various pieces of evidence from the different analyses together, in order to offer answers to the broad research questions posed. I discuss the different contributions that the various chapters make to answering the research questions, and compare the results from the various analyses. I also offer possible explanations for the overall findings.

Finally, Chapter 8 ends the thesis by concluding the main findings.

2 Background

2.1 Introduction

In this chapter, I provide a more detailed discussion of the issues that were introduced in Chapter 1. Additionally, this chapter aims to provide a justification for the importance of the research questions addressed, and of the methods which are used to investigate them.

The chapter starts by reviewing research on the distinction between the /ɪ/ and schwa, including how this links with claims about the variability of schwa. It then moves to a detailed discussion of the main studies which have examined the question of whether schwa is targetless. This is followed by a more general discussion of the issues in investigating targetlessness and phonetic variation in schwa. The difference in phonetic quality between non-final and word final schwa is discussed, including how this might be explained in light of the previous discussion. Finally, this chapter concludes with a summary of the issues raised in this chapter, and how the analyses in the rest of the thesis build on the literature discussed in this chapter.

2.2 Schwa versus /ɪ/

Descriptions of Received Pronunciation (henceforth RP) English tend to describe it as having one of three possible vowel qualities in unstressed syllables: /ɪ/, /ʊ/ or /ə/ (Gimson, 1962; Roach, 1983). However, Roach (1983) notes that /ʊ/ is less common and only tends to appear following /j/. The vast majority of the commentary on variation between unstressed vowel qualities is referring to /ɪ/ and /ə/, suggesting that this is the main locus of variation.

The vowels /ɪ/ and /ə/ are said to contrast in certain words e.g. *abbot* and *rabbit* (Wells, 1982), *accept* and *except; affect* and *effect* (Gimson, 1962). Despite this, /ɪ/ and /ə/ are, in a sense, not seen as equal. Whilst the /ə/ vowel quality is seen as unique to unstressed syllables, the /ɪ/ quality is reported to be present in both stressed and unstressed syllables (e.g. Roach, 1983). Similarly, Gimson (1962) reports the relationship between the /ɪ/ and /ə/ as asymmetrical, suggesting where /ɪ/ occurs it can be replaced by /ə/, but not vice versa. Therefore, whilst traditional accounts have tended to suggest a categorical and potentially contrastive difference between /ɪ/ and /ə/, they have also implied that there is a degree of variation in their use.

Whether unstressed /ɪ/ and /ə/ are distinct has also been observed to differ across varieties (Wells, 1982). In varieties with this distinction *‘a massi̲ve cloud’* and ‘*a mass o̲f cloud*’ will be pronounced differently (Wells, 1982), or *Leni̲n* and *Lenno̲n* word medially, but in other varieties these pairs will be homophones. Wells (1982) suggests that in Southern Hemisphere varieties there is no contrast available, in General American the distinction is possible but is often not made, but in the majority of British varieties there is clear opposition between these two vowels.

There have also been reports that suggest the use of /ə/ could be increasing at the expense of /ɪ/. In a study of 20 RP speakers, Gimson (1984) found that /ə/ was the majority form in many unstressed syllables, where previously /ɪ/ would have been expected e.g. in words ending in *–ity*, such as *uni̲ty.* This was based on auditory categorical judgements. This same trend is also reflected in how dictionary transcriptions have changed over time. Fabricius’s (2002) survey of pronouncing dictionaries published between 1917-2000 shows a marked move towards /ə/ variants.

Whilst such descriptions suggest that there is interesting variation between /ɪ/ and /ə/, these findings mainly come from categorical and impressionistic judgements, rather than more fine grained phonetic data. There has been very little research that quantifies this apparent difference between /ɪ/ and /ə/, although research on phonetic variation in schwa alludes to the fact that the distinction between these two vowels may not be as clear cut as such descriptions suggest. Research has consistently shown that schwa covers a large range on the backness dimension (Kondo, 1994; Bates, 1995; Flemming and Johnson, 2009), and /ɪ/ has also been suggested to be similarly variable in this way (Bates, 1995). In addition, non-final schwa has often been found to be realised as fairly high (Kondo, 1994; Flemming and Johnson, 2007; Bekker, 2014), occupying a similar height to /ɪ/. Such findings suggest that there could be a large degree of overlap between /ɪ/ and /ə/, and the difference between them may not be as robust as some of the descriptions suggest. Descriptions of variation in unstressed vowels have also implied that it is categorical; that speakers always aim for either /ɪ/ or /ə/. This idea implies that all other vowel differences are completely neutralised, and does not consider the idea that there could be any other intermediate variants. These ideas are based more on native speaker intuition than empirical evidence, and there is reason to suppose that there could be more fine-grained variation in unstressed vowels.

One study that has aimed to quantify variation between unstressed /ɪ/ and schwa in British English is Fabricius (2002). Fabricius uses interview data from 8 speakers of RP, and focuses on investigating claims of ‘KIT/schwa shift’, which is the hypothesis that schwa variants have increased at the expense of /ɪ/. The study focuses on the vowel in suffixes e.g. *horses*, *batted*, examining whether the /ə/ vowel is used at the expense of the more historically more common /ɪ/. Formant measurements from speakers’ -*es*/-*ed* tokens are compared with measurements of unambiguous KIT and schwa vowels e.g. *big*, *Oxford*. Fabricius finds that speakers’ formant values in suffixes are closer to the KIT than schwa values, which confirms that the shift has not taken place in suffixes for this group of speakers. However, only tokens of schwa that were close to the average values for /ɜ/ were used i.e. only mid central schwas. As will be seen, various studies have shown that (Kondo, 1994; Bates 1995; Flemming and Johnson, 2007; Lilley, 2012; Bekker, 2014) non-final schwa can and often is produced with a much higher realisation. Since the vowels in the suffixes which Fabricius investigated are word medial, if a schwa was used here it would be expected that it would often be realised as a fairly high vowel. Therefore the fact that speakers are shown to have -*es*/-*ed* values which were closer to the reference KIT than schwa vowels is not unexpected, as schwa is not consistently realised as a mid central vowel in this position. Because of the lack of an appropriate reference comparison it is hard to know how KIT- or schwa-like the participants’ vowels were. This highlights the importance of having an appropriate way to compare vowels in such cases.

Flemming and Johnson (2007) also compare different types of unstressed vowel, in American English. Although it was not the main focus of their study they compared the vowels in suffixes in words like *roses* with unambiguous schwas in words such as *suggest*. They found that there was no difference between these two sets of words. As they note, a difference would not be expected in American English. They note that empirical investigation of this variable in British varieties of English is needed.

As will be seen in the following sections, schwa quality has been reported by some to be variable, and even targetless. These claims are in a sense at odds with what is found in general descriptions of English, which describe /ə/ as being a quality distinct from others in unstressed syllables. There is therefore a need for empirical research to look into the detail of how these different vowel qualities are distributed, and to examine the phonetic reality of /ɪ/ and /ə/ in unstressed syllables. Given the reported variability of schwa, and also /ɪ/ in some cases, it is of interest to see to what extent this difference is actually phonetically realised, and in what ways.

2.3 Previous studies on targetlessness in schwa

There have been a number of studies which have looked at variability in schwa. This section reviews in detail the main studies which have focused specifically on the issue of targetlessness. This is in order to provide an overview of the variety of theoretical viewpoints on this issue, and the methods that have been used. I start with Browman and Goldstein (1992) who suggests that schwa has a target, and finish with Bates (1995), who suggests that schwa is fully targetless.

Browman and Goldstein (1992) - schwa has a neutral target

Browman and Goldstein (1992) collected articulatory X-ray data from pellets on the tongue dorsum during schwa production. Schwas were produced by one speaker of American English in nonce words in the frame /pV_1pəpV_2pə/, where the first schwa was the measured vowel. As the frame only contained bilabial consonants, it was expected that the articulation of schwa would only be influenced by the surrounding vowels. They therefore aimed to test whether there was an independent contribution of schwa or whether the articulation of schwa could be predicted according to the flanking vowels. In order to do this, they performed multiple linear regression analyses on schwa, with the articulatory positions of the adjacent vowels, and an independent factor for schwa as variables. They concluded that schwa has an independent target, due to the fact that the articulation of schwa was better predicted when an independent schwa factor was included in the model. This means that the articulation of schwa was not possible to explain simply as a function of the tongue movement between adjacent vowels.

In order to explore this idea further, they tried simulating schwa as a straight trajectory between each of the adjacent vowels, in order to see if this would lead to a schwa being perceived. They found that in most cases this process produced a schwa which was similar to the actual schwa that had been produced. For example, for the utterance /pipəpɑpə/, producing a vowel halfway between [ɑ] and [i] was perceived as a schwa. However, although this worked well in asymmetrical vocalic contexts, the problems came when the surrounding vowels were the same, particularly when they were high vowels. When the utterance /pipəpipə/was produced and the schwa was produced with the same quality as /i/ it was unsurprisingly heard as /i/. They therefore argued that schwa cannot be phonetically neutral or targetless as its articulation could not always be predicted by its surrounding context. Instead, they argued that it has a ‘neutral’ target in terms of its position in the vowel space, defined as an articulatory target that is the mean positon all of the other vowels in the vowel space. They base this claim on the fact that they found that the tongue position of schwa was very similar to the average tongue position of the full vowels that they measured.

This study thus provides some evidence that schwa may have an articulatory target, and is not completely assimilatory from context. However, these results should perhaps be interpreted with caution given that they come from only one speaker. In addition, Flemming (2009) points out a major caveat with the study in that the nonce words that they used could easily have been interpreted as two words rather than one (pV_1pə pV_2pə/), and so the schwa that the speaker was producing may have been equivalent to a word final, rather than non-final schwa. As will be explored further in section 2.6, some, including Flemming himself, argue that the mid central target is a property of word final schwa and not non-final schwa. Flemming suggests that this could be the reason an articulatory target was found in this experiment. This study therefore provides some evidence that schwa can have an articulatory target, but it does not quell the suggestion that non-final schwa could be targetless.

Kondo (1994) - “schwa is targeted in F1 but targetless in F2”

Kondo (1994) conducted an experimental study with three speakers of British English that manipulated the effects of both consonants and vowels on schwa. The preceding and following phonetic context was varied symmetrically between the consonants /p t k/ and the vowels /ɪ æ u/ around the indefinite article *a*, e.g. *pick a kitten*, *wrap a package*. The phonetic context was treated as a categorical variable according to the identity of the adjacent consonants and vowels, and the formant values of schwa were taken as the F1 and F2 values averaged over the whole of the segment. ANOVAS were used to assess the influence of phonetic context on the F1 and F2 of schwa, and the proportion of F1 and F2 variance in schwa that could be predicted from the phonetic context. It was found that both the consonant and vowel and their interaction had a significant effect on the F2 value for all 3 speakers. However, there were no phonetic context variables that had an effect on the F1 of all three speakers. When both the surrounding vowel and consonants were considered, the proportion of explained variance in the F2 of schwa was high (over 0.8 for all three speakers). The effect of the surrounding consonants, however, had much more of an effect on the F2 than the vowel. Kondo also examined both the F1 and F2 formant trajectories of schwa, and found that the F1 of schwa deviated towards a higher F1 value at its midpoint, therefore showing a movement away from its phonetic context. The F2 trajectories, by contrast, were found to be much more linear, with no deviation in the F2 trajectory away from the influence of the surrounding context.

Kondo uses these findings as evidence to argue that schwa has an F1 target but is targetless in F2. However, there are a number of problems with this conclusion, both with regards to the methods used to investigate the question of targetlessness, and the interpretation of the evidence. A large part of the argument for the F2 targetlessness of schwa rests on its apparent predictability by phonetic context. However, the F2 measurement was taken as the average over the whole of the schwa, rather than the midpoint. This means that it included measurements from the very start and end of the schwa, which will of course be affected by adjacent consonants. Measuring F2 in this way clearly maximises the probability that schwa will be found to be highly dependent on context. In addition, one of the pieces of evidence that is used to show that the F2 of schwa is affected by context is the fact that a high proportion of variance in the F2 of schwa can be explained by phonetic context. However, this was a highly controlled experiment, and the only thing that was varied was the phonetic context, so it is not surprising that variation in the F2 of schwa is predictable by phonetic context. All vowels are affected by their surrounding phonetic context to a degree; so in a controlled experiment like this it is likely that such a result could be found with any vowel. However, this particular measure is not compared with any other vowels so it cannot be argued that schwa is more predictable from context than other vowels.

Secondly, the conclusion that schwa is targeted in F1 but not F2 is flawed. Firstly, in practical terms it is hard to interpret since people cannot control these formants fully independently of each other. However, even if we take the conclusion to be that schwa has a height target but not a backness target, the methods used are not appropriate to show that. The findings show that phonetic context is a good predictor of F2 but not F1. However, this does not mean that only F2 is affected by phonetic context. It simply shows that the effect on F2 varies more across phonetic contexts than it does for F1. That phonetic context is not very predictive of F1 does not necessarily mean that F1 is unaffected; it may simply mean that the effects of context on F1 are more uniform across phonetic contexts (Ladefoged, 2014). As all of the adjacent consonants examined were stops, uniformity in the effect of consonants on F1 is expected.

Flemming (2009) - schwa is highly predictable from the context

Flemming's study on the phonetic variation of schwa is somewhat similar to Kondo's. Flemming varied the phonetic context of schwa within the nonce frame /$bV_1C_1əC_2V_2t$/, so that speakers produced nonce words such as /budəgit/. Flemming created a predictive model based on the assumption that schwa formant values move linearly between the surrounding consonants, with the surrounding consonants themselves being affected by adjacent vowels. Like Kondo, Flemming's study again reports high R^2 values for F2 (ranging from 0.73 to 0.86 for each speaker), but less high values for F1 (0.54-0.72). Flemming also shows that the F2 of schwa is a lot more variable than F1, with most schwas being fairly high, but values being found across the whole of the front/backness dimension of the vowel space.

Like Kondo, Flemming also finds differences between the dynamics of the F1 and F2 trajectories of schwa. Whilst linearity is reported in the F2 trajectories, deviation from linearity is found in the F1 trajectories. When schwa is between two high vowels the F1 of schwa raises during the segment. This result would not be found if schwa completely assimilated to its context, since this change involves schwa moving away from its phonetic context. This is similar to the finding by Browman and Goldstein (1992) that if schwa assimilates to the height of two neighbouring high vowels then it is not perceived as a schwa. As Flemming points out, even if schwa does not have a vowel quality target, it must still have a minimal height target, in order to be perceived as a vowel rather than as a consonant. Thus it would be expected that even if schwa were targetless, there should generally be at least a minimal rise in F1 in between two consonants. Nevertheless, Flemming's results suggest the height target to be more than the minimum requirement for a vowel as the F1 of schwa was found to be higher than for their neighbouring /i/ vowels. Flemming thus concludes that schwa is not a maximally high vowel.

Flemming makes a clear distinction between non-final and word final schwa. Word final schwa is identified as a less variable and lower vowel. The less variable nature of word final schwa is argued to be due to it being longer, and needing to contrast with other unstressed vowels, namely the happY vowel, in the same position. Flemming argues that where there is no pressure to contrast with other vowels then the pressure to minimise effort dominates, and that this explains the variability of non-final schwa. Since there is no motivation for speakers to contrast non-final schwa with any other particular vowel it is articulated with minimal effort, which means minimal movement between adjacent consonants.

Unlike Kondo, Flemming does not claim that these results demonstrate targetlessness in F2, but just that the F2 of schwa is highly predictable from context. Though this is true, again, without a comparison of this finding to other vowels, this does not tell us whether schwa is exceptionally predictable from context.

Bates (1995) - schwa is completely targetless and unspecified

Bates conducted an in depth analysis of the variability and predictability of the monophthongs of English in a single speaker of southern British English. Bates found relatively high variability for schwa in F1 and F2, as well as linear interpolation between neighbouring segments for both F1 and F2. This was argued to be evidence of targetlessness in both height and backness, with schwa argued to be an “empty time slot” (p. 37), and its phonetic realisation dependent on context. Unlike some of the aforementioned studies, Bates compared the measures used to other vowels, so it is possible to see the degree to which schwa is exceptional in this regard. Although schwa was found to be the most variable according to context there was also substantial variation between the other vowels in their contextual variability. /ɪ/ and /ʊ/ also had comparatively high variability to schwa in F2, as did /ɪ/ in F1. /ɪ/ is reported to be particularly similar to schwa in these measures. This leads Bates to conclude that both schwa and /ɪ/ are unspecified for tongue position. As Bates notes, there is indeed a general link between the contextual variability of the vowels, and how peripheral they are within the vowel space. It may therefore be that the high variability of schwa is related to its non-peripheral vowel quality rather than a unique phonological underspecification. The idea that both schwa and /ɪ/ are unspecified is particularly odd in that it poses the question of how it can actually be said that these vowels are different from each other at all. Barry (1998) also criticises Bates for classifying the consonantal environment based on the formant measurements at the onset and offset of the vowel. Barry argues that, since schwa is so short, it is therefore unsurprising that the onset and offset are highly predictive of its midpoint. This is also likely to be why /ɪ/ was found to be highly predictable according to its context, as it was the second shortest vowel that Bates observed.

2.3.1 Summary

This section has described in detail the main studies which have specifically examined the issue of targetlessness in schwa. All of the studies examined looked at elicited speech. This included data from read sentences (Bates, 1995), short phrases (Kondo, 1994), and nonce words (Browman and Goldstein, 1992; Flemming, 2009). Browman and Goldstein analysed articulatory measurements and the remaining three studies used formant measurements. Although there are some methodological differences between the studies, there is some similarity in the way that they assess targetlessness. All four studies examine schwa’s predictability according to phonetic context, as part of their assessment of targetlessness. However, due both to variation in the methods used and in the interpretations of the results, the conclusions from the studies are different. Browman and Goldstein argue that schwa has a neutral target, Flemming that schwa is highly predictable from context if not necessarily targetless, Kondo concludes that schwa is targeted in F1 and not F2, and Bates that schwa is completely targetless. The following section examines some of the general issues in assessing targetlessness, and explains how the work in this thesis builds on the aforementioned studies.

2.4 Issues in the investigation of targetlessness

The preceding section reviewed some of the main studies which look specifically at the issue of targetlessness in schwa. As was seen, however, they come to differing conclusions about the nature of the target of schwa. In the rest of this section I review some of the main issues involved in assessing targetlessness, and how these issues inform the way in which targetlessness and phonetic variation in schwa is examined within this thesis.

2.4.1 Type of speech used

Although they used a range of methodologies, there is an element of similarity between the studies reviewed in section 2.3 in that they all use elicited speech, making use of either nonce words or read speech. Browman and Goldstein (1992), Kondo (1994), and Flemming (2009) all conducted experimental studies where the words elicited were taken from a set of limited contexts. This, of course, allows for a high degree of control over the data, but may not be reflective of the way that’s schwas are produced in natural speech. With the case of nonce words in particular, the speakers’ interpretation of the target vowel may be ambiguous. As will be explored in more detail in section 2.6, the position in which schwa appears in a word can affect its phonetic realisation, so where the speaker’s interpretation of a word is uncertain this is a problem. Although Kondo used real words, the range of schwas elicited was also limited as they were all from the word '*a*', which is the only word containing schwa with a morpheme boundary on either side, so the degree to which the findings can be generalised to schwa in other word types is unknown. In addition, Lilley (2012), although not looking specifically at the issue of targetlessness, found that schwas in articles were slightly backer than other schwas. Similarly, Gick (2002) also found that one participant articulated schwas in lexical and function words differently. Both of these findings suggest that one should not simply assume that schwas from one particular word/category are representative of all schwas in general.

In addition, all of the studies described above rely on small numbers of speakers and data to reach their conclusions. There is a lack of research into the way schwa varies in spontaneous speech, particularly in non-final contexts. This thesis therefore makes use of corpora with large amounts of speakers and data. A range of variables that naturally occur in spontaneous speech are used in order to better understand phonetic variation in schwa, and how this relates to the claim of targetlessness.

2.4.2 Comparison with other vowels

A targetless vowel is one that not does not have an articulatory target of its own, beyond what is easiest to produce in its phonetic environment. A targetless vowel is therefore one that maximally assimilates to context. However, we know that all vowels assimilate to context to a degree. For example, Moon and Lindblom (1994) show that, given the right conditions, any stressed vowel will also show assimilation to its environment. In order to show that targetlessness or extreme assimilation to context is a property unique to schwa, it is vital to compare it to other vowels. The same measures that show high variability or assimilation in schwa must be able to also show that other vowels are *less* variable. As noted, in some of the measures that Kondo and Flemming use to show high variability in schwa there is no comparison with other vowels. Finding that a high proportion of variance in schwa can be explained by phonetic context, for example, is meaningless if that finding is not compared with other vowels.

Comparison with other vowels also serves as a way of making sure that the methods used are suitable for finding evidence of targetlessness. If the methods used do not clearly show a vowel quality target in other vowels then it is likely that they are not a suitable method of analysis. For example, if the study is conducted such that linear formant trajectories are used as evidence of targetlessness in schwa, then other vowels should be expected to show clear deviation in formant trajectories towards a target. In particular, comparing schwa to other vowels can shed light on what properties of schwa may be unique and which may be properties that are shared with other similar vowels. Bates (1995) found, for example, that other non-peripheral vowels were also similar to schwa in terms of variability. To address this issue, Chapter 6 of this thesis compares schwa to the central and lowered KIT vowel in NZE, which is a stressed vowel that is similar to schwa in terms of its in quality, and therefore serves as an ideal baseline for comparison.

For the above reasons, therefore, an important part of the analysis in this thesis will be the inclusion of a range of other vowels for comparison. This is important to a) make sure any claims about the high variability / context dependency of schwa are unique to schwa, and b) to check that the methods used to test for a target are suitable methods for doing so, and can provide evidence for a target in vowels known to have one.

2.4.3 Maximal assimilation has different predictions for F1 and F2

A possible motivation for targetlessness could be for the vowel to be produced with the least effort possible (Flemming, 2009). It would thus be expected not to deviate substantially from the surrounding phonetic context. The predictions for a ‘least effort’ vowel without a unique vowel target are, however, quite different for F1 and F2. Some of the methods used in the studies discussed above are problematic in that they appear to analyse F2 and F1 in the same way. F2 is expected to be highly variable in a targetless vowel because of assimilation to the surrounding contexts. It is also expected that this variation is predictable according to the phonetic context. Indeed this was found and provided as evidence for targetlessness in F2 in Kondo (1994) and in Bates (1995). However, although there are exceptions (Lindblom, 2002), on the whole surrounding consonants are expected to uniformly lower F1. This means that the expectations for F1 in a targetless vowel are different from those for F2 in two ways. Firstly, if schwa is targetless, great variability in F1 is not actually expected, since in the majority of phonetic contexts producing a high vowel should be less effort than producing a low vowel. Indeed, as lower vowels are generally longer (e.g. Lehiste, 1970) this suggests that they require more effort to produce. Therefore, very little F1 variation is expected in a targetless vowel. It further follows that F1 is also expected to be less predictable than F2 according to phonetic context. This means that F2 being less contextually dependent than F1 is not good evidence for schwa being targeted in height and not backness (e.g. Kondo, 1994).

If a targetless schwa is expected to be a high vowel that is very variable in backness it follows that a targeted schwa would behave differently. Many of the phonetic studies described above find that schwa is indeed often fairly high, and variable in backness (Kondo, 1994; Bates, 1995; Flemming and Johnson, 2007; Flemming, 2009). However, the existence of such productions does not necessarily indicate targetlessness. Such productions could also be used in environments where the conditions encourage increased assimilation to context. One example of such an environment is where vowels are of very short durations. Consequently, there is a need to distinguish between where such productions are caused by other factors, and where this maximal assimilation is caused by lack of a phonetic target. If a vowel has a phonetic target, there will be some situations where the influence of phonetic context is less and so the vowel moves towards the target. Thus if schwa has a non-high target then it is expected that in environments where it is less assimilated to context it would be produced as a lower vowel. It therefore follows that a targeted schwa would be more variable in F1 than a targetless schwa, as the vowel height would vary between productions that are more assimilated to context at longer durations, and less assimilated to context at shorter durations. A targetless schwa on the other hand should mostly be realised as high. Therefore, although Bates argues that high variability in the F1 of schwa indicates it has no F1 target, in fact this high variability points more towards it having a height target.

In spite of the fact that the predicted effects of being targetless on F2 and F1 are different, this does not mean that these two dimensions should be considered separately, as the above studies generally seem to imply. Both measurements need also to be considered together in terms of the positioning of schwa over the whole vowel space to circumvent problematic conclusions such as ‘schwa is targeted in F1, but targetless in F2’. In light of the different expectations for F1 and F2 then it is expected that if there is a non-high schwa target then lower schwa productions are also likely to be less variable in F2. Where there is more assimilation to context, higher productions that are more variable in F2 are expected. Lilley (2012) found this type of variation in a study of American English schwa. Lilley found that there is more variability in the F2 of higher schwas. This finding confirms that the link between F2 and F1 is important to consider. Indeed, the majority of speech sounds are perceived using a combination of multiple acoustic cues, not just one cue in isolation (Holt and Lotto, 2010, Kirby, 2010). In terms of how vowels are organised and heard within the vowel space, it is the combination of height/F1 and backness/F2 which allows vowels to contrast with each other.

2.5 Finding a schwa target

If schwa is completely assimilatory to context it is predicted to show great variability according to phonetic context. However, this does not mean that variability or predictability according to context necessarily implies targetlessness. It is possible for a vowel to have a phonetic target which it does not often reach. As a degree of variability is expected, rather than looking for evidence of targetlessness, it makes more sense to look for signs of independence in schwa (Barry, 1998). What is needed is to examine circumstances where schwa may vary if it has a target. By using the wealth of data available in corpora, this can be done by making use of the natural variation in spontaneous speech. This thesis will do this by making use of three main variables: formant trajectories, vowel duration, and the birth year of speakers. The way that these different variables can be used to examine the issue of targetlessness is explained in the following section.

2.5.1 Formant trajectories

Vowel formant trajectories have often been analysed in search of a target (Kondo, 1994, Bates, 1995, Flemming, 2009), the idea being that if schwa has a target, some deviation in the trajectory should be found, as opposed to a linear movement from the preceding to following context. With regards to F1, in most cases this will mean F1 rising up to the midpoint of the segment, before falling again. However, as Flemming points out, even if schwa does not have a unique phonetic vowel quality target, it must still have a minimal height target in order to be perceived as a vowel rather than as a consonant. Thus, it would be expected that even if schwa was targetless, there should generally be at least a minimal deviation to a higher F1 in between two consonants. Therefore, in order to provide evidence of a phonetic target, it must be shown to move to a height lower than the minimum required for a vowel.

2.5.2 Vowel duration

Whilst variation across the vowel trajectory shows variation within individual tokens, examining variation across schwas of different durations involves comparison across tokens. This is important as it may be that evidence of targetedness cannot be found across all schwas on average, but only in certain conditions; in this case in schwas of longer durations. This hypothesis stems from Lindblom's (1963) undershoot theory. The phenomenon of vowel undershoot (Lindblom, 1963) occurs when, as the duration of a vowel decreases, the articulators have less time to reach their target position. This results in the articulatory and acoustic properties of shorter vowels being further away from their targets, as Lindblom demonstrates in Swedish. The same phenomenon is also demonstrated with a range of English vowels in Moon and Lindblom (1994). Although the effects of short duration can be overcome (Lindblom et al, 1990) it becomes increasingly difficult to produce a high F1 as the duration decreases (Flemming, 2004). As the data used in this thesis is from conversational speech, many of the schwa tokens measured are even shorter than the vowels from the studies above. This thus increases the likelihood of formant undershoot, if schwa has a target. It is also possible that schwa is of such short duration that on average there may not be any deviation in the trajectory, which is why examining variation across schwa tokens may help identify evidence of a target. Variation in schwa according to duration has been examined in Barry (1998), who examined variability in German schwa. The methodology was quite different from the one employed in this thesis as an experiment was used where speakers were asked to deliberately modify their speaking rate. However, the results are broadly in line with what would be expected if schwa has a target, in that in the condition where speakers were asked to speak slowly they produced slightly lower schwas that varied less in F2. This thesis, on the other hand, explores how schwa productions vary according to the naturally varying durations of vowels in spontaneous speech.

2.5.3 Change over time

Unstressed vowels have generally been neglected in research on language change. It is routine for vowels without full stress to be excluded from analysis in variationist and language change research (e.g. in Labov, 1994). However, examination of how schwa varies over time can shed light on some of the debates about the nature of schwa.

If schwa is completely predictable from phonetic context then it follows that the only way its realisation could change is if there were changes in its phonetic context. Such changes could occur if there were random changes in the words schwa occurs in. Of course, random changes of this type would only change the average position of schwa, not its target. The analysis in this thesis, however, takes phonetic context into account, such that even if there are random changes in the phonetic contexts that schwa finds itself in, this would be controlled in the results. In addition, any random changes in the words schwa appears in are be taken into account within random effects in the mixed effects regression models used in this thesis (which means that words that change a lot in terms of their frequency cannot skew the results). If the phonetic value of schwa is only affected by its immediate phonetic context then changes in the production of other individual sounds are unlikely to affect its overall value in any noticeable way. If there is a genuine change in schwa then its effect will appear even after the phonetic context and random fluctuations in word usage are controlled for.

The hypothesis that schwa is targetless therefore does not accord with schwa changing over time in any unified way. If its realisation is purely determined by context then there is no inherent target which could change. Therefore, regardless of the changes in other individual vowels, it should remain unaffected. As a consequence, we would not expect it to change in response to changes in other vowels or take part, for example, in chain shifts.

Bates (1995) argument that schwa is targetless is extended to contexts where schwa tends to be lower, such as when it is utterance final, or at longer durations. Bates suggests that the more mid position occupied by pre-pausal schwa is still phonetically neutral as it is caused by the tongue moving to its natural rest position at the end of an utterance. Importantly, these ideas would also not predict change in schwa over time, as any analysis of schwa that describes its realisation purely in terms of ease of articulation is not compatible with it changing. Importantly, Bates's ideas about word final schwa mean that this would also not be predicted to change over time, since the natural rest position of the tongue would not be expected to change over time.

Browman and Goldstein's (1992) main claim is that schwa is not targetless, but is still neutral, in that it is articulated with a tongue position which is the same as the average tongue position of all other vowels. In this analysis the articulation of schwa is not predictable from its phonetic context, but it is still predictable from all of the other vowels in the language. This type of target could indeed change. However, as it is not a completely independent target, it would only change in response to changes in other vowels in the language. This analysis implies that schwa could change phonetically but remain neutral in terms of the overall vowel system, if other vowels in the vowel space also change. This means that schwa would only change in specific circumstances. If the overall vowel space of the language changes, such that the mean position of all vowels changes, then schwa would in turn be expected to change. However, this hypothesis does not imply that schwa would change in response to individual vowels if the shape of the vowel space in general was not changing.

Flemming's (2009) argument that schwa can be influenced by the pressure to maintain contrasts with other vowels allows for the potential of schwa changing over time. Flemming, however, seems to suggest that there would only ever be pressure to maintain contrasts with other unstressed vowels. This implies that there would only be change in schwa if there was change in other contrasting unstressed vowels, meaning that schwa should be unaffected by any changes within stressed vowels. This argument only makes sense if it is the case that stressed and unstressed syllables are completely distinguishable based on factors other than vowel quality. If vowel quality plays any part at all in distinguishing schwa from different stressed vowels, then there is no reason why schwa should only be influenced by the position of unstressed vowels. Indeed, Cohen Priva and Strand (2020) take the view that it may be important for schwa to contrast with other vowels in quality in order to indicate that it is an unstressed vowel. If this is the case we may well expect that schwa could show change over time, in response to changes in other vowels.

The key point is that any account of schwa that claims that it lacks an independent target, or that its realisation is purely motivated by ease of articulation, is essentially incompatible with the idea that schwa could change over time. Therefore, examining whether schwa changes over time is a good opportunity to test these claims about schwa. If schwa was found to undergo change this would be evidence against it being phonetically targetless, and would suggest that its realisation is not purely motivated by ease of articulation.

Unstressed vowels have, however, been oft neglected in theories of sound change. For example, Labov's (1994) theory of vowel change describes vowels as operating separately within two subsystems, tense vowels and lax vowels. The central idea is that chain shifts in vowels only take place between vowels from the same subsystem, meaning that lax vowels only change in response to other lax vowels, and tense vowels in response to other tense vowels. Tense vowels are said to raise along a peripheral track, and lax vowels to fall along a non-peripheral track. Such an idea that vowels move strictly within these subsystems suggests that unstressed vowels would be unlikely to take part in the same shifts as stressed vowels. The idea that vowels only change within strict subsystems bears some resemblance to Flemming's implication that schwa would only change in response to other unstressed vowels. Schwa is not mentioned at all Labov's theory so it is not clear under what circumstances it would be expected to change, or whether it is simply not expected to change at all.

There has, however, been very little empirical research into change in schwa over time. Non-final schwa in particular has been neglected in variationist research. Whilst there has been research relating to non-final schwa, most of it has specifically focused on categorical variation between it and other vowel qualities. For example, studies have looked at the alternation between /ɪ/ and schwa in unstressed syllables (e.g. Fabricius, 2002), and the alternation between schwa and full vowels (e.g. Mesthrie, 2017). However, research focusing on changes and variation within schwa itself is sparse. This thesis therefore fills an important gap.

Leach (2018) is a rare example of a study which examines gradient variation in unstressed vowels. Leach does not look at schwa in particular, but examines the variation in unstressed vowels in words such as *private* in Stoke-on-Trent in the UK, where the local accent can have quite a high front realisation. The study examines whether there has been any change over time in the realisation of these vowels, although it finds no evidence of change. The realisation of these unstressed vowels is found to be much more constrained by linguistic variables such as phonetic context, rather than social variables such as gender and topic. Leach argues that this is due to this particular variable having quite a low level of social awareness. This may be a reason why unstressed vowels are often neglected in research more generally, as variation in their realisation may not be as noticeable. Nevertheless, the question of how unstressed vowels vary and interact with other vowels is an important one.

There has been more focus on variation in word final schwa. Whilst few studies look specifically at changes in these vowels, various realisations have been reported in different varieties e.g. a low [ɐ] in Tyneside (Wells, 1982; Watt and Allen, 2003) and London (Tollfree, 1999), and low and back [ɒ] in Sheffield and Manchester (Beal, 2008). To the extent that these different transcriptions represent genuine differences between varieties, if word final schwa can vary in this way we might also expect to find evidence of it changing. Kiesling (2005) is a rare example of research which looks specifically at changes in word final schwa. Kiesling finds possible lowering and backing of word final schwa over time in the English of Australians of Greek descent. The extent of lowering and backing is also linked to the length of the segment with longer segments being lower and backer. Indeed, duration is a factor that will be also examined in the analysis in this chapter.

Ramsammy and Turton (2012) suggest that the realisation of schwa could be linked to other vowels. They examine the realisation of word final schwa in Manchester, which is stereotyped to have a low back schwa (Beal, 2008). They quantify how low and back both schwa and happY are using the Euclidean distance between F1 and F2. Where speakers have a large difference between F1 and F2 this indicates the vowel is relatively high and front, and where the Euclidean distance is low this indicates a lower and backer vowel. They find a correlation between individual speakers' values for this measurement between happY and schwa. That is, speakers with a low/backed schwa also tend to have a low/backed happY vowel, suggesting a possible link between the two vowels. They suggest that this pattern may indicate a chain shift between the two vowels, although as they do not provide any age or time related data there is no evidence specifically for change. Nonetheless the idea that the realisation of schwa could be related to the realisation of other vowels is an important one, and one that will be explored in this thesis.

2.6 Word position differences in schwa

Much of the research discussed so far has treated schwa as if it is one phonological unit. This is in line with the way in which schwa is transcribed, with both word final and non-final schwa usually being denoted by [ə]. Many of the studies described above (e.g. Kondo, 1994) also simply refer to schwa and do not make a division between word final and non-final schwa. However, descriptions of English pronunciation have long noted schwa in word final position (e.g. *comma*) to have a lower quality than in non-final environments (e.g. *mature, breakfast*) (Jones, 1914; Gimson, 1962; Lass, 1986). More recent instrumental studies have confirmed this difference, making a division between canonical mid central schwa word finally, and non-final schwa, which tends to be higher and more variable in backness (Flemming and Johnson, 2007; Lilley, 2012, Bekker, 2014).

Flemming and Johnson (2007) examined the difference between word final and non-final schwa spoken by 12 speakers of American English, in read sentences. They found that non-final schwa is higher and more variable in F2 than word final schwa. In their vowel plots non-final schwa is shown to exhibit an average height that is roughly in between /i/ and /ɪ/ and spans the entirety of the front/back dimension of the vowel space. By contrast, they noted that word final schwa is a mid vowel on average, and is less variable in F2. They found that this pattern extends to stem final schwas which are not word final (e.g. *Rosa's*). It should be noted, though, that around this mid value, there was a wide variation in the height of their stem final schwas, with some being as low as /ɑ/ and some as high as /ɪ/. Nevertheless, they argued that this overall difference between non-final and word final schwa shows that there is a fundamental distinction between them.

Other studies since have reported similar findings. Bekker (2014) examined the same issues in South African English, using acoustic measurements of vowels from 27 speakers reading word lists. Bekker also found that word final schwas were lower than non-final schwas, transcribing them as [ɜ] and [ɘ] respectively. Bekker suggests that this high vowel quality in non-final position has been phonologised, based on low standard deviations in the formant values of non-final schwa. Lilley (2012) examined schwa realisation in American English, predicting the variability of schwa using a range of variables in hidden Markov models. They found, like Flemming and Johnson, that there is a division between word final and non-final schwa, with word final schwa being lower and backer. However, unlike Flemming and Johnson, they found that this is just a word position effect rather than a stem position effect. This means that when other variables are considered, schwa in words like *Rosa's* actually patterns with non–final rather than word final schwa.

A number of explanations have been proposed for the difference between word final and non-final schwa. Flemming and Johnson (2007) suggest that this difference may be due to the fact that word final schwa contrasts with other unstressed vowel qualities: / əʊ/ as in *motto,* and /i/, as in *city*, whereas in non-final position it does not contrast with any other unstressed vowel qualities. Their argument is that shorter higher vowels require less effort than longer lower vowels, so speakers will produce a higher schwa where there is no need to contrast the vowel with /i/, but will produce a lower vowel when there is, in order to keep these vowels distinct. This hypothesis is based on Lindblom's hyper- and hypoarticulation (1990) theory that speech is a balance between economising on articulatory effort (hypoarticulation) and the need to be understood/maintain contrasts between sounds (hyperarticulation). Flemming and Johnson's study is of American English, where they claim that, in most accents, there is no distinction between /ə/ and /ɪ/ in unstressed syllables. In most varieties of British English, however, this is not the case (Wells, 1982) e.g. the minimal pair *Lennon* and *Lenin*. Therefore such an explanation would not make sense for most varieties of British English, as here schwa it is claimed to contrast with another unstressed vowels (eg. Fabricius, 2002).

Furthermore, although in many varieties of English non-final schwa may not contrast with other unstressed vowel qualities, it still needs to be distinct from other non-final stressed vowels. Flemming's (2009) claim is that non-final schwa can be realised as short and variable in quality since it does not contrast with any other unstressed vowels. This would only make sense if duration and vowel quality themselves make no contribution to the perception of stress. However, this has not been found to be the case (e.g. Fry, 1955; Klatt, 1976; Zhang and Francis, 2010). For example, Fry (1955) shows that modulations in duration alone can cause listeners perception of stress placement to change, so it is clear that duration is used to help mark stress within a word. Zhang and Francis (2010) also show that both vowel quality and duration influence how speakers perceive lexical stress. Given that these cues contribute in themselves to stress, it is somewhat circular to say that schwa can be produced with a lack of effort because it is unstressed.

Bates (1995) argues that non-final and word final schwa can both be regarded as targetless vowels and only have different realisations because of the differing manifestations of coarticulation in the environments which they occur in. Whereas non-final schwa is coarticulated with the sounds around it, word final schwa, when before a pause, can be seen as coarticulating with the pause that follows it and thus moving to a rest position. However, some studies have cast doubt on the idea that schwa is equivalent to the rest position. For example, Gick (2002) compares the articulatory position of word final schwa with the rest position for 4 speakers of American English, using X -ay data. Gick suggests that schwa and rest positions are not the same thing due to schwa having a more retracted tongue root.

Another explanation is that the differences between word final and non-final schwa are simply due to the phonetic side effects of non-final schwas being shorter in duration and because they have a following phonetic environment, since both of these factors are liable to increase the influence of coarticulation with surrounding segments. It could be that both types of schwa have the same target, but because of these factors they simply surface differently. The duration explanation alone cannot provide the full answer as to why any such differences may exist, as it would still require an explanation of why word final schwa is longer than non-final schwa. However, this difference may not be something which is unique to schwa but stems from the more general process of domain final lengthening (e.g. Rakerd et al, 1987), where vowels are longer before a syntactic boundary or pause. Where some studies have looked only at words in isolation (e.g. Bekker, 2014), these differences between final and non-final schwa may have been exaggerated due to phrase final lengthening. Although word final schwa will not always occur before a pause, word-specific effects could also cause these effects to remain somewhat even where word final schwa occurs before another word (e.g., as shown in Sóskuthy et al, 2018)

2.7 Summary and contribution of the thesis

Schwa has been demonstrated to be a highly variable vowel, with some arguing that it has no inherent target of its own, and that its phonetic realisation is fully dependent on its surrounding context. Phonetic differences between word final and non-final schwa have also been found, with some arguing that they are fundamentally distinct from each other. However, there is a lack of research which explores these issues in spontaneous speech, or considers change over time. Despite the reported variability in schwa, there are also reports that it is distinct from /ɪ/ in unstressed syllables, although most accounts of this variation are impressionistic.

In this thesis I conduct an acoustic analysis of schwa in two different varieties of English using large amounts of corpus data. I thus redress the issue of the lack of analyses of schwa in spontaneous speech. Rather than using highly controlled data, as in the studies discussed in section 2.3, I am thus able to make use of the natural variation in speech in order to learn about the behaviour of schwa. Like the aforementioned studies in section 2.3 I examine the predictability of schwa according to phonetic context. However, for the reasons stated in section 2.5 (p. 38), in terms of the assessment of targetelessness there is more of a focus on looking for positive evidence of a target through the way that schwa varies according to a number of variables. As described in section 2.5 I examine how schwa varies across its formant trajectory, according to its duration, and also how it changes over time. Following the research discussed in section 2.6, I also consider both word final and non-final schwa in the analyses, in order to see whether they pattern in the same way as each other. In addition to an in depth examination of the realisation of schwa, the thesis also explores how the difference between schwa and unstressed KIT is phonetically realised, thus building on the more impressionistic literature which has looked at this distinction in the past.

The thesis includes three main analyses of the behaviour of schwa. In Chapter 3 I look specifically at the difference between unstressed KIT and schwa in Derby in England. The analysis explores to what degree there is a clear a consistent difference between unstressed KIT and schwa, and looks at how the difference is phonetically realised. Chapter 4 offers an in depth look at variability in schwa in Derby. I examine its variability and predictability, and also how it varies according to its formant trajectory and its duration. An important part of this chapter is the comparison to a range of vowels, which as noted in section 2.4.2 is crucial to assess the meaning of any patterns found in schwa. Chapter 6 also looks at the issue of targetlessness by analysing the realisation of schwa in New Zealand English. Again, this chapter explores the variability and predictability of schwa, and how it varies according to its duration. In addition, this chapter provides another angle by looking at change in schwa over time, and comparing it to changes in the KIT vowel over time. Examining whether schwa changes over time is yet another way in which its potential targetless status can be assessed (section 2.5.3) and also addresses the dearth of research into changes in unstressed vowels. As the NZE analysis contains a large amount of tokens it was necessary to use automated measurement. Because unstressed vowels are rarely included in such analyses Chapter 5 provides a methodological contribution by assessing how automated measurement can be used effectively when dealing with unstressed vowels. Altogether, the findings from the various chapters contribute to a better understanding of how schwa is produced in natural speech, and in particular whether it has a phonetic target.

3 Variation between /ə/ and /ɪ/

3.1 Introduction

This chapter examines the distinction between /ə/ and unstressed /ɪ/. This chapter primarily addresses the third main research question (p. 22) which asks whether schwa is distinct from other vowels in unstressed syllables. This question is addressed through the analysis of unstressed vowels from 26 speakers from Derby in the UK. Most varieties of British English are traditionally described as having one of three possible vowels in unstressed syllables: /ɪ/, /ʊ/ and /ə/ (Gimson, 1962; Roach, 1983). However, the vast majority of the commentary on variation between unstressed vowel qualities refers to /ɪ/ and /ə/, suggesting that this is where the main locus of variation is. The first goal of this chapter is therefore to ascertain the extent to which there exists two clearly separate /ɪ/ and /ə/ categories in unstressed syllables.

In British English, other than /ɪ/, /ʊ/ and /ə/ (Gimson, 1962; Roach, 1983), vowel distinctions are said to be neutralised (Flemming and Johnson, 2007). Such descriptions imply categorical variation between discrete categories. However, as described in section 2.2, most descriptions of the quality of different unstressed vowels have tended to be based on auditory and impressionistic judgements (e.g. Gimson, 1984). There has, however, been very little research investigating the empirical reality of these different vowel qualities and how they are distributed in the speech of individuals. The work in this chapter thus addresses this gap by providing an in depth analysis of the actual phonetic realisation of these vowels. The second goal of the chapter is therefore to investigate whether there is complete neutralisation to these two vowel qualities or whether there are any additional vowel quality differences.

There is a slight conflict between the impressionistic literature which transcribes and suggests a clear and categorical difference between /ɪ/ and /ə/ (Gimson, 1962; Roach, 1983), and the more instrumental literature on non-final schwa which has tended to suggest that schwa is a very variable, if not targetless, vowel (Kondo, 1994; Bates, 1995; Flemming, 2009). Given that schwa has been suggested to be highly variable (Flemming, 2009; Bates, 1995) over the whole of the vowel space, and additionally /ɪ/ has been said to be very variable across the F2 dimension (Bates, 1995), there is every reason to suspect that the relationship between these different vowel qualities may not be as straightforward as implied in standard descriptions of unstressed vowels. The idea that schwa is a targetless vowel (cf. Bates, 1995) does not suggest they there would be a clear distinction between schwa and other unstressed vowel qualities. Exploring the distinction between different unstressed vowel qualities thus also feeds into debates about whether schwa is targetless.

The findings in this chapter are also important from a methodological point of view, as the results contribute to decisions about which tokens should and should not be included in the schwa analyses in Chapters 4-6. The difference between /ɪ/ and /ə/ is examined by comparison of unstressed vowels with different spellings. The reasons why the variable of spelling is used are explained in section 3.4.

The chapter starts by explaining the history of the difference between /ɪ/ and /ə/ and how their use has changed over time (3.2). There is then a discussion of the particular difficulties with categorising unstressed vowels into different vowel categories (3.3). The chapter then moves on to explain the reasons why the variable of spelling is used for these means (3.4). This is proceeded by an explanation of the methodology used in the analysis of the Derby Corpus, (3.5) followed by a description of the research questions and predictions for this chapter (3.6). This is followed by the analysis of the data (3.7) and finally a discussion of the results (3.8).

3.2 Historical background

There is a long and complex relationship between /ɪ/ and /ə/ in unstressed syllables. The process of the reduction of full vowels in unstressed syllables goes back to the 1400s, with variation between /ɪ/ and /ə/ existing in some varieties since then (Beal, 1999; Lass, 1999). However, Lass (1999) suggests that at around the time of the 15-1600s there may not have actually been one unified schwa vowel but simply weaker allophones of the equivalent stressed vowels.

By the 1700s the use of fully reduced variants appears to be much more common (Lass, 1999; Beal, 1999). It is clear from Beal's review of the work of dictionary writers, however, that phonological reduction was still very much a change in progress. For example, the work of some writers in this period suggests that it is still possible to have an /ɛ/ vowel in the *–es* suffix, which would not be possible today. Although today the use of schwa in unstressed syllables is generally unremarkable, according to Beal, in the 1700's, it is not without comment. For example, Beal notes that the writer Sheridan was criticised by his contemporaries for suggesting that the vowels in *culpable* and *tavern* can be pronounced the same. These kind of comments suggest that the use of reduced vowels was still somewhat salient, and also implies that full vowels also occurred in the same words.

The process of vowel reduction is now phonologically complete in certain contexts, in that there are words, such as *about*, where only the reduced vowel /ə/ is possible. However, variation between /ɪ/ and /ə/ in unstressed syllables is still present, with both pronunciations considered acceptable in many words (Wells, 1990) e.g. *before*, *remember*. In addition, it has been suggested that there is a still an ongoing change away from /ɪ/ towards /ə/, known as KIT/schwa shift (Gimson, 1984; Fabricius, 2002). For example, Jones (1956) reports that the use of /ə/ is increasing. Gimson (1984) also suggests that the use of /ə/ may have increased at the expense of /ɪ/, basing this conclusion on auditory judgments of the speech of 20 RP speakers. Similarly Fabricius' (2002) survey of pronouncing dictionaries 1917-2000 shows a marked move towards /ə/ variants.

3.3 Difficulties of /ɪ/ and /ə/ categorisation

For several reasons, observing the distinction between /ɪ/ and /ə/ is difficult. As seen in section 3.2, the use of /ɪ/ and /ə/ has historically been very variable. Furthermore, reports that there could be an ongoing change from /ɪ/ to /ə/ also complicate matters. In addition, pronunciation dictionaries show that it is normal, even within the same word, for either vowel quality to be used. Furthermore, although standard descriptions imply that there is categorical variation between the two vowel qualities it is possible that the variation between them could be gradient. Indeed, Ladefoged and Johnson (2014) note that the use of the symbols /ɪ/, /ə/ or /ʊ/ could actually mask the fact that intermediate variants may be used. In addition, the relatively high variability reported for both /ɪ/ and /ə/ (e.g. Bates, 1995) suggests there could be a high degree of overlap between the two vowel qualities.

Clearly in this case automatically ascribing individual lexical items to an /ɪ/ or /ə/ category based on auditory judgement or native speaker intuition is not appropriate. Nor would it be appropriate to decide *a priori* that certain phonetic realisations are attributable to either category. Instead, the novel approach of examining unstressed vowel realisation through the lens of different spellings is used. This approach was chosen for several reasons which are laid out in the following section.

3.4 The relationship between spelling and unstressed vowel quality

There appears to be a link between spelling and expected vowel quality in unstressed syllables. Roach (1983) suggests that /ɪ/ in unstressed syllables is most often represented by the spellings <e> and <i>. Similarly, Jones (1956) reports that /ə/ tends to be spelt with most letters with the exception of <i>. The same pattern is also found in dictionaries (e.g. Wells, 1990); <i> tends to represents /ɪ/, <a>, <o> and <u> tend to represent /ə/, and <e> often represents both. Historically, this also seems to be the case. Beal's (1999) review of writers in the 1700s shows that in this time period there were also intuitions about the relationship between spelling and vowel reduction. For example, from the writers Beal reviews it seems that <u>, <o> and <a> were more likely to be seen as having schwa-like pronunciations than <e> or <i>. Crucially, all of these descriptions are based on auditory judgements rather than instrumental analyses but they provide a sense of native speaker instincts about when these different vowels occur.

Ladefoged and Johnson (2014) suggest that it is likely that reduced vowels are historically derived from full vowels. They exemplify this with the fact that are many reduced vowels with non- reduced cognates e.g. *political* vs *politics*. Where this is the case it of course means the spelling will often provide a link to the historical full vowel quality in unstressed syllables. This link can also be seen in words where there is present day variation between schwa and a full vowel e.g. *direction*, where both [aɪ] and [ə] are possible pronunciations. In stressed vowels there is clearly a link between spelling and vowel quality. It therefore makes sense that, if reduced vowels are historically derived from full vowels, that there could also be a link between spelling and vowel quality in unstressed vowels.

Van Bergem (1993) and Bates (1995) both make a distinction between phonetic and phonological vowel reduction (referred to as lexical and acoustic reduction by Van Bergem). The process of phonetic vowel reduction is where vowels are produced with less extreme and more coarticulated articulations in situations with less stress and /or shorter durations. This is a process which affects all vowels to some degree (e.g. Lindblom, 1963; Bates, 1995; Proctor et al, 2015). Phonological vowel reduction, on the other hand, is when a reduced vowel becomes the target vowel. For example in the word *around* there is phonological reduction as only a schwa is possible. We can assume that phonologically reduced vowels are historical derived from full vowels, likely through the phonologisation of phonetic reduction. The spelling of the vowel therefore provides a clue as to what the historical full vowel may have been. It is possible that there may not have been complete neutralisation of all full vowel qualities, and minor differences based on the historical full vowels could still remain.

This chapter therefore uses the variable of spelling to investigate vowel quality differences in unstressed vowels. Although this approach is not perfect, it is useful for two main reasons, as seen above. Firstly there is a clear relationship between the spelling of an unstressed vowel, and native speaker instincts about its vowel quality. Secondly the spelling of an unstressed vowel in general is related to its historical full vowel pronunciation. Looking at unstressed vowel through the lens of spelling thus allows us examine the phonetic reality of unstressed vowel differences using a variables which seems to be meaningful for speakers now and is also meaningful historically. The fact that spelling ties into the historical pronunciation of the vowel also allows us to examine to what extent full vowel quality differences were fully neutralised in unstressed syllables.

As far as I am aware Lilley (2012) is the only study to specifically look at spelling differences in unstressed vowels. Lilley uses hidden Markov models to assess the effects of a number of predictors on the F1 and F2 of unstressed vowels in American English. Unstressed vowels were grouped in two ways, whether they were relatively low, mid or high; and whether they were relatively front, central or back. Spelling was found to be a significant predictor, with spellings that represent low or back vowels (i.e. <a>, <o>and <u>) being more likely to be classed as low and back. However, a major problem with this study is that it includes all unstressed vowels, including those in words where a pronunciation with a full vowel may also be possible. This is a problem as there a words such as *advice* where either a phonological reduced or full pronunciation is possible. In such words, even if a vowel is produced which sounds reduced it cannot be known whether the speaker was aiming to produce the reduced vowel or whether they were in fact producing a phonetically reduced variant of the full vowel. This is a problem because in cases where speakers are simply producing a phonetically reduced version of a full vowel, it is expected that their pronunciation will be influenced by the full vowel that are aiming to produce, which of course is related to spelling. Therefore, an important part of the work in this chapter is that only tokens where only a reduced vowel is possible are included in order to prevent this problem. This means that where there are spelling differences these are genuine differences within unstressed vowels.

In this research the variety being examined is that spoken in Derby. Derby is a city in the north of England. Like most British dialects we would expect there to be an /ɪ/ and /ə/ distinction available in Derby. There is little known about unstressed vowels in Derby specifically. Foulkes and Docherty (1999) transcribe the vowel in the keyword horsES with /ə/, rather than the more commonly used variant in such suffixes in British English: /ɪ/. This was an auditory impression rather than based on any empirical investigation, and only explicitly refers to one very specific context where an unstressed vowel might occur, that of suffixes. It will thus be interesting to see the extent to which there appears to be a /ɪ/ and /ə/ distinction in Derby, and whether there is a general tendency for speakers from Derby to use more schwa-like variants in positions where in other varieties of British English there would generally be something more akin to /ɪ/.

3.5 Methodology

This section describes the data measurement and preparation steps that were taken in analysing the data from the Derby corpus (Milroy et al, 1997). This data is used in the analyses in both Chapters 3 and 4, so sections 3.5.1-3.5.4 apply to the analysis that was performed in both chapters. Section 3.5.5 describes the data filtering and modelling that are specific to the analyses in this chapter. Section 4.3 describes the data filtering and modelling steps that are relevant for the analyses conducted in Chapter 4.

3.5.1 The data

The unstressed vowels measurements were taken from 31 speakers from a corpus of speakers from Derby (Milroy et al, 1997), as shown in Table 3.1. For reasons explained in section 3.5.5, only data from 26 speakers is included in the analysis presented here in Chapter 3. All of the data used is from spontaneous speech data which consists of dyadic conversations between speakers of the same age and class.

Table 3.1 *Speaker frequencies by class, gender and age*

Middle class				Working class			
old		young		old		young	
male	female	male	female	male	female	male	female
4	4	4	3	4	4	4	4

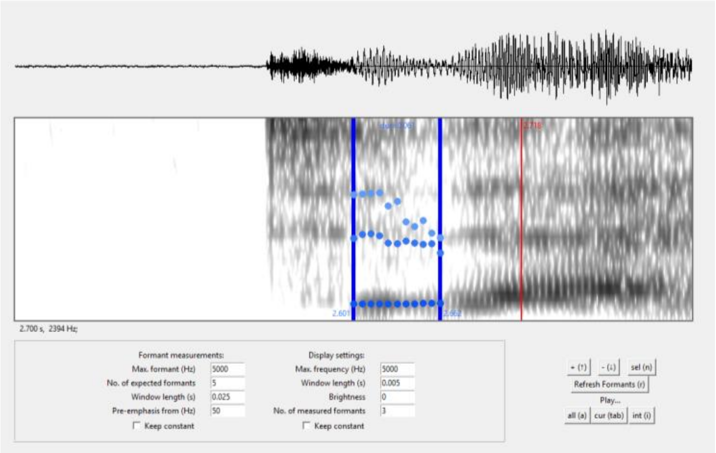
3.5.2 Token extraction

Extraction of tokens from the corpus was performed using the software LaBB-CAT (Fromont and Hay, 2012). This software allows the user to search an aligned corpus on a number of layers using regular expressions. Phonemic labelling in the corpus was based on CELEX (Baayen et al, 1995). All instances of non-final lexically unstressed vowels labelled as /ə/, /ɪ/ or /ʊ/were extracted from the corpus, and word finally all tokens labelled as /ə/ were included. The decision to include non-final vowels with all three transcriptions was made because, as explained in Chapter 1, the aim was to avoid any initial assumptions about which lexical items would have which vowel qualities. Assigning vowels to an /ə/, /ɪ/ or /ʊ/ group solely based on the CELEX transcription would also not be straightforward, given that there were many words in the corpus where more than one transcription was given. In addition, part of the focus of the thesis was to examine variation between different unstressed vowel qualities. Indeed, this is the focus of the current chapter. Therefore, rather than decide from the outset which tokens to include in the schwa analyses in Chapters 4 and 6, the criteria for token inclusion is informed by the analysis in the current chapter. Note that in word final position only tokens labelled as /ə/ were included as in word final position /ɪ/ is used to represent the happY vowel, which was not the focus of analysis.

Only lexically unstressed vowels were extracted, meaning all tokens were in words of 2 syllables or longer, and thus potentially reduced vowels in function words such as *of* and *them* were not included. In total, 9105 lexically unstressed tokens were extracted.

3.5.3 Data measurements

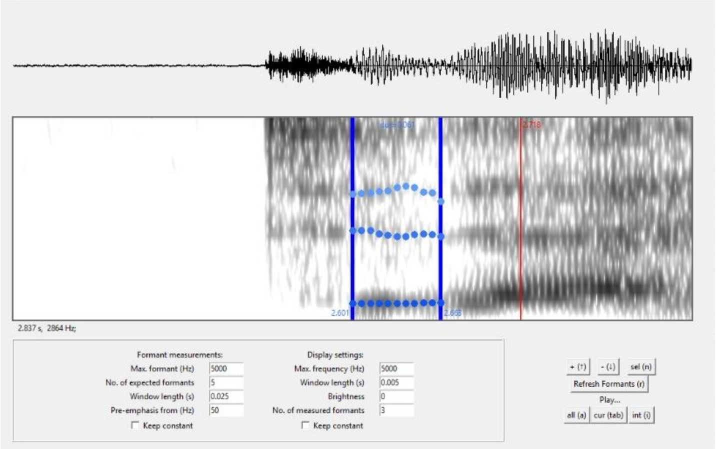
The first and second formants were extracted and obtained using the program *Formant editor* (Sóskuthy, 2014). Within each extract, the boundaries of each token were found manually. F1 and F2 measurements were initially taken automatically, but all were manually checked and corrected where necessary. The process of manually editing formant measurements in Formant editor are shown in Figures 3.1 and 3.2.



Formant measurements:
Max. formant (Hz) 5000
No. of expected formants 5
Window length (s) 0.025
Pre-emphasis from (Hz) 50
☐ Keep constant

Display settings:
Max. frequency (Hz) 5000
Window length (s) 0.005
Brightness 0
No. of measured formants 3
☐ Keep constant

Figure 3.1 *Uncorrected schwa formant measurements in today*



Formant measurements:
Max. formant (Hz) 5000
No. of expected formants 5
Window length (s) 0.025
Pre-emphasis from (Hz) 50
☐ Keep constant

Display settings:
Max. frequency (Hz) 5000
Window length (s) 0.005
Brightness 0
No. of measured formants 3
☐ Keep constant

Figure 3.2 *Corrected formant measurements in today*

11 measurements were taken at equal intervals for each token i.e. at 10% intervals. The duration of each vowel was also recorded. In the majority of the analyses conducted, however, only the vowel midpoint is examined, although the formant trajectory analysis in section 4.4.3 uses these 11 measurements from across the vowel.

After the process of listening to the data, manually locating the vowel boundaries, and manually checking formant measurements, only 4,980 unstressed vowel measurements were recorded. This was because there were only 4,980 tokens where an unstressed vowel could be found and measured on the spectrogram.

For the remaining 4125 tokens, formant measurements were unable to be recorded for various reasons. These tokens were of the following types:

- 550 voiceless tokens , where no measurement could be obtained e.g. *support*, *today*
- 1,332 tokens where no unstressed vowel was realised. 851 of these were words where the vowel was simply not produced e.g. *probably* as [prɒbli], *perhaps* as [praps]. 481 were words where the whole syllable containing the transcribed schwa was not produced e.g. speakers only uttering the second syllable of *because.*
- 118 tokens where an unstressed vowel was perceived but there was no visible vowel on the spectrogram which could be measured
- 2,125 tokens for other miscellaneous reasons not unique to unstressed vowels. These included the word not appearing at all in the extracted speech, atypical speech e.g. singing, not being able to measure because of speaker overlap.

3.5.4 Data preparation

3.5.4.1 Classification of consonantal environment

The phonetic context was classified based on the adjacent consonants. The effect of the adjacent vowel was not included as a variable. This was because previous research has shown that the effects of adjacent vowels are negligible when the effect of adjacent consonant is also considered (Kondo, 1994; Bates, 1995; Lilley, 2011). The added complexity which would have been added to the statistical models of the data through the addition of adjacent vowels was therefore not justified.

In order to produce models that converged, adjacent consonants were pooled into 3 groups based on their relative effect on the backness of schwa, following initial observations of the effects of different consonant types, and the types of effects that have been found previously on schwa (Kondo, 1994, Bates, 1995, Flemming, 2009., Lilley, 2011). Velar consonants are split into 2 different groups based on whether the vowel they were adjacent to was front or back. For example, the velar preceding schwa in the word *longer* is classed as a back velar, as it is preceded by the back vowel /ɒ/; whereas in the word *Nicholas* the velar preceding the first schwa is classed as front as it is preceded by the vowel /ɪ/. This is because velars tend to be coarticulated with a neighbouring vowel. For example, Kondo (1994) found there to be an interaction between adjacent velars and vowels on the F2 of schwa. Observation of the data confirmed this to be an appropriate division, and made more sense the placing velars into one category. This is shown in the comparison of velar and coronal consonants in Figure 3.3 below.

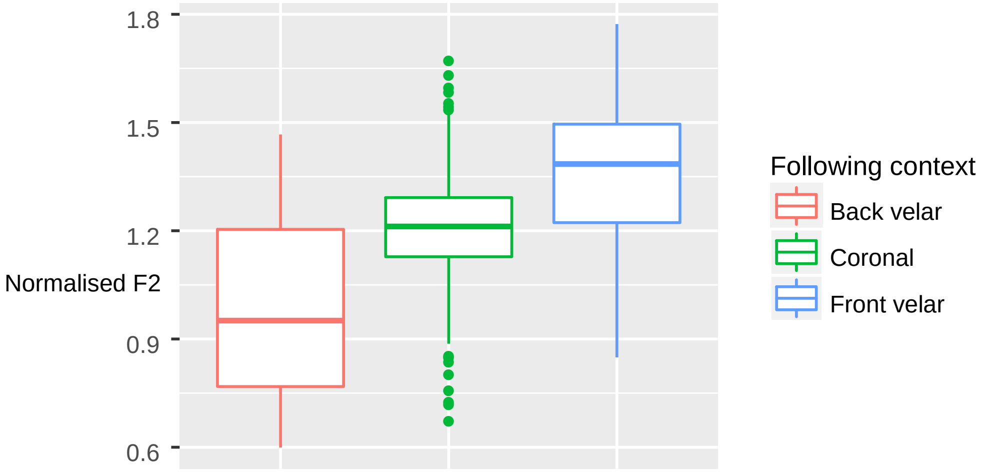


Figure 3.3 *Boxplot comparison of effects of following context on normalised F2 of non-final schwa*

The phonetic contexts were thus as follows:

- F2 lowering consonants- containing /w/, /r/, /l/, labial obstruents and nasals, and back velars
- Coronal consonants- containing dentals and alveolar obstruents
- F2 raising consonants- containing /j/, postalveolars, and front velars.

Consonants have been grouped together based on their relative effects on F2, rather than their specific articulations. As there is a correspondence between F2 and vowel backness, the consonants will henceforth be referred to in this way:

- F2 lowering consonants as backing consonants
- Coronal consonants as central consonants
- F2 raising consonants as fronting consonants

This is intended as a descriptor of the acoustic effects on vowels, rather than as an articulatory descriptor of the consonants themselves.

3.5.4.2 Normalisation

Formant measurements were normalised, using the modified Watt and Fabricius method (Fabricius et al, 2009). The normalisation was carried out using the online version of the vowel normalisation and plotting suite (Thomas and Kendall, 2007). In order to normalise the data, speakers’ TRAP and FLEECE vowels were also measured. These measurements were taken from word list data, as being a more formal situation, it was thought that these would yield more extreme formant values, and would be more likely to represent the extremes of the vowel space. All tokens of these vowels in the word lists were extracted.

Note that the normalised values are only used for presentation when data from all speakers is plotted together. Elsewhere, raw Hz values used. In the modelling of the data the random differences between speakers are taken into account using random speaker intercepts. Neither of the analyses of data from the Derby corpus use variables that differ between speakers (e.g. year of birth, class etc) so it was felt it was not necessary to use normalised data, but preferable to use Hz as it is a more meaningful unit of measurement.

3.5.4.3 Data filtering

The sample of vowels was further restricted to cases where the vowel would be unambiguously reduced, as many words can be pronounced alternatively with a reduced or full vowel. This meant that tokens where a full vowel was also possible e.g. *November, advice* were excluded. The decision to exclude a word was made if any of these three things was true:

- The author had heard productions using a full vowel e.g. *Yorkshire ,* where /ɪə/ is also possible
- An alternative vowel was given in the Longman pronunciation dictionary (Wells, 1990).
- The vowel was clearly pronounced with a full vowel on at least one occasion within the corpus, e.g. *July* pronounced as /dʒuˈlaɪ/.

Tokens were also filtered out if:

- They occurred in common suffixes such as *–ed*, *-es*, and *-ing*
- The token was a stem final but word medial schwa e.g. *sugars*. This was so that the non-final schwa group of vowels only included vowels that couldn't also occur in a word final position in order to be able to make a clear comparison between word final and non-final schwas.

These steps removed a further 590 tokens from the dataset, leaving an unstressed vowel dataset of 4390 tokens. It is likely that many of the tokens that were excluded were produced with a reduced vowel. However, for the purposes of the research questions, it was felt more important to only include unambiguous cases of phonological schwa, in order to be sure that the vowel speakers were 'aiming' for was a reduced vowel. In words where an alternative pronunciation is possible, even if speakers produce something that is heard as a schwa, it is possible that they could be producing a phonetically reduced version of a full vowel.

3.5.5 Data analysis

The data analysed in this chapter is all taken from unstressed vowels in non-final position e.g. *office*, *before*, *arrive.* The aim in this chapter is to compare the realisation of the spellings <a>, <e>, and <i>. As this chapter includes comparisons of data amongst individual speakers, only speakers with at least 10 tokens of each of these spellings were included in the analysis. This resulted in 5 speakers being excluded from the final analysis, meaning that in this chapter the analysis is from a total of 26 speakers. The final analysis included 584 <i> tokens, 792 <a> tokens, and 805 <e> tokens.

In the data analysis conducted in this chapter, mixed effects linear regression models are used to assess the significance of differences between spelling on both F1 and F2. Where significant effects are reported these are derived from likelihood ratio model comparisons with and without the effect in question.

All models have random intercepts for word, and random slopes by speaker for spelling. The three levels of spelling are <i>, <a> and <e>. In addition, the F2 models also have phonetic context as a fixed effect, and random slopes by speaker for phonetic context. The levels of phonetic context are back, central and front, and these were coded as described in section 3.5.4.1. Initially both the F1 and F2 models included the effect of phonetic context. However, phonetic context did not have a significant effect on F1, and model fit improved when phonetic context was not included. This concurs with findings elsewhere which have found that phonetic context is less predictive of the F1 of schwa than F2 (Kondo, 1994; Bates, 1995; Flemming, 2009). As a result, phonetic context was not included in the final F1 model. The final models used for both F1 and F2 are shown below.

F1 model structure:

F1~spelling+(1+spelling|speaker)+(1|word)

F2 model structure:

F2~spelling+preceding context+following context+(1+spelling+preceding context+following context|speaker)+(1|word)

3.6 Research questions

The following analysis focusses on comparing unstressed vowels represented with the spellings <a>, <i> and <e>. Unfortunately, there were not sufficient numbers of <o> and <u> tokens to make for a meaningful analysis.

As explained in section 3.4 there is a link between the orthographical representation of an unstressed vowel and its association with certain vowel qualities. In summary, <a> is generally said to be associated with unambiguous /ə/, <i> with unambiguous /ɪ/, and <e> is seen to be more variable. Comparison of these three vowel qualities therefore provides a way of comparing the tokens that have these expected vowel qualities. <a> and <i> are used as proxies for /ɪ/ and /ə/, as if there is a difference between two such vowel qualities then we would expect these two spellings to differ phonetically. In addition, the fact that <e> is said to be more variable provides a way of exploring whether there is any additional variation beyond a categorical difference between /ɪ/ and /ə/.

The analyses within this chapter are broken into two separate questions. Below I describe how each of these research questions relates to the broader aims and questions of the thesis as set out in Chapter 1.

Question 1: How is the difference between /ɪ/ and /ə/ realised in unstressed syllables?

Why this is relevant to the main research questions:

One of the main aims in this thesis (see Chapter 1) is to assess whether /ə/ is distinct from other vowels in unstressed syllables. As has been discussed in section 3.4, the majority of the variation in unstressed syllables is said to be between /ɪ/ and /ə/ (Gimson, 1962; Roach, 1983) so it is important to examine how this difference is phonetically realised. As explained in section 3.3, categorising unstressed vowels into different vowel qualities is difficult and so examining these differences through the lens of spelling is a useful way of approaching this problem. By comparing the phonetic realisations of <i> and <a> (used as proxies for /ɪ/ and /ə),this chapter is not only able to address the question of whether there are differences between /ə/ and other vowels in unstressed syllables, but on what phonetic dimensions these distinctions manifest. In addition to looking at these differences over the whole speaker sample this chapter also examines the difference between /ɪ/ and /ə/ in individual speakers. This allows for an assessment of how consistently /ɪ/ and /ə/ are differentiated in unstressed syllables.

Predictions:

If /ɪ/ and /ə/ are phonetically distinct in unstressed syllables it is expected that the <a> and <i> tokens would be clearly separately distributed along at least one phonetic dimension (F1 or F2 in this case). If this is a real and consistent difference we would expect this difference to show both over the whole data set and in the distributions of the individual speakers.

Question 2: Is there categorical reduction to /ɪ/ and /ə/ in unstressed syllables?

Why this is relevant to the main research questions:

This question concerns whether all vowels tested fall clearly into either an /ɪ/ or /ə/ category, or alternatively whether there is additional variation. Although most previous commentary on unstressed vowel variation has focussed on /ɪ/ and /ə/ (see sections 2.2 and 3.3), it is important to consider the possibility that there could be further vowel quality differences in unstressed syllables. By examining whether there are any additional vowel quality differences the work in this chapter is able to address the degree to which differences between full vowels have neutralised in unstressed syllables.

For this question the main focus is on the <e> tokens. This is because these spellings are generally said to be more variable, both as a whole class and also within individual words. For example, in the Longman pronunciation dictionary (Wells, 1990) a large number of words spelt with <e> have both /ɪ/ and /ə/ listed as variants. It is therefore expected that the <e> spellings will be where the most variation is present. <e> spellings are therefore used as a proxy for variable /ɪ/ and /ə/, with the intention of seeing what kind of variation exists between /ɪ/ and /ə/. They are thus compared to <a> and <i>, as these are used as the proxies for /ɪ/ and /ə/.

In the context of targetlessness this question is also important in terms of understanding the degree to which /ə/ differentiates itself from other vowels. Although the work in this chapter does not speak directly to the issue of targetlessness, the degree to which it is different from other vowel qualities is of course a related issue. This is because, in general, being extremely variable and targetless is not conducive to contrasting with other vowels.

Understanding how /ə/ may or may not differ from unstressed vowel qualities is also important from a methodological point of view for the work in later chapters, in terms of understanding which tokens are appropriate to include in the schwa analyses.

The predictions for the patterning of <e> follow directly from the answer to question 1. Therefore the analysis for question 1 is first presented, which compares the <a> and <i> tokens. The predictions for question 2 are then described in section 3.7.2.1.

3.7 Results

This analysis first addresses question 1, focussing on the comparison of <a> and <i> tokens. In the second part of the analysis the <e> tokens are also compared in order to address questions 2. Within each of the two main sections the patterns are first examined over the whole of the data, and then the data is analysed across individual speakers.

Raw Hz values are shown in the plots by individual speakers, and in the models raw values are also used, where variation in the formant values of individual speakers can be controlled through random intercepts. Where there are plots of the data across all speakers, normalised measurements are shown (Fabricius et al, 2009).

3.7.1 How is the difference between /ɪ/ and /ə/ realised in unstressed syllables?

As noted in section 3.6, if there is a phonetic distinction made between /ɪ/ and /ə/ in unstressed syllables, then the distributions of <i> and <a> are expected to differ in at least one of F1 or F2. Additionally, if this difference is in line with the standard transcriptions of [ɪ] and [ə] it would be expected that the <a> tokens would have a higher F1 and lower F2 than the <i> tokens.

3.7.1.1 Results from all data

The F1 density distributions for <i> and <a> (Figure 3.4) are very similar overall. The difference between <a> and <i> in F1 was modelled with random intercepts of word with random slopes by speaker for spelling. There is a significant difference between the F1 of <a> and <i> ($p<0.0001$) in the direction that was predicted, as the <a> tokens have a higher F1 than <i>. This is, however, a fairly small difference, as the estimated difference between the spellings was only 28.64 Hz. In addition, the distributions are largely overlapping, as shown density distributions in Figure 3.4 below.

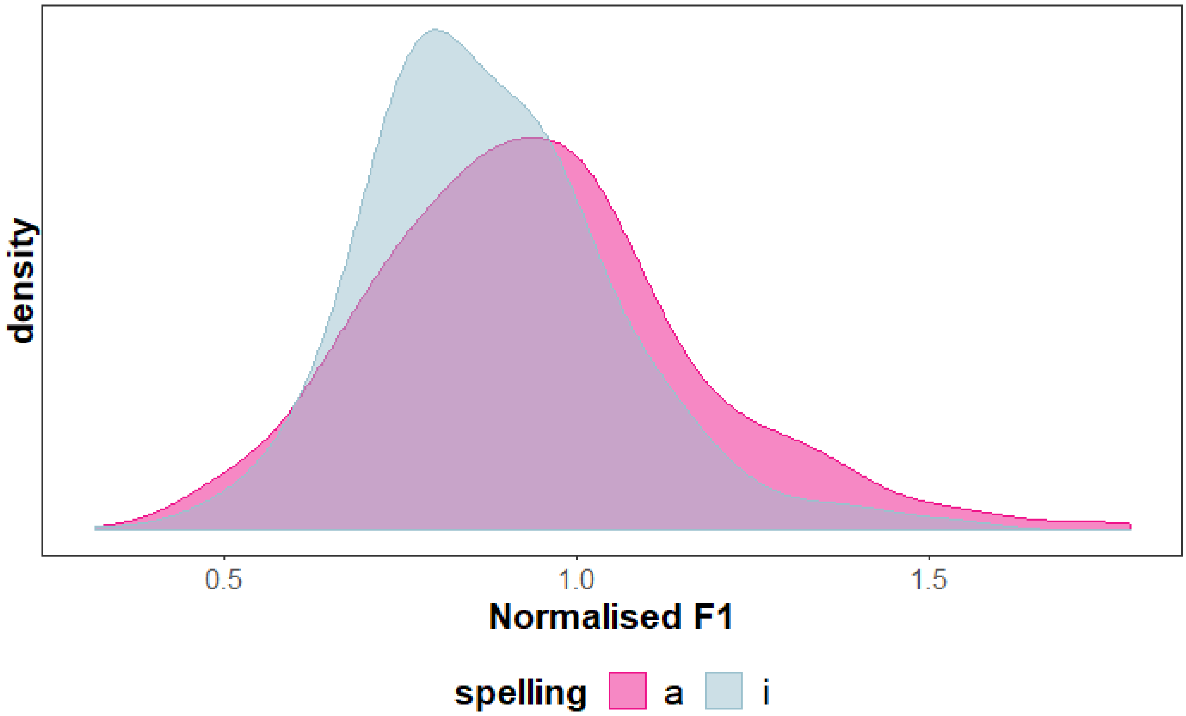


Figure 3.4 *Normalised F1 density distributions for <i> and <a> over all data*

Table 3.2 *Summary of fixed effects on F1 (Hz) of <a> and <i> (reference level=<a>)*

	Estimate	Std. error	*t*-value	*p*-value
(Intercept)	473.48	12.69	37.325	< 0.0001
spelling.<i>	-28.64	8.1	-3.535	< 0.001

Figure 3.5 below shows the density distribution for the F2 data for both <a> and <i>. As would be expected if /ɪ/ and /ə/ have clearly distinct vowel qualities, the <i> spellings have a higher F2 than <a>. Although there is a reasonable degree of overlap between the two, they look to be clearly separate distributions, with the peaks of the density distributions clearly separated for the two spellings. The spread of the data is also similar for both groups, as evidenced both in the density distributions and the standard deviation for both groups, in Table 3.5.

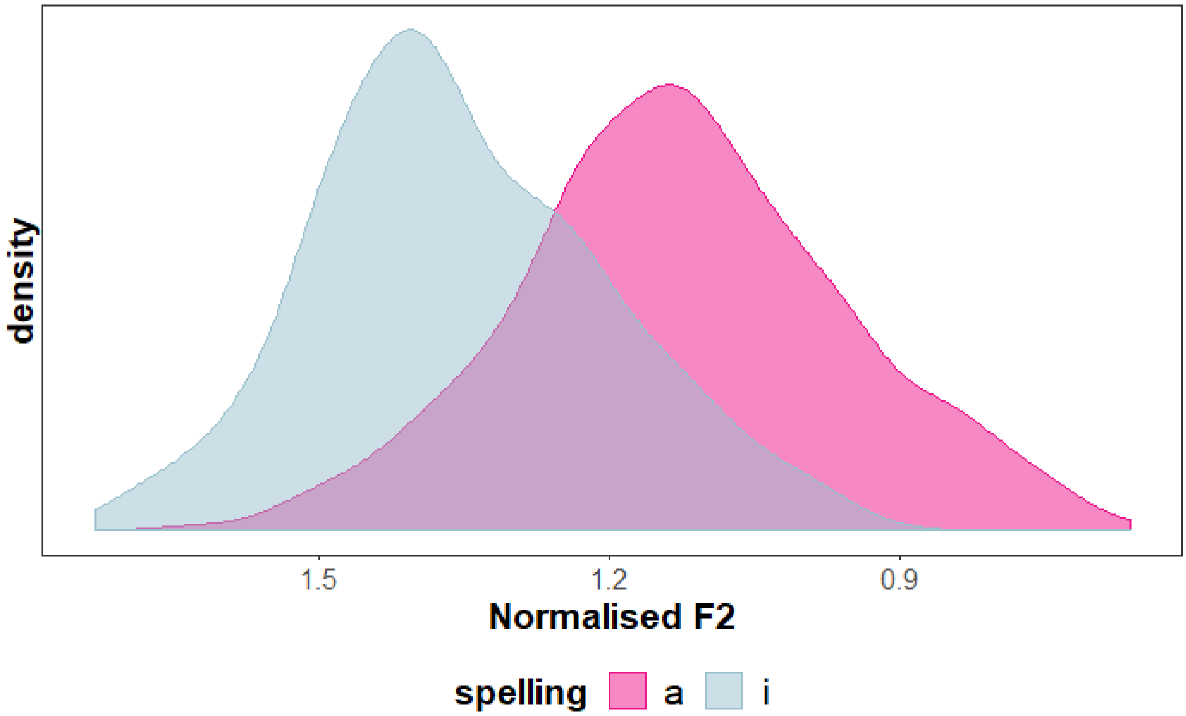


Figure 3.5 *Normalised F2 density distributions for <i> and <a> over all data*

The difference between <a> and <i> in F2 was modelled including fixed effects of spelling and phonetic context, with random intercepts of word and random slopes by speaker for spelling and phonetic context. In terms of F2, the difference is again in the expected direction, with <i> having a higher F2, as was predicted ($p<0.0001$).In comparison to F1 the difference is much larger, with <i> tokens overall being estimated to be 289 Hz higher than the <a> tokens. As the modelling takes into account the effect of phonetic context, this suggests that this is a genuine difference and not one that exists simply because <i> and <a> happen to be in different phonetic environments.

Table 3.3 *Summary of fixed effects on F2 (Hz) of <a> and <i> (reference levels: spelling=<a>, preceding context=back, following context=back)*

	Estimate	Std. error	*t*-value	*p*-value
(Intercept)	1469.48	30.33	48.457	<0.0001
Preceding context. central	158.11	15.73	10.049	<0.0001
Preceding conext.front	260.72	23.83	10.942	<0.0001
Following context.central	102.83	25.24	4.074	<0.001
Following context.front	189.94	31.32	6.065	<0.0001
Spelling. <i>	289.33	23.1	12.523	<0.0001

Overall, these results reflect what was predicted if <i> and <a> are representative of a separate /ɪ/ and /ə/ category, as <a> is found to have a lower F2 and a higher F1.The distribution of the tokens over the whole vowel space is shown in Figure 3.6. It is particularly clear here that these groups of vowels differ mainly on the F2 dimension, where they are clearly separable. Although <i> and <a> are significantly different in F1 it is clear from Figure 3.6 that this is very small difference. What is more striking is the fact that <a> is spread over a wider range in terms of F1. The standard deviations for <i> and <a> are shown in Tables 3.4 and 3.5 below. Whilst the spread of data across F2 is broadly similar for <a> and <i> in terms of F1 the range for <a> is clearly larger.

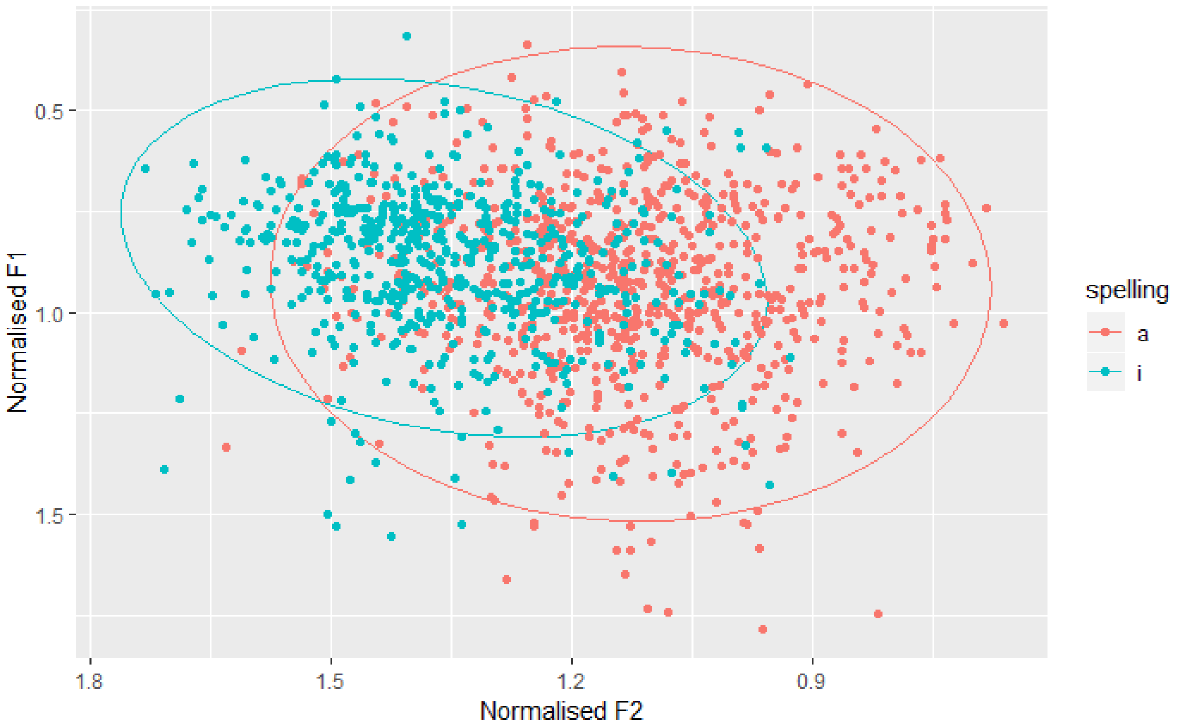


Figure 3.6 *Normalised vowel plot of all <a> and <i> tokens*

Table 3.4 *<a> and <i> F1 standard deviation and means*

	Norm. F1 mean	F1 (Hz) mean	Norm. F1 SD	F1 (Hz) SD
<a>	0.98	478	0.23	121
<i>	0.88	448	0.18	99

Table 3.5 *<a> and <i> F2 standard deviation and means*

	Norm. F2 mean	F2 (Hz) mean	Norm. F2 SD	F2 (Hz) SD
<a>	1.12	1613	0.17	254
<i>	1.35	1956	0.15	278

In spite of the lack of a large difference in F1, <a> and <i> are clearly different in F2, with the <i> tokens clearly being more front overall. It was predicted that if /ɪ/ and /ə/ are phonetically distinct in unstressed syllables that it would be found that the <a> and <i> tokens would be clearly separately distributed along at least one phonetic dimension. Therefore, over the whole dataset, this prediction is borne out.

3.7.1.2 Results from individual speakers

This section examines the differences between <a> and <i> for individual speakers in order to see how consistent the differences found over the whole set are across speakers.

As seen in Figure 3.7, across individual speakers, there is not a clear difference in F1 between <a> and <i>. Many of the speakers have a slight difference in the height of the two groups of tokens (e.g. speakers 19, 20, 26). However, at the individual level this is also only ever a very small difference. For all speakers the F1 distributions of <a> and <i> almost entirely overlap.

Figure 3.8 shows the F2 differences for individual speakers. It is clear that the F2 difference which is found over the whole data is also found for the individual speakers. For speaker 15 the distributions are mostly overlapping. However, this is the only speaker for whom this is the case. For all other speakers, the F2 distributions of <a> and <i> are clearly different. This difference is bigger for some speakers than others (e.g. speaker 4 has particularly separated distributions), but importantly all but one speaker have clearly separate distributions.

In summary, unstressed vowels represented by <a> and <i> are distinctly different from one another. It is, however, primarily only on the F2 dimension where the distributions are strongly distinct, with a clear difference in F2 overall, and for all but one individual speaker. Although there is a significant F1 difference overall this is a fairly small difference and is not consistent over individual speakers. As <a> and <i> are taken as proxies for /ɪ/ and /ə/, this thus suggests that these vowels are phonetically distinct in this group of speakers, although this is primarily on the backness rather than height dimension. /ɪ/ is clearly fronter as was predicted, but only slightly higher.

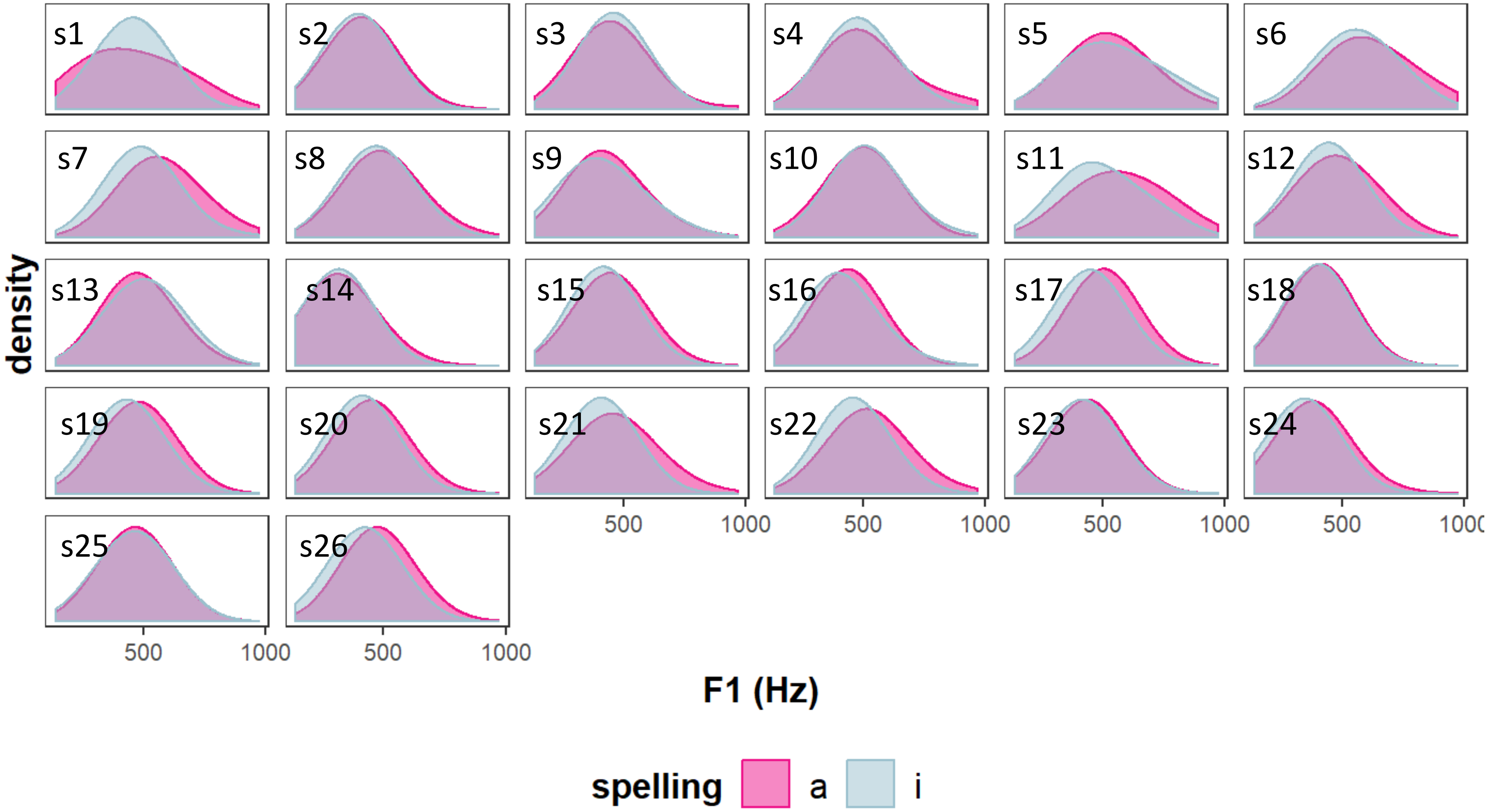


Figure 3.7 F1 (Hz) density distributions for individual speakers for <a> and <i>

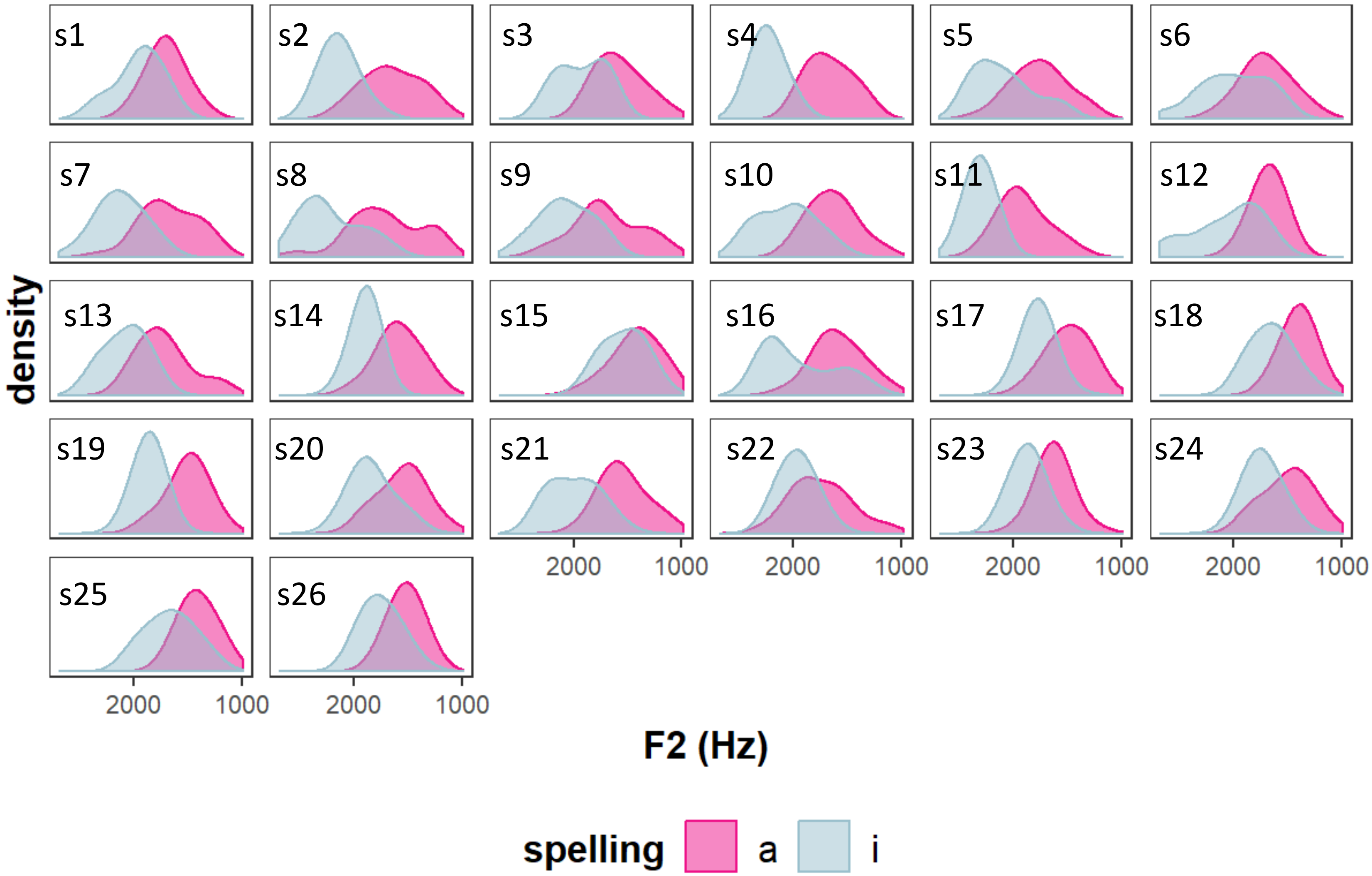


Figure 3.8 *F2 (Hz) density distributions for individual speakers for <a> and <i>*

3.7.2 Is there categorical reduction to /ɪ/ and /ə/ in unstressed syllables?

Section 3.7.1 established that there is a difference between <a> and <i> and that this is primarily on the F2 dimension. This therefore gives a basis for comparison with the <e> spellings. To recap section 3.4, <e> spellings are analysed here as in dictionaries they are found to be variably transcribed with /ɪ/ and /ə/. A central questions of this chapter is whether unstressed vowel qualities vary categorically between /ɪ/ and /ə/. Therefore, the question of how the <e> tokens compare phonetically to <a> and <i> is of interest.

Although there are overall differences between the F1 of <a> and <i>, the distributions are largely overlapping. Therefore, observation of F1 values would not be a good way to differentiate between the more <a> and <i> like <e> tokens. Since F2 values are the main differentiator between <i> and <a>, it is F2 which will be the focus here.

3.7.2.1 Predictions

The focus of this section is on how unstressed vowels represented with <e> compare with those represented with <a> and <i>, and whether they categorically vary between two distinct realisations. Since <i> and <a> are being used as proxies for /ɪ/ and /ə/, this comparison allows us to see to what extent unstressed vowels vary categorically between /ɪ/ and /ə/.

If there is categorical variation between /ɪ/ and /ə/, there are different ways in which this could manifest itself. It is possible that the <e> tokens could pattern exclusively with <a>, which would indicate that they are produced categorically with /ə/. Conversely, they could pattern exclusively with <i> which would indicate a more categorical /ɪ/-like pronunciation. These possibilities would be evidenced by the density distribution of <e> largely overlapping with either <a> or <i>, and with <e> not being significantly different in F2 to just one of <a> and <i>. If <e> patterns consistently with either <a> or <i> in this variety we would expect this pattern to also be consistent in the distributions of individual speakers. If this is the case it can be concluded that variation between /ɪ/ and /ə/ is categorical.

The variable use of /ɪ/ and /ə/ in dictionaries, however, suggests that <e> spellings may be more variable. It is therefore of interest to see the degree to which their realisation varies between speakers. Through comparing individual speaker's productions of <e>, it can be seen whether speakers <e> vowels consistently pattern with their <a> or <i> productions or whether their <e> realisations are more variable.

3.7.2.2 Results across all data

Figure 3.9 below compares <e> (dotted distribution) with <i> (blue) and <a> (pink) across all speakers. It shows that <e> is intermediate overall, suggesting that unstressed vowels do not pattern exclusively with either /ə/ or /ɪ./ Comparing the three spellings confirms that the F2 of <e> is indeed significantly different from both <i> and <a> (Table 3.6). It is significantly lower in F2 than <i> ($p<0.01$) and significantly higher in F2 than <a> ($p<0.0001$). This is as would be expected given that <e> is suggested to be variable in unstressed vowel quality. In this variety at least, this does not suggest that the <e> tokens are produced with either a categorical /ɪ/ or /ə/ like pronunciation. However, Figure 3.9 does not tell us whether the intermediate distribution of <e > derives from a mixture of tokens that pattern with <a> and <i>, or whether <e> is genuinely intermediate between <a> and <i>. It also does not tell us whether this intermediate distribution is consistent amongst speakers, or whether it stems from variation in the realisation of <e> across individual speakers.

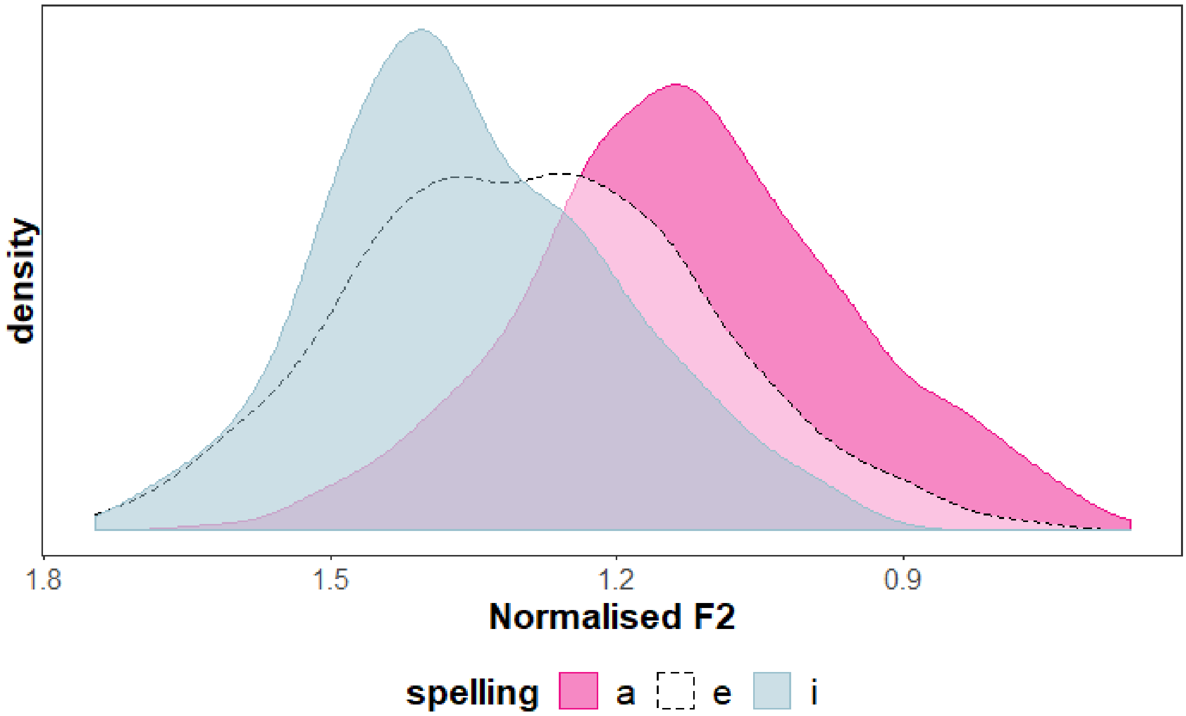


Figure 3.9 *Normalised F2 density distributions for <a>, <e> and <i> over all data*

Table 3.6 *Summary of fixed effects on F2 (Hz) of <a>, <i> and <e> (reference levels: spelling=<e>, preceding context=back, following context=back)*

	Estimate	Std. error	*t*-value
(Intercept)	1704.08	37.06	45.98
Spelling. <a>	-230.33	24.96	-9.229
Spelling. <i>	59.56	22	2.708
Preceding context. central	152.82	14.77	10.349
Preceding context. front	244.9	21.3	11.497
Following context. central	100.11	20.2	4.956
Following context. front	191.83	29.73	6.452

In Figure 3.9 the distribution of <e> looks slightly bimodal, although the dip in the middle of the distribution is very small. Figure 3.10 shows the combined <a> and <i> distributions over all of the data. As shown in 3.7.1, <a> and <i> clearly occupy two separate distributions. This is reflected in the fact that in Figure 3.10 there is strong bimodality in the data. If the <e> distribution shown in Figure 3.7 is simply a mixture of /ɪ/ and /ə/, we would expect its distribution to be more similar to the combined distribution of <i> and <a> as shown in Figure 3.10. However, the distribution in Figure 3.10 is much more clearly bimodal than <e>

in Figure 3.7. This suggests that the intermediate distribution of <e> may not simply be a mixture of /ɪ/ and /ə/ tokens.

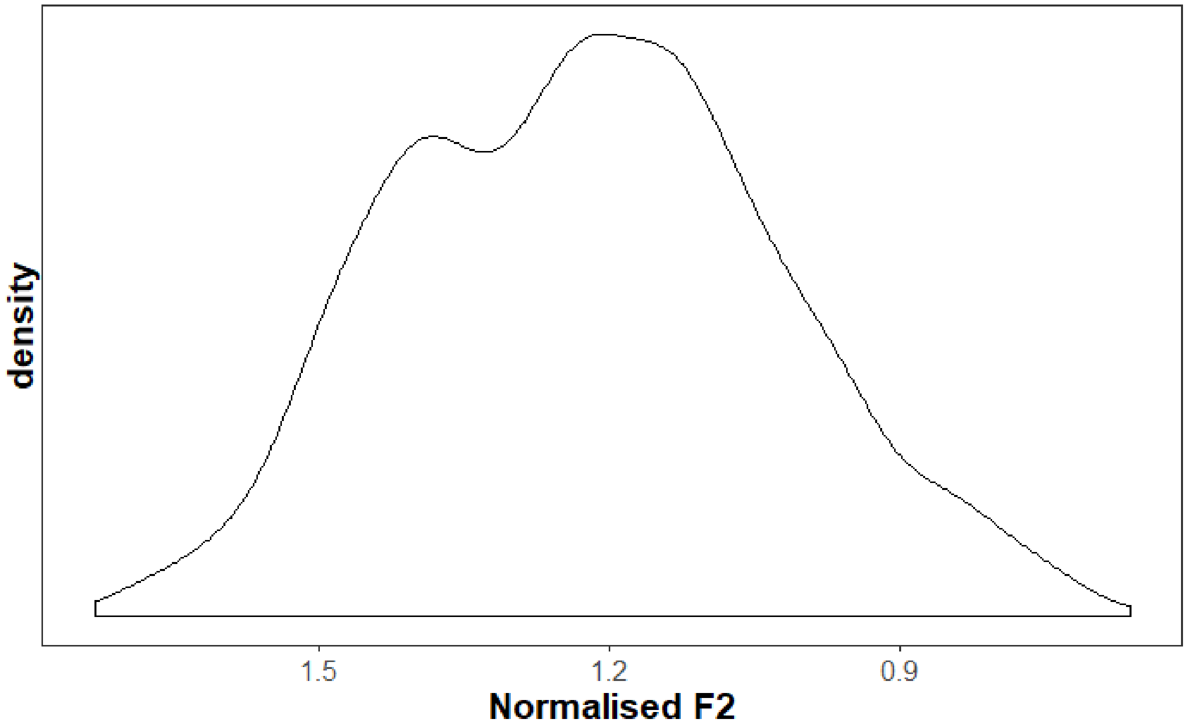


Figure 3.10 *Combined normalised F2 density distribution of <a> and <i>*

Table 3.7 *F2 means and standard deviations from all spellings*

	Norm. F2 mean	F2 (Hz) mean	Norm. F2 SD	F2 (Hz) SD
<a>	1.12	1613	0.169	254
<i>	1.35	1956	0.152	278
<e>	1.29	1865	0.181	308

3.7.2.3 Results from individuals

The fact that <e> is intermediate overall leads to the question of whether this is a pattern that is consistent amongst speakers or whether this overall intermediate pattern is as a result of variation between individual speakers.

When the F2 distribution for individual speakers is examined (Figure 3.11) it is clear that there is much interspeaker variation in terms of the position of <e> in relation to <i> and

<a>. Some speakers have an <e> distribution that patterns with <a>, suggesting a categorical schwa-like pronunciation, and some have an <e> distribution that patterns with <i> which suggests a more categorical [ɪ] pronunciation. However, some speakers (highlighted in yellow) have an intermediate <e> distribution, like that which was seen in the overall data (section 3.7.2.2). It is therefore clear is that the pronunciation of <e> tokens is much more variable between speakers than <i> or <a>.

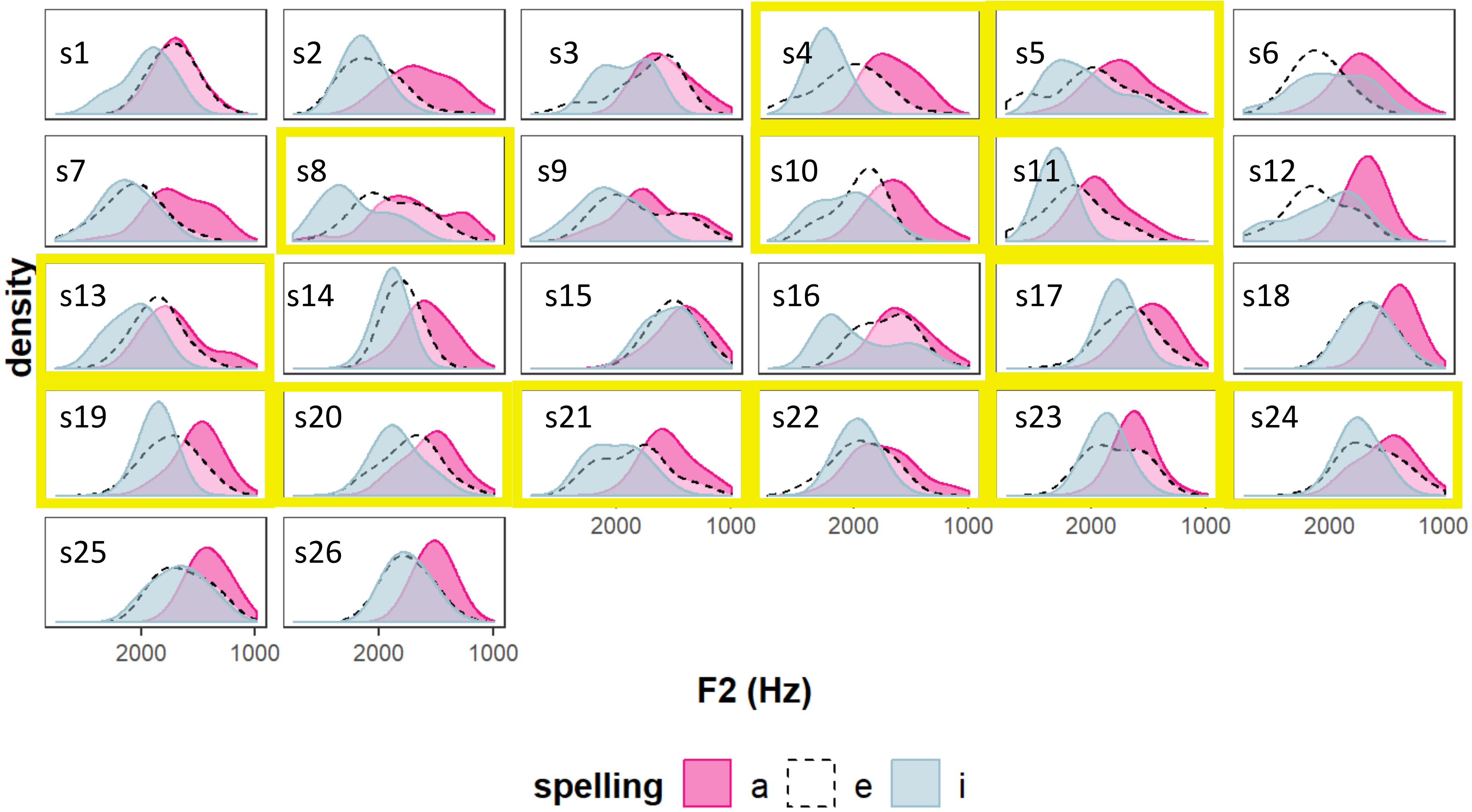


Figure 3.11 *F2 (Hz) density distributions for individual speakers for <a>, <i> and <e>*

There are some speakers who have <e> patterning categorically with <i> or <a>, but for roughly half of the speakers (in yellow) <e> appears to be intermediate. Again, from the data presented here it is not possible to know conclusively whether these intermediate distributions represent a mixture of tokens like <i> and <a> or whether the <e> tokens are genuinely intermediate. However, we do not see clear bimodality for the<e> distributions from most of these speakers, which suggests that for some speakers at least, these <e> tokens could be genuinely intermediate. In addition, the spread of data of <e> compared to <i> and <a> is not consistently higher, which is what would be expected if the distribution was the result of a mixture of/ɪ/ and /ə/ tokens. It therefore seems possible that, for at least some of the speakers, not all unstressed vowels categorically reduce to /ɪ/ and /ə/. The intermediate distribution of <e> overall and the distributions for some individual speakers thus indicates that other intermediate variants may be possible.

3.7.3 Summary of results

Taking the <i> and <a> spellings as proxies for /ɪ/ and /ə/ we see that there is a clear difference between these vowel qualities, as evidenced both in the overall pattern and in the speech of individuals. This is a difference that is clearly present in F2, with <i> having a consistently and significantly higher F2 which is as would be expected if the <i> spellings represent /ə/, and the <a> spellings /ə/. However, the difference between <a> and <i> in F1 is minor and is less consistent between speakers. In this way what is found differs from traditional ideas about the difference between /ɪ/ and /ə/. Although /ɪ/ and /ə/ are found to be distinctive, in the data presented here this difference is primarily a difference in backness and not in vowel height.

The results from <e> present a more complex picture**.** <e> tokens were examined as it was thought that this group of spellings would be the prime locus of variation between /ɪ/ and /ə/. It was clear overall that <e> did not pattern with either <i> or <a> and was intermediate. Compared to the relationship between <a> and <i>, where speakers were very uniform, there was much more variation in terms of the relative position of <e>. Some speakers' <e> tokens patterned with <a>, and for some they patterned with <i>. There were many speakers, however, that did not have their <e> tokens pattern neatly with either group. Many speakers had this more intermediate pattern. Importantly, for these speakers it was not necessarily the case that the <e> tokens were clearly split between two categories. For some speakers, these tokens looked genuinely intermediate, as evidenced in unimodal distributions and a similar spread of data to <i> and <a>. We cannot definitively say that these speakers have an intermediate variant between /ɪ/ and /ə/. However, the variable and often intermediate status of the <e> tokens suggests that the relationship between /ɪ/ and /ə/ is not straightforward, and we have not been able to find evidence that all unstressed vowel tokens fall clearly into an /ɪ/ or /ə/ category.

3.8 Discussion

One of the main research questions in this thesis was to discover whether schwa is distinct from other vowels in unstressed syllables. This chapter aimed to address this overarching question, breaking the findings of the chapter into two separate smaller questions. This section will discuss the findings related to each of these individual questions and how they relate to the overall goal of this chapter. It will then explore some of the possible reasons for the findings.

3.8.1 How is the difference between /ɪ/ and /ə/ realised in unstressed syllables?

The purpose of this question was to ascertain to what degree there exists two distinct vowel qualities present in unstressed syllables. This was examined through the lens of spelling, based on the fact that dictionary transcriptions and commentary on the distribution of these vowel qualities suggested a link between spelling and vowel quality. The hypothesis was that if there is a distinction between /ɪ/ and /ə/ in unstressed syllables then this should be reflected in there being a phonetic reality to the intuitions about these vowel qualities.

<a> and <i> were found to be significantly and consistently different from each other, which supports the idea that, at least in this variety of English, there are at least two distinct vowel qualities in unstressed syllables. <i> was found to be a fronter and higher vowel than <a>, which is what was predicted if the spellings were representative of contrasting /ɪ/ and /ə/ vowel qualities. Taking both F1 and F2 into account, the two sets of spellings occupy clearly separate distinctions. This suggests that the intuitions about the different vowel qualities that occur in unstressed syllables and their association with different spellings (cf. Jones, 1956; Roach, 1983) appear to be correct, at least for this variety of English. In this chapter the spellings <i> and <a> were used as proxies for /ɪ/ and /ə/, so the difference between these spellings therefore provides empirical evidence of the observed phonetic difference between /ɪ/ and /ə/ in British English (e.g. Jones, 1956; Well, 1982; Gimson, 1984). However, it is important to note that contrary to what the commonly used transcriptions suggest, this difference was primarily one of backness rather than height. This means, for example, that Fabricius's (2002) findings (see p. 27 for full details) may have been misleading. Fabricius's findings that suffixes have not shifted to schwa in RP are based on an assumption that schwa is a mid central vowel. However, the findings here show that, in reality, there may not always be much of a height difference between /ɪ/ and /ə/.

The finding of /ɪ/ and /ə/ as distinct vowel qualities also further supports the idea of schwa as a vowel with a target, rather than it being a targetless vowel. The fact that /ə/ is distinguishable from /ɪ/ based on a lower F2 suggests that it is not targetless on this dimension. /ɪ/ is realised as a more front vowel and /ə/ as more central. It has been suggested that /ə/ may surface as a central vowel due to the averaging out of the effects of different coarticulatory processes (Van Bergem, 1993). Of course, it would be possible for /ə/ to be targetless and /ɪ/ to be targeted and for the distributions to consequently look different. However, if this was the case we would have expected the <a> vowels to be substantially more variable in F2 than <i>, which is not what we find. Instead, we find that <a> and <i> have a similar level of F2 variability. This therefore does not suggest that the <a> tokens are representative of a targetless vowel. This therefore provides indirect evidence for the possibility, taking the <a> spellings to be representative of schwa, that schwa may not be a targetless vowel. This data therefore does not support the claims made by Kondo (1994) and Bates (1995) that schwa is targetless in F2.

3.8.2 Is there categorical reduction to /ɪ/ and /ə/ in unstressed syllables?

The focus of this question was on the comparison of <e> spellings with <a> and <i>. The focus here was on <e> due its supposed status as a spelling that is commonly variable between /ɪ/ and /ə/ (Jones, 1956; Roach, 1983). It was established that <a> and <i> are clearly different and can be assumed to generally represent /ə/ and /ɪ/ respectively. Following this, it was then possible to compare the <e> spellings to these groups in order to see how the vowel quality of this supposedly variable group compared.

The most important finding here was the intermediate position of <e> in relation to <a> and <i>. This was seen both overall and also in many of the individual speakers. Indeed, compared to the variation between <a> and <i>, the relative position of <e> was much more variable across speakers. Together these findings suggest that, at least in Derby, vowel quality variation in unstressed syllables is more complex than just a simple and categorical opposition between /ɪ/ and /ə/. The findings within this chapter therefore chime with the suggestion by Ladefoged and Johnson (2014) that the use of the symbols /ə/ and /ɪ/ could actually mask the fact that other intermediate variants may also be used. Whilst previous work on variation in unstressed vowels (Gimson, 1984; Fabricius, 2002) has implied that tokens fall into either into a /ɪ/ or /ə/ category, the reality is that this distinction may not quite be quite so straightforward. Indeed, the existence of a three way spelling difference is surprising given that differences between full vowels are often said to be neutralised in unstressed syllables (cf. Flemming and Johnson, 2007).

From the data analysed here, whether the intermediate pattern in <e> is a result of it being genuinely intermediate of <i> and <a>, or whether it simply appears as intermediate as a result of being a mixture of <a> like and <i> like tokens is inconclusive. Although, for some speakers <e> patterns categorically with <i> or <a>, there are many speakers where <e> is intermediate.

We do not see any clear bimodality in the data overall, or for the majority of individual speakers, which is suggestive that the <e> spellings could be genuinely intermediate. However, this alone does not prove that the intermediate distribution is as a result of <e> having a vowel quality intermediate of <a> and <i>. It is possible that this intermediate distribution is made of a mixture of <a> and <i> like tokens, as it is possible for the combination of two phonetically overlapping categories to create a unimodal distribution (Bermúdez-Otero and Trousdale, 2012; Dinkin and Dodsworth, 2017). For example, Turton (2014) shows that, with regard to l-darkness, what initially looks like an intermediate category can be the result of a mixture of the two extreme ends of the spectrum. There are reasons to suspect, however, that this is not the case here for two main reasons. Firstly, in comparison to the <e> tokens, the combined <i> and <a> distribution in this case was actually clearly bimodal, and secondly the spread of the data for the <e> group was comparative to the <i> and <a> groups.

3.8.3 Explaining intermediate <e>

Since it is highly likely that some speakers produce <e> in a way that is genuinely intermediate of <a> and <i>, it is worth considering how such a pattern could be explained. The three way F2 difference between <i>, <e> and <a> corresponds to the vowel qualities these spellings would generally represent, namely /ɪ/, /ɛ/and /æ/. The reason for the three way spelling difference that is found cannot be determined from the data here, but there are at least three potential explanations. These are: influence of orthography, phonetic analogy with related forms e.g. *prefer*/*preference* and residual differences between historical full forms. Each of these explanations will be discussed in turn.

3.8.3.1 Influence of orthography

The first possible reason for the three way spelling difference is that speaker's production of reduced vowels is directly influenced by the orthography. This would mean that the way speakers pronounce these vowels would be influenced and related to the general pronunciation of that spelling as a full vowel. Taft and Hambly (1985) provide some evidence that speakers' representation of unstressed vowels could be related to the spelling. They asked speakers to say whether a phoneme string appeared in a given word.

Speakers were more likely to make a mistake and suggest that a phoneme string appeared in a given word if the spelling in the word suggested it i.e. they were more likely to think the word *validity* contained the phoneme string /væl/, than /vɔl/. Speakers also made more errors when the target word contained a reduced vowel. However, this was obviously quite an artificial task and, and the study did not examine the influence of spelling on speaker's actual productions.

It is also somewhat difficult to see exactly how a spelling influence would work in practice in speech which is not read. If this explanation were correct it is unclear whether this would mean that speakers were actually aiming for the full vowels related to the spelling. However, this seems unlikely in that if that were the case we would expect that on some occasions the words would be heard with the full vowel that was being targeted, but this is not the case with any of the words here. All of the tokens used here were chosen for the very reason that an alternative pronunciation with a full vowel was not possible. In addition, as the data analysed in this chapter are from spontaneous rather than read speech it seems unlikely that speakers' productions would be directly influenced by orthography.

One study which suggests that the reason for such effects may not be a direct effect of spelling on speakers is Lilley (2012). Lilley examined the effects of spelling on unstressed vowel production and found that spelling was related to vowel quality in the same way as in this study i.e. that spellings that generally represent front full vowels would be realised as fronter. In addition, Lilley also examined the effect of a variable labelled as 'underlying phoneme'. The underlying phoneme of a vowel was labelled based on an alternative full vowel pronunciation in that word if there was one, or on the equivalent stressed full vowel in a related word e.g. the underlying phoneme of *apply* would be labelled as /æ/ based on the fact that this is the vowel in *application*. Both variables had an effect but in model comparison underlying phoneme was a slightly better predictor. The fact that underlying phoneme had a stronger effect than spelling suggests that in such cases it may actually be that the effect of spelling is an indirect effect of the vowels spellings are associated with, rather than an effect of the spelling itself. However, one issue with this study was that words with alternative full pronunciations were included e.g. *advice*. There is therefore a possibility that the effect of underlying phoneme may have been caused by speakers actually aiming to produce that underlying phoneme, and simply producing a phonetically reduced version of that full vowel.

3.8.3.2 Phonetic analogy

The second possible explanation for the intermediate position of <e> is that when producing a reduced vowel speakers display underlying connections with other related words e.g. *prefer*, *preference*; *atomic*, *atom*. This means that through the association with the related word, and the general relationship between spelling and vowel quality, spelling will have an indirect effect on their pronunciation. Unfortunately, in the data here it was not possible to test this particular hypothesis. Out of the tokens used for this analysis, there were only 125 tokens out of 2181 that had an equivalent cognate word where the vowel was stressed. However, this fact in itself suggests that this explanation is therefore unlikely to be correct. It is not the case that all unstressed vowels have such cognates, so if this was the reason for the effect then it would only actually apply to a subset of the unstressed vowels. Therefore, the fact that we see this pattern here suggests that this is unlikely to be the explanation.

In addition, findings in Taft and Hambly (1985) also suggest that this might not be a valid explanation. When they asked speakers if a phoneme string appeared in a given word, they asked participants both about words that had a related word containing the full vowel in the phoneme string e.g. *validity,* where *valid* has a full vowel, and about words where there was no related word containing the vowel in the phoneme string e.g. *lagoon*. Participants were just as likely to say the word contained the phoneme string on both cases. If speakers make connections between reduced and full vowels in cognate pairs it would be expected that they would be more likely to claim a phoneme string was in a word where there was a related cognate. Although there are some problems with relating this study to production, as noted above, it does provide some evidence against the idea that speakers associated reduced vowels with their full vowel cognates.

3.8.3.3 Residual differences between historical forms

The third possible explanation for the intermediate positon of <e> is that there has not been complete neutralisation of historical full forms. This would mean that, although reduction has taken place, residual differences reminiscent of historical full vowel forms remain. Therefore the spelling differences found may be an indirect effect of the fact that historically these spellings would have represented different vowels. Although there is not an exact correspondence between spelling and vowel quality in stressed syllables, it can be assumed that the spellings of phonologically reduced vowels is a general indicator of their historical pronunciation (Cf. Ladefoged and Johnson, 2014).

We can see how historically different full vowels may have ended up with slightly different reduced vowel qualities, through examining the synchronic process of phonetic reduction. All vowels are subject to a degree of phonetic reduction. In some cases this reduction had been phonologised, such that speakers aim for a reduced vowel, such as schwa e.g. *a̲rrive*. However, in other vowels there may be a continuum of variation between the citation form and the most phonetical reduced form. For example, in the word *di̲rection*, depending on the variety, the first vowel could be produced with [aɪ], [ə], or something in between. In terms of this gradual reduction process, different full vowels will move in different directions and therefore through slightly different areas of the vowel space. This can be seen in synchronic studies of reduction patterns in different vowels (Van Bergem, 1993; Proctor et al, 2015). Van Bergem manipulates a range of different stress, accent and speech style conditions in Dutch for different vowels. Acoustically, all vowels lower in F1, unless already a high vowel, and the F2 range for the vowels becomes smaller. Similarly, Proctor et al (2015) demonstrate the different direction of articulatory movements needed for reduction in different vowels.

These synchronic patterns of phonetic reduction are therefore likely to mirror the gradual diachronic processes of reduction which has led to phonologised reduction in some cases, where producing a reduced vowel is the target. If phonologically reduced vowels have arisen from a process of phonetic reduction, it thus makes sense that there may still be residual differences from the historical full vowel. It is not, however, being suggested that the three way difference found here is as a result of speakers actually aiming for /ɪ/, /æ/ and /ɛ/, or that they are aiming for anything other than a reduced vowel. The fact we find a three way difference, however, suggests even in phonologically reduced vowels there are still residual subphonemic differences, which are indicative of the historical full vowel. It is thought that this third explanation is the most likely to be the cause of the patterns that we see, and thus the common view (i.e. Flemming and Johnson, 2007) that full vowel differences have been completely neutralised in unstressed syllables may not be correct.

3.9 Conclusion

The main research question addressed by this chapter was whether schwa is distinct from other vowels in unstressed syllables. This question was approached by examining variation between /ɪ/ and /ə/ in unstressed syllables, through comparing unstressed vowels represented with different spellings. This chapter has shown that, although spelling is sometimes dismissed as not relevant to speech production, it can be a useful tool for investigating variation in certain situations.

The analysis within this chapter was broken up into two smaller questions. The first of these asked how the difference between /ɪ/ and /ə/ is realised in unstressed syllables. It was found that <a> and <i>, used as proxies for /ɪ/ and /ə/, clearly represent two different vowels that differ mainly in F2. This difference was consistent among speakers. This shows that there is a clear difference between at least two unstressed vowel qualities in this group of speakers. However, this difference is not as robust as has often been claimed, existing primarily in backness rather than height.

The second part of the analysis asked whether there was categorical reduction to /ɪ/ and /ə/ in unstressed syllables. This question was approached by examining how vowels represented with <e> patterned. The <e> spellings did not neatly pattern with <i> or <a>, having an intermediate position overall, and differing widely in realisation between speakers. It was shown that for some speakers <e> was intermediate between <a> and <i>. This is a surprising finding in any account that claims full neutralisation to /ɪ/ and /ə/ in this set of vowels (e.g. Flemming and Johnson, 2007). It was therefore argued that, atleast for some speakers in this variety, the unstressed vowels did not vary categorically between /ɪ/ and /ə/. It was argued that this three way pattern could be caused by residual differences between historical full vowel forms. Whatever the exact reason for the three way spelling difference, it suggests that variation in unstressed vowels is more complex than the simple and categorical opposition between /ɪ/ and /ə/ that is sometimes assumed.

It is important to keep in mind that the data in this chapter is only from 26 speakers from Derby. It therefore remains to be seen whether these findings would be found in English more widely or whether they are specific to Derby and related varieties of English. There is some reason to believe that there may be some variation within British English in this respect. The horsES vowel is said to be /ə/ in Derby, where many other varieties may be more likely to be transcribed with /ɪ/ (Foulkes and Docherty, 1999). Although the vowels in suffixes were not examined in this chapter, the vowel used in such suffixes could still be related to other unstressed vowels spelt with <e>. It could well be that other varieties which are said to have the fronter vowel [ɪ] in this position are also likely to have a fronter vowel in unstressed <e> vowels in general. In such varieties therefore it is possible that we may see a more categorical pattern with <e> patterning more exclusively with <i>. The large amount of inter speaker variation we see in Derby could therefore possibly be connected to variation between a more local form /ə/ and a more RP form /ɪ/. Because of the lack of research that examines the phonetic realisation of non-final unstressed vowels this is an open question. Indeed, the rich individual variation found in this chapter hints at why variation in unstressed vowels is worth more investigation.

Notwithstanding this individual variation, the key finding from this chapter is that there is a consistent difference between /ɪ/ and /ə/ across speakers. When discussing the potential for targetlessness in schwa it is thus important to be aware that it is not the only possible realisation for unstressed vowels. The continued importance of spelling in the realisation of unstressed vowels suggests that the historical differences between these vowels may not have completely neutralised. The problem with some of the more controlled studies of targetlessness in schwa that use nonce words (e.g. Browman and Goldstein, 1992; Flemming, 2009) is that they divorce schwa from their historical context and origins. That there is a continued clear difference between unstressed /ɪ/ and /ə/ is not suggestive of a targetless schwa, especially given that the level of variability between /ɪ/ and /ə/ was similar. Of course, this chapter did not directly address the issue of targetlessness. In order to assess this question a more in depth examination of variability in schwa is needed. This is the focus of the following chapter.

The differences in spelling suggests that unstressed vowels spelt with <i> and <e> should not be included in analyses of schwa as they pattern differently. In particular, the fronter average realisations of these vowels do seem to correspond with native speaker intuitions, as shown in pronunciation dictionaries. Consequently, such tokens will therefore not be included in the further analyses of schwa in this thesis. Therefore, the findings of this chapter directly inform the methodology in the rest of the thesis.

4 Synchronic variation in schwa in Derby

4.1 Introduction

This chapter presents an analysis of variation in schwa in speakers from Derby (Milroy et al 1997). The first of the main research questions (p. 22) addressed by this chapter is the question of whether there is evidence of a phonetic target in schwa. This question is addressed in two parts. The first part of the analysis examines the variability and predictability of schwa according to its phonetic context. As discussed in section 2.5, however, high variability and predictability alone cannot be used as measures of targetlessness, since it is possible for a vowel to have a target which it does not often reach. For this reason, the second part of the analysis looks at whether positive evidence of a target in schwa can be found. Unlike the analysis of NZE in Chapter 6, there was no particular reason to suppose that there would be any change in schwa in Derby. Therefore a detailed synchronic analysis of variation in schwa is presented. Variation in schwa is examined according to its formant trajectory, in order to see if it shows deviation towards a target. I also investigate how schwa varies according to duration, with the aim of discovering if there is any indication of schwa moving towards a target as it gets longer.

The second research question that this chapter addresses is the issue of whether word final and non-final schwa are phonologically distinct. In order to answer this question, for all of the analyses performed, data from non-final, pre-consonantal word final, and pre-pausal schwas are compared. This three way comparisons means that it is possible to separate the effects of word position from the effects of the following environment.

Throughout the chapter, the analysis of schwa is also compared to a range of stressed vowels. As discussed in section 2.4.2, this is crucial in order to check that the methods used to assess targetlessness are suitable for doing so, and also to be able to position the realisation of schwa within the wider vowel space.

The chapter starts by introducing the particular questions which are answered within this chapter, and explaining how these will be addressed (4.2). Section 4.3 then describes the methodological choices that were made. The main body of the analysis is presented in 4.4, followed by a discussion of the overall results in 4.5, and finally in 4.6, there is a summary of the contribution of this chapter.

4.2 Research questions

The overall aim of this chapter is to address whether schwa has a phonetic target. Within this broad aim, I focus on two key questions. I first examine whether schwa is more variable and predictable than other fully stressed vowels. As discussed in sections 2.3 and 2.4.3, examining the predictability and variability of schwa has often been used as a way examining the issue of targetlessness is schwa. As discussed in section 2.5, however, variability and predictability do not necessarily imply targetlessness or full context dependence, as it is possible for a vowel to have a target which it does not often reach. It is therefore imperative for any investigation of targetlessness to also look at circumstances where vowels may vary if they have a target. The second part of the analysis therefore moves on to examining whether there is any positive evidence that schwa has a target. I use two methods to do this. I firstly assess whether schwa shows a deviation in its trajectory towards a target. I secondly examine whether the durational variation in schwa points towards it having a phonetic target. I consider evidence from both F1 and F2 measurements in the answer to these questions. Following the discussion in section 2.6, schwas from non-final, pre-consonantal word final, and pre-pausal position are considered. This three way comparison allows for an assessment of the degree to which any differences between word final and non-final schwa are due to word position itself, or whether they are due to the differing following environment. Over the following pages, I set out the predictions and expectations for the data for each of the research questions.

Question 1: Is schwa more variable and context dependent than other vowels?

Why this is relevant to the main research questions:

As was outlined in the introduction (p. 21) a targetless vowel is one with no independent articulation of its own, beyond the minimum needed to be perceived as a vowel. This thus implies that its production will be highly dependent on its phonetic context. Importantly, in terms of variability, maximal assimilation to context has different implications for F1 and F2, as discussed in section 2.4.3. To recap this argument, as the influence on F1 should not vary very much between adjacent consonants, high context dependency predicts low F1 variability. However, the effects that adjacent contexts have on F2 varies much more, and therefore high context dependency predicts high F2 variability. Both the contextual predictability and general variability of the vowels is analysed.

Importantly, the answer to this question is intrinsically linked to whether schwa has a target. If schwa does not have target independent of its phonetic context then it follows that its vowel quality will be highly predictable from phonetic context, and thus also highly variable in F2. This is not necessarily true vice versa, as it is possible that a vowel could be highly predictable according to its phonetic context, but yet still show evidence of movement towards a target. Being highly context dependent and variable in F2 is thus a prerequisite for schwa to be targetless, but cannot in itself prove targetlessness.

Predictions:

Assessment of schwa's predictability according to phonetic context will be addressed through analysis of the effect of phonetic context on its F2 at its midpoint. As explained above, the effects of phonetic context on F1 are expected to be more uniform generally, so this part of the analysis focuses on F2. If schwa is highly dependent on phonetic context we would expect that both the preceding and following phonetic context would have a significant effect on the F2 of schwa at its midpoint.

Unlike the studies discussed in section 2.3, the analysis in this chapter is based on natural speech. Therefore, due to the natural variation in speech, it is not expected that even a targetless schwa would be completely predicable according to context. It should, however, show more dependence on phonetic context than other vowels. An explained in section

2.4.2 it is crucial for analyses of targetlessness in schwa to include comparison to other vowels, and this type of comparison forms an important part of the analysis in this chapter.

In terms of the variability of schwa, the predictions are different for F1 and F2. We would expect the F2 of a targetless schwa to be highly predictable from phonetic context, and thus more variable overall than other vowels. By contrast, we would expect the effects of phonetic context on F1 to be more uniform and thus for the F1 of a targetless schwa to be consistently low. If schwa is targetless we would therefore expect a relatively low variability and standard deviation in F1. The variability of schwa will be assessed by comparing its variability to other vowels along F1 and F2 and also the whole vowel space.

Question 2: Is there any evidence that schwa has a phonetic target?

The focus of question 1 was on the overall predictability and variability of schwa at its midpoint. As was explained in section 2.5, however, it is possible for a vowel to have a phonetic target which it does not often reach, if it is often found in environments that encourage high assimilation to phonetic contexts (e.g. at very short durations). It is therefore imperative to go beyond simply looking at variability and context dependency when assessing targetlessness, and to look for positive evidence of a phonetic target. Consequently, the second part of the analysis focusses on finding evidence of a phonetic target in schwa through its variation according to two key variables. The first of these is variation according to formant trajectory. The second of these is variation in schwa according to duration. The way in which these variables will be utilised to answer this question are each explained in turn.

2a: Does schwa show movement towards a target across its trajectory?

Why this is relevant to the main research questions:

Since a targetless vowel is expected to show maximal assimilation to phonetic context, the examination of vowel formant trajectories is a way of examining whether a given vowel shows any movement towards a phonetic target. Examination of formant trajectories is one of the methods that was used to assess targetlessness by both Kondo (1994) and Flemming

(2009), as described in sections 2.3 and 2.5.1. All vowels are affected by phonetic context to a degree and this will be most apparent at the beginning and end of a vowel. However, it is expected that if a vowel has a phonetic target that it will show a deviation away from its phonetic context in its formant trajectory. A targetless vowel on the other hand would be expected to show a linear formant trajectory, moving straight from the preceding to following context. The exact patterns in the formant trajectories that we would expect to see again differ according to F1 and F2 and are described below.

Predictions:

Surrounding consonants are expected to have a uniform lowering effect on F1, so a maximally assimilatory schwa should have a low F1. If schwa has a phonetic target, we would therefore expect to see a deviation in its F1 trajectory away from its consonantal context to higher F1 values during its trajectory. That said, in a targetless vowel there may not always be complete linearity in its F1 trajectory, as even a targetless vowel still needs to be recognised as a vowel rather than a consonant (Flemming, 2009). In terms of vowel height, therefore, the hypothesis is that a targeted schwa will show evidence of aiming for a height which is lower than the minimum requirement to be perceived as a vowel.

In terms of F2, whether schwa is targeted will be determined by examining the trajectories of schwa in symmetrical phonetic contexts, as was done by Kondo (1994), meaning where the preceding and following contexts are phonetically the same. This is because in asymmetrical contexts, whilst the presence of a deviation or a curve in the trajectory is likely to indicate a target, the absence of a deviation would not be evidence of targetlessness. This is due to the fact that in such cases it would be ambiguous whether the articulators were indeed moving straight from the preceding to the following context, or whether they had passed through the target on the way between the two surrounding contexts.

The hypothesis is that if schwa is targeted then in symmetrical phonetic contexts there should be a deviation in its F2 trajectory, showing a movement away from the phonetic context, as shown in Figure 4.1. If schwa is targetless then in such symmetrical contexts the F2 trajectory should be linear with no deviation to a target. This follows the idea that if schwa is completely targetless in front/backness then its F2 value should be determined by a combination of its preceding and following context, as the articulators should move straight from the position of the preceding context to the following context, without aiming for a target in between, as was found by Kondo (1994) and Flemming (2009). In addition, these patterns will be compared to stressed full vowels for control purposes. If linear formant trajectories are to be used as evidence of phonetic targetlessness in schwa it is expected that stressed full vowels would not have linear formant trajectories in symmetrical contexts. As explained in section 2.4.2, it is necessary for this to be the case in order to show that linearity of formant trajectories is a valid method for identifying targetlessness. Therefore, unlike some of the previous studies of schwa (see section 2.3) a key part of the interpretation of the findings will depend not only on what is seen in schwa, but how it compares to other vowels.

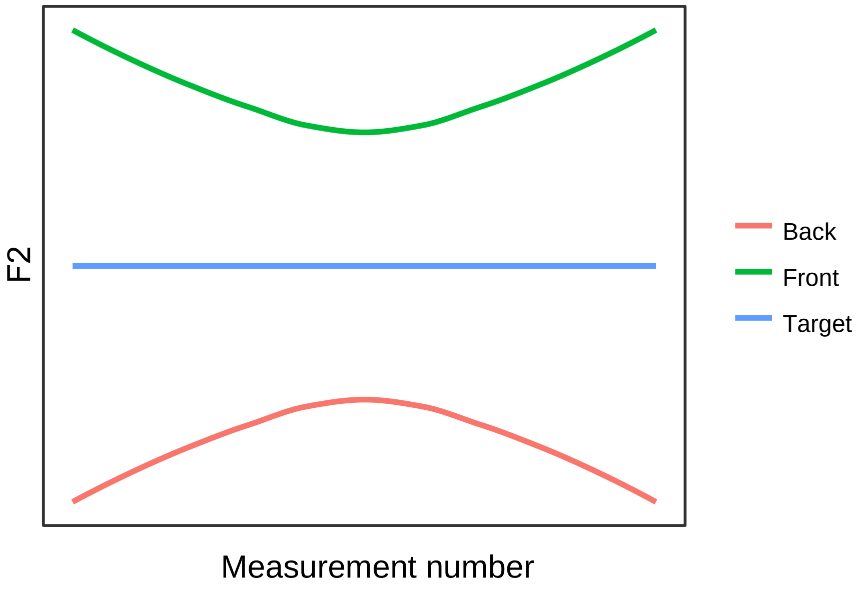


Figure 4.1 *Hypothetical F2 trajectories by context in relation to target*

2b: Does schwa show movement towards a target at longer durations?

Why this is relevant to the main research questions:

As explained in section 2.5.2, a vowel could appear to be very dependable on phonetic context, if it often appears in an environment which encourages a high degree of assimilation to context. Short durations are one such environment. Indeed, all vowels show increased influence of phonetic context at shorter durations (Moon and Lindblom, 1994). Schwa is a very short vowel (e.g. Bates, 1995), and it is therefore possible that the reason it can appear highly dependent on phonetic context could be related to this. In this way, simply looking at averages across all schwas, such as its average formant trajectory, may not always be able identify signs of targetedness if they are not present in the average vowel. Duration is therefore an important variable in assessing targetless, as examining the way in which schwa varies according to its duration will show to what degree schwa's behaviour is constrained by its short duration, and whether it shows more independence from phonetic context at longer durations.

This analysis will examine the effects of duration on the midpoint of schwa. Following Lindblom's (1963) undershoot hypothesis it is proposed that the higher degree of contextual assimilation ascribed to schwa by some may be as a result of undershoot because of short durations rather than targetlessness. The hypothesis is that if schwa has a target then as it gets longer the influence of phonetic context will lessen, and it will become closer to this phonetic target. This because at longer durations the articulators will have had longer to the reach the target, and longer to move away from the effects of the surrounding context. Again, following the discussion in section 2.4.3 the predictions for the data are slightly different for F1 and F2 and are explained below. The way in which schwa is expected to vary over the whole vowel space according to duration is also explained.

Predictions

In terms of F1, it is expected that at longer durations schwa will have a higher F1 as it has more time to move away from its phonetic context. If schwa were to show a stronger relationship with duration than the short high vowels KIT and FOOT, this would be particularly strong evidence of it moving to a height which is lower than the minimum required to be a vowel.

With F2, different phonetic contexts will have different effects that may pull schwa in different directions in relation to its target. It is therefore expected that duration will have different effects on schwa depending on the phonetic context. Consequently it is predicted that if schwa has a phonetic target then there will be an interaction between phonetic context and duration, as duration should have a different effect on schwa depending on the phonetic context. Figure 4.2 demonstrates the kind of hypothetical relationship which is expected here.

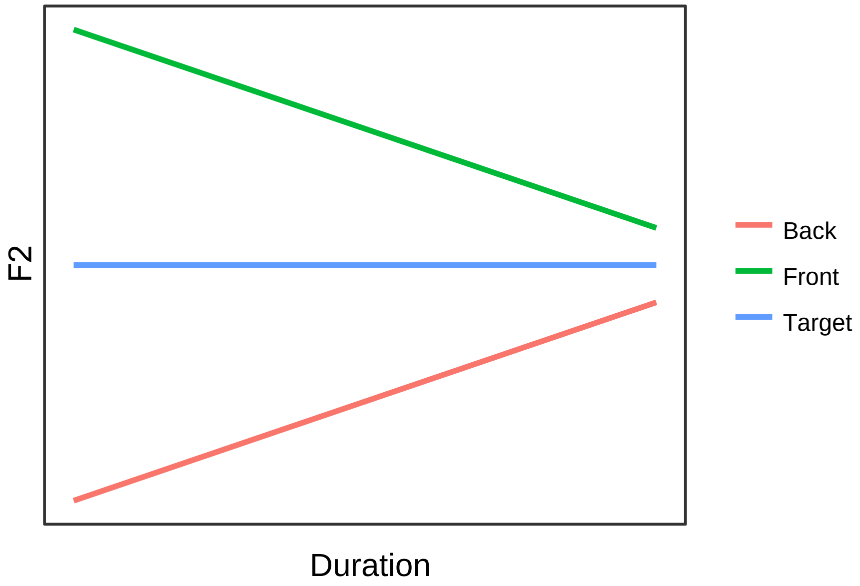


Figure 4.2 *Hypothetical interaction between duration and context on F2 in relation to target*

Conversely if schwa is targetless we would expect no such interaction and no effect of duration in any phonetic context, as the F2 values should be predictable from phonetic context regardless of duration. As a control measure, we would also expect the stressed vowels to show an effect of duration to prove that this method is suitable for detecting targetless.

If schwa shows evidence of a phonetic target in terms of how it is affected by duration, this would suggest that the reason it has been regarded as targetless by some (e.g. Kondo, 1994; Bates, 1995) could simply be a side effect of its short duration rather than inherent dependence on phonetic context.

In terms of overall variability across the vowel space, it is also expected that duration would have an effect. The hypothesis here is that if schwa is targeted then longer schwas should be a) higher in F1, and b) less varied in F2, as the influence of phonetic context decreases. Conversely, if schwa is targetless then the hypothesis would be that schwa is unchanged according to duration, and should be a fairly high vowel which is variable in backness at all durations. Figures 4.3 and 4.4 below show what the distribution of tokens over the vowel space should look like according to these different scenarios. Targetless schwa is expected to look like Figure 4.3 regardless of duration, whereas targeted schwa may look more like 4.3 at shorter durations, but is expected to look more like Figure 4.4 as duration increases.

Although F1 and F2 analyses will be run separately on the data, considering the relationship between F1 and F2 is particularly important in the assessment of targetlessness. Unlike some analyses before (see section 2.3 p.31 and section 2.4.3 p.37-8) this thesis does not take the view that it is possible for a vowel to be targeted on one dimension and not on the other. Therefore, if the predictions by these hypothetical plots are borne out this would not only provide evidence for schwa being a targeted vowel, but also for the importance of considering F1 and F2 in combination.

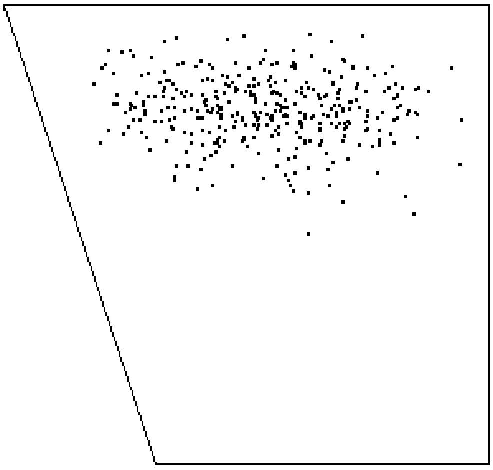


Figure 4.3 *Hypothetical targetless schwa/short schwa*

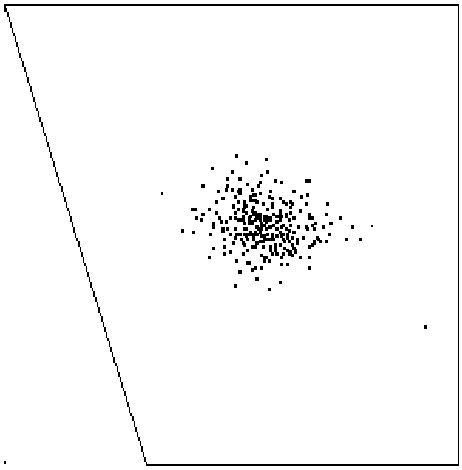


Figure 4.4 *Hypothetical targeted schwa*

4.3 Methodology

In the analysis in this chapter the data preparation steps taken in sections 3.5.1-3.5.4 all apply here. This section therefore describes the additional data that was used in this chapter, and the additional data filtering steps that were undertaken. Not that the majority of the analyses use formant measurements taken at the midpoint of the vowels, excepting the formant trajectory analysis in 4.4.3.

In this chapter measurements from all 31 speakers whose data were extracted are used. This chapter uses schwa data from non-final schwas, pre-consonantal word final schwas, and pre-pausal schwas. Following the findings in Chapter 3, unstressed vowel tokens spelt with <e> and <i> are not included in the analysis. The findings in Chapter 3 showed that these unstressed vowels may variably be part of a separate /ɪ/ lexical set, and that they had fronter realisations than vowels spelt with <a>. In order to focus on unambiguous schwa vowels these vowels were therefore excluded.

In addition, a range of stressed vowels were also measured and analysed, in order to compare the results of the schwa analyses with vowels that unambiguously have a phonetic target. As stressed vowels were just measured for comparison purposes and are not the main focus of this research, the data for the stressed vowels was taken from a smaller subset of the speakers. This subset was designed to be fairly representative of the overall data set, so 2 of each gender and age combination were used.

Like the unstressed vowels analysed (see section 3.5.2) stressed vowels were also only included if they were in a word of two or more syllables, in order for a valid comparison with schwa. The stressed vowels chosen are grouped in terms of Well's lexical sets (1982). The NURSE vowel was chosen as its target is generally said to be mid and central, phonetically similar to [ə]. KIT and FOOT were chosen as these vowel qualities are said to be possible to contrast with schwa in unstressed syllables (Gimson, 1962). In addition, as non- final schwa has previously been found to be a fairly high vowel with wide ranging backness (e.g. Kondo, 1994, Flemming, 2009), its vowel quality is expected to be somewhat overlapping with KIT and FOOT. These vowels were also chosen as they are relatively short and high, so are potentially the most similar to what a targetless schwa could look like. The selection of stressed vowels was chosen in order to ensure a mix of lengths, heights, and front and back vowels. In addition, therefore, THOUGHT and TRAP were also chosen on the basis of being back and low respectively. The final numbers of tokens analysed from all vowel categories are listed in Table 4.1.

Table 4.1 *Token numbers and examples*

Lexical set	Examples	Number of tokens
KIT	kitchen, assistant	326
FOOT	Cooker, woman	180
NURSE	Journey, dessert	119
THOUGHT	Forty, afford	193
TRAP	Passage, kebab	212
Pre-consonantal word final schwa	Formula one, other problems	1130
Pre-pausal schwa	Partner , Africa	347
Non-final schwa	Today, barracks	1514

Two types of model are used in this chapter. In section 4.4.2 the focus is on the predictability of the vowels from phonetic context, and this is assessed using mixed effects linear regression models. Sections 4.4.3 and 4.4.4 assess how the vowels vary though their formant trajectory, and according to their duration respectively. In these sections, GAMMs are used in order to be able to capture any potential non-linearity in the formant trajectories, or in the effect of duration on formant values.

All GAMMs include random intercepts for speaker and word. Where differences between vowels and phonetic context are examined within a model these are also included as random slopes by speaker, where doing so allowed for model convergence. An important feature of GAMMs is that in addition to random slopes and intercepts they also allow for random smooths, which is particularly important in analysing formant trajectories. All of the modelling of formant trajectories includes random smooths by each individual token. This means that the random variation in the shape of the formant trajectories across each individual token is accounted for in the modelling.

As there are many large models used in this chapter, in most cases the model effects and full model structures are not shown in the text here, but can be found in the appendix. Instead, I show the main important trends graphically. Where the significance of an effect is important the coefficients are in the text.

4.4 Results

This analysis addresses each aspect of the two research questions in turn. I first address the two aspects of research question one; variability and then predictability according to context. The analysis then moves to look at the two measures that were used to find evidence of a phonetic target (question two); first examining variation in schwa across its trajectory, and then by duration.

4.4.1 Variability

Figure 4.5 shows the comparative spread of F1 values compared across all vowels. We can see that the average F1 of non-final and pre-consonantal word final schwa is very similar to that of FOOT and KIT, showing that on average they both have a fairly high realisation. A low F1 is what is expected in a targetless schwa. However, the shape of the distribution in pre-consonantal schwa is quite different from KIT and FOOT. As can be seen in both Figure 4.5, and the standard deviations in Table 4.2, the F1 values for schwa are more spread than for KIT and FOOT. This is as was predicted if schwa shows movement towards a target in certain situations. Pre-pausal schwa is on average lower than other schwas and also shows a high degree of F1 variability. Interestingly, the other vowel which shows a high degree of F1 variability is TRAP. TRAP has a fairly high proportion of tokens which are realised with fairly low F1s. This shows that even vowels with lower targets, such as TRAP, can be realised with fairly low F1 values. As this data is from spontaneous speech it is clear that a degree of phonetic reduction is expected in all vowels, including stressed vowels. Therefore the fact that schwa is on average fairly low in F1 does not necessarily indicate targetlessness. Rather, the high level of F1 variability in schwa suggests that it may vary between productions which are more affected and less affected by phonetic reduction. Indeed, the high F1 variability of schwa is suggestive of a vowel quality target which is lower than the minimum requirement to be perceived as a vowel.

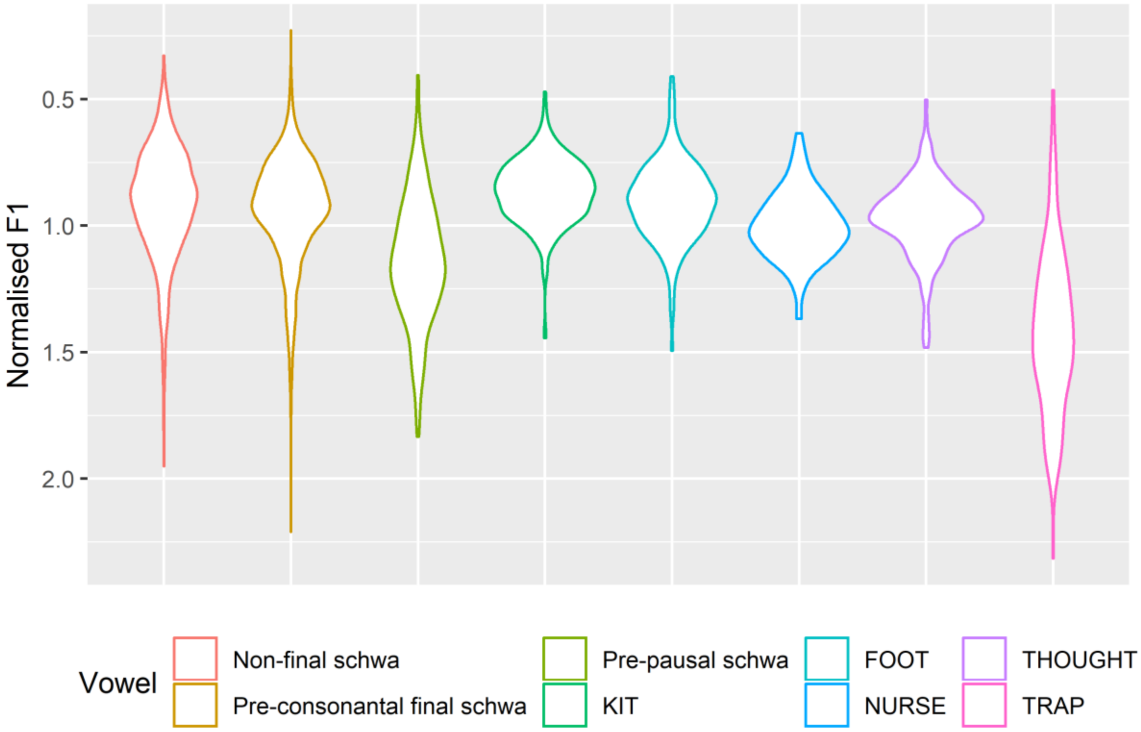


Figure 4.5 *Violin plot of normalised F1 by vowel*

Table 4.2 *Normalised F1 means and standard deviations for all vowels*

Vowel	Mean	Standard deviation
Non-final schwa	0.91	0.22
Pre-consonantal word final schwa	0.94	0.21
Pre-pausal schwa	1.14	0.25
KIT	0.86	0.12
FOOT	0.91	0.15
NURSE	1.00	0.13
THOUGHT	0.97	0.15
TRAP	1.41	0.31

In terms of F2 we find that, as would be expected, all three schwas are on average roughly central within the vowel space, as seen in Figure 4.6. Around this average value, however, there is a high level of F2 variability for schwas which are not before a pause. As can be seen in the standard deviations though, the amount of variability in schwa is not exceptional as KIT and particularly FOOT also have similarly high standard deviations in their F2 values Table 4.3). This suggests that there is a link between being phonetically high and F2 variability. In pre-pausal schwa, which is a lower vowel, there is less F2 variability.

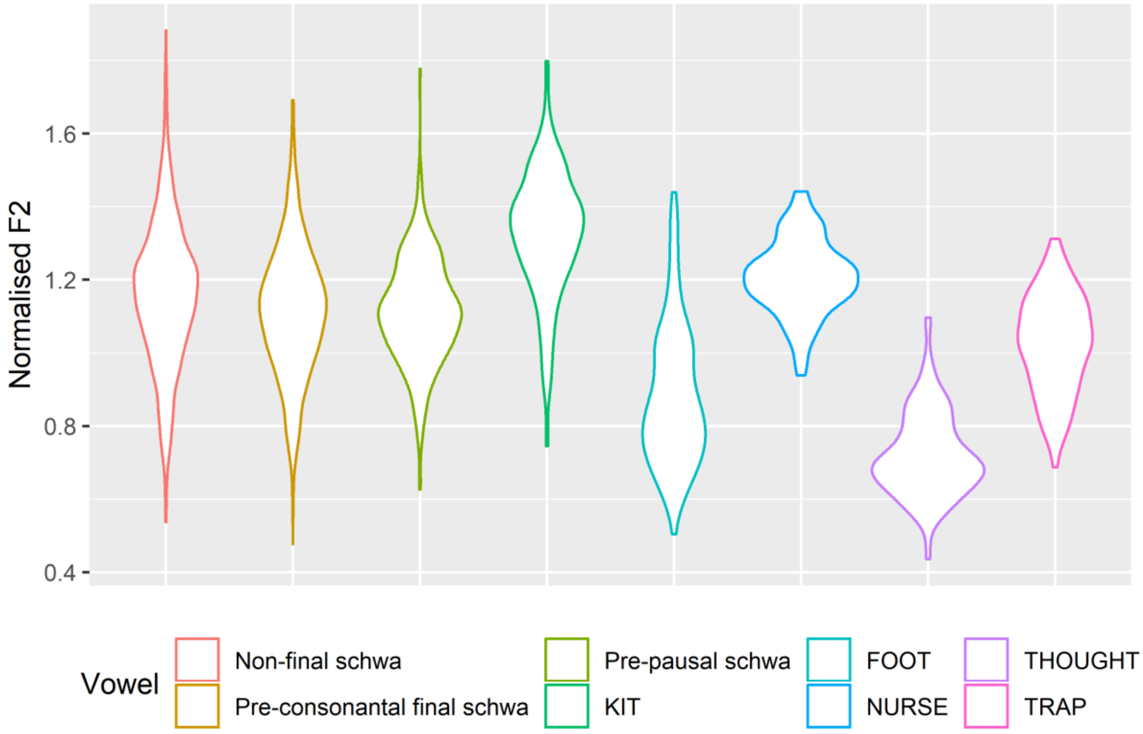


***Figure 4.6** Violin plot of normalised F2 by vowel*

***Table 4.3** Normalised F2 means and standard deviations for all vowels*

Vowel	Mean	Standard deviation
Non-final schwa	1.16	0.20
Pre-consonantal word final schwa	1.11	0.18
Pre-pausal schwa	1.11	0.14
KIT	1.34	0.17
FOOT	0.88	0.19
NURSE	1.20	0.10
THOUGHT	0.72	0.12
TRAP	1.02	0.14

Figure 4.7 shows the spread of the tokens over the whole vowel space. Comparing all of the vowel plots, it is clear that both non-final and pre-consonantal schwa are the most spread over the whole vowel space. Whereas there are other vowels that show high F1 variability (TRAP) or high F2 variability (KIT and FOOT), the variability in pre-consonantal schwa is striking because of the large variability on both dimensions. As was argued in section 2.4.3, however, high F1 variability is not actually expected in a targetless vowel. Therefore, the fact that schwa has a relatively high F1 standard deviation is actually evidence against it being targetless, as we would expect a maximally assimilatory schwa to have consistently low F1 values. If the spread of tokens is compared with the predictive Figures 4.3 and 4.4 on in section 4.2, many of the pre-consonantal schwas do indeed occupy the range in the Figure 4.3, which shows what would be expected in a targetless schwa. Notably though, the fact that in Figure 4.7 both non-final schwa and pre-consonantal word final schwa also show many tokens which are lower is evidence against pre-consonantal schwa being targetless.

Pre-pausal schwa is actually more similar to the predictive Figure 4.4 which shows a hypothetical prediction for targeted schwa, as it is both lower in the vowel space and less spread in F2 than pre-consonantal schwa. Importantly, here the division in the realisation of schwas is between pre-consonantal and pre-pausal schwa, and not between word final and non-final schwa.

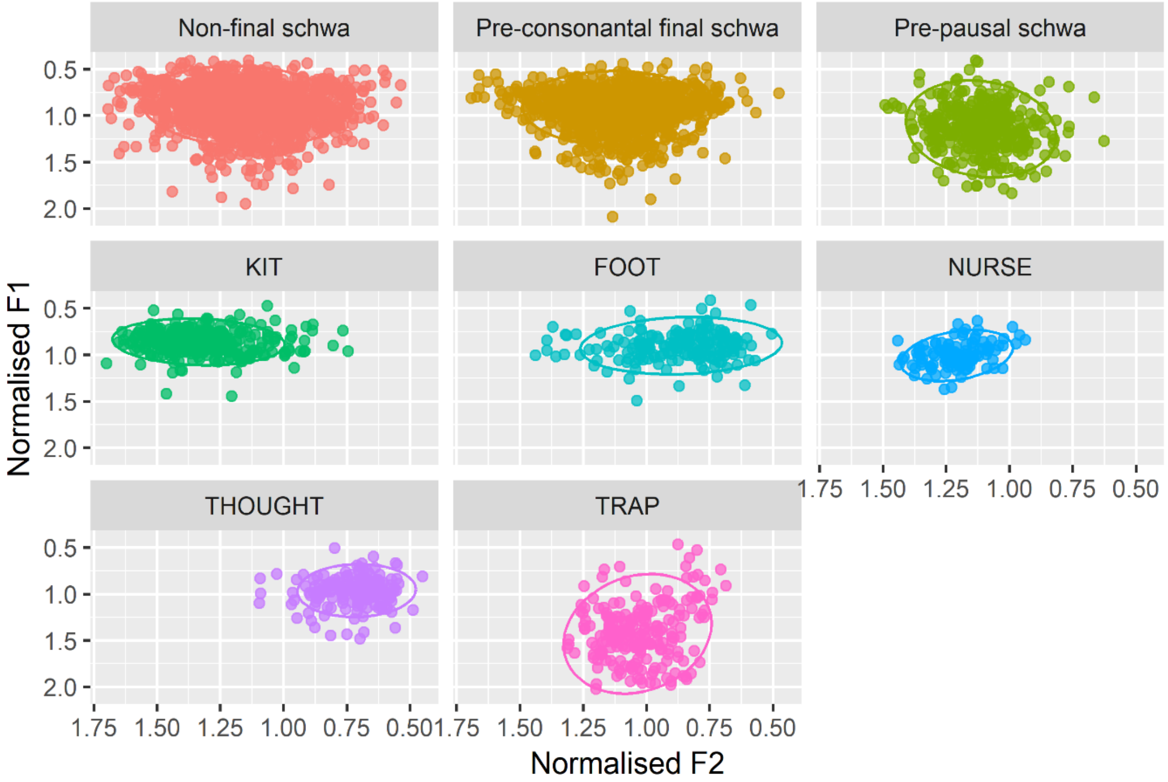


Figure 4.7 *Normalised vowel plots by vowel*

In summary, the F2 variability in schwa is high, as is expected if it is targetless. However, the F1 variability is also high, which is not what is expected from a targetless schwa. The way in which the variability of schwa is constrained by phonetic context and duration will be explored through the rest of this chapter.

4.4.2 Predictability

The analysis in this section assesses to what degree the variability in the F2 of the vowels is predictable by phonetic context.

In order to test the effect of phonetic context on the vowel midpoints, mixed effect linear regression models were run on the data. Models were run individually on each vowel (models 4.1-4.8). The models included the fixed effects of the preceding and following phonetic context, with random intercepts for speaker and word. Where significant effects are reported these are derived from likelihood ratio model comparisons with and without the effect in question.

For all of the vowels other than pre-pausal schwa, there were three levels of front, central and back for both the preceding and following environment. For the pre-pausal schwas these three levels just occurred for the preceding environment. For pre-consonantal word final schwa the following environment was from the consonant that began the following word.

Phonetic context has a clear and significant effect on the F2 of all three schwa types, as shown in Figures 4.8-10 below. Both the preceding and following context have a significant effect on the F2 of non-final and pre-consonantal word final schwa (all p<0.0001). Although pre-pausal schwa does not have a following context, its F2 is significantly affected by its preceding context (p<0.0001). The exact coefficients are shown in Table 4.4. There are some slight differences between the different schwa types. The size of the effects are biggest for non-final schwa, and smallest for pre-pausal schwa. It is perhaps to be expected that context may have slightly smaller effect on pre-pausal schwa as it longer than the other schwas, so the midpoint will be further away from the adjacent segment. However, what is clear is that all three schwas types show clear effects of phonetic context.

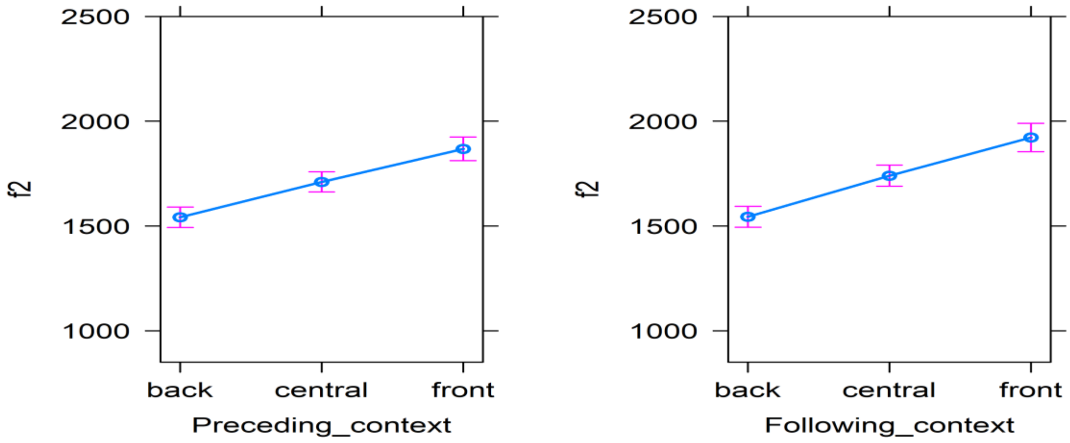


Figure 4.8 *Effect of phonetic context on F2 (Hz) of non-final schwa*

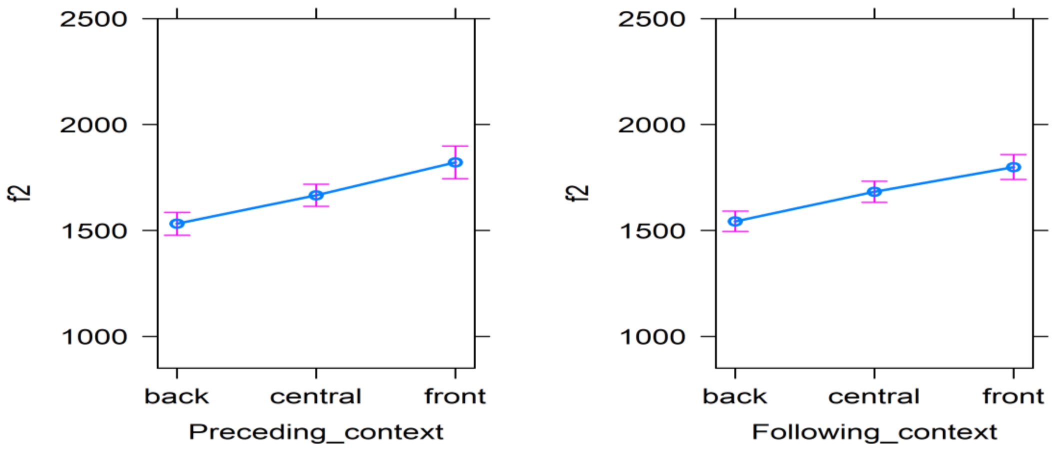


Figure 4.9 *Effect of phonetic context on F2 (Hz) of pre-consonantal word final schwa*

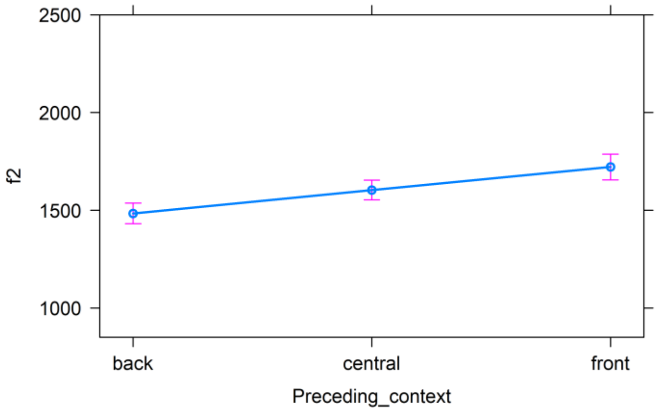


Figure 4.10 *Effect of phonetic context on F2 (Hz) of pre-pausal schwa*

Figures 4.11-4.15 show the results of the same models for the stressed vowels. All of the stressed vowels tested are also significantly affected by phonetic context, although not all of them show a significant effect of both the preceding and following context. Both the preceding (p<0.0001) and following context (p<0.01) have a significant effect on FOOT. It is also the case that both the preceding (p<0.0001) and following context (p<0.001) have a significant effect on TRAP. However, the F2 of THOUGHT and NURSE is only significantly affected by the preceding context (both p<0.001), and KIT by the following context (p<0.0001). Even though there are some significant effects of phonetic context on stressed vowels, as Figures 4.11-15 shows these coarticulatory effects are not as consistent as for schwa. Whereas for schwa, there is always a clear three way difference between back, central and front context, for many of the full vowels this is not the case.

It is also apparent that for TRAP, THOUGHT and NURSE the size of the differences between the environments is smaller than for any of the schwa types. The relatively small effect sizes for these vowels can be seen 4.11-15, and also in the comparatively small estimates that these vowels have (shown in Table 4.4). It is perhaps expected that the F2 of these vowels may be less predictable from phonetic context, as it was shown in section 4.4.1 that their F2 is less variable than schwa generally. KIT and FOOT, on the other hand, show some slightly bigger effects of phonetic context. The size of the effect of the following context on KIT is comparable to that of schwa, and the effect of the preceding context on FOOT is comparable to schwa. It is perhaps expected that KIT and FOOT would show a higher degree of variation according to phonetic context than the other stressed vowels, as they also show a higher degree of overall F2 variability (seen in section 4.4.1). It is, however, important to note that even where the effect sizes are reasonably large for KIT and FOOT the standard error is considerably higher than for any of the schwa types. This is seen in Figures 4.11 and 4.12 in the larger standard error bars, and also is also apparent in Table 4.4. This suggests that phonetic context is a better predictor of the F2 of schwa than of KIT or FOOT, and indeed any of the other stressed vowels.

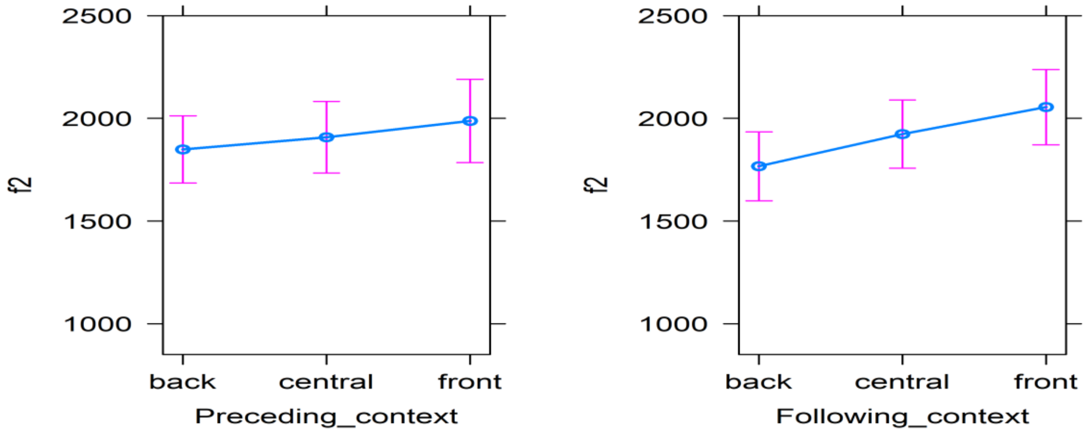


***Figure 4.11** Effect of phonetic context on F2 (Hz) of KIT*

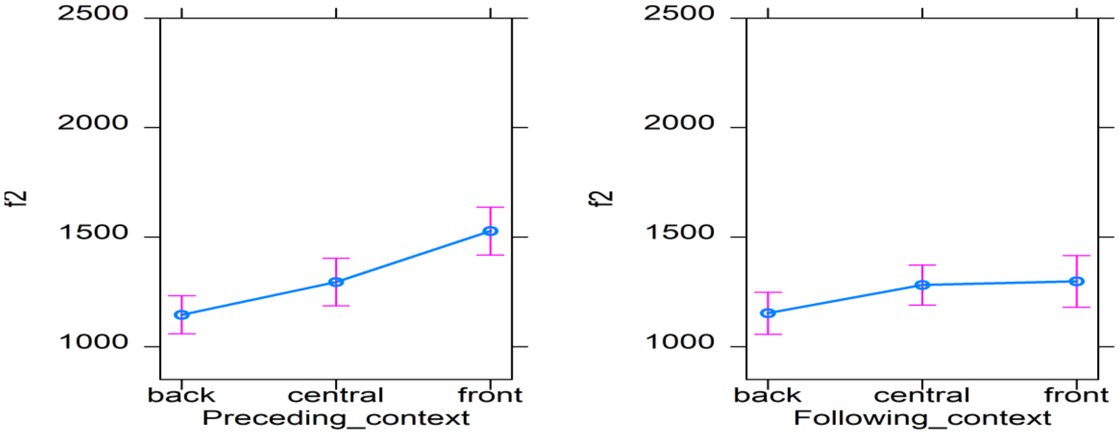


Figure 4.12 *Effect of phonetic context on F2 (Hz) of* FOOT

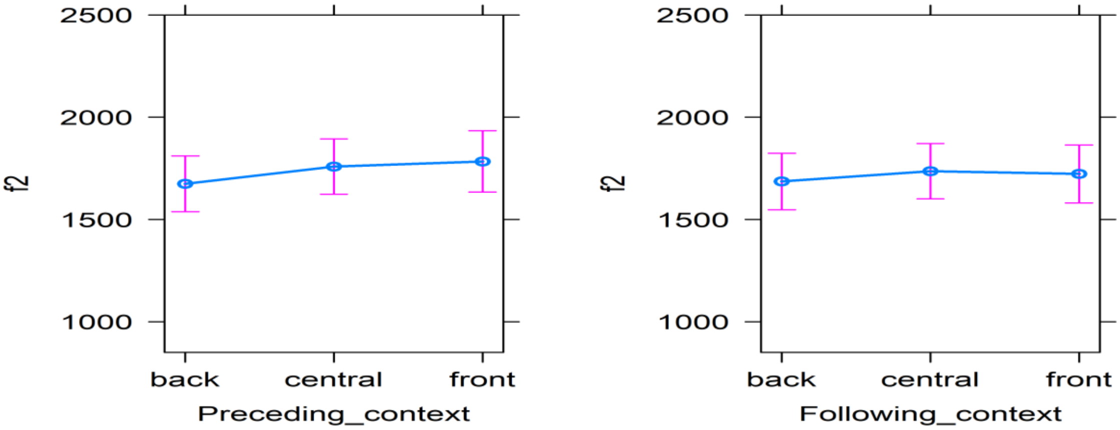


Figure 4.13 *Effect of phonetic context on F2 (Hz) of* NURSE

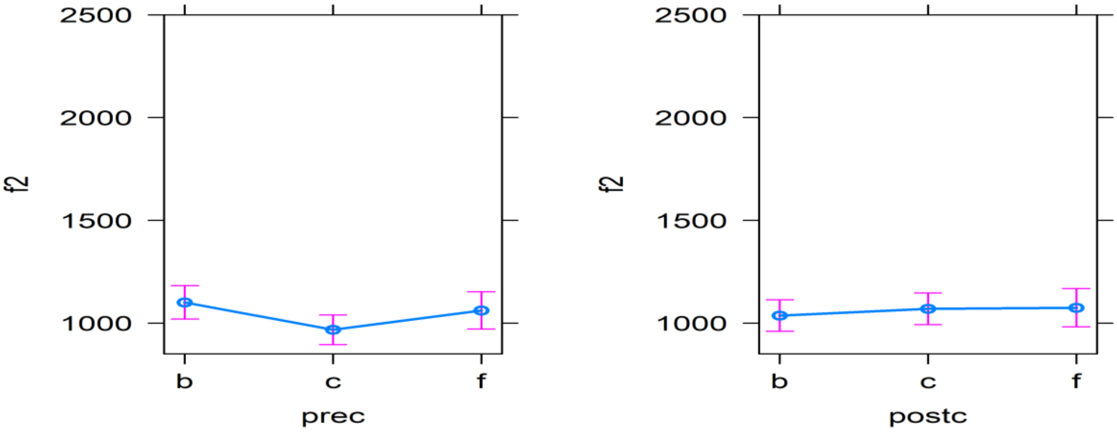


Figure 4.14 *Effect of phonetic context on F2 (Hz) of* THOUGHT

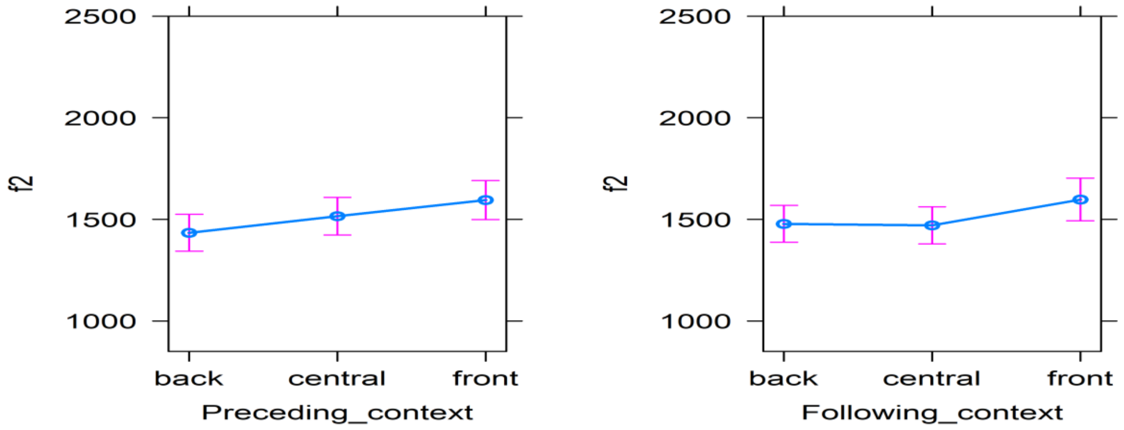


Figure 4.15 *Effect of phonetic context on F2 (Hz) of* TRAP

Table 4.4 *Fixed effects of phonetic context for all vowels*

	Estimate	Std. error	*t*-value
Non-final schwa			
(Intercept)	1787.74	26.69	66.98
Preceding-back	-168.59	13.95	-12.083
Preceding-front	157.43	19.76	7.966
Following-back	-195.74	18.67	-10.484
Following-front	182.03	29.67	6.135
Pre-consonantal final schwa			
(Intercept)	1729.39	27.84	62.12
Preceding-back	-134.24	24.11	-5.569
Preceding-front	155.51	36.66	4.242
Following-back	-138.96	12.98	-10.709
Following-front	116.63	21.57	5.406
Pre-pausal schwa			
(Intercept)	1603.73	25.66	62.498
Following-back	-119.73	21.31	-5.618
Following-front	117.89	29.02	4.063
KIT			
(Intercept)	1957.8	90.28	21.685
Preceding-back	-58.84	41.91	-1.404
Preceding-front	79.25	73.75	1.075
Following-back	-156.91	37.58	-4.176
Following-front	130.94	53.36	2.454
FOOT			
(Intercept)	1336.02	61.32	21.788
Preceding-back	-149.34	43.56	-3.428
Preceding-front	232.39	56.34	4.125
Following-back	-128.5	39.54	-3.25

Following-front	16.57	49.72	0.333
NURSE			
Intercept	1774.52	68.97	25.729
Preceding-back	-83.92	22.1	-3.797
Preceding-front	25.31	38.68	0.654
Following-back	-50.12	24.64	-2.034
Following-front	-13.19	29.78	-0.443
THOUGHT			
Intercept	1077.37	44.12	24.417
Preceding-back	-133.39	32.66	-4.084
Preceding-front	-38.89	41.65	-0.934
Following-back	32.93	28.6	1.152
Following-front	37.64	37.59	1.001
TRAP			
Intercept	1500.738	48.082	31.212
Preceding-back	-81.282	20.504	-3.964
Preceding-front	79.589	25.833	3.081
Following-back	7.858	19.489	0.403
Following-front	127.575	32.777	3.892

In summary, phonetic context has a significant effect on all the vowels examined here, which serves as a reminder that schwa is not unique in being affected by its surrounding context. Even the midpoint of the longer vowels is affected by phonetic context to a degree. Rather, it is the extent to which schwa is affected by context that differs. Its midpoint F2 value clearly differs more according to phonetic context than the stressed vowels measured here. However, there does appear to be a general relationship between the length of the vowel and the degree of influence of phonetic context. Other than schwa, it is the next shortest vowels KIT and FOOT (see 4.4.4.1 for a full comparisons of duration values between vowels) which show the highest degree of influence of phonetic context. It is perhaps unsurprising that the longer vowels show less effect of phonetic context on their midpoints, as the longer the vowel the further the midpoint is from either adjacent consonant. Therefore it seems that schwa is highly variable according to phonetic context, although this evidence alone does not suggest that it is targetless.

4.4.3 Variation by formant trajectory

In this section the variation in formant values across the vowel trajectories is examined. In order to find evidence of a phonetic target, deviations in the formant trajectories are expected. It was predicted that if vowels display evidence of a phonetic target that they will show movement away from their consonantal context during their trajectory. In the case of vowels that are surrounded by two adjacent consonants this means that they are expected to show movement away from their contexts towards the middle of the trajectory. This is the case with most of the vowels tested here. Pre-pausal schwa, however, only has a preceding consonantal context. In this case, it is expected that the influence of phonetic context will diminish through the schwa, as the influence of the preceding consonant decreases.

As the consonantal context should have a uniform lowering effect on F1 on average, the F1 results are modelled on all tokens from each vowel together. It is expected that if there is evidence of a phonetic target, that F1 values should rise as vowels move away from their consonantal context.

As different phonetic contexts are expected to have different effects on F2, the models for F2 consider the effects of phonetic context. In terms of F2, it is also expected that vowels will show a deviation away from their consonantal context. However, in terms of F2, the movements over the formant trajectory are expected to vary depending on the phonetic context.

4.4.3.1 F1

Figure 4.16-4.18 below shows the F1 formant trajectories for the three schwa types (models 4.9-4.11). All three show a significant change in F1 over the trajectory (non-final schwa: edf=6.525, F=11.276, $p<0.0001$; pre-consonantal word final schwa: edf=5.527, F=5.725, $p<0.0001$, pre-pausal schwa: edf=5.88, F=21.038, $p<0.0001$). The non-final and pre-consonantal word final schwas are both surrounded by adjacent consonants. They both show movement away from their phonetic context through the trajectory, showing lower F1 values at their onset and offset that raise towards the middle of the trajectory, although for word final schwa in particular, this effect is very small. In pre-pausal schwa, where there is not a following consonant, this effect is much clearer. Here, it can be seen in Figure 4.17 that the F1 of the vowel is lowest at its onset and raises thorough the first half of the formant trajectory.

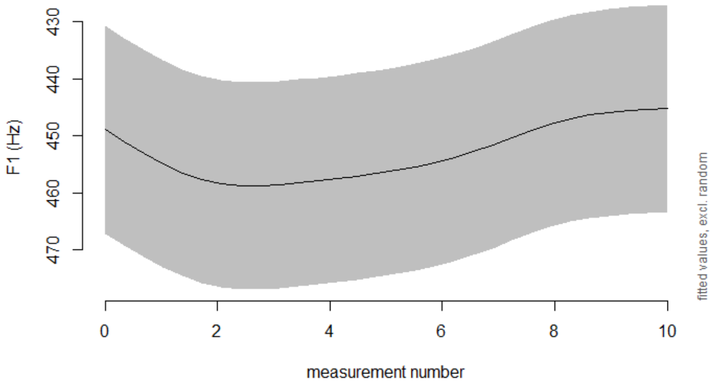


Figure 4.16 *F1 (Hz) trajectory of non-final schwa*

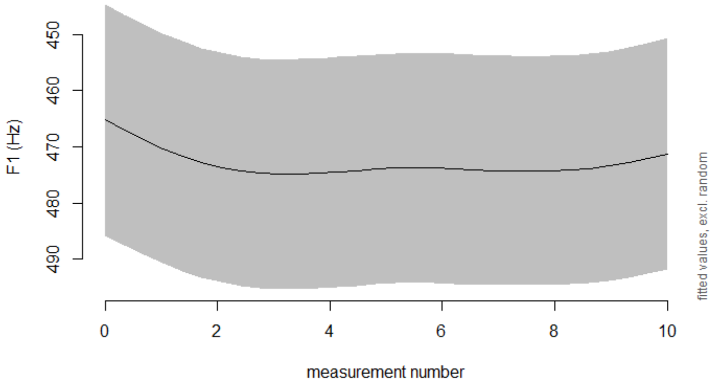


Figure 4.17 *F1 (Hz) trajectory of pre-consonantal word final schwa*

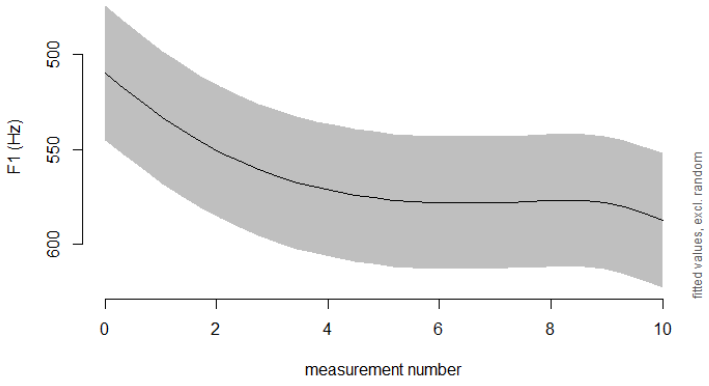


Figure 4.18 *F1 (Hz) trajectory of pre-pausal schwa*

All the stressed vowels also show significant changes in their F1 over the course of their duration (Figure 4.19-23). In most cases the pattern shown is the same as for pre-consonantal schwa. The general pattern is that F1 is lower at the onset and offset, where the influence of the consonant is greater, and the F1 raises towards the middle of the vowel. THOUGHT is the only vowel that does not show this pattern at all. The F1 of THOUGHT actually raises throughout its trajectory.

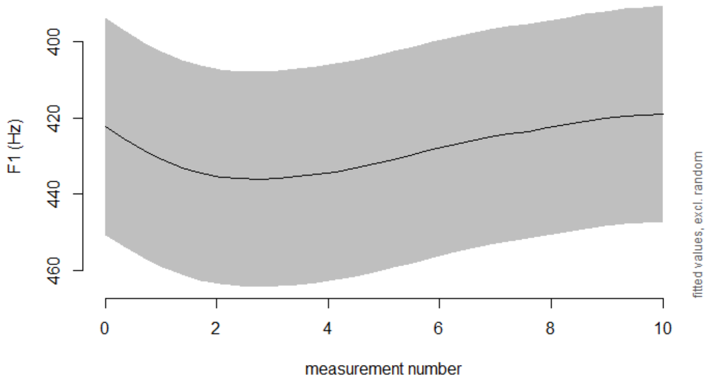


Figure 4.19 *F1 (Hz) trajectory of KIT*

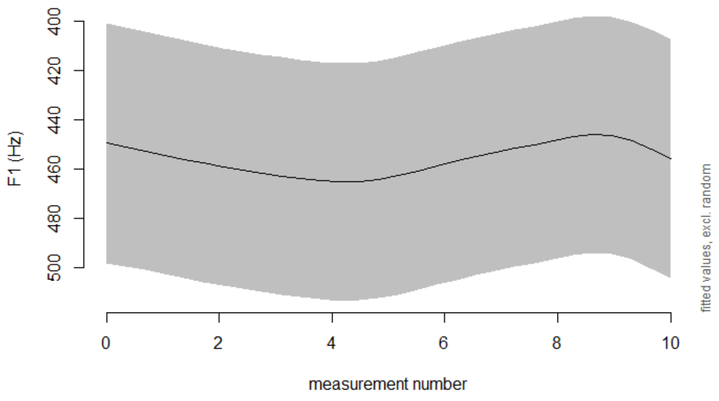


Figure 4.20 *F1 (Hz) trajectory of* FOOT

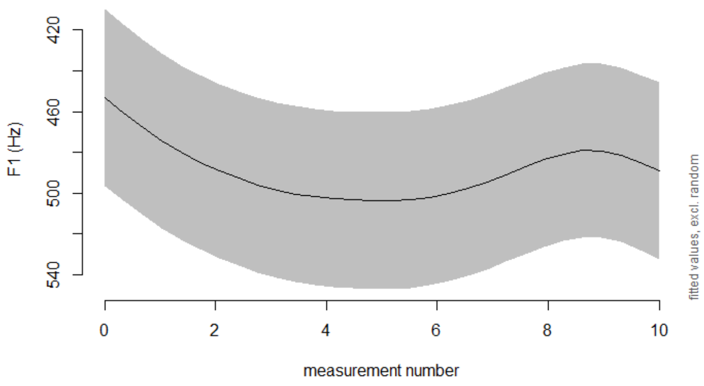


Figure 4.21 *F1 (Hz) trajectory of* NURSE

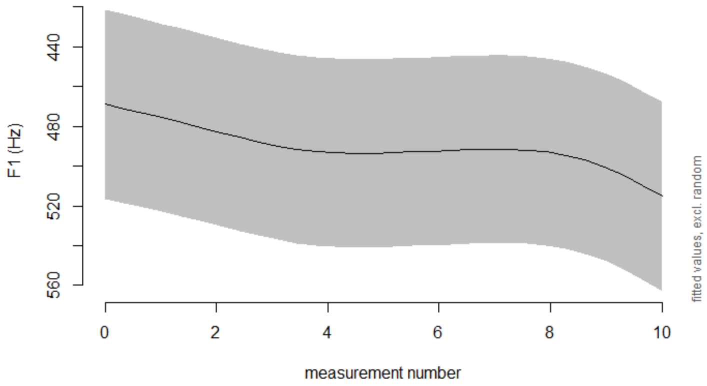


Figure 4.22 *F1 (Hz) trajectory of* THOUGHT

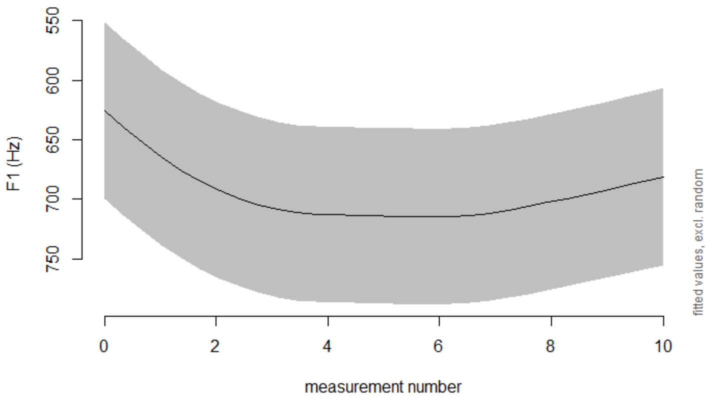


Figure 4.23 *F1 (Hz) trajectory of non-final* TRAP

Despite the significant changes in all of the vowels across their trajectories, it is important to note that in most cases the variation across the trajectories is not particularly large. The changes shown for non-final schwa (Figure 4.16), pre-consonantal word final schwa (Figure 4.17), and KIT (Figure 19) are particularly small. Figures 4.16-4.23 show that the majority of the vowels have very large confidence intervals around the predicted F1.

Figures 4.24-4.27 compare the trajectories of the three schwa types with KIT within the height of the vowel space. It is particularly clear here that many of the changes across the trajectories are fairly small. This is particularly the case for the pre-consonantal schwas (Figure 4.24 and Figures 4.25), and for KIT (Figure 4.27). Although they all show a deviation in their F1 trajectories, these are subtle changes and do not represent large movements over the whole vowel space.

On average, KIT has the lowest F1 out of the vowels measured (as shown in section 4.4.1). Although both schwas show a deviation in their F1 trajectory, as can be seen below, there is no drastic difference between their trajectories, and that of KIT. Indeed, the height of non-final schwa and KIT is very similar over the whole of the trajectory. Therefore the evidence from these F1 trajectories does not show that the pre-consonantal schwas are not phonologically high, or that the aim in terms of height is lower than the minimal requirement for a vowel.

Pre-pausal schwa (Figure 4.26) clearly shows greater change over its trajectory than either non-final or pre-consonantal final schwa. Whilst non-final schwas and pre-consonantal final schwas are fairly high in the vowel space, pre-pausal schwas are more mid. However, this is only the case in the later part of their trajectory. In the earlier part of the trajectory, where pre-pausal schwa is closer to the preceding consonant, it is actually a lot more similar to the other schwas. This suggests that part of the reason for the difference between the height of pre-consonantal and pre-pausal schwa may be the fact that pre-pausal schwa is not surrounded by two consonants.

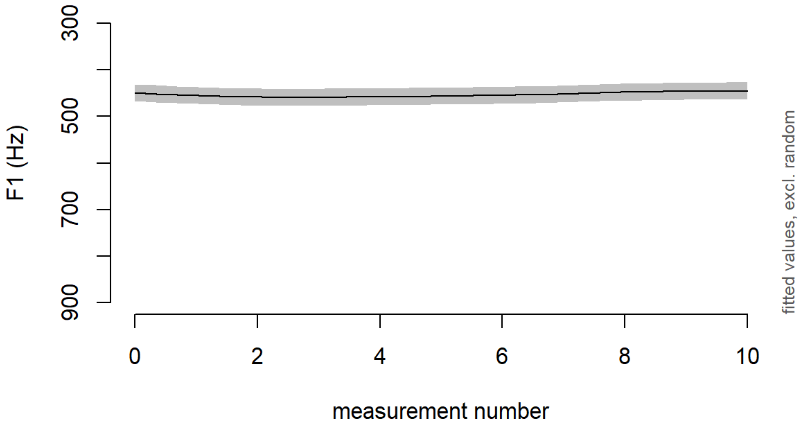


Figure 4.24 *F1 (Hz) trajectory of non-final schwa over whole vowel space*

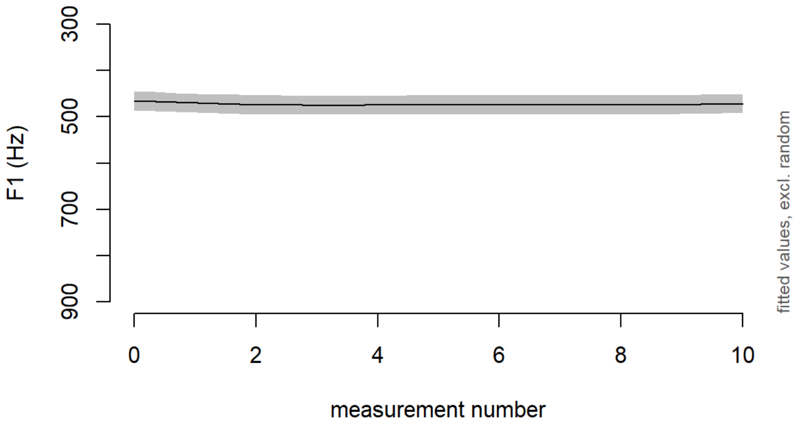


Figure 4.25 *F1 (Hz) trajectory of word final pre-consonantal schwa over whole vowel space*

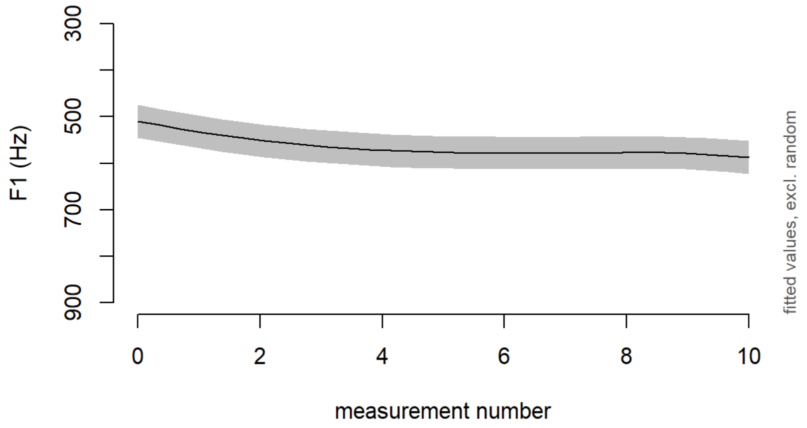


Figure 4.26 *F1 (Hz) trajectory of pre-pausal schwa over whole vowel space*

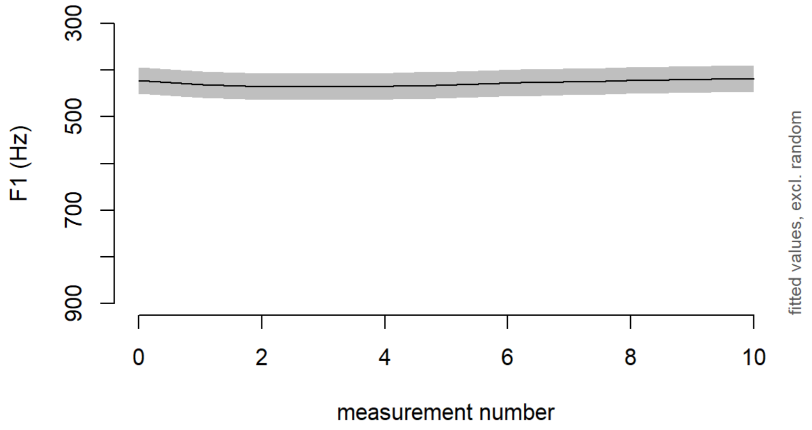
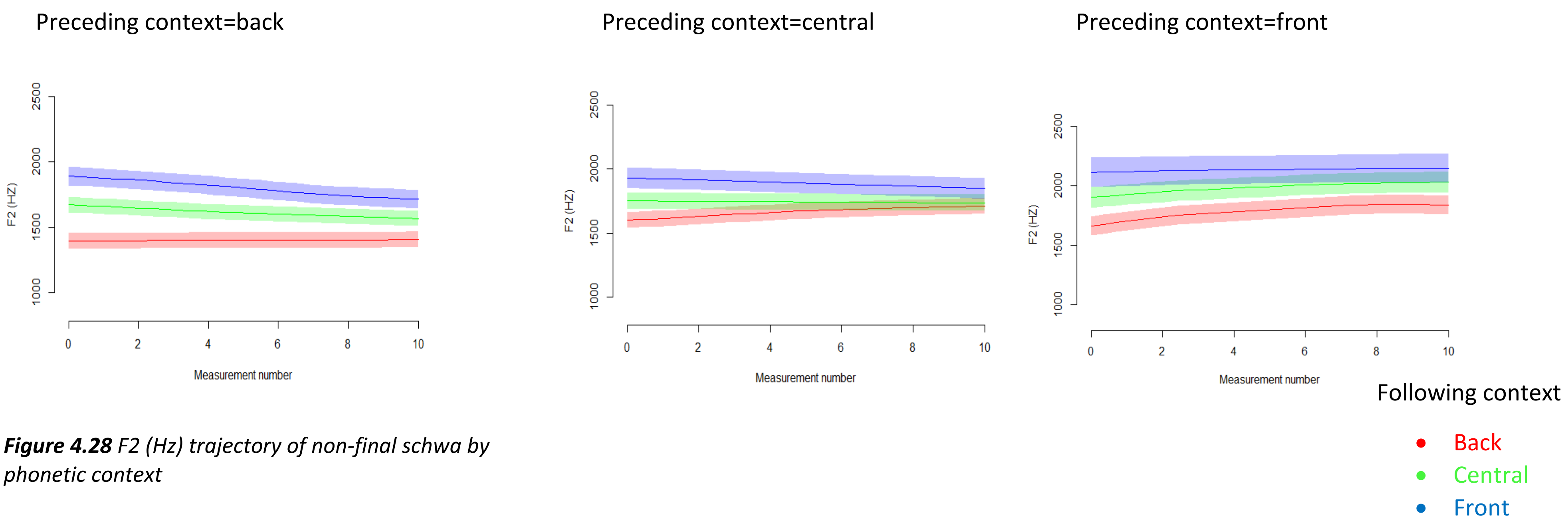


Figure 4.27 *F1 (Hz) trajectory of* KIT *over whole vowel space*

4.4.3.2 F2

In this section the way that F2 formant trajectories differ according to phonetic context is examined. In order to show evidence of a phonetic target, a deviation in the F2 trajectory in at least some environments is expected.

Figures 4.28 and 4.29 below show how the trajectories of non-final and pre-consonantal word final schwa differ according to each individual environment combination. The overall picture is very similar for both schwa types. The effects of the preceding and following context are clear, with the different environments consistently patterning in the way in which we would expect. Within each preceding environment the following environments pattern in the following way: the back contexts are the lowest in F2, followed by central and then front environments. Vice versa, within every following environment the preceding contexts also pattern in the expected order. This is true for both types of schwas.


Preceding context=back

Preceding context=central

Preceding context=front

Figure 4.28 *F2 (Hz) trajectory of non-final schwa by phonetic context*

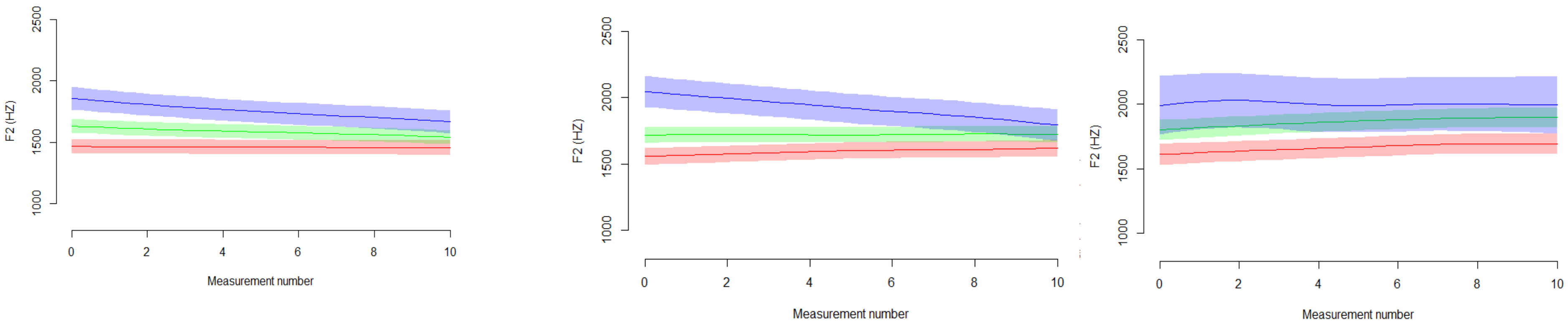


Figure 4.29 *F2 (Hz) trajectory of word final pre-consonantal schwa by phonetic context*

The formant trajectories in Figures 4.28 and 4.29 do not display any evidence of an F2 target in the way that was predicted. Both non-final and word final schwa are the same in this way, despite the following environment for word final schwa occurring over a word boundary. For the most part, the trajectories are almost completely linear, clearly influenced by the preceding and following environment. The trajectories all show a clear movement from the preceding to the following context. For example, where the preceding environment is back and the following environment is front both schwas show raising of F2 values over the course of the trajectory. Most importantly, in symmetrical contexts the trajectories are linear and there is no evidence of deviation in the trajectory towards a target. The fact that trajectories for pre-consonantal schwa from symmetrical back, central and front environments are all linear clearly does not provide any evidence of an F2 target for pre-consonantal schwa.

Figures 4.30 and 4.31 shows the formant trajectories for the stressed vowels, for comparison. As explained in section 4.2, it is the trajectories of vowels in symmetrical phonetic contexts that are the most informative in terms of finding evidence of a target. Therefore, the stressed vowel trajectories are only shown in symmetrical contexts. Figures 4.30 and 4.31 are derived from models which compared the differences in the shape of the trajectories across the vowels. A model was run comparing the vowels in back contexts (model 4.17), and in central contexts (model 4.18). Figure 4.30 shows the trajectories for stressed vowels in symmetrical back contexts, and Figure 4.31 shows the trajectories for schwa in symmetrical central contexts. Symmetrical front contexts were not compared as there were too few vowels within this context across the stressed vowels for any meaningful analysis. Each stressed vowel is plotted with non-final schwa, for comparison.

Figures 4.30 and 4.31 show that, like pre-consonantal schwa, not all stressed vowel show deviations in their F2 trajectories either. Figure 4.30 shows that In back contexts there is a clear positive deviation in the trajectory for KIT, and a negative deviation for FOOT and THOUGHT. In symmetrical central contexts, there is a clear negative deviation for FOOT and THOUGHT. TRAP and NURSE, the more central vowels in terms of average F2 values, like pre-consonantal schwa, do not show any deviation in their F2 trajectories.

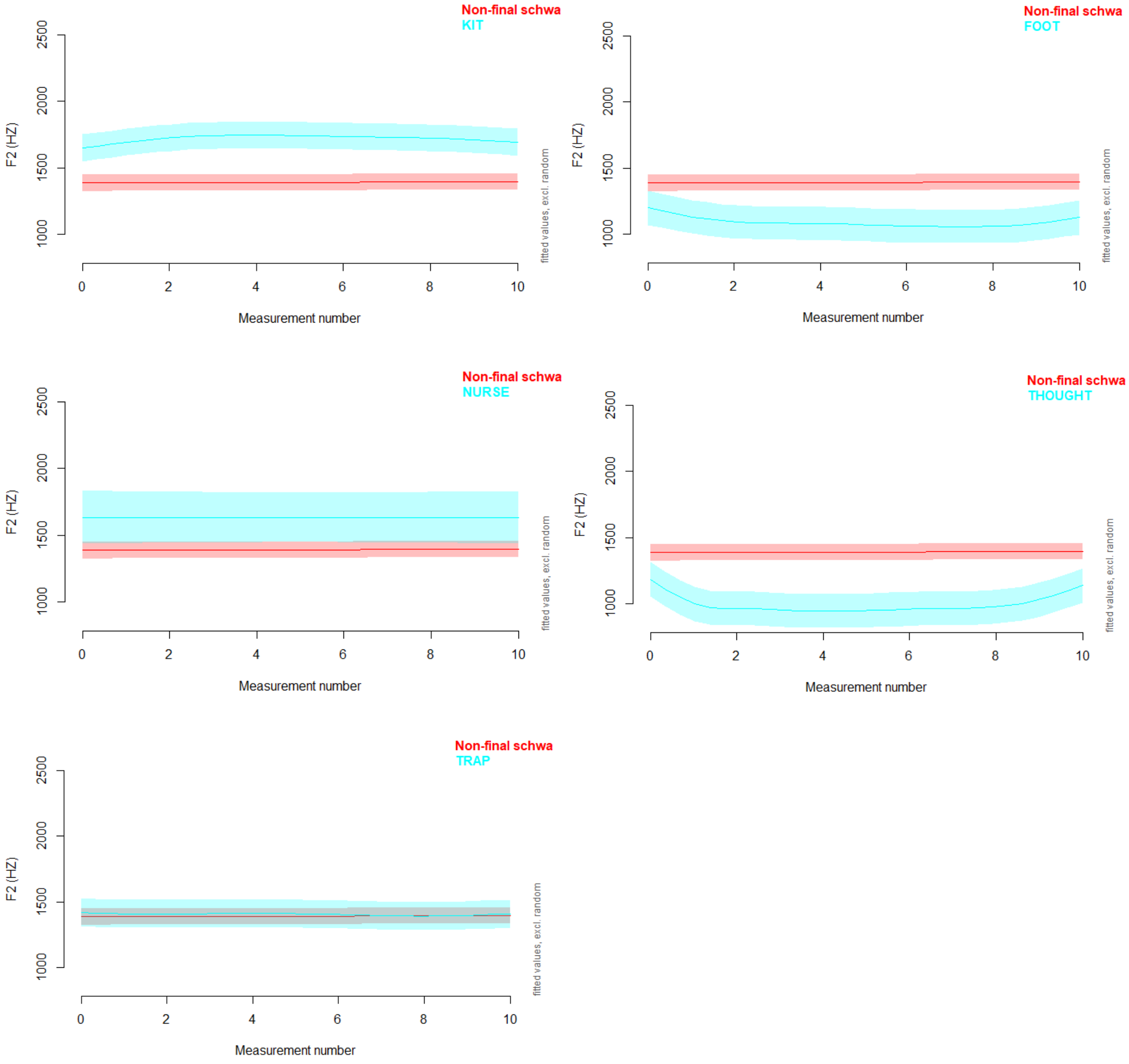


Figure 4.30 *F2 (Hz) trajectories in back contexts*

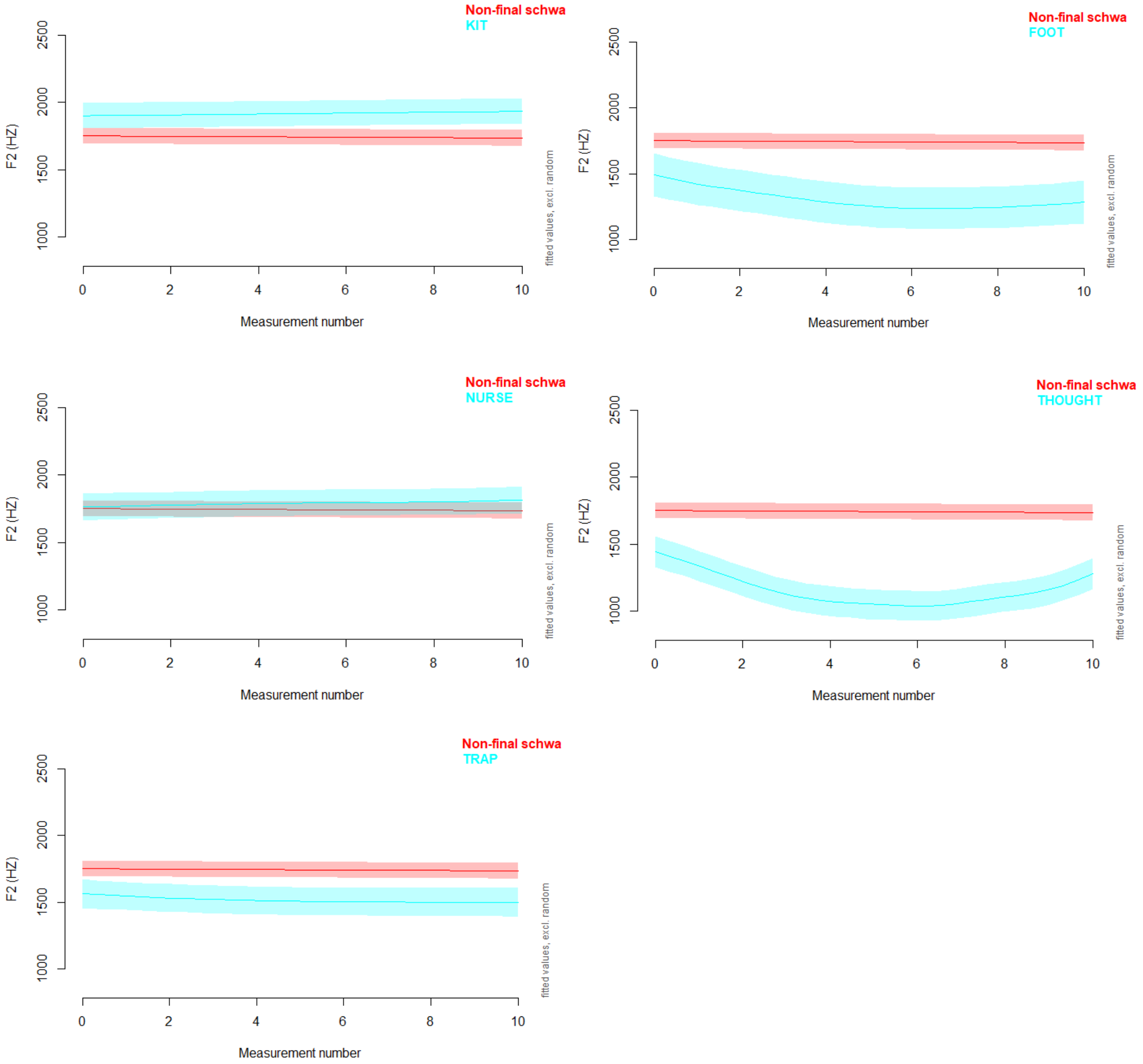


Figure 4.31 F2 (Hz) trajectories in central contexts

It is therefore the case here that the vowels that have greater deviation in trajectories are the vowels that have more extreme F2 values. It is not the case that all stressed vowels have strong deviations in their trajectories, as TRAP and NURSE also do not show any evidence of deviation in their F2 trajectories. In this way, schwa is not exceptional in showing no clear deviation towards a target. Schwa's behaviour in this respect may be to do with it being a relatively central vowel rather than having a unique targetless status. Although the effects of phonetic context on schwa are clear, it may be that they do not cause sufficient variation in its production to impede communication. For example, even when non-final schwa is in back contexts it is not as low in F2 as THOUGHT or even FOOT, as seen in Figure 4.31 . It may therefore be that the contextual variation in schwa does not cause it to diverge so drastically from its target that it needs to deviate away from the effects of phonetic context. Either way, because of the lack of consistent deviations in the trajectories of the stressed vowels, the lack of deviation in the F2 formant trajectories for schwa cannot be taken as evidence of targetlessness.

So far, the F2 trajectories in vowels surrounded by two consonants have been examined. However, the behaviour of pre-pausal schwa also has the potential to be revealing. Indeed, the fact that pre-pausal schwas are both longer and unaffected by a following consonant mean it is more likely that evidence of any target could be found here. Although pre-pausal schwas all have the same following environment, they can still be grouped by preceding consonantal context. In section 4.4.2, it was established that preceding context has a significant effect on the F2 value of the midpoint of pre-pausal schwa. If pre-pausal schwa shows evidence of a phonetic target it is therefore expected that there should be some movement in its trajectory in at least one of the phonetic contexts, from the beginning to the end of the segment. If evidence of a phonetic target is found it should be such that at the end of the vowels there is less difference between the three phonetic contexts then at the beginning i.e. the influence of the preceding context on the schwa should decrease through the duration of the vowel.

Figure 4.32 shows the trajectories for pre-pausal schwa by preceding environment (model 4.19). While in central contexts it has an unchanging trajectory through the segment, schwas preceded by front contexts decrease in F2 through the trajectory, and those preceded by back contexts increase in F2. This therefore has the result that there is on average less difference between the environments at the end than at the beginning of the tokens. This difference in the shape of the formant trajectories from different phonetic contexts is shown in the fact that the difference smooth between back and central contexts is significant (F=31.436, edf=1, $p<0.0001$) and also that the difference smooth between front and central contexts is significant (F=9.869, edf=4.107, $p<0.0001$). This pattern therefore provides evidence for this group of schwas having a target, and suggests that it is likely that the target is roughly at the average F2 value for those schwas preceded by central contexts, since the average formant value here is stable over the trajectory.

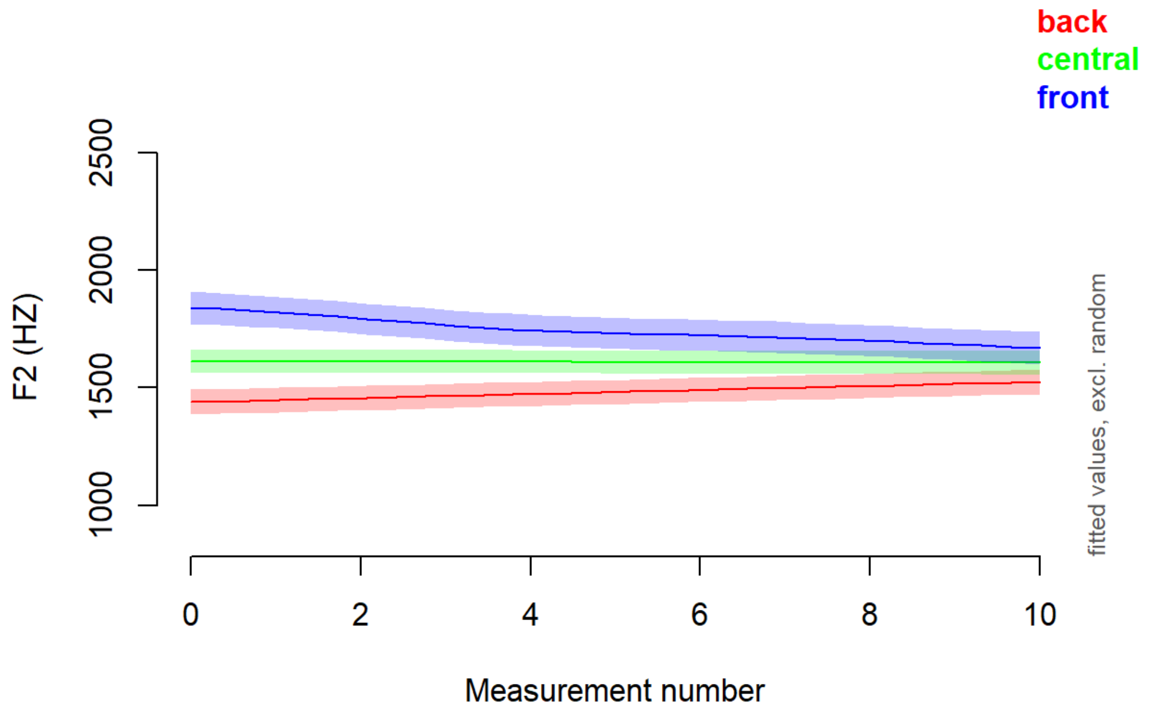


Figure 4.32 *F2 (Hz) trajectories in pre-pausal schwa by preceding context*

4.4.3.3 Summary

In summary, both the F1 and F2 formant trajectory analyses show clear evidence of a phonetic target in pre-pausal schwa. For pre-consonantal schwa the evidence is less clear. Like most of the other vowels tested, pre-consonantal schwa shows a deviation in its F1 trajectory, although this deviation is not very big. There is no evidence shown of deviation in the F2 trajectories of pre-consonantal schwa. However, as this was also the case for the other more central stressed vowels, this does not provide evidence of targetless.

4.4.4 Variation by duration

This section examines variation in the vowel midpoints according to their duration. The rationale behind this analysis is that longer vowels have more time to move away from their phonetic context and thus more time to reach a phonetic target. It is therefore predicted that if schwa has a phonetic target that it will be closer to this target at longer durations. On the other hand, if schwa is a targetless vowel it is not expected to vary according to duration. As for the formant trajectory analysis, these analyses are conducted slightly differently for F1 and F2. For F1 it is expected that schwa would have a higher F1 at longer durations, but for F2 the effect of duration is expected to vary according to phonetic context. Therefore, the F1 results are shown across all tokens for each vowel, but for F2 the effect of phonetic context is considered.

This section starts by examining the overall duration differences between the vowels. It then moves to an analysis of how F1 varies according to duration, and then how F2 varies according to duration. Finally, the effect of duration on schwas position across the whole vowel space is considered.

4.4.4.1 Overall duration differences

Figure 4.33 shows the differences in duration for all of the vowels. These results are as would be expected. The phonologically long vowels NURSE and THOUGHT are the longest. This is followed by TRAP, which, although phonologically short, we would expect to be slightly longer since it is a low vowel (Lindblom, 1963). The high vowels FOOT and KIT are the next shortest stressed vowels, with KIT being the shortest stressed vowel.

In terms of schwa, there is a large difference between the duration of pre-pausal schwas and the pre-consonantal schwas, with the duration of pre-pausal schwas being most similar to that of FOOT. It is clear that this lengthening effect applies to schwas in pre-pausal position only, rather than word final schwas in general, as Figure 4.33 shows that pre-consonantal word final schwas pattern with non-final schwas. Non-final schwa and pre-consonantal word final schwa are very similar to each other in duration, and are the shortest vowels measured.

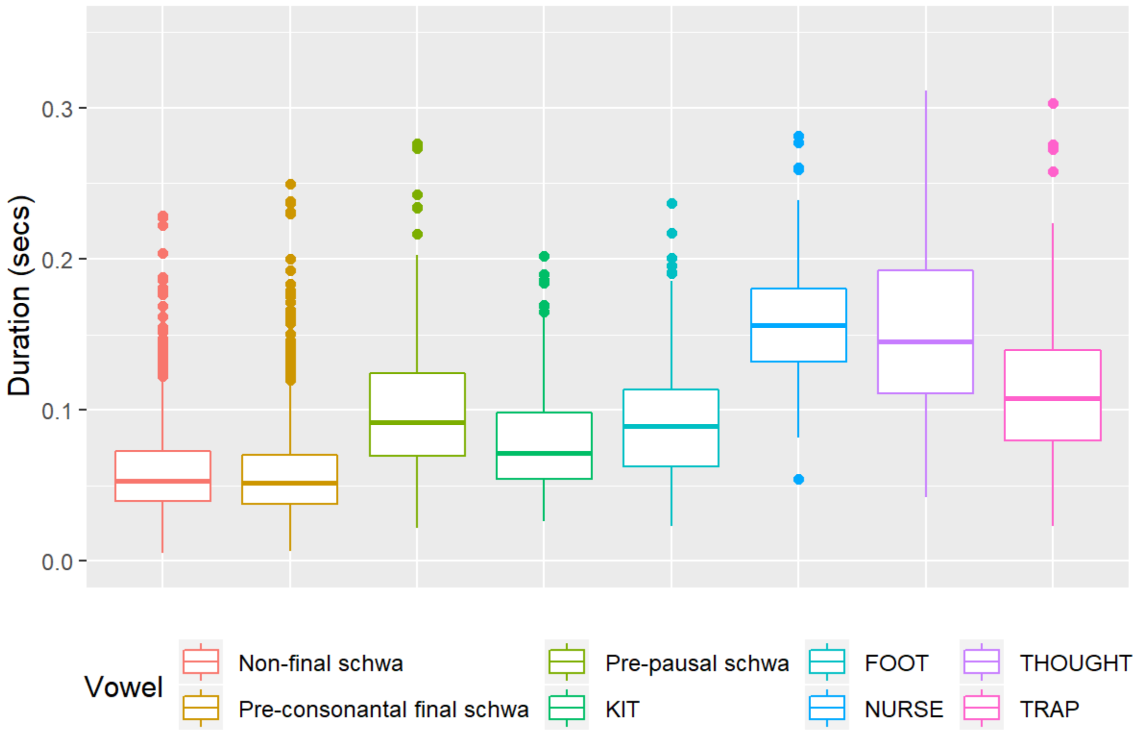


Figure 4.33 *Boxplot of duration by vowel*

4.4.4.2 F1

Figure 4.34 below shows the relationship between F1 and duration for all of the vowels. The values for the three schwas are shown compared with a different stressed vowel in each panel. These figures are derived from a model (model 4.20) that compared the effect of duration on F1 across the different vowels. Duration has a clear effect on the F1 of all three schwa types. As shown in the Figure below, as predicted, all three schwa types have an increasingly high F1 as they get longer. It is clear also that there are F1 differences between the schwas. Pre-pausal schwa is the lowest in height, followed by pre-consonantal word final schwa, and then by non-final schwa.

TRAP also shows the same relationship between duration and F1 as schwa, with longer TRAP vowels having a higher F1. This makes sense since TRAP is a low vowel and it is difficult for the articulators to produce a low vowel at very short durations. This relationship is, however, not found for all of the stressed vowels. Importantly, KIT does not show this relationship. The fact that schwa shows this effect of duration, and KIT does not means that schwa is more different in height to KIT at longer durations than at short durations. The difference between the effect of duration on KIT and non-final schwa is reflected in the fact that the difference smooth for KIT was significant (edf=1, F=5.239, p<0.05), showing that the duration smooth is a different shape for non-final schwa and KIT. This suggests that schwa has a height target which is genuinely lower than the minimum requirement for a vowel. It also suggests that schwa is phonologically a lower vowel than KIT.

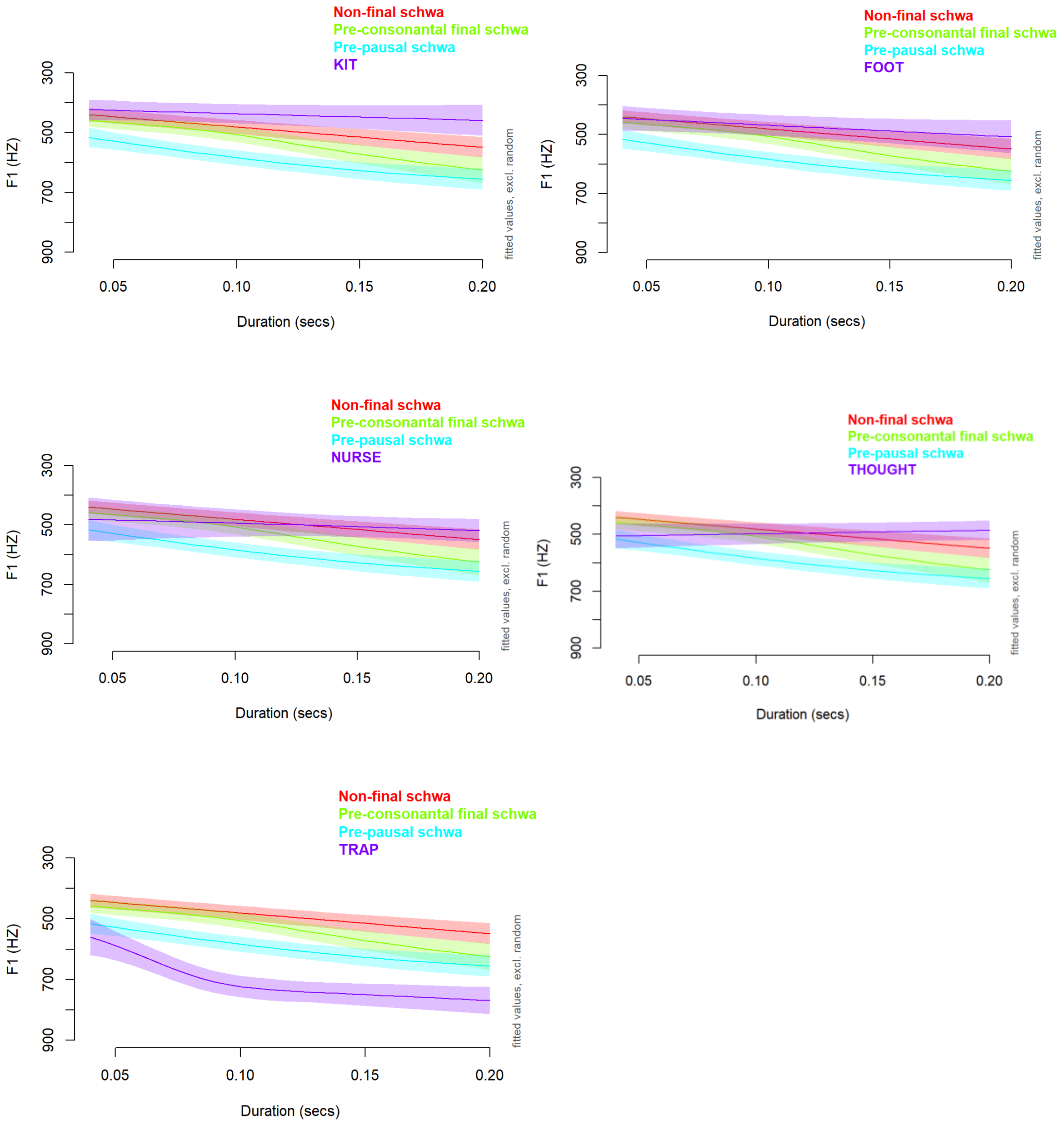


__Figure 4.34__ F1 (Hz) by duration across vowels

4.4.4.3 F2

For the analysis of the effect of duration on F2, models were run on each vowel individually in order to see if there was an interaction between phonetic context and duration. The prediction was that, if there is evidence of a phonetic target, vowels from different phonetic contexts should be affected differently by duration (see Figure 4.2).

There was no significant interaction between phonetic context and duration for non-final schwa (model 4.21) or pre-consonantal word final schwa (model 4.22). Neither of these schwas show any clear changes in their F2 according to duration, and the difference between phonetic contexts is fairly constant over duration. This therefore does not provide evidence of an F2 target for schwa.

With regards to pre-pausal schwa (model 4.23), however, there is an interaction between F2 and preceding phonetic context. Duration has a slightly different effect on pre-pausal schwa depending on its preceding phonetic context. This has the result of schwas from different phonetic contexts being more similar to each other at longer durations. Pre-pausal schwas in back contexts have a higher F2 at longer durations. This is reflected in the fact that the difference smooth for back contexts in significant (edf=1, F=4.706, p<0.05), meaning that pre-pausal schwas are affected differently by duration in back and central contexts. Again, this shows evidence of a target in pre-pausal schwa.

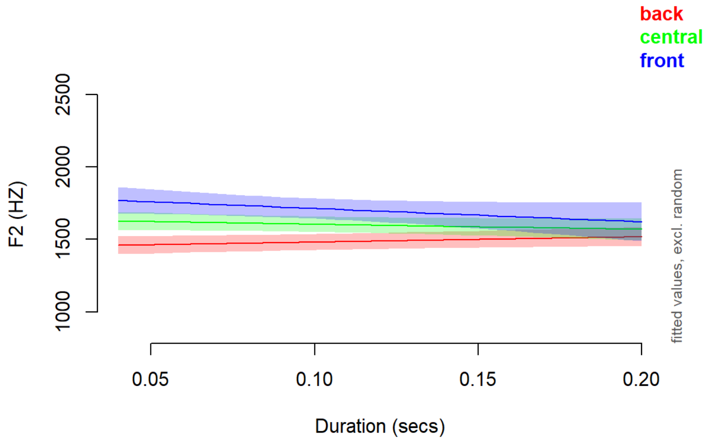


Figure 4.35 *F2 (Hz) by duration and preceding context for pre-pausal schwa*

Models were also run on the stressed vowels in order to see if they showed an interaction between duration and F2. It was found that, like for pre-consonantal schwa, there was no clear interaction between phonetic context and duration for any of the stressed vowels. In the case of TRAP (model 4.28) and NURSE (model 4.28), duration had no clear effect on their F2 values at all. For KIT (model 4.24), FOOT (model 4.25), and THOUGHT (model 4.27), however, duration had an overall effect on F2 values. THOUGHT (edf=1.658, F=3.86, p<0.05) and FOOT (edf=1.97, F=3.719, p<0.05) both had significantly lower F2 values at longer durations, as reflected in the significance of the overall smooth for duration. This shows that, as would be expected, these vowels are further away from their targets at shorter durations. Figures 4.36-4.38 below show the overall effects of duration on KIT, FOOT and THOUGHT.

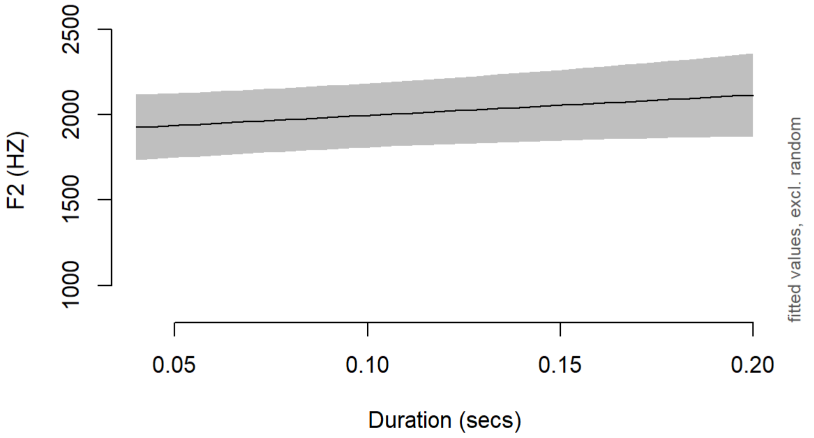


Figure 4.36 *F2 (Hz) by duration for KIT*

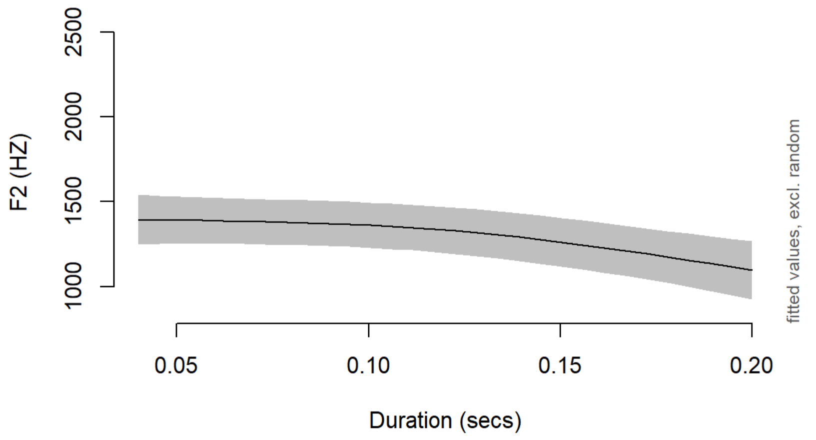


***Figure 4.37** F2 (Hz) by duration for* FOOT

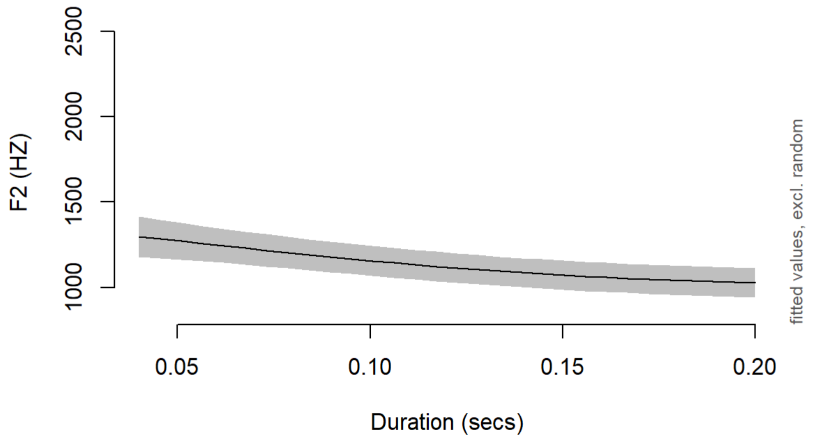
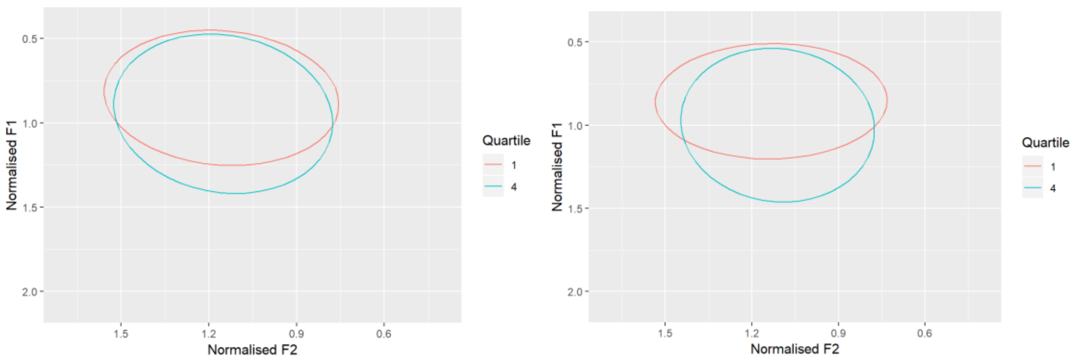


***Figure 4.38** F2 (Hz) by duration for* THOUGHT

Similarly to the pattern that was found in terms of the relationship between formant trajectories and F2, it is therefore also the case here that the vowels where F2 changes according to duration are the vowels that have more extreme F2 values. It is not the case that all stressed vowels show a relationship between F2 and duration, as NURSE and TRAP show no such relationship. In this way, therefore, pre-consonantal schwa is not unique in its lack of relationship between duration and F2. It therefore cannot be argued that the lack of a relationship between F2 and duration in pre-consonantal schwa is evidence of it having a targetless status.

4.4.4.4 Overall vowel space

In section 4.2 I set out the prediction that if schwa has a target then, over the whole vowel space, we would expect the variation in F2 to decrease according to duration, and F1 to increase. To test this, for each schwa type, schwas that were in the first and last quartile of duration values are compared. The plots below show the overall differences in the position of long and short schwas in the vowel space. As was apparent from the analyses in section 4.4.4.2, all three schwas are clearly lower in the vowel space at longer durations (4th quartile) than at short durations (1st quartile). There is also a slight decrease in the F2 variability of pre-consonantal schwas at longer durations.


Non-final schwa

Word final pre-consonantal schwa

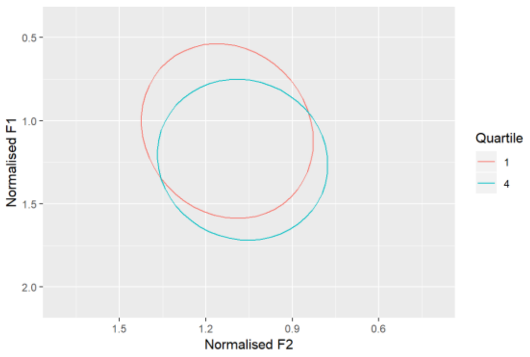

Pre-pausal schwa

Figure 4.39 *Normalised vowel plots of schwa by duration*

This is as was predicted if schwa has a target. Therefore, despite there being no interaction found between phonetic contexts and F2 in pre–consonantal schwa, this change in the variability of F2 with increased duration suggests an F2 target for pre-consonantal schwa. If schwa is targetless it would be expected that it would be extremely variable in F2 regardless of duration.

Although the differences between the first and last quartile are small in some cases they still show the predicted movement away from phonetic context as duration gets longer. The increase in F1, coupled with the decrease in F2 variability at longer durations clearly suggests a vowel that it moving away from the influence of phonetic context as it gets longer.

4.4.4.5 Summary

The duration analyses conducted in this section have shown that both pre-consonantal and pre-pausal schwas have a higher F1 when longer, thus showing clear evidence of a height target in schwa. In terms of F2, only pre-pausal schwa showed any change over duration, with schwas in different phonetic environments being more similar at longer durations. However, there were also some stressed vowels that did not show a relationship between duration and F2, so the lack of a relationship in pre-consonantal schwa was not taken as evidence of it being targetless. In addition, the overall variability in the F2 of pre-consonantal schwa was found to decrease at longer durations, which is as was predicted in a targeted schwa. This means that pre-consonantal schwa still showed the overall relationship that was predicted; longer pre-consonantal schwas are both lower vowels and less variable in backness, showing a move away from phonetic context at longer durations, and thus evidence of a phonetic target.

4.5 Discussion

In section 4.2 the main research questions for this chapter were proposed, broken into smaller hypotheses about what would be expected from the data according to whether schwa is targetless or has a target. The various pieces of evidence for these two questions will be pulled together here so the answers can be assessed. This will then be followed by an analysis of these findings in relation to the overall research questions.

4.5.1 Is schwa more variable and context dependent than other vowels?

The overall vowel quality of schwa is highly variable, and is spread over a larger area of the vowel space than other vowels. High F2 variability is indeed expected in a targetless schwa. However, this high F2 variability is not unique to schwa, since comparably high levels of variability (as shown in standard deviations) are found for KIT and FOOT. KIT and FOOT are the closest to schwa in terms of vowel height and duration. Therefore it could be that this high variability is a feature that is due to either being phonetically high or their short duration, rather than something which is unique to schwa, as has often been claimed (Kondo, 1994; Bates, 1995). Much of the variability seen over the vowel space is also caused by F1 variability. This therefore is evidence against a targetless schwa. In a maximally assimilatory vowel, it would be expected that phonetic contexts should generally cause low F1 values, and therefore low F1 variability would be expected (see section 2.4.3 and 4.2).

It was found that the F2 value of schwa was indeed significantly affected by its phonetic context, and midpoint and trajectory analysis showed these effects to be systematic and have an effect throughout the whole of the schwa. Although there were clear effects of phonetic context on stressed vowels, schwa displayed the most consistent systematic effects, from both preceding and following contexts. In general, the size of the effects on the midpoints was also larger for schwa than for any other vowel.

4.5.2 Is there any evidence that schwa has a target?

This question was answered by use of two different methods; examining whether there was a relationship between formants and duration at the schwa midpoint, and by looking at the whole trajectory of schwa to see if there was any movement towards a target.

In terms of F1, the trajectory of all schwa types deviated towards a higher value. However, for pre-consonantal schwa this was no more so than for FOOT or KIT, so this did not provide clear evidence of a phonetic target which is lower than the minimum required for a vowel. However, the fact that F1 significantly increases with duration for schwa, and not for the high full vowel KIT, suggests that schwa is unlikely to be phonologically high. This provides support for schwa having a height target that is lower than a high vowel and thus has not being maximally assimilatory to context in F1.

In terms of F2 the results were slightly more complex. The first hypothesis was that if schwa has an F2 target then there should be an interaction between phonetic context and duration on F2 at the midpoint, since schwas from different phonetic contexts would have to move in different directions to get towards a target. The second hypothesis was that when looking at the whole of the trajectories, there should be some deviation in the trajectory towards a target, rather than being a straight interpolation between the preceding and following environment. Pre-pausal schwa showed evidence of a phonetic target using both of these methods. The influence of preceding phonetic context on pre-pausal schwas decreased over the vowel trajectory, and also decreased as pre-pausal schwas got longer. For pre-consonantal schwa, neither hypothesis was confirmed, as neither method revealed any evidence of an F2 target. However, these methods also did not unanimously show evidence for all stressed vowels having a target. Only the THOUGHT, FOOT and KIT vowels showed any significant effects of duration on F2. Additionally, although no effect of phonetic context was found for any of the specific phonetic environments in pre-consonantal schwa, it was found that there was slightly less variability overall in the F2 of longer schwas, as evidenced by a lower standard deviation in F2. That this difference is in the predicted direction offers tentative support for the hypothesis that a schwa has a target.

In this data, word final schwas that were immediately followed by another word did not show any evidence of an F2 target and behaved in much the same way as non-final schwa, so there is no evidence that these two categories of schwa should be considered as phonologically different. The only clear evidence of an F2 target for word final schwas was found when it occurred in pre-pausal environments. It seems that the division in realisation is therefore one of following environment i.e. a pause or a following consonant, rather than of word position. It follows that if word final schwas that precede a pause have targeted F2 then so too might word final schwas that are not followed by a pause. The fact that word final schwas, which are demonstrated as having a target in a pre-pausal environment, are still variable in F2 before a phonetic environment means that it may be possible that a schwa with a target could still be realised without showing evidence of that target in certain environments.

The results from the trajectory analysis were broadly similar to the durational findings, in that unanimous evidence of an F2 target was not found for all the full vowels. Both methods used only showed clear evidence of a phonetic target for FOOT, THOUGHT and KIT, but not for NURSE or TRAP. Note that in both these methods it is the vowels with the more extreme F2 values that show evidence of having a target, and conversely those with more central F2 values where that did not show evidence of an F2 target. In light of these results from full vowels, therefore, it is perhaps not so surprising that schwa also does not show evidence of an F2 target, as it is also a fairy central vowel on average. It is therefore possible that schwa could have a target, and that the lack of clear positive evidence for this is to do with it having relatively central F2 target , rather than being because of a targetless property unique to schwa.

Across the different analyses for F1 and F2, these results present a somewhat mixed picture, with some results being as predicted if schwa has a unique target and others being as would be predicted if it does not. As explained in sections 2.3 (p.31) and section 2.4.3 (p.37-8) though, it does not make sense to suggest that a vowel can be targeted in F1 and targetless in F2. Pre-pausal schwa shows clear evidence of a target in both F1 and F2. Direct evidence for a target in pre-consonantal schwa, however, has only been found in the F1 analysis. The decrease in the F2 variability of pre-consonantal schwa at longer durations, however, provides tentative evidence for an F2 target. Although no direct evidence was provided for an F2 target, the results by no means provide evidence that schwa is targetless, given that the methods used also do not provide unanimous evidence for a target for all of the full vowels tested either. Taken together, therefore, these findings provide tentative evidence for a unique phonetic schwa target.

4.5.3 General discussion

On balance, it has been suggested that evidence supports the idea that schwa has a unique phonetic target, rather being maximally assimilatory to context. This is despite the fact that not all of the predictions for a targeted schwa were borne out; pre-consonantal schwas were shown to have linear F2 formant trajectories, and there was no interaction between phonetic context and duration on F2. These findings are, however, not necessarily incompatible with the idea that schwa has a vowel quality target. Unanimous evidence for vowel quality targets in stressed vowels was also not found using these methods. There was a pattern in these findings in that for both measures it was the vowels with more extreme F2 values that showed the patterns that were initially predicted to be found in targeted vowels. Although evidence towards a target may not have been found for schwa using these measures, it is not the case that vowel targets will just be one point in phonetic space, or just one very specific articulation. Within in any speech category there is a range of acceptable variation where the vowel will still be perceived as the intended category (Kirby, 2010). It may thus be that the coarticulatory effects of neighbouring consonants simply do not remove the more central vowels sufficiently far enough away from their targets to be unacceptable realisations of these vowels. For example if the effects of context on the F2 of schwa still generally produce F2 values that are within schwa's normal F2 range, then a deviation in the formant trajectory would not be seen, nor would there be an interaction between duration and context.

The findings in this chapter corroborate claims from other studies that non- final schwa is phonetically higher than word final schwa (Flemming, 2009; Lilley, 2012; Bekker, 2014) as non-final schwa was on average realised as a high vowel, and word final pre-pausal schwas were lower. In spite of this, these findings do not support the idea that this is because non-final schwa is produced with less effort from speakers, due to less need for contrast maintenance in a non-final environment. If this was the case we would have expected all word final schwas to pattern together whether they precede a pause or a consonant. In reality, pre-consonantal word final schwa patterned with non-final schwa. This suggests that the difference between word final and non-final schwa is actually a phonetic effect due to different following environments (i.e. a pause or a consonant). Vowels are generally shorter when preceding another sound, as supposed to a pause (Klatt, 1975). In addition, duration and vowel height are known to be linked, with higher vowels generally being shorter (e.g. Lindblom, 1963). It is likely that both the decreased duration and lower F1 in non-final schwas can be attributed to the effect of the following consonant. Therefore, differences between non-final and word final schwa can be explained by phonetic effects that occur elsewhere and are not unique to schwa, without any need to appeal to notions of lack of effort or any special targetless schwa property.

4.6 Conclusion

This chapter aimed to investigate whether schwa has a unique phonetic target, or whether it is a targetless vowel, where the aim in production is to produce a vowel which maximally assimilates to phonetic context. The chapter also addressed the question of whether non-final schwa is phonologically different to word final schwa. These questions were addressed through an analysis of the overall variability and predictability of schwa, and of its variation according to formant its formant trajectory and duration. Throughout, the findings were compared to a range of stressed vowels, in order to make sure the methods used were valid methods for assessing targetless (see section 2.4.2).

Non-final schwa was found to on average to be realised with a fairly high realisation, and to be more predictable from phonetic context than other vowels. However, it was not found to be exceptionally variable in F2. Evidence for a vowel height target which is lower than the minimum requirement for a vowel was found in a stronger positive relationship between F1 and duration than was found for KIT. Although no evidence for an F2 target was found from either analysis of trajectories, or the interaction between duration and phonetic context, there was also a lack of unanimous evidence for a phonetic target in stressed vowels using these methods. This meant that the lack of evidence from non-final schwa could not be taken as evidence of targetlessness. In addition it was found that non-final schwas had a lower level of F2 variability when longer.

Pre-pausal schwas were found to have F2 trajectories showing a movement towards a target, although word final schwas that preceded consonants did not. Generally pre-consonantal word final schwas pattern with non-final rather than pre-pausal schwa. The similarity between pre-consonantal word final schwa and non-final schwa show that the division often made between word final and non-final schwa is unlikely to be phonological or because of a lack of effort in production of non-final schwa (e.g. Flemming, 2009). Instead it was proposed that the differences between these two groups can be explained as phonetic consequences of their differing following environments. There is no need to suggest that non-final schwa is targetless to account for the difference.

In summary, these results provide support for the idea that schwa has a target. Direct evidence for a height target was demonstrated through the relationship between duration and F1, as well as indirect evidence of a likely F2 target in terms of a decrease in F2 variability at longer durations. The results also suggest that differences between non-final and word final schwa may be an artefact of their environment, rather than intrinsic to word position itself. These results show the need for adequate comparisons of other vowels to be used as a control for before making any claims of targetlessness. These same issues will be further explored in Chapter 6, where the changing position of the KIT vowel in NZE provides a particularly useful stressed vowel comparison to schwa.

5 Using automated data for schwa analysis

5.1 Introduction

This chapter presents an analysis and comparison of unstressed vowel measurements collected using both manual and automated means. The aim of the chapter is to assess whether automated measurements are suitable to use for analysis of unstressed vowels, and to find out what steps can be taken to improve the quality of automated data. This aim is addressed through comparing sets of measurements of the same tokens, collected manually and automatically. I then compare datasets that have been filtered in different ways in order to see how the automated data can best be suitably prepared for analysis. The data in this chapter is taken from the Derby Corpus (Milroy et al, 1996). The manual data analysed is manual in both segment boundary location, and in formant measurement. By contrast, the automated data uses automatically located segment boundaries found through forced alignment in LaBB-CAT (Fromont and Hay, 2012), and also automated formant measurement using praat (Boersma and Weenink, 2011).

The motivation for this chapter arose from the need to use automated data in the analysis of the Origins of New Zealand English Corpus (Gordon et al, 2007) in Chapter 6. So far both analyses in this thesis have used data from the same Derby corpus (Milroy et al, 1997) where, with only 4980 relevant unstressed tokens, it was possible to gather measurements manually. However, in using a corpus as big as the ONZE corpus (228,498 relevant unstressed tokens) using automated data is the only way to get through such a large amount of data within a reasonable time frame. However, the process of analysing schwa tokens in the Derby corpus caused some concerns about how appropriate the use of this method would be for the measurement of unstressed vowels. In addition, there is little indication from previous research about whether automated methods are appropriate for unstressed vowel analysis. This chapter therefore compares manual and automated measurements from the Derby corpus data in order to inform the methodology used in the analysis of the ONZE data in Chapter 6.

This chapter starts by reviewing the general issues with automated measurement (5.2), including the particular potential issues faced in the automated extraction and measurement of unstressed vowels. The methodology is then explained (5.3). This is followed by the analysis, which is split into two main parts. In 5.4 the focus is on comparing the automated and manual dataset. In 5.5 the focus moves to investigating how different exclusion criteria can be used to improve the automated data. This is then followed by a discussion of the implications of the findings (5.6).

5.2 Background

In the measurement of vowel formants, it is possible to automate both the location of the vowel boundaries, and the measurement of the formants themselves. Evanini et al (2009), for example, focus on the issue of formant measurement. They examine the use of automatic formant tracking and conclude that it is an appropriate tool to use for acoustic analysis. The other key element of gaining fully automated formant measurements is in locating the boundaries of each segment, which is done through forced alignment. There are several different forced aligners available including LaBB-CAT (Fromont and Hay, 2012), FAVE (Rosenfelder et al, 2014) and MAUS (Schiel et al, 1999). Some of the main ways in which these aligners differ from each other are in what toolkit they use, and whether they use pre-trained models, or models based on the speech data itself.

Several studies have investigated how successful different forced aligners are in accurately segmenting speech. Typically this is done by comparison to manual alignment. The alignment in this thesis is performed using LaBB-CAT. LaBB-CAT uses the HTK toolkit (Young et al, 2006), and uses train/align models, meaning that the models are trained on the speech data itself. It has been suggested that, in many circumstances, train and align models may be better than pre-trained models. Fromont and Watson (2016) look at how a range of factors affect alignment. They find that where there is more than 5 minutes of speech from a participant, that using train and align models produces better alignment than pre-trained models. Similarly, Gonzalez et al (2020) compare the success of a range of aligners at locating vowel boundaries. They find that LaBB-CAT produces higher quality alignments than

FAVE and MAUS. They suggest that a reason for this could be because LaBB-CAT uses train and align, whereas FAVE and MAUS use pre-trained models. In fact, in terms of locating vowel onsets they find that LaBB-CAT is often not significantly worse than manual alignment. Overall, however, they find that none of the aligners tested are as consistent as manual alignment. This echoes the sentiments of Mackenzie and Turton (2020), who also suggest that manual checking is needed to correct the worst errors of manual alignment.

Clearly, there is a trade-off between efficiency and accuracy and it may often be worth sacrificing a small amount of accuracy in data measurements in order to be able to gather a large amount of data in a reasonable time frame. Indeed, there is an argument that if enough data is collected then even having a few large errors in the data should not harm the outcome of the analysis (e.g. Evanini et al, 2009). Unfortunately, however, some of the studies that have looked specifically at automated measurements of vowels (Evanini et al, 2009; Gonzalez et al, 2020) have only looked at stressed vowels. Generally, this is not a problem given that much of the work in the field also tends to exclude unstressed vowels. However, given that unstressed vowels are the focus of this thesis, it is not a given that such findings will be applicable to the data analysed here.

5.2.1 Potential issues with automated measurements of unstressed vowels

Whilst a degree of noise is expected in any set of automated measurements, there are reasons to suspect that the use of automated measurements of unstressed vowels could be particularly problematic. Firstly, unstressed vowels tend to be of particularly short duration, particularly when before a consonant, as shown in this thesis (Chapters 3 and 6) and elsewhere (Bates, 1995; Flemming, 2009). This means that there is a smaller margin for error in boundary placement compared to longer vowels. In addition, findings have found increased speech rate (Bailey, 2016; Mackenzie and Turton, 2020) to be correlated with decreased accuracy of aligners. Although this is not a direct measure of the effect of vowel duration, speech rate and vowel duration are clearly related. This therefore suggests that there could be decreased accuracy for particularly short segments. Small variations in boundary placement of shorter vowels could result in a relatively high proportion of the target segment not being measured, or conversely a high proportion of what is measured could contain acoustic material that is in fact produced primarily in the articulation of neighbouring segments. The shorter duration of unstressed vowels can in itself make it hard to measure them accurately. Indeed, on manually measuring the formants in the Derby data there were many occasions where a schwa could be perceived, but there was no obvious place on the spectrogram which could be labelled as such. For example, this sometimes occurred where the schwa was adjacent to two voiced segments.

The other problematic aspect of measuring unstressed vowels is that they often do not appear where transcriptions suggest that they should. As LaBB-CAT works by automatic alignment of a segment-based transcription with the spectrogram, this means that regardless of whether the sounds in the transcription are actually present on the spectrogram, it will attempt to find and measure them. It is common for a potential schwa to be elided in certain words (Lacosto and Connine, 2002; Davidson, 2006; Polgardi, 2015) e.g. *properly* as /prɒpli/, *factory* as /faktri/. In the Derby data, there were many tokens where a vowel was almost never produced e.g. *different* as [dɪfrənt]. A related issue is where schwa occurs as a voiceless segment (Heselwood, 2007). This often occurs after voiceless consonants (e.g. in *today* and *support)*. In such cases, although schwa can be perceived, there is no voicing apparent on the spectrogram, and instead what is perceived as schwa is in fact a continuation of the aspiration in the preceding plosive, or lengthening of the preceding fricative. Alternatively, in such cases it could also be that the phonology of the listener is simply predicting a schwa, even where there is no acoustic evidence for it. In both of these cases, although there are no voiced unstressed vowel tokens to measure, as there is a vowel in the transcription the automated measurements will still be returned. However, these measurements will reflect something else in the spectrogram rather than the target token. This is a particular problem and not one which has been addressed as much in the literature. This is an issue because, even if it is the case that the alignments are successful where a segment exists, there could still be issues if there are many occasions where the target token is not present.

As Foulkes et al (2018) note, relying solely on automated measurements without some level of manual checking can be a particular problem when the object of interest varies subtly and is related to lots of other factors. This is the case with this particular study, as I am focussed on the interacting effects of a number of factors on unstressed vowels. Where large amounts of data may not contain the target token this could be a particular problem, and would raise doubt on the idea that if you have enough data, noise can be overlooked. In such cases, it may not matter how much data there is overall, as if a large enough proportion of the data is not from the target token, the outcome of the analysis will still be affected. An excessive amount of random noise in the data is likely to muddy any subtle effects that might be found. In addition, if these false measurements are not randomly distributed across the variables of interest the data could be substantially skewed, as a result of this systematic noise.

This chapter thus compares sets of measurements collected using manual and automated methods from the Derby corpus. Unlike some of the other research on this issue I do not compare exact formant measurements or boundaries. Instead, I focus more broadly on how the results of analysis are affected as a result of the method of analysis.

5.3 Methodology

The unstressed tokens used for the analysis in Chapters 3 and 4 were found through searching the corpus using LaBB-CAT, before manually segmenting the tokens. In the initial search that was conducted there were 9,105 unstressed tokens that were relevant to my research. After the process of listening to the data, manually locating the vowel boundaries, and manually checking formant measurements, only 4,980 tokens were included in the final dataset. This was because there were only 4,980 tokens where an unstressed vowel could be found and measured on the spectrogram.

For the remaining 4125 tokens formant measurements were unable to be recorded for various reasons. These tokens were of the following types:

- 550 voiceless tokens , where no measurement could be obtained e.g. *support*, *today*
- 1,332 tokens where no unstressed vowel was realised. 851 of these were words where the vowel was simply not produced e.g. *probably* as [prɒbli], *perhaps* as [praps]; 481 were words where the whole syllable containing the transcribed schwa

was not produced e.g. speakers only uttering the second syllable of *because.* Clearly this is a transcription error. However, it is of course only possible to notice such transcription errors through manual checking.

- 118 tokens where an unstressed vowel was perceived but there was no visible vowel on the spectrogram which could be measured
- 2,125 tokens where the token could not be measured for various reasons not unique to unstressed vowels. These included the word not appearing at all in the extracted speech, atypical speech e.g. singing, not being able to measure because of speaker overlap.

This chapter is broken into two following sections of analysis. The first section compares the manual data with the full set of automatically extracted data (9105 tokens), and also a reduced automatic dataset. Whilst the full automatic data contains all of the data extracted automatically, the reduced automatic data only contains the tokens which were also used in the manual analysis (4980 tokens). This three way comparison allows for an assessment of to what extent the difference between the manual and full automatic data is due to measurement/alignment differences, and to what extent it is due to the different tokens which are included. The analysis in this chapter focusses on measurements taken at the vowel midpoint.

The second section of this chapter additionally compares the outcome of editing the full automatic data, using various exclusion criteria.

5.4 Comparison of automated and manual data

In this section three datasets are compared: manual, reduced automatic (including exactly the same tokens as the manual data) and full automatic (including all extracted and measured automatic data).

This section starts by comparing the ranges and distribution of the raw F1 and F2 measures and the way that these combine over the whole of the vowel space across the three datasets. It then compares the results of patterning by word position across these three datasets as word position was shown to have a clear effect on the formant values of unstressed vowels (Chapter 4). It then compares the effects of additional variables between the manual and reduced automated data only.

5.4.1 Statistical comparisons of full automatic, reduced automatic and manual datasets

Tables 5.1 and 5.2 compare the basic statistics for F1 and F2 across the three datasets. Unsurprisingly, the reduced automated measurements are more similar to the manual data than the full automatic data is. As expected the full automatic data has noticeably larger standard deviations in both the F1 and F2 data. The reduced automatic data also has a noticeably bigger standard deviation than the manual data in its F1 measurements.

***Table 5.1** F1 statistics across the three datasets (Hz)*

	mean	median	standard deviation	IQR	N tokens
manual	464	445	114	121	4980
reduced automatic	477	449	157	127	4980
Full automatic	513	459	224	157	9105

***Table 5.2** F2 statistics across the three datasets (Hz)*

	mean	median	standard deviation	IQR	N tokens
manual	1720	1703	308	409	4980
reduced automatic	1738	1718	306	385	4980
Full automatic	1744	1724	326	401	9105

The greater spread of measurements in F1 compared to F2 is apparent in the violin plots in Figures 5.1 and 5.2, suggesting that the automatic system of measurements had more problems in measuring F1 than F2. Whilst the main body of the distribution is similar over the three datasets there are many more clearly erroneous measurements for the automatic data, particularly for the full automatic data. Notably there are many more overestimates than underestimates for F1.

This shows the initial quality of the raw automated measurements before any data cleaning. The existence of outlying tokens is not necessarily a major problem, as obvious errors should be able to be removed with some data cleaning and through removing obviously problematic tokens.

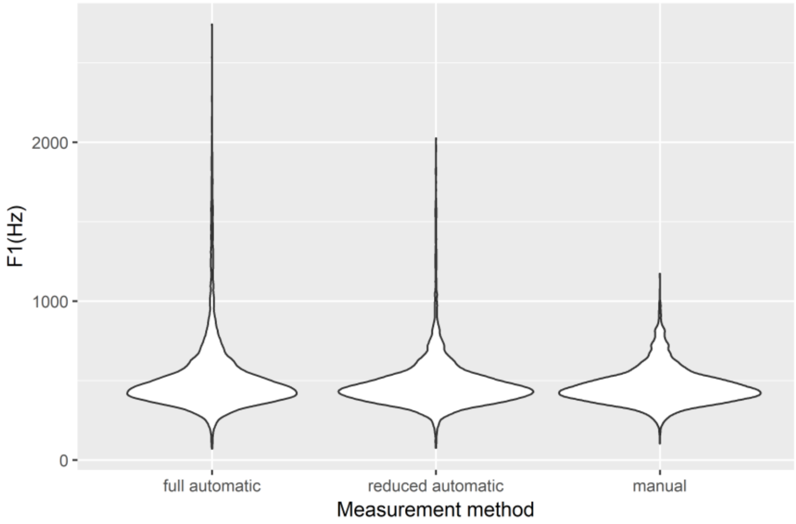


Figure 5.1 *Violin plot for F1 (Hz) by dataset*

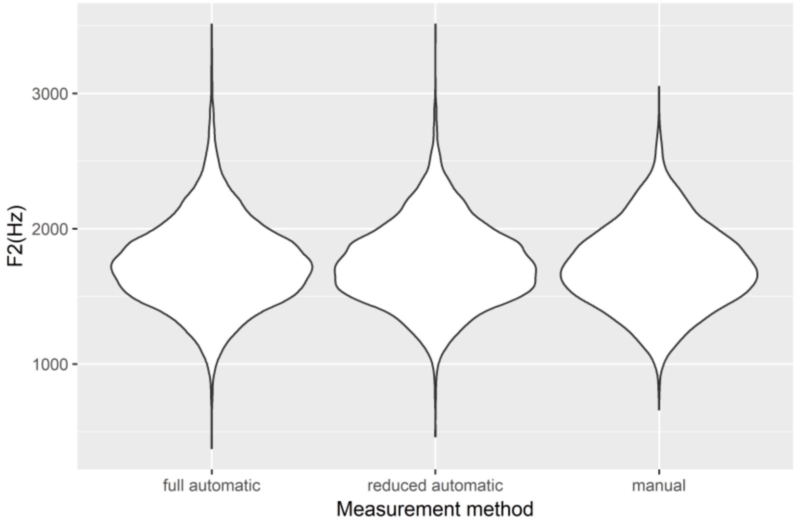


Figure 5.2 *Violin plot for F2 (Hz) by dataset*

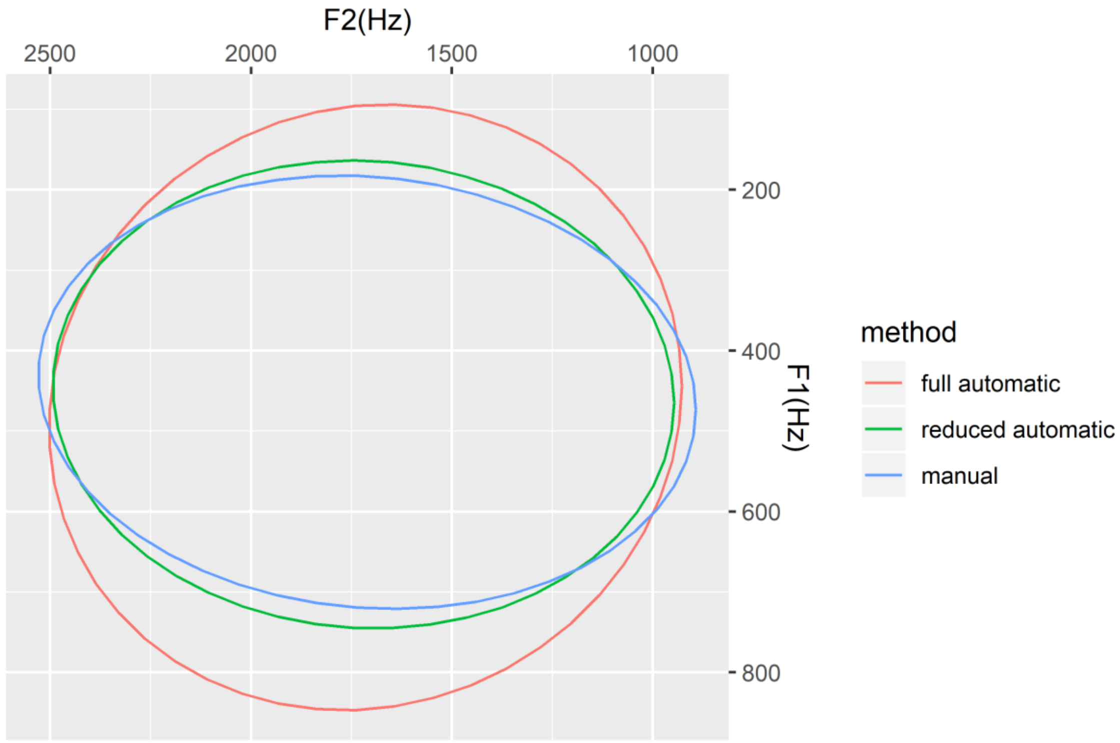


Figure 5.3 *0.99 ellipses of unstressed vowel data (Hz) by dataset*

Figure 5.3 shows how these F1 and F2 measurements combine over the whole vowel space. Over the whole vowel space, there is not a great difference between the reduced automatic and manual datasets. The full automatic data is also very similar on the F2 dimension. However, the positioning of the full automatic data is markedly different on the F1 dimension, reflecting the fact that there appeared to be more serious errors in the F1 than F2 measurements.

If we look at the same graph for individual speakers (Figure 5.4) we can see that the vowel space distributions for the reduced automatic and manual data are remarkably similar for most speakers, considering no data cleaning has taken place. For many of the speakers the ellipses for the manual and reduced automatic data overlap almost completely e.g. speakers, 3, 4, 10 and 19). For the majority of speakers the full automatic data is, however, quite different (e.g. speakers 10, 11, 25, 26). Note that in the full automatic data there is a slight positive correlation between F1 and F2, which suggests that where there is a problem with the measurement of F1 there was likely to also have been a problem in the measurement of F2. This suggests that in many cases F2 may have been measured instead of F1, and F3 instead of F2.

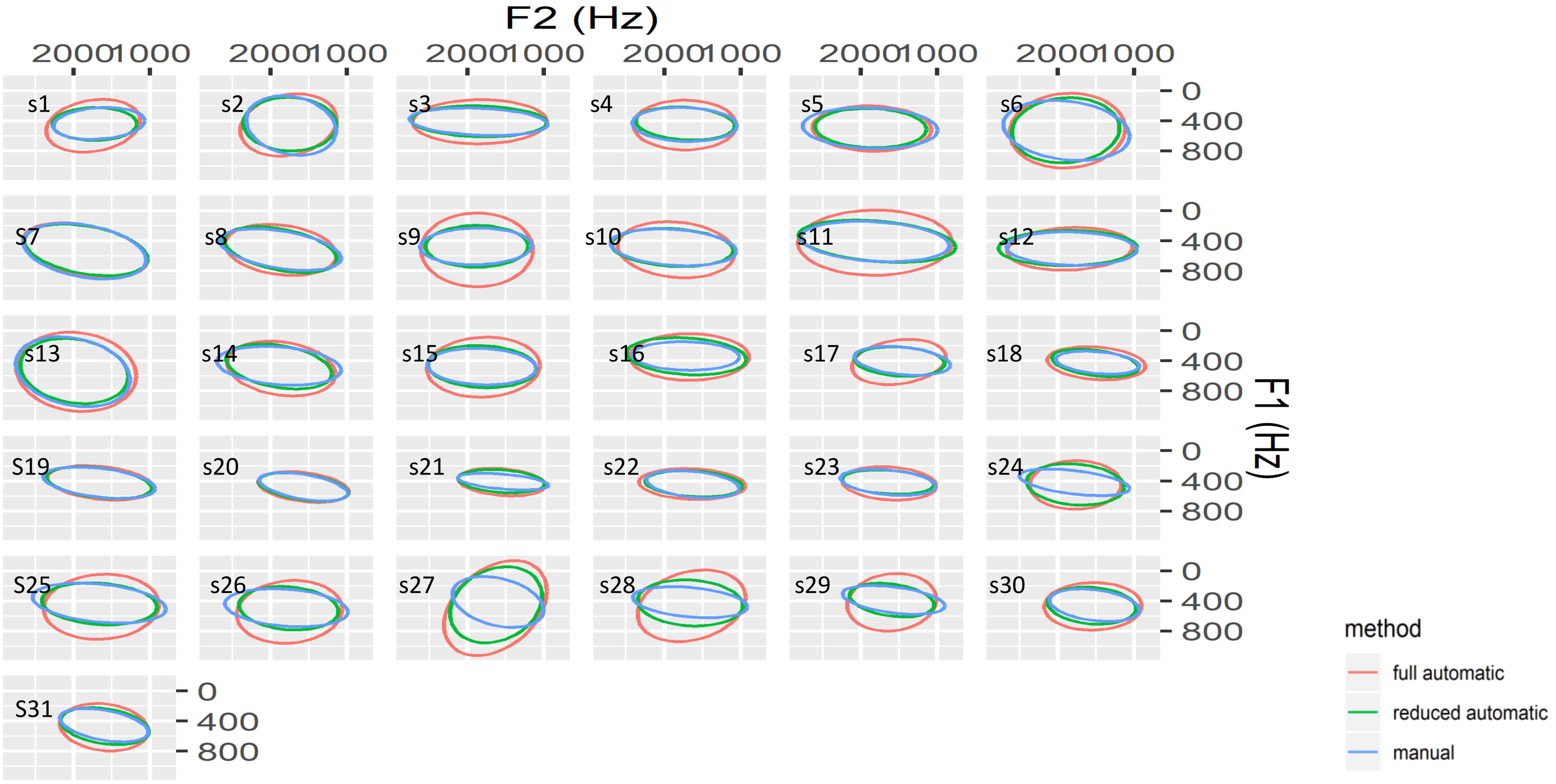


Figure 5.4 *0.99 ellipses of unstressed vowel data (Hz) by dataset and speaker*

Of course, it is not just important that automatic measurements take up the same space in the vowel space, but that meaningful patterns that are found in manual analysis are able to be replicated. I therefore also examine how well these three datasets show the pattern of word final schwas having a higher F1 and lower F2 than non-final schwa, as this was a clear pattern found in the original analysis in Chapter 4, and this is also a pattern consistently found amongst many other studies on schwa (Flemming and Johnson, 2007; Lilley, 2012, Bekker, 2014). Note that the analyses here are slightly simplified compared to Chapter 4. Here all non-final vowels are compared with all word final vowels altogether, rather than additionally considering the effect of following phonetic context. The ellipses in Figure 5.5 show that the reduced automatic data set is remarkably similar again to the manual dataset. Although there is a difference between word final and non-final schwa in the full automatic data, the datasets are not as separate as for the other two datasets, and in F1 in particular the distributions for the two word positon groups are very similar.

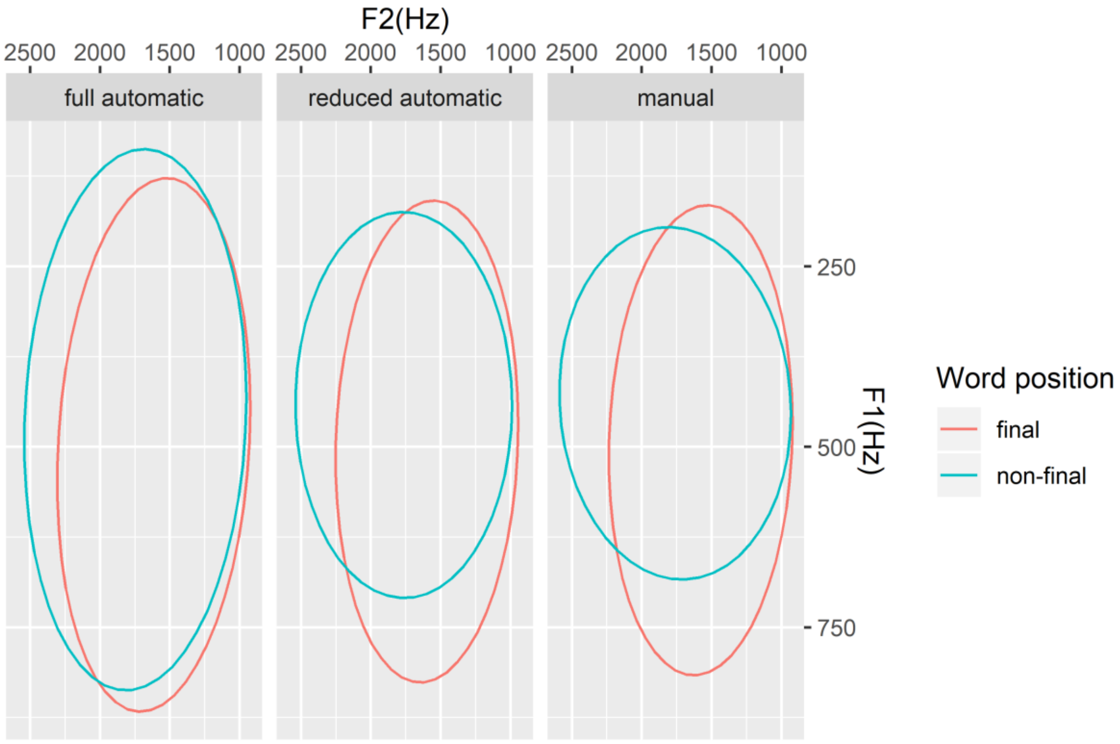


Figure 5.5 *0.99 ellipses of unstressed vowel data (Hz) by dataset and word position*

A mixed effects model with word position as a fixed effect and random effects for speaker and word was modelled for each of the datasets. Tables 5.3 and 5.4 show the differences between the coefficients for the three sets of measurements.

Table 5.3 *Summary of fixed effects for word position on F1 (Hz) (reference level is word final)*

	Estimate	Std. error	*t*-value
manual			
(Intercept)	511.582	9.648	53.02
Word position. Non-final	-65.521	4.762	-13.76
reduced automatic			
(Intercept)	510.851	9.366	54.542
Word position. Non-final	-46.784	6.048	-7.735
Full automatic			
(Intercept)	533.18	10.14	52.577
Word position. Non-final	-34.94	7.5	-4.658

Note that for all three datasets there was a significant F1 difference despite the huge amount of noise in the full automatic data. However, the full automatic coefficients are markedly different from those in the reduced dataset, with the estimate (or size of difference) smaller, the standard error bigger, and the *t* value (or strength of the effect) smaller compared to the reduced automatic dataset. The reduced automatic data in itself has a smaller estimate, bigger standard error, and smaller strength of effect compared to the manual data. These differences are not surprising given that the reduced automatic dataset contains the same tokens as the manual dataset, and that the full automatic data contains exactly the same measurements and in addition lots of extra tokens as well. It shows clearly that the inclusion of tokens which would not be included in a manual analysis makes a big difference. Even though we do find some effect in the full automatic data, as was noted before, this height difference based on word position was a particularly robust effect, so it might be expected that it can withstand some noise in the data. Other more subtle effects, effects with fewer tokens in each group, or effects created by the interaction of different variables, may well not show up at all with this level of noise in the dataset.

Table 5.4 *Summary of fixed effects for word position on F2 (Hz) (reference level is word final)*

	Estimate	Std. error	*t* -value
Manual			
(Intercept)	1620.74	27.37	59.21
Word position. Non-final	152.33	15.00	10.16
reduced automatic			
(Intercept)	1631.45	25.67	63.55
Word position. Non-final	159.97	13.96	11.46
Full automatic			
(Intercept)	1655.76	24.42	67.81
Word position. Non-final	135.68	12.34	11.00

In comparison, the results in Table 5.4 show that the measurement of F2 was much less of a problem in the automated analysis than F1. Across all three datasets the *t* value is very similar, and both automated datasets actually have a smaller standard error than the manual data. As for F1, the full automatic data does have a slightly smaller estimate than the manual data. Overall, however, there is not much of a discrepancy between the automated and manual data in terms of F2.

5.4.2 Comparison of reduced automatic and manual datasets

As the reduced automatic data was matched by token with the manual data, it was also possible to compare some additional effects between the reduced automatic and manual data. Table 5.5 below shows the comparative effects of phonetic context on the F2 of schwa. Reassuringly again, we find these effects in the automatic data as well as the manual data. Again we find the pattern of the automatic data having lower estimates, higher standard errors and a lower *t* value compared to the automated data, suggesting that even though the tokens are exactly the same that there is still slightly more noise in the automatic than manual measurements.

Table 5.5 *Summary of fixed effects for the effect for phonetic context on F2 (Hz) (reference levels=back)*

	Estimate	Std. error	*t*-value
manual			
(Intercept)	1428.41	26.99	52.92
Preceding: Central context	158.83	12.52	12.69
Preceding: Front context	312.97	18.63	16.8
Following: Central context	175.73	16.45	10.68
Following: Front context	381.46	29.93	12.74
reduced automatic	Estimate	Std. error	*t*-value
(Intercept)	1532.09	26.61	57.574
Preceding: Central context	78.69	14.53	5.416
Preceding: Front context	207.81	21.77	9.548
Following: Central context	134.98	17.76	7.599
Following: Front context	292.16	31.53	9.267

In addition to formant data, the duration of each vowel was also collected, both manually and automatically, and in the original analysis this was found to be an important variable in some cases. We can therefore also examine whether the automatically derived duration measures show the same patterns as when boundaries are found manually. One pattern that exists is that schwas before a pause are longer than word final schwas that are not pre-pausal (Chapter 4). Again, this finding is also found in the reduced automated dataset, with a similarly strong effect for both the manual and automated data, as shown in Table 5.6.

Table 5.6 *Summary of fixed effects for the effect of a following pause on duration (reference level=pre-consonantal)*

	Estimate	Std. error	*t*-value
manual			
(Intercept)	0.059936	0.001944	30.83
Pre-pausal	0.040484	0.002267	17.86
reduced automatic			
(Intercept)	0.067416	0.003354	20.1
Pre-pausal	0.062933	0.003732	16.86

5.4.3 Conclusions

The comparison of the reduced automatic and manual datasets shows that in terms of the measurement of tokens where an unstressed vowel is actually present, the automated measurement performs well, and gives very similar results to the manual measurements. There is remarkable consistency in the overall positioning of vowels in the F1-F2 space, which shows that, whilst there may be slight variations in in individual measurements, on the whole the same patterns are found. Reassuringly, this was also shown to be the case when the results of mixed effects modelling were compared for these two data sets. Where significant effects were found in the original analysis they were also found in the automated data. There were some slight differences, and the *t* value tended to be smaller for the automated data. This can likely be attributed to the fact that the automated data set contains more noise. This in itself, however, is not a big problem. It shows that when using automated, rather than manual data, model results are simply likely to be more conservative in terms of the size of effects shown, and the likelihood of a result reaching statistical significance. Being cautious in the interpretation of small effects is not a bad thing, so this is not an especially negative result.

With regards to the analysis of automated unstressed vowels, the main problem is not in the measurements themselves, but in identifying which tokens should be included in the analysis. Although the analysis with the reduced automated dataset is successful in replicating findings and patterns from the manual data, it was only possible to select these tokens for analysis, based on the manual analysis. Clearly this is an unnatural scenario, and normally in a fully automated analysis this would not have been the case. Therefore, on the one hand, the analysis is reassuring in suggesting that, at least for midpoint and duration measures, automated methods can be used successfully for analysis of unstressed vowels. However, without having the benefit of knowing which tokens to include from previous manual analysis, the data set would have contained a large proportion of false tokens, where the object of measurement was not actually an unstressed vowel token. Whilst a degree of noise is inevitable in any set of measurements, especially if automated, it is obviously not a good idea to have such a large proportion of a dataset to be false tokens. This therefore leads to the question of how, when working with such a large dataset as

ONZE, the dataset can be reduced to a set of tokens which is more reflective of those tokens that would be selected in a manual analysis.

Some erroneous measurement can of course be removed using outlier removal. However, this will not necessarily get rid of all problematic tokens. Not all falsely measured tokens will necessarily look unusual in their raw formant measurements. In particular, since unstressed vowel measurements vary over a wide range, false tokens may not necessarily stand out as looking unusual.

It is therefore possible that it may be necessary to make additional exclusions to the data, based on linguistic knowledge of the phenomena being examined. This next section therefore compares the effects of using different combinations of exclusion criteria in order to replicate more closely the tokens and measurements in the manual dataset.

5.5 Using exclusion criteria to replicate manual data

As the above comparisons between the full and reduced automatic dataset show, much of the problem with the full automated dataset lies in the inclusion of tokens that should have been excluded, rather than any problem with the measurements themselves. This is apparent from the good performance of the reduced automatic dataset in replicating the results of the manual data set. This is somewhat reassuring, in that it shows that when measuring the same valid set of tokens that the manual and automatic measurements are comparable. However, this is an artificial situation in that it was only possible to generate the reduced automatic dataset because of the filtering work that was done manually. It is vital to find steps that can bring fully automated datasets closer to a pool of tokens that are more similar to what would be selected manually, without recourse to individual token selection.

This section thus focusses on comparing the effects of reducing a fully automated data in different ways. There are two main parts to this comparison. The first (5.5.1) examines to what extent different exclusion criteria can bring the set of tokens closer to what is included in the manual analysis. It explains which token types are being targeted in the exclusions and compares their effects on the data. The second section (5.5.2) uses statistical comparisons of the formant values and results associated with each method, in the same way as was done in section 5.4.

5.5.1 Replicating the set of tokens in the manual data set

5.5.1.1 Overall approach

Although only 4,980 tokens were included in the manual analysis of unstressed vowels in Derby, 9,105 were included in the full automatic dataset. This section looks at the extent to which different exclusion criteria reduce the automated data to a set of tokens which are more closely matched with the data in the manual analysis.

When examining the tokens included in the automated datasets we can classify them in terms of true positives, false positives, true negatives and false positives. The positives are the tokens that were included in the automated dataset, and the negatives are those that were excluded in the automated dataset. Whether a particular token is true or false depends on whether its inclusion or exclusion aligns with whether it was included/excluded in the manual analysis. Table 5.7 below summarises these categories.

Table 5.7 *Correspondences between manual and automatic data*

	True	False
Positive	Included in manual analysis Included in automated analysis	Not Included in manual analysis Included in automated analysis
Negative	Not Included in manual analysis Not Included in automated analysis	Included in manual analysis Not Included in automated analysis

Since the full automatic analysis included 9,105 tokens as opposed to the 4,980 tokens in the manual data this means that there were lots of tokens that were in the full automatic and not in the manual data. If we take the original full automatic data then all of these tokens can be categorised as positive since they were all included in the dataset. As only 54% of these tokens were also in the manual analysis, we can regard the full automatic data as being made up of 54% true positives and 46% false positives. The aim of using exclusion criteria is to minimise the proportion of false positives and instead turn them into true negatives, as false positives contained in the analysis will introduce noise and perhaps skew the data. However, no exclusion criteria will be perfect and so it is inevitable that exclusions will include false negatives as well as true negatives. Therefore the aim is also to minimise the number of false negatives i.e. the number of tokens that are wrongly excluded. In excluding data there can be a trade-off between the number of false positives and negatives. More stringent measures may decrease the number of false positives, but in doing this may also wrongly exclude much data and thus also have a high number of false negatives. Conversely, less stringent measures may not have so many false negatives, but also wrongly include much data in the analysis. These two aspects of the exclusion criteria are combined into an accuracy measure. Accuracy assesses the proportion of individual tokens whose inclusion/exclusion aligns with the decision in the manual analysis. Put simply, it is the proportion of all 9,105 tokens that are either true positives or true negatives.

The distribution of the excluded tokens in the manual analysis was not random; rather there were a few common reasons why tokens were excluded from the final manual dataset (see section 5.3). Therefore knowledge of these patterns means that is possible to aim to target exclusions to specific groups of tokens which would be more likely to be excluded in manual analysis. The most common linguistic reasons for exclusions were that the vowel was voiceless, and that the vowel was fully deleted i.e. there was no audible or acoustic trace of it. I therefore excluded tokens based on criteria aimed to target voiceless or deleted tokens. The effects of these two linguistic criteria are compared with exclusions based on outlier removal, as well as various combinations of these three criteria. The aim of any exclusion criteria is to include the tokens that were included in the manual analysis. These will be referred to as *target inclusions*. The other aim is to exclude tokens which were excluded in the manual analysis. These will be referred to as *target exclusions*.

For each of the exclusion criteria assessed, four measures are reported:

1) The proportion of target inclusions which are true positives, as opposed to false negatives. This assesses the strength of the criteria at minimising false negatives, and maximising the amount of good data which is included
2) The proportion of target exclusions which are true negatives, as opposed to false positives. This assesses the strength of the criteria at minimising false positives, and maximising the amount of false data which is excluded
3) Accuracy - a combination of the above two measures; the proportion of the overall dataset which are either true positives or true negatives.
4) The proportion of the remaining data which is also included in the manually filtered data set, or the proportion of positives which are true positives. This is the proportion of the actual data analysed which is likely to be good data.

Table 5.8 below shows what the baseline measures are for the full automatic data. They show that without any exclusions 54% of the full automatic data matches the manual data set, and as such 54% of the data is accurate. Therefore the aim of the exclusion criteria is to increase both of these figures.

Table 5.8 *Replication of manual tokens in full automatic dataset*

% target inclusions which are true positives	100
% target exclusions which are true negatives	0
% accuracy	54
% remaining data also in manual data	54

5.5.1.2 Targeting likely deletion contexts

There is a general consensus about when a schwa can be deleted lexically. It has been suggested that it can happen if the preceding consonant is less sonorant than the consonant that follows it, and the vowel that follows it is also unstressed (Zwicky, 1972; Dalby, 1986; Patterson et al 2003) e.g. the medial vowel e.g. *memory*. When deciding which were potential contexts for deletion these contexts were therefore all included.

The above definition refers to word medial contexts. There are also other contexts where a schwa may be deleted which do not fit these phonological criteria. They occur mainly in the first syllable of words e.g. in words such as *about*, *around*, *because*. Therefore these were also coded as possible deletion contexts if either, based on native speaker judgment, it was thought a possible site for deletion, or if they had a variant listed in the Longman dictionary (Wells, 1990) without the vowel. This was based on inspection of individual lexical items.

5.5.1.3 Targeting likely voiceless contexts

The contexts where voiceless schwa occurs most often is in syllables with other voiceless sounds (Davidson, 2006; Glowacka, 2001; Heselwood, 2007), e.g. *potato*. The effect of excluding tokens in syllables with voiceless consonants was compared for open and closed syllables, in order to be more targeted in what tokens would be best to exclude in order to maximise true rather than false negatives. Table 5.9 below shows the effect of excluding tokens from voiceless open and closed syllables. As can be seen, it is only in open syllables where there is any improvement in accuracy, or in the proportion of the remaining data which is also in the manual data. Although voiceless tokens do occur in the closed syllable environment, they are rare, so excluding tokens from this environment actually produces more false than true negatives. Based on this comparison potentially voiceless tokens were therefore only excluded when in an open syllable following a voiceless consonant. This is because exclusions in this context improve both accuracy, and the proportion of the remaining data which is the manual data, in relation to the full automatic data.

Table 5.9 *Replication of manual tokens across different voiceless exclusion contexts*

	In open syllable	In closed syllable
% target inclusions which are true positives	84.9	77.5
% target exclusions which are true negatives	20.6	5.3
% accuracy	58.2	47.1
% remaining data also in manual data	58.1	51.1

5.5.1.4 Outlier removal criteria

Measurements were marked as outliers using the two following criteria:

1) Vowels were marked as outliers based on the formant values in Hillenbrand et al (1995). Tokens were classified as outliers if either the first or second formant measurement occurred below the 1st or above the 99th percentiles, calculated separately for males and females, to exclude acoustically implausible measurements.
2) Measurements were marked as outliers based on the individual speaker's measurements. Tokens were excluded if they had measurements marked as an outlier for either F1 or F2. Measurements were classified as outliers using the procedure built into R's (R Core Team, 2020) boxplot () function, which identifies outliers as data points that are more than 1.5 IQR higher/lower than the upper/lower quartiles.

5.5.1.5 Comparison of exclusion criteria

Table 5.10 below shows the comparative effects on the data by excluding different combinations of tokens. It compares the effects of exclusions based on targeting likely deletion contexts, voiceless contexts, and outliers, and also different combinations of these criteria.

Table 5.10 *Replication of manual tokens for different exclusion contexts (D=deletion, V=voiceless, O=outlier, E=excluded)*

	DE	VE	OE	DVE	ODVE
% target inclusions which are true positives	85.3	84.9	85.2	72.6	62.6
% target exclusions which are true negatives	36.9	26.1	24.8	53.9	63.3
% accuracy	62.9	58.2	57.8	64.3	62.9
% remaining data also in manual data	62.0	58.1	57.7	65.6	67.3

Although the highest accuracy for any measure is 64.3% (DVE), all measures offer some improvement on the accuracy of the full automatic data (54%). Although the accuracy measure does not vary hugely between the different criteria, there is somewhat of a trade-off between the amount of true positives and true negatives. In the datasets using only one exclusion criteria the correct identification of true positives is much better than the correct identification of true negatives. When two measures are combined (DVE) the retention of true positives worsens and the identification of true negatives improves. This pattern continues when all three exclusion measures are combined (ODVE), such that the proportion of correctly identified true positives and true negatives becomes more similar.

In terms of the effectiveness of exclusions based on linguistic factors, DE is more accurate than VE, but accuracy improves most when both measures are combined in the DVE dataset. By combining the deletion and voiceless exclusion criteria in DVE over half of the target exclusions are able to be recognised as true negatives. These factors result in the proportion of true positives included in the remaining data being higher than either DE or VE in isolation.

In the outlier removal dataset, the accuracy and identification of true negatives is low compared to DVE. It is noteworthy that only 24.8% of target exclusions are recognised as true negatives, whilst 75.2% remain as false positives. This is important as it shows that the majority of problematic tokens cannot be excluded with outlier removal. The automated measurements may produce measurements that appear normal. However, these measurements may be from tokens where it was not appropriate to take a measurement. Outlier removal also produces false negatives. These false negatives are tokens that were deemed acceptable for analysis in the manual filtering process, and where it was possible to take a manual measurement. However, given that the measurements in the automated data are different, such tokens could of course still have outlying measurements, which may be problematic in the analysis of the data. Therefore the labelling of these tokens as false negatives is a slight misnomer in that it would not actually not be desirable to include these outlying measurements in a final dataset. Despite this though, recording false negatives here is still informative insofar as it demonstrates the amount of tokens that were included in the manual analysis that weren't able to be included because of outlying measurements, suggesting faulty measurement. This was the case for 13.8% of target inclusions. According to Table 5.10, though, it is the failure to eliminate false positives which is the greater problem with using outlier removal alone.

Despite the limitations of outlier removal in excluding relevant tokens, I do not suggest that outlier removal should not be used, as this would clearly not be sensible. Given that the main problems with using outlier removal are in what it does not exclude (false positives), rather than what it does (false negatives), it makes sense to use outlier removal, but in addition to use other exclusion measures in order to eliminate those false positives which outlier removal does not. The real question is therefore whether we can improve upon outlier removal alone when using it in combination with other linguistic measures.

When we look at the numbers for ODVE we can see that additionally using DVE in combination with outlier removal is beneficial. It raises the proportion of target exclusions which are identified as true negatives from 26.1% to 63.3%. Although the additional tokens excluded also bring the trade-off of having a decrease in true positives and an increase in false negatives, the overall effect of using this additional criteria is beneficial. This is reflected both in a 5.1% increase in accuracy, and a 9.6% increase in the proportion of the remaining data which are true positives.

These numbers reported in this section refer to the comparative strengths of exclusion criteria in producing a dataset that is more aligned with the manual data in terms of the tokens included. Of course, although the two issues are interlinked, the real test of the success of the exclusion criteria is in the data itself, and the quality of the data analysis which can be produced. In addition, none of the exclusion criteria are completely successful in replicating the tokens from the manual analysis. The highest accuracy of any criteria is 64.3% (DVE), and the highest proportion of the remaining data which is also in the manual data is 67.3% (ODVE).

Although there is some variation between criteria in terms of the success at eliminating false positives and retaining true positives in the data, not all individual tokens will have an equal effect on the data. Therefore the following section (5.5.2) assesses how the changes made affect the statistical patterns in the data. The following section compares DVE, OE, and ODVE in terms of the statistical patterns in the actual measurements. DVE has been chosen as in terms of token selection this was the most successful in terms of accuracy. OE has been chosen as it is important to see what the effects of outlier removal alone are. ODVE has been chosen as it was the most successful in terms of the percentage of the remaining data which was also in the manual data. We will therefore be able to compare an exclusion measure which uses linguistic criteria alone, with outlier removal alone, and a measure which uses a combination of the two.

5.5.2 Statistical comparison of exclusion criteria datasets

This section compares the manual, reduced and full automatic data with three additional automated datasets. These are:

- Full automatic with outliers excluded (OE) (n=7,344)
- Full automatic with those in targeted likely voiceless and deletion contexts excluded (DVE) (n=5,517)
- Full automatic with outliers, targeted likely voiceless tokens, and targeted likely deletion contexts excluded (ODVE) (n=4,529)

Where possible, I show statistics comparing all six datasets but for clarity of visualization some of the graphics do not include the reduced automatic and full automatic datasets, as these graphics were already included in section 5.4.1.

Table 5.11 shows the F1 statistics compared across the 6 datasets. In terms of F1, in some ways the OE data behaves as was expected. In terms of both the mean and IQR the F1 reduces to get closer to that of the manual data, although it still does not quite reduce to this level. Curiously, the median actually rises slightly. What is very striking is that the standard deviation reduces to a level that is even lower than that of the manual data. This can be explained by the fact that all of the extreme values have been excluded, whereas in the manual data, where each token was judged individually, some of the tokens with extreme values would have remained. In every measure, though, the data becomes closer to the manual after the outlier removal.

Table 5.11 *F1 statistics across the six datasets (Hz)*

	mean	%diff to manual	median	%diff to manual	standard deviation	IQR	N. tokens
manual	464	0	445	0	114	121	4980
reduced automatic	477	2.8	449	0.9	157	127	4980
OE	486	4.74	464	4.27	106	127	7344
DVE	492	6.03	456	2.47	181	143	5517
ODVE	483	4.09	463	4.04	102	122	4629
Full automatic	513	10.56	459	3.14	224	157	9105

The main differences between the F1 of the full automatic dataset and the manual and reduced automatic data is that the full automatic F1 data has a notably higher mean, standard deviation, and IQR. The median is not notably different (never more than 14 Hz). This suggests that the major differences are caused by extreme values. All three of the exclusion datasets reduce the mean, standard deviation, and IQR to a level which is nearer to that of the manual and reduced automatic data. However, this reduction is notably greater for the measures which involve outlier removal. The differences between OE and ODVE in these figures is negligible.

Table 5.12 *F2 statistics across the six datasets (Hz)*

	mean	%diff to manual	median	%diff to manual	standard deviation	IQR	N. tokens
manual	1720	0	1703	0	308	409	4980
reduced automatic	1738	1.04	1718	0.88	306	385	4980
OE	1709	0.63	1699	0	280	377	7344
DVE	1755	2.03	1738	2.06	313	394	5517
ODVE	1728	0.47	1717	0.82	281	376	4629
Full automatic	1744	1.4	1724	1.23	327	401	9105

As noted in section 5.4.1, there was not as much issue in the measurement of F2 as there was for F1 in the full automatic data. Therefore the need for change in the F2 measures was not as pertinent as for F1. It is still, however, important to know what changes the exclusion criteria make to F2. Again, both the measures involving outlier removal result in smaller measures of dispersion. Here, both measures of dispersion reduce beyond the level of the manual data set. Note, however, that the reduced automatic also has a similarly low IQR, so this lower IQR is actually likely to be related to the difference in measurement method, rather than because of differing tokens in the dataset. Again, there is some indication that DVE performs worse than either OE or ODVE in that DVE causes the measures of central tendency to rise and get further away from the manual data.

Below we examine the spread of the data visually. The violin plots in Figure 5.6 show that the extreme values have been removed for OE. The ODVE data does not look very different to OE and similarly DVE does not look very different to the full automatic data. The upper tail for DVE is, however, not quite as long as for the full automatic data so it appears that this method of exclusion removes some of the most extreme high F1 values.

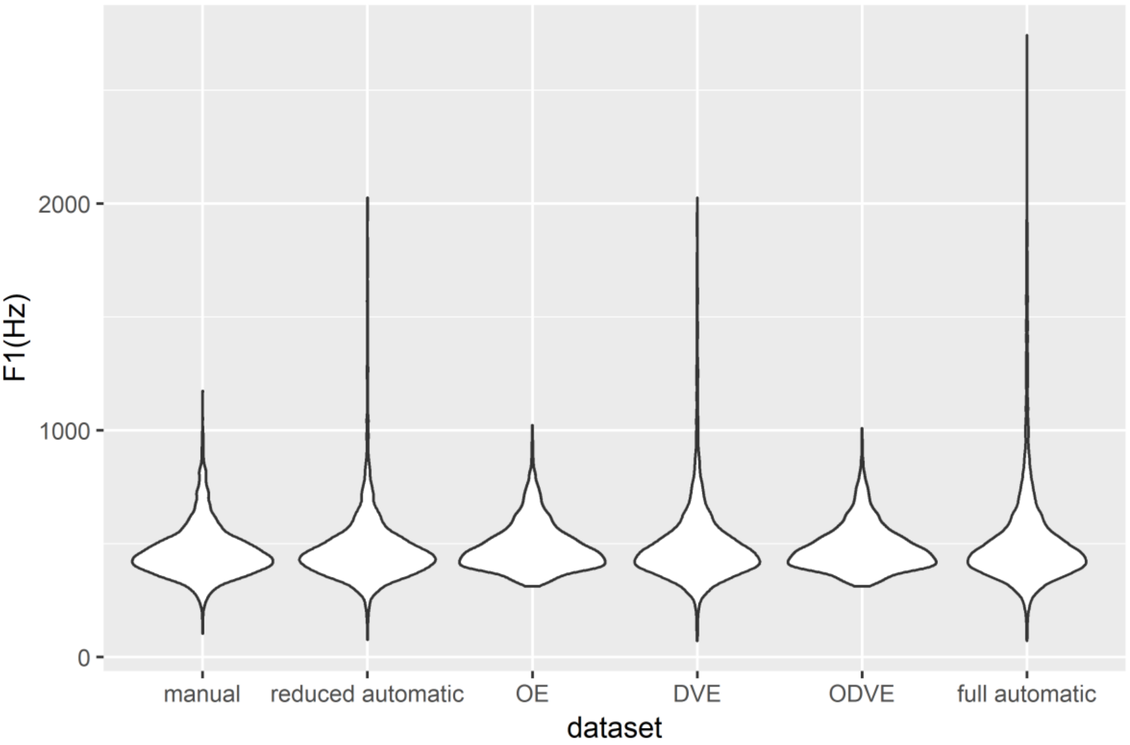


Figure 5.6 *Violin plot for F1 by exclusion dataset*

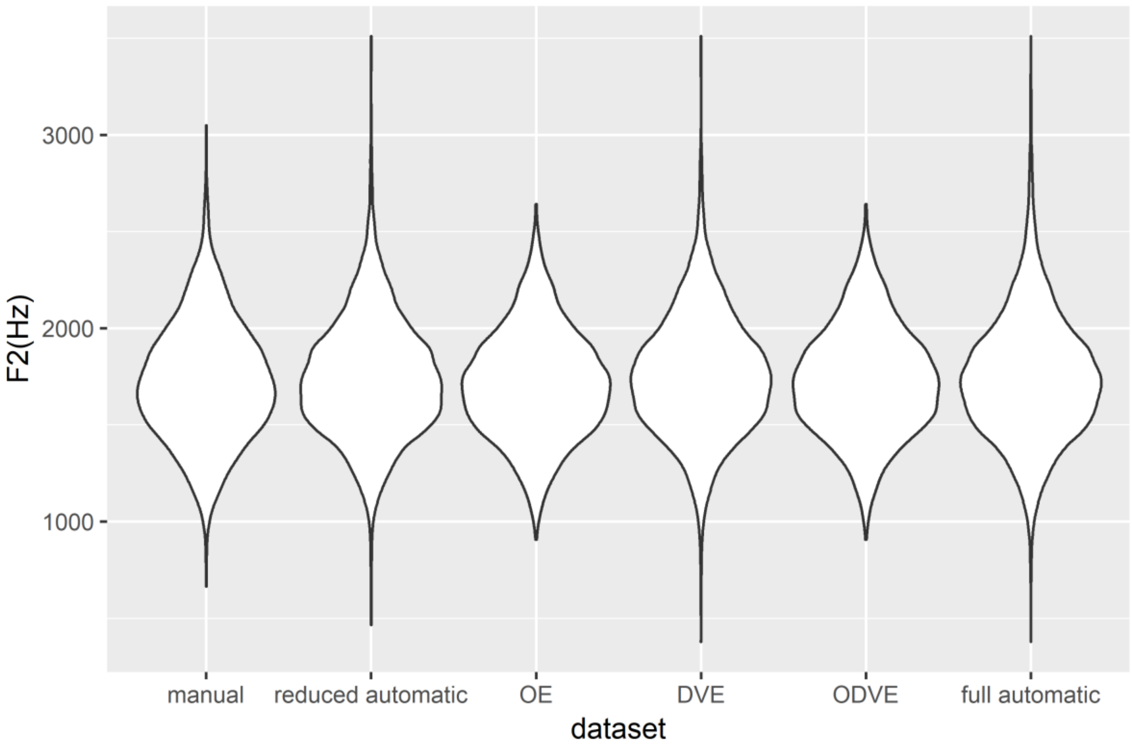


Figure 5.7 *Violin plot for F2 by exclusion dataset*

For F2, as shown in Figure 5.7, none of the exclusion methods make the data look drastically more like the original manual data than full automatic data, as might have been indicated by the mixture of results in the basic statistics, with no particular method improving the data across all measures. Whilst ODVE and OE remove the outliers, the F2 outliers in the full automatic are not as extreme as for F1.

In summary, there are clearly more problems in the raw F1 than F2 in the full automatic data. The measures involving outlier removal, as would be expected, reduce the dispersion in the data. All combinations of exclusion criteria produce an improvement in F1 measures but this is more so in OE and ODVE than DVE.

The ellipses in Figure 5.8 show the comparison between the distributions of the edited automated datasets with the manual data across the whole vowel space.

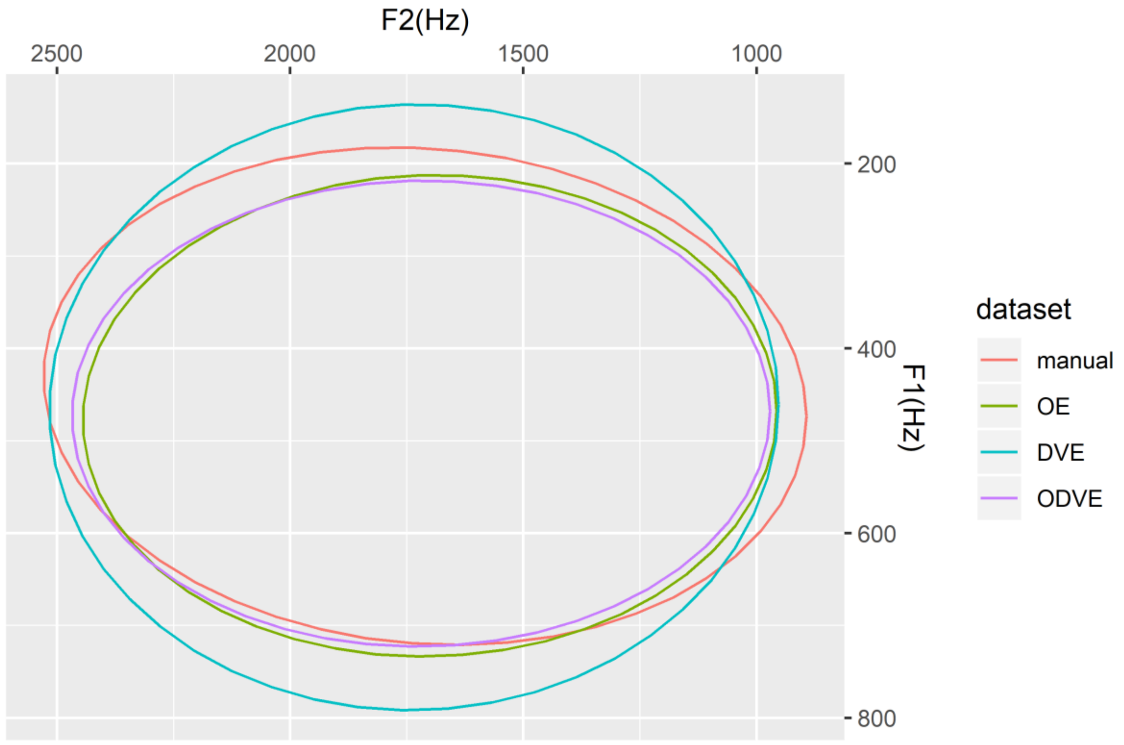


Figure 5.8 *Ellipses of unstressed vowel data by exclusion dataset*

It is again apparent that the outlier removal dataset occupies a smaller area of the dataset than the manual data. OE and ODVE are again very similar. However, the range of data points for DVE is still much larger over F1, and does not offer much improvement over the full automatic data (compare with Figure 5.3).

Figure 5.9 shows the same ellipses for individual speaker. Again, ODVE and OE are very similar, covering a very similar vowel space for the vast majority of speakers. There is clearly some variation between speakers. For some speakers (e.g.7, 19, 20), even the full automatic data was comparable to the manual data, and there was therefore not much room for improvement after exclusion. For many of the speakers there is not a discernible difference between the manual data and any of the automated datasets. Where there is a clear difference between the manual data and one of the exclusion methods the majority of the time it is DVE which is most different in vowel space from the manual data (e.g. speakers 24, 27 and 28). Again, this therefore suggests that overall all of these exclusion methods improve upon the full automatic data. This is truer for some speakers than others, but overall the datasets which include outlier removal are superior.

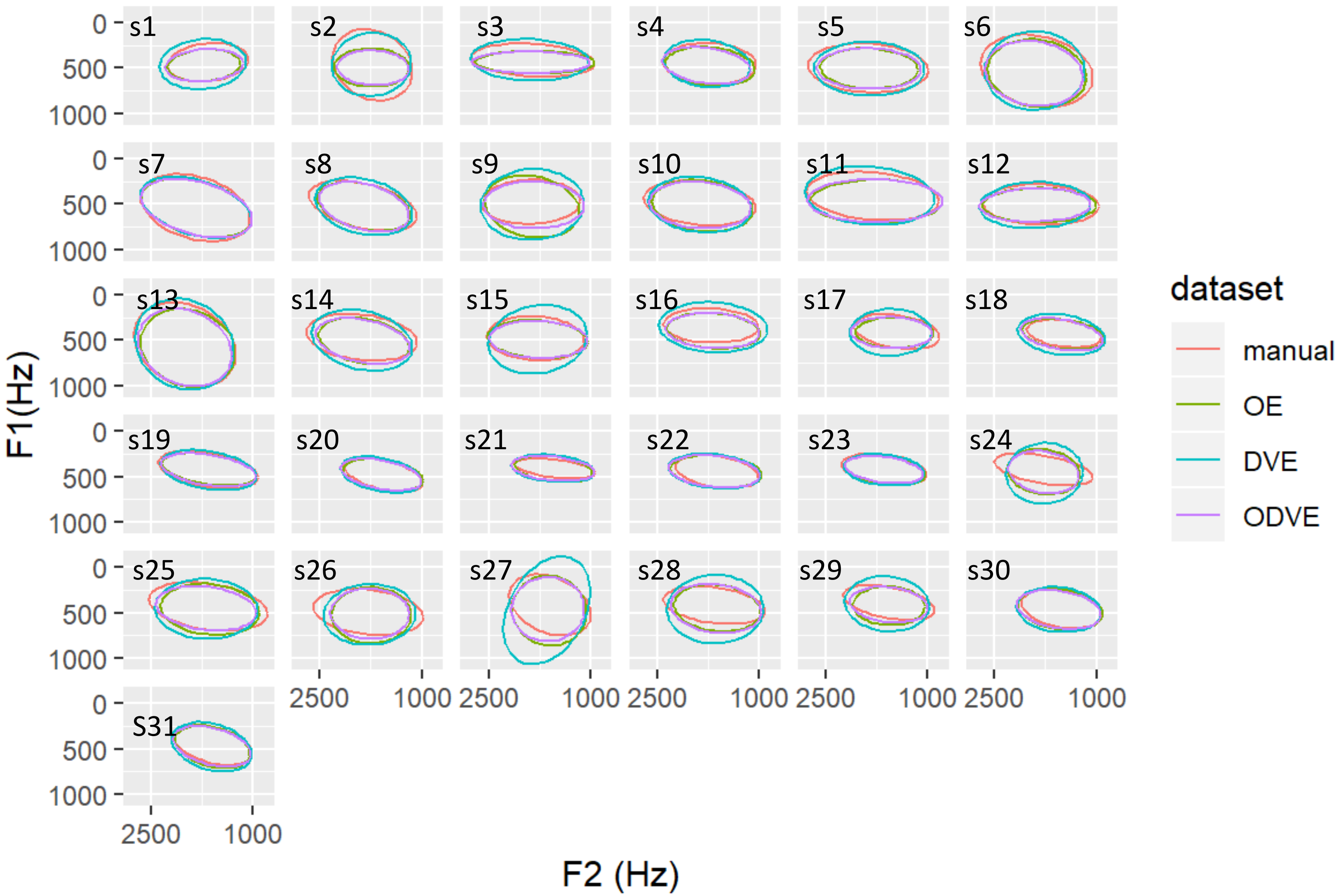


Figure 5.9 *0.99 ellipses of unstressed vowel data by exclusion dataset and speaker*

So far I have only compared the F1 and F2 values over the whole data sets. What is more important is seeing how the datasets replicate the patterns found in the manual data. Of course, in the majority of linguistic analysis it is the patterns in the data which are the object of interest, rather than the raw measurements. The ellipses in Figure 5.10 below show the overall differences between word final and non-final schwa over the vowel space. The manual data shows a clear difference between the word positions in both F1 and F2. Word final vowels have on average a higher F1 and a lower F2. Again, it is F1 which is the main problem with the full automatic data. Whilst the F2 difference is reasonably clear in the full automatic data, the graph only shows a marginal F1 difference. Therefore, whilst there is room for improvement upon the F1 and F2 differences shown in the full automatic data, it is the F1 that requires most improvement.

So far, it has been shown that the measures involving outlier removal are more successful in replicating the distribution of measurements and the shape of the data over the vowel space. Here we see a different pattern. It is actually the datasets involving linguistic exclusions (DVE and ODVE) that show more improvement, showing the difference between the two word positions more clearly. All three of the exclusion criteria offer an improvement over the full automatic data to a degree. However, the separation of the two word position groups is less clear when just outlier removal is used rather than ODVE and DVE. This exemplifies the fact that outlier removal does not remove all problematic tokens. Measurements can look normal in their formant values, but if that measurement is from a problematic token, or from a word where no schwa vowel exists at all, then these erroneous measurements will muddy the patterns in the data. Without using any additional linguistic criteria, these false positive tokens remain in the data. It is striking how much clearer the difference between the word positions becomes, simply by using linguistic exclusion criteria (DVE). Without recourse to outlier removal, exclusion of potentially deleted or voiceless tokens still removes the most extreme F1 values in non- final vowels. ODVE looks similar to DVE in terms of the positional differences, but of course the boundaries for word final and non-final groups are less extreme. Again this is has much more of an effect on F1. This results in ODVE being less stretched along the F1 dimension than DVE. This can be seen in the improvement in the shape of the non-final vowel space, becoming more spread in F2 than F1, as in the manual data.

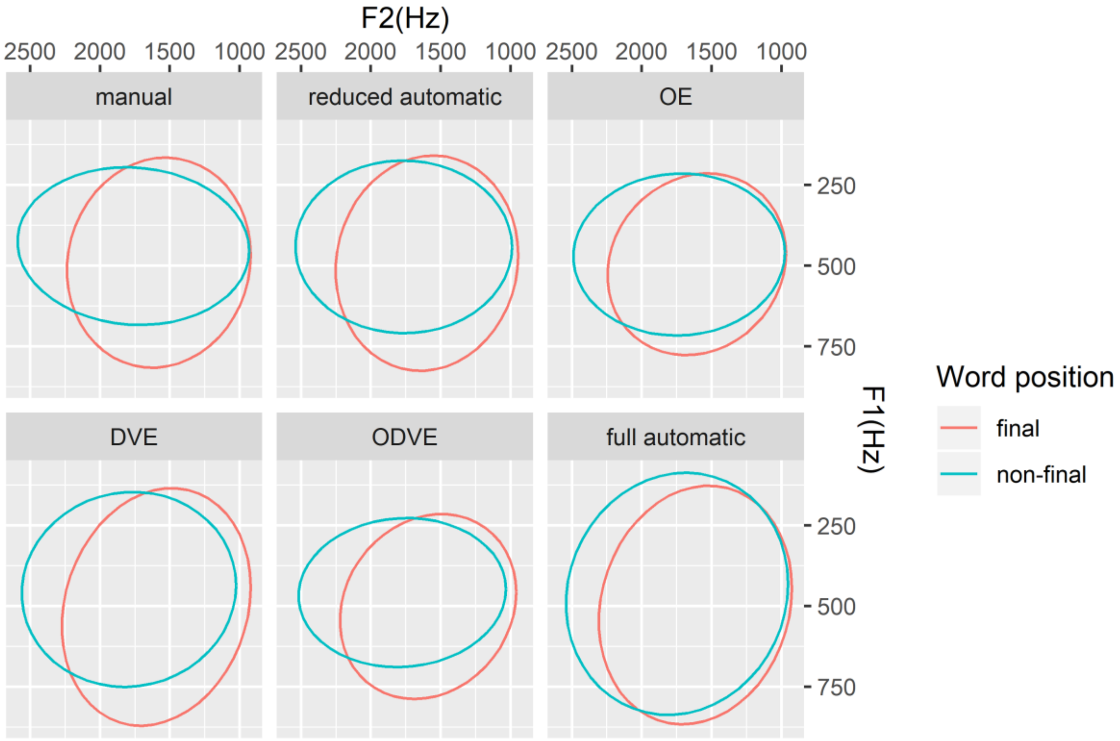


Figure 5.10 *0.99 Ellipses of unstressed vowel data by dataset and word position*

Figure 5.10 showed the whole areas covered in the vowel space by both word positons across all the dataset. However, this only shows the shape of the whole distribution; it does not identify the degree of variation within it. To do this, and thus to get a real sense of the strength of the patterns across the datasets, we can compare the statistics within mixed effects regression models. Again I compare models from each dataset, using the same model structure as described in section 5.4.1.

Table 5.13 *Summary of fixed effects for word position on F1 (Hz) (reference is word final)*

	Estimate	Std. error	*t*-value
Manual			
(Intercept)	511.582	9.648	53.02
Word position. Non-final	-65.521	4.762	-13.76
reduced automatic			
(Intercept)	510.851	9.366	54.542
Word position. Non-final	-46.784	6.048	-7.735
With OE			
(Intercept)	511.070	8.992	56.837

Word position. Non-final	-36.193	3.685	-9.821
With DVE			
(Intercept)	534.563	9.464	56.482
Word position. Non-final	-57.572	6.165	-9.338
With ODVE			
(Intercept)	520.273	9.107	57.13
Word position. Non-final	-51.804	4.091	-12.66
Full automatic			
(Intercept)	533.18	10.14	52.577
Word position. Non-final	-34.94	7.5	-4.658

In comparing the full automatic data with the manual data (section 5.4.1) we saw that the F1 estimate was substantially lower and the standard error substantially higher than in the manual data. Consequently, because of both of these things *t* was substantially lower. The same was also true of the reduced automatic data, although not to the same degree. In improving upon the full automatic data the focus is therefore on increasing the estimate, decreasing the standard error, and thus increasing *t*. The results are shown in Table 5.13. In the modelling, as in Figure 5.10, it is clear that the different types of exclusion criteria make different improvements and changes to the data. When outlier removal alone is used there is a clear improvement in the standard error, but not in the estimate. Indeed, the standard error reduces to a level which is below that of the manual data. This change is also reflected in Figure 5.10, where extreme values are removed, but there is not much of a change in the difference between the two groups.

In DVE the estimate is a clear improvement on the full automatic data, although it does not quite reach the level of the manual data. By contrast, the improvement in the standard error is only slight. This is the same pattern that is reflected in Figure 5.10 above, where we can see that the word position groups are more clearly separated than the full automatic data, but that the data is very spread because of the lack of outlier removal. By using both the outlier removal and the linguistic criteria exclusions, ODVE improves the data in the ways that both OE and DVE improve the data. ODVE combines clear improvements in both the estimate and standard error and consequently ODVE has the most improvement in *t*. It is worth noting that although the standard error for ODVE is not quite as small as for OE, it is actually still smaller than what was found in the manual data.

It is clear from both Figure 5.10 and from the statistics in Table 5.13 that both the outlier removal and linguistic exclusion criteria have an important contribution to make in terms of data cleaning. Through the removal of different sets of false positives from the data they both improve the data in different ways. Indeed the most improvement is seen when these two types of criteria are combined in ODVE. Importantly, all of the compared datasets are of varying sizes, due to having different amounts of tokens excluded. It is notable then that despite being the smallest dataset ODVE is still the most similar overall to the manual data, and still has the strongest effect of word position. Not only is this the smallest dataset overall, but it also contains the smallest number of true positives. This suggests that it was worth removing a higher proportion of true positives from the data for the sake of also removing a higher proportion of false positives. The relative strength of ODVE compare to the other automated datasets, particularly full automatic, at replicating the patterns in the manual data suggests that it isn't necessarily the case that if a dataset is large enough then the noise in the data will be drowned out by other tokens. It is the smallest dataset here which is the most successful at replicating the patterns of the manual data.

Table 5.14 *Summary of fixed effects for word position on F2 (Hz) (reference is word final)*

	Estimate	Std. error	*t*-value
Manual			
(Intercept)	1620.74	27.37	59.21
Word position. Non-final	152.33	15.00	10.16
reduced automatic			
(Intercept)	1631.45	25.67	63.55
Word position. Non-final	159.97	13.96	11.46
With OE			
(Intercept)	1634.47	24.90	65.64
Word position. Non-final	126.73	11.43	11.09
With DVE			
(Intercept)	1633.61	24.93	65.52
Word position. Non-final	170.57	13.17	12.95
With ODVE			
(Intercept)	1622.10	25.39	63.89
Word position. Non-final	160.11	12.81	12.50
Full automatic			
(Intercept)	1655.76	24.42	67.81
Word position. Non-final	135.68	12.34	11.00

As could be seen in section 5.4.1 there is less overall discrepancy between F2 in the manual and automated datasets. The similarity of the statistics between the reduced automatic and manual data also suggested that there was less of a problem with measurement error. Here, Table 5.14 shows that the standard error in the full automatic data is actually smaller than it is in the manual data, and this contributes to it also having a bigger *t* value. The estimate is, however, slightly smaller than in the manual data. Despite this slight difference, overall there was not so much of a need for improvement in the automated data in F2. There is little difference in the standard error amongst the three edited datasets. There is, however, slightly more difference in terms of the estimates. From the original level in the full automatic dataset OE decreases the estimate and DVE increases it. ODVE has an estimate intermediate of OE and DVE, and it is the ODVE estimate which is the most similar to the estimate in the manual data. The *t* values of these three datasets are all slightly higher than in the manual data. Overall all of these small differences are fairly inconsequential, as there was not such a problem in the F2 values of the original full automated data.

5.6 Discussion

Overall, the results are reassuring, both in using an automated methods for analysis of unstressed vowels, and in general. This is especially the case given that there were a number of reasons to suspect that unstressed vowels could be particularly problematic to analyse using automated methods.

Overall, the ODVE automated dataset replicated the manual data the best. This shows the importance of using a range of measures in data cleaning. Indeed, ODVE is the most successful as it combines the strength of outlier exclusion and the linguistic exclusion criteria. These results suggest that depending on the object being studied, there may a limit to how much the data can be improved through outlier removal alone, and so additional exclusion criteria may need to be used. It is therefore recommended that the linguistic criteria specific to the given variable of investigation be also considered in data exclusion.

Here, the process of conducting the manual analysis contributed to influencing which exclusion criteria were considered for the automated analysis. Normally, a researcher would not have gathered the same set of measurements first manually, before collecting them using automatic means. However, it is still possible to make an informed decision about likely problematic tokens without manually checking the whole dataset. By measuring a small subset of the data manually, one can become aware of the most common reasons for exclusion of tokens from analysis. If there is a predictable pattern or a specific environment when these exclusions are likely to occur, then it may be potentially beneficial to target such environments for manual checking, or to exclude them from any automated analysis. For example, here tokens where schwa was voiceless or deleted were able to be targeted because they tended to occur in specific environments.

Here, the effects of exclusion criteria were checked by comparing them to the full automatic and manual data. Clearly, in a normal situation, the benefits of any exclusions cannot be checked by comparison to a full manual analysis. However, there are still potential ways of checking whether exclusions have had a positive effect on the data. Datasets with exclusions can still be compared to the full automatic data in order to see the effects of exclusions. The effects of exclusion criteria can be checked by looking at how they change the various statistics and patterns in the data, as was done here. The variable which is the focus of investigation can clearly not be compared in this way, as without prior analysis it will not be known how it should be expected to affect the data. However, it is possible to compare statistics for well-established patterns which would be expected, and check which exclusions increase the clarity of these patterns.

The relative success of the different exclusion criteria combinations can be related to the way in which they aligned with the manual data in terms of the true/false positives and negatives they identified. In endeavouring to improve upon the original full automated dataset there is a balance to be struck between eliminating false positives from the data, without eliminating too many true positives. The most successful dataset in terms of the statistical comparisons is ODVE, which is also the dataset where the most tokens were removed from the original full automated dataset. This meant it removed both the most true and false positives. This shows that, in some circumstances, it may be worth removing more true positive tokens and working with a smaller dataset if it also means removing more false positive tokens. In many ways, what matters most is the quality of the data that is kept, and it is false positives which affect this. In this way ODVE is also the best set of exclusion criteria, as it is the dataset with highest proportion of true positives.

Clearly, it would be possible to remove too much data, even if it fulfilled the ultimate aim of increasing the proportion of the data which is made up of true positives. If a dataset becomes too small, any improvement in this proportion may be counteracted by a reduction in statistical power. In reality, how much data one is prepared to exclude will generally depend on how big the dataset it is to begin with. When the initial dataset is smaller it may be less desirable to use exclusion measures which result in the removal of a large number of tokens. In these circumstances, however, it may be possible, and indeed more beneficial, to manually segment and measure tokens. Typically, however, datasets used for automated analysis tend to be large enough that it would be worth using a more cautious approach to token inclusion. It is better to have a smaller dataset where one can be more confident that the tokens analysed were what was aimed to be analysed, rather than a larger dataset filled with erroneous tokens.

5.7 Conclusion

The aim of this chapter was to ascertain whether automated methods are suitable for the analysis of unstressed vowels. The chapter has shown that automated measurement works well for unstressed vowels when it measures the target tokens. Overall, the problems with the automated data were due more to errors in token inclusion than measurement or segmentation error. When exclusion criteria was used, the quality of the automated data was improved to a level that made it suitable for analysis. There was, however, a limit to the improvement in the data that could be made with outlier removal alone. To best improve the data, additional linguistic exclusion criteria had to be used. The results favour a cautious approach to data inclusions, and show that it can be beneficial to be more liberal in which data is excluded in order to have a smaller but better quality dataset. These results are used to inform the data filtering steps used in the preparation of the ONZE data in the following chapter (Chapter 6). Overall, the results are reassuring and show that with the use of appropriate exclusion criteria automated analysis is suitable for unstressed vowel analysis.

6 Change in Schwa in New Zealand English

6.1 Introduction

This chapter presents an analysis of schwa in New Zealand English, using data from speakers whose birth years span a period of around 120 years. The primary aim of this chapter is to address the question of whether schwa has a phonetic target. As in Chapter 4, this question is partly addressed through an examination of the variability and predictability of schwa, and by examining whether it shows movement towards a target at increased durations. In addition, this chapter also examines the issue of targetlessness in schwa from a different angle - that of the year of birth of the speaker - and thus examines changes in schwa over time.

Whether schwa shows any change over time has implications for whether it can be considered targetless. These arguments are explained in detail in section 2.5.3. To recap these arguments, if a vowel is targetless, or its realisation is motivated purely by ease of articulation then it should not show change over time. Therefore examining how the realisation of schwa differs across the timespan of the ONZE Corpus is another way of investigating claims of targetlessness.

Throughout the analyses in the chapter data from both non-final and pre-pausal schwa is compared, meaning that the findings in this chapter also contribute to addressing the question of whether there is a phonological difference between non-final and word final schwa.

Section 6.2 explains why NZE in particular is relevant to the overall goals of the thesis. It also provides an overview of previous research and commentary on unstressed vowels in NZE. Section 6.3 sets out the specific research questions answered in this chapter, and explains how they contribute to answering the broader aims of the thesis. The data preparation and method of analysis used are then described in section 6.4. The analysis of the data is then presented in 6.5, and this is followed in 6.6 by a discussion of the results in reference to the questions set out in section 6.3.

6.2 Background

Unstressed vowels have generally been neglected in research on language change. It is routine for vowels without full stress to be excluded from analysis in variationist and language change research (e.g. Labov, 1994). However, examination of how schwa varies over time can shed light on some of the debates about the nature of schwa. This section reviews the specific changes that have occurred in NZE which make it a particular interesting variety in which to investigate changes in schwa. I then take a brief look at the observations that have been made previously about unstressed vowels in NZE.

6.2.1 The relevance of New Zealand English

The variety of English in focus in this chapter in New Zealand English. There is a well-established sound change in NZE involving the short front vowels. TRAP and DRESS have raised, and KIT has lowered and backed. These changes have been demonstrated in a wide range of studies using impressionistic (Easton and Bauer, 1998. Trudgill et al 1998, Gordon et al, 2004), acoustic (Watson et al, 1998a; Watson et al 2000, Langstrof, 2006) and articulatory methods of analysis (Watson et al 1998b). There is now a consensus that these changes are the result of a push chain, whereby TRAP raised first, causing DRESS to raise, and then causing KIT to consequently lower and back (Hay et al, 2015). Evidence for this is found, for example, by Gordon et al (2004) who look at speakers born between 1851 and 1910 in the Mobile Unit subcorpora of what is now the ONZE corpus. They find little centralisation of KIT in this period, but a considerable number of raised variants of TRAP and considerable DRESS raising over the period, suggesting that the order of the change was TRAP then DRESS and then KIT. They also find correlations between the raising of TRAP and DRESS, and of the raising of DRESS and centralisation of KIT for individuals, providing further evidence for the existence of a chain shift. Langstrof (2006) confirms this view in an acoustic analysis of TRAP, DRESS and KIT in the Intermediate Archive, where Langstrof finds speakers who have raised TRAP and DRESS but not centralised KIT.

The reason why these changes are of interest in terms of schwa is because of the lowering and backing of KIT, as this change means that it has moved towards the position of schwa. Indeed, modern NZE KIT is now sometimes transcribed with a central symbol. For example, Watson et al (1998a) use [ə] to describe KIT; Gordon et al (2004) describe KIT as "realised towards [ɘ]"; and Warren (2018) transcribes it as /ɘ/. These kind of descriptions clearly suggest encroachment on schwa. Note that the [ɘ] symbol used by Gordon and Warren is slightly higher than the [ə] symbol normally used to describe schwa. Watson et al (1998b) also suggest that KIT may not have lowered as much as backed, as they find that speakers still use compressed lips to produce it. However, given that analyses elsewhere in this thesis (Chapters 3 and 4) have found that non-final schwa can often be fairly high, such pronunciations would still represent an infringement into the position of schwa. It is thus of interest to see whether is any change in schwa also in NZE.

6.2.2 Unstressed vowels in New Zealand English

There has not been much research in NZE that looks specifically at changes in schwa, although there has been commentary on the perception that KIT has become similar to schwa. Trudgill et al 1998 write that "in modern New Zealand English… / ɪ / and / ə / are not distinct" (p. 37). Bauer and Warren (2008) suggest that COMMA and horsES are not distinct. However, this is a slightly ambiguous statement, in that it is not clear if they are indicating a lack of difference between KIT and schwa in general or only in unstressed syllables, or between word final and non-final unstressed vowels. Hay et al (2008) state that KIT and schwa are very similar in NZE, writing that "there is almost no audible difference between KIT and the neutral vowel schwa" (p. 23). They also provide anecdotal evidence that speakers may not always perceive a difference between these vowels, reporting that linguistics students in New Zealand tend to use /ɪ/ to transcribe most unstressed syllables.

Interestingly, they report that this transcription does not tend to be used for word final schwa, suggesting that this vowel at least may still be perceived as different from KIT.

Nokes and Hay (2012) report on perceptions that NZE may have become more syllable timed, and unstressed vowels more fully articulated. They find, using the ONZE corpus, that durational variability between adjacent syllables has decreased, which they say does not suggest that unstressed syllables have become more fully articulated. Instead they suggest that it may be the changed realisation of KIT which has caused this perception, since its centralisation may have caused unstressed vowels to be less differentiated from stressed ones. Their analysis did not, however, look specifically at differences between stressed and unstressed syllables, and they note the need to more explicitly compare the quality of stressed and unstressed syllables in NZE.

Gordon et al (2004) investigated changes in a group of unstressed vowels which they dub the *rabbit* vowel. Although related to the analysis in this chapter, their focus is different: these vowels have an unstressed KIT in varieties which distinguish between KIT and schwa in unstressed syllables i.e. where *rabbit* and *abbot* do not rhyme. By contrast, the analysis here focuses on unambiguous schwa, in words like *abbot*, which would never have been produced with the KIT vowel. Gordon examined realisation of the *rabbit* vowel in speakers from the Mobile Unit corpus, using auditory analysis of whether or not speakers used centralised variants. They find that the use of centralised variants is greater in the younger speakers. However, they do not comment on whether there has been any change in words like *abbot* where there would never have been a KIT vowel present. These are the vowels which are the focus of this chapter.

6.3 Research questions

Here follows a description of the particular questions which are of interest with regard to NZE, and why they are relevant to the overall research goals of the thesis. More specific empirical predictions relating to the data modelling of F1 and F2 are set out within the results section.

Question 1: Does schwa change over the period?

Why this is relevant to the main research questions:

The articulation of a targetless vowel would be determined by its phonetic context alone, rather than a need to maintain contrasts with other vowels. Therefore, if schwa does not have a target and its position is purely determined by context, then it follows that its realisation should be unaffected by the realisation of any other individual vowel quality, and should not be involved in vowel shifts. The centralisation and lowering of KIT in NZE towards a more schwa-like position offers an opportunity to see if schwa has any involvement in such vowel shifts. If schwa is a vowel without an independent target, then it should be unaffected by this change in KIT.

Predictions:

If schwa is found to change along with the changing KIT vowel, this would be evidence to indicate that schwa is not a targetless vowel, and that it can participate in sound change.

Question 2: How similar are schwa and stressed KIT throughout the period?

Why this is relevant to the main research questions:

The lowering and backing of KIT has led to it moving closer to the positon of canonical schwa i.e. a mid central positon. In addition to its positon in the vowel space, schwa is also thought to differ from other vowels in further ways, such as being more predictable according to its phonetic context, and shorter. In Chapter 4 the behaviour of schwa was compared to certain full vowels, but none of these vowels are directly comparable to schwa in that they are at different positions in the vowel space. Therefore, the changing position of stressed KIT in NZE is a unique opportunity to compare schwa with a stressed vowel which may be more similar in its position in the vowel space. Importantly, it is clear that KIT is not a targetless vowel in NZE, because of its clear role in the front vowel shift, and because stressed vowels are generally seen as unambiguously having phonetic targets. It will therefore be of interest to compare schwa to the behaviour of a vowel which is similar in the vowel space but which is known to be targeted.

Schwa and KIT will be compared across the period, in order to see how similar they are in overall F1 and F2, and in how they are affected by phonetic context and duration. As the position of stressed KIT moves to become lower and more central, it will be seen whether it also becomes more similar to schwa in the way it is affected by phonetic context and duration. This will tell us something about which properties of schwa are unique to schwa, and which are simply related to its position in the vowel space. Whether and in what ways schwa and KIT remain different by the end of the period will tell us in what ways schwa is different from stressed vowels, if at all.

Predictions:

If, as KIT changes in vowel quality, it is found to become more like schwa in terms of variability according to phonetic context, this would suggest that predictability from context is not unique to schwa, and instead is likely to be an effect of the part of the vowel space it occupies. This would provide evidence against the argument that schwa is exceptionally variable according to context, and would thus also be evidence against it being targetless. If KIT remains different from schwa in contextual predictability it would suggest that a high level of variability according to phonetic context is unique to schwa.

More generally, whether and to what degree KIT and schwa remain distinct will tell us in what ways it is important for schwa to contrast with stressed vowels, and which of its properties are unique to its lack of stress.

6.4 Methodology

6.4.1 The data

The data is taken from the ONZE corpus (Gordon et al, 2007), and includes speakers from the three sub-corpora of the Mobile Unit, the Intermediate Archive, and the Canterbury Corpus. This data covers a birth year range of 130 years, and so is ideal for examining language change. The data includes speakers born in a range of locations in New Zealand. There is some variation in formality level in some of the earlier recordings, but the more recent recordings in the Canterbury Corpus are of more casual speech. The initial birth year span that was examined covered 1851-1987. However, the analysis presented here uses data from a slightly narrower range. This was because there were fewer speakers at either extreme of the age range. When modelling the data this affected the estimated smooths, sometimes producing sudden reversals of trends at the beginning or end of the time period, and with large confidence intervals at either end of the period. These sudden reversals in turn affected the rest of the estimated smooth, causing wigglier curves than necessary. The range of speakers was thus reduced to those born between 1864 and 1982, as there were only 3 speakers born before this range, and only 8 speakers born after this point. This meant that a total of 557 speakers were used in the analysis. This includes 63 speakers from the Mobile Unit, 81 speakers from the Intermediate Archive, and 413 speakers from the Canterbury Corpus.

Measurements of F1, F2 and duration were taken automatically using forced alignment in LaBB-cat (Fromont and Hay, 2012), which extracts F1/F2 measurements using Praat (Boersma and Weenink, 2011). The initial dataset included all monophthongs in the corpus[1], before data filtering processes were used in order to select the relevant data for the analysis in this chapter. The analysis in this chapter focusses on midpoint measurements.

6.4.2 Token selection

The dataset used included all monophthongs in the corpora. These were labelled automatically in LaBB-CAT. Most of the vowel segment labels were left unchanged and were not checked manually, as the majority of monophthongs were simply used for normalisation purposes rather than being part of the analysis. Those vowels that were labelled as /ə/, /ɪ/, and /ʊ/ had their labels manually checked as these were the focus of the analysis. The words within these groups were manually assigned to either an unstressed vowel group, a stressed KIT group, or a stressed FOOT group. Unstressed vowels were defined as those that were unambiguously lexically unstressed. This definition and categorisation of unstressed

---

[1] I thank James Brand and NZILBB for generously sharing this automatically extracted vowel formant dataset with me

vowels follows that used in the Derby analysis (see 3.5.2), where vowels labelled as /ə/, /ɪ/, and /ʊ/ were all combined into one unstressed vowel group. At this point, some tokens were not assigned to any of these groups and were excluded from the dataset for a number of reasons, including:

- Function words, e.g. *his* labelled as /həz/. Many of these were monosyllabic, and whilst they were labelled as unstressed within the corpus, these vowels are not lexically unstressed, as they can also be produced with a full vowel, and would contain a stressed vowel in citation form
- Vowels that could also be part of a diphthong/triphthong, e.g. *Ireland* transcribed as /aɪələnd/
- Vowels that could also be articulated as syllabic nasals, e.g. *different* transcribed as /dɪfɹənt/
- Tokens that were not part of a lexical item, e.g. conversation fillers such as *hmm* transcribed as /həm/
- Incomplete words
- Wrongly labelled words
- Instances of vowels labelled as /ɪ/ that corresponded to the happY vowel were removed as these were not part of the analysis.

After these exclusions the whole dataset consisted of 787,239 tokens, including 159,001 unstressed vowel tokens and 75,532 KIT tokens. Outliers were then removed based on two criteria:

3) Vowels were marked as outliers based on the formant values in Hillenbrand et al (1995). Tokens were classified as outliers if either the first or second formant measurement occurred below the 1st or above the 99th percentiles, calculated separately for males and females, to exclude acoustically implausible measurements.
4) Measurements were marked as outliers based on the individual speaker's measurements for each vowel. Tokens were excluded if they had measurements marked as an outlier for either F1, F2 or duration. Measurements were classified as outliers using the procedure built into R's (R Core Team, 2020) boxplot () function, which identifies outliers as data points that are more than 1.5 IQR higher/lower than the upper/lower quartiles.

After outlier removal, 638,964 tokens remained, meaning that outlier removal removed 19.8% of the data. There were 126,485 unstressed vowel tokens and 67, 088 KIT tokens left. Thus 20.3% of unstressed vowel tokens, and 15.1% KIT tokens were marked as outliers and removed.

6.4.3 Normalisation

One of the key variables in the analysis was year of birth. As this is a variable which varies across speakers it was necessary to normalise the data, in order to be able to compare data across all speakers. Therefore, following outlier removal, the F1 and F2 measurements were normalised following the Lobanov (1971) method using the *vowels* package in R (Kendall and Thomas, 2018).

Although the Fabricius and Watt (2009) method was used for the normalisation in Chapters 3 and 4, Lobanov normalisation was chosen instead here. This was because the Fabricius and Watt method requires that particular vowels be used, normally TRAP and FLEECE, to calculate the corners of the vowel space. Given that we know that there has been a major shift in the TRAP vowel in NZE it would not have been appropriate to use the TRAP vowel as a means of normalising the vowel space. In addition, FLEECE has also been found to show some slight change over the period (Maclagan and Hay, 2007). In general, because many vowels in NZE are known to have undergone a lot of change over the period, it was thought better to use a system of normalisation that didn’t rely on the measurements of only two vowels, but which took into account the whole vowel space.

Lobanov has been shown by Flynn (2011) to be an effective method of vowel normalisation. In a comparison of 20 normalisation methods, Lobanov was found to be the third most effective, and one of the two best procedures that are widely available. Furthermore, there are instances of other research using the ONZE corpus that have opted for Lobanov normalisation. These include Gordon et al’s analysis of the speakers from the Mobile Unit, Brand et al (2019), and Langstrof (2006). Langstrof looks at data from the Intermediate Archive, and compares the effects of Lobanov with two other normalisation methods, and finds Lobanov to be the most suitable for analysis of the data.

Figure 6.1 shows the 99% confidence ellipses for all vowel tokens. It is clear that the data for the male and female speakers are much more comparable after the normalisation, and cover almost exactly the same range.

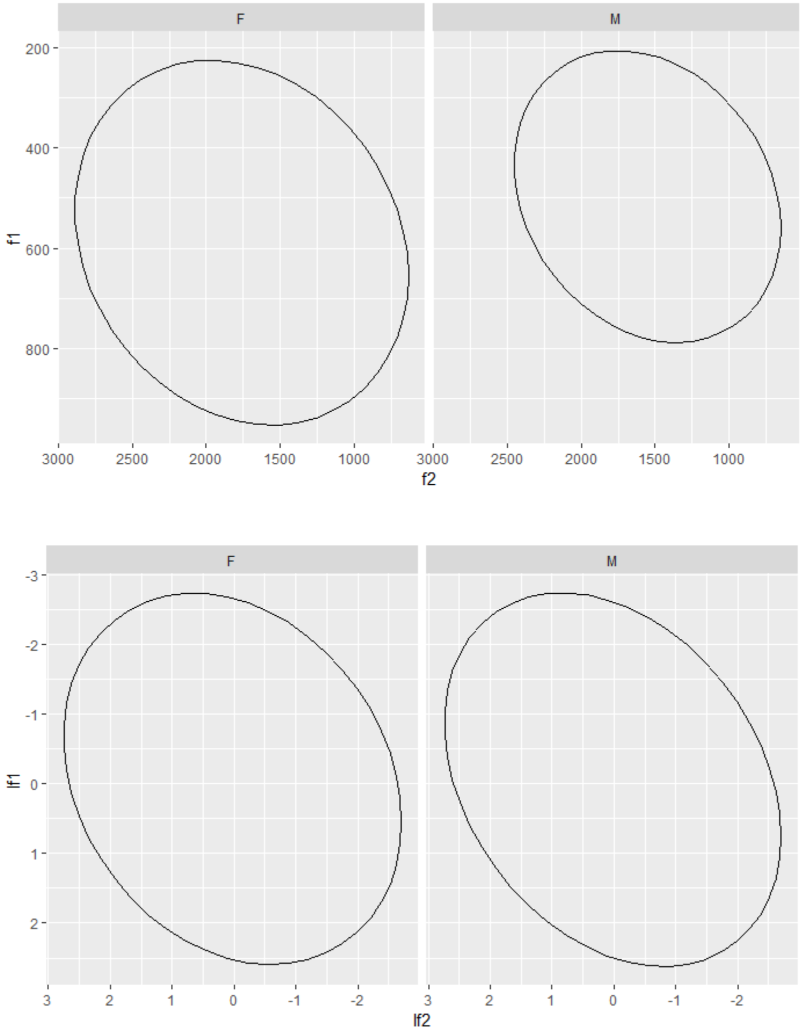


Figure 6.1 *0.99 ellipses by gender for raw data (top) and normalised data (bottom)*

6.4.4 Data filtering

In order to achieve the final groups of vowels ready for analysis, the 126,485 unstressed vowel tokens and 64,065 KIT tokens were further filtered. The different filters that were applied are listed below. For each filter it is reported how many tokens and what percentage of tokens were removed. Note these steps are not cumulative, and many tokens fitted into more than one of the categories which were filtered out, so these numbers do not add up to the total amount of data that was filtered out.

Across all of the vowels, tokens were filtered out if:

- Speaker birth year was before 1864 and after 1982 as they contained few speakers, and this eliminated statistical artefacts in the fitted smooths that were due to outliers at the extremes of the age range. This removed 8,400 (1.3%) of the tokens from the whole dataset.

For KIT only, tokens were filtered out if:

- The token was in a monosyllabic word. This was because one of the aims of this chapter was to directly compare schwa and KIT and how similar they are at different points within the period. Since all of the schwa tokens were included on the basis of being lexically unstressed this means that they were all in multisyllable words. Therefore keeping the same criteria for KIT made for a more like for like comparison. This removed 34,050 tokens (53.1% of the KIT tokens)

For unstressed vowels only, tokens were filtered out in the following contexts:

- The token was in a context where deletion was possible e.g. *prob<u>a</u>bly* (22,778, 18% removed), or where it was possible for the unstressed vowel to be voiceless, e.g. *t<u>o</u>day*, (21,345, 16.9% removed). In total this removed 38,652 tokens (30.06%). These exclusions follow directly from the analysis performed in Chapter 5, and the full details about how such contexts are defined are provided in section 5.5.1.
- The token was represented with the spellings <e> or <i> out (35,548 tokens, 28.1%). This is because it is possible for schwa to show categorical variation with other vowel qualities in unstressed syllables, namely /ɪ/. In Chapter 3 it was shown that both the <e> and <i> spellings had more front realisations, and thus excluding such tokens is consistent with the analysis of schwa in the Derby corpus in Chapter 4
- The token was in a common suffix such as *–ed*, *-es*, and *-ing* in line with the analyses in Chapters 3 and 4(8,900 tokens, 7%)
- An alternative pronunciation of the token with a full rather than reduced vowel was possible e.g. *<u>a</u>dvice* as [advaɪs] rather than [ədvaɪs]. The methods used to assign words to this category are given in full in Chapter 3 (4,427 tokens, 3.5%)

- The token was a stem final but word medial schwa e.g. *sugars,* so that the non-final schwa group of vowels only included vowels that couldn’t also occur in a word final position e.g. (5,984 tokens, 4.7%). These tokens were also excluded in Chapters 3 and 4

This left 26,626 KIT tokens (41.6% of all KIT tokens), and 46,218 unstressed vowel tokens to analyse (36.5% of all unstressed vowel tokens).

As in the Derby analysis, the schwas were initially split into three groups for analysis: non-final, pre-consonantal word final, and pre-pausal. In the following analysis, however, only non-final and pre-pausal schwas are included. As in the analysis of the Derby data in Chapter 4, pre-consonantal schwas were more similar to non-final than pre-pausal schwa. However, in terms of their overall behaviour they were intermediate between pre-pausal and non-final schwas, and thus did not contribute any extra information to answer the research questions stated at the beginning of this chapter. Consequently, it was thought better to keep the focus on pre-pausal and non-final schwa, as here the differences between them were more extreme.

The analysis used throughout the rest of the chapter focuses on three main groups of data, labelled as:

1) KIT - stressed tokens of the KIT vowel
2) Non-final schwa - unstressed vowels that are not in word final position
3) Pre-pausal schwa - unstressed vowels which occur at the end of a word and before a pause or hesitation.

6.4.5 Modelling

As in the earlier analyses of formant trajectory and duration in Derby, the data is modelled using generalised additive mixed models. This type of model allows for random effects to be modelled as in linear mixed effect models. Where it differs is that it also allows for nonlinear relationships to be modelled. These fitted relationships are known as smooths. For example, in the following models the effect of birth year on F1 and F2 is modelled as a non-linear relationship. This is important because in a vowel change a continuous change at a constant rate throughout the period would not be expected. Rather, it is expected that the vowels change more in some parts of the period than others. Modelling the data using GAMMs allows these types of relationships between variables to be captured more accurately.

The main focus of this chapter is on the effects of birth year, duration, and their interaction across different vowels and phonetic contexts, so the basic structure of all the models analysed is the same. The main variation between different models is in the dependent variable (whether it was F1 or F2), and the difference variable. The difference variable was either the vowel (possible levels were KIT, non-final schwa, and pre-pausal schwa), or the phonetic contexts (possible levels were back, central, and front). Therefore the models either focus on comparing differences between vowels, or differences between phonetic contexts within one vowel only. As the models analysed in this chapter have large summaries, the full code used to run each model and the model summaries are given in the appendix.

The basic model structure used was:

```
Normalised F1/F2 ~ difference variable
+ s (birth year, bs="ad")
+ s(birth year, by=difference variable, bs="ad")
+ (duration, bs="cr")
+ s(duration, by=difference variable, bs="cr")
+ ti(birth year, duration)
+ ti(birth year, duration, by=difference variable)
+ s(speaker, bs="re")
+ s(word, bs="re")
+ s(speaker, difference variable, bs="re")
```

Each component of the model is described in more detail below:

- Difference variable. The difference term is a parametric term which simply represents the average difference between the levels of the difference variable along the entire smooth – in other words, a difference in the height (but not necessarily the shape) of the smooths. Note that this is encoded separately from shape differences in the smooth. This means, for example, that it is possible for the overall F1 of two vowels to be different, but the difference in the shape of their smooths according to birth year could be the same. This would mean that the vowels were different heights overall, but changed in the same way according to birth year.

- A smooth by birth year. This is the shape of the smooth for birth year, which represents potential non-linear effects of birth year on the dependent variable. The birth smooths are fitted with adaptive smoothers. This means that the amount of wiggliness in the smooth varies as a function of birth year, to reflect the fact that there will be more change in certain parts of the period than others.
- A difference smooth by birth year. This captures differences in the shape of the smooths by birth years, between the different levels of the difference variable. For example, if birth year has a different effect on the F2 of KIT and schwa, this would be encoded here. These are also fitted with adaptive smoothers.
- A smooth by vowel duration. This captures potentially non-linear effects of duration on vowel formants. This was modelled using cubic regression splines.
- A difference smooth by vowel duration. This captures the differences in the shape of the duration effect across the different levels of the difference variable.
- A tensor product interaction between birth year and duration. This models how the effect of duration varies over birth year. This type of interaction captures potentially non-linear interactions between the two variables. In practice, this means that the effect of birth year on duration may be different at different points of the birth year continuum.
- A difference tensor product interaction between birth year and duration- This term captures how the difference variable impacts the change in the effect of duration over historical time (i.e. birth year) – for instance, how duration effects change differently over time for different vowels.
- Random intercepts by speaker.
- Random intercepts by word.

- Random slopes by speaker for the difference variable. This models random variation in the effect of the vowel or phonetic contexts on the formant values across speakers.

Although the analyses in this chapter are quite different to those conducted in the Derby Corpus (Chapters 3 and 4) in that they make use of different data and different variables, there are similarities with the structure of the GAMMs used in Chapter 4. Where duration is modelled this is modelled as a non-linear term using cubic regression splines in both chapters. In addition, in both chapters (see section 4.4.4) the difference in the effect of duration is compared across different vowels as a difference smooth by duration. The random effects structure is also similar across both chapters, as all GAMMs used have random intercepts for speaker and word. In addition, all of the GAMMs that model differences between vowels also have random slopes by speaker for the variable of vowel.

As has been explained, in some models the difference variable refers to the type of vowel, and sometimes it refers to phonetic context. The phonetic context variable was coded into three levels of back, central, and front. These labels were only applied to the vowels that were not before a pause, as these did not have a following segmental context.

Initially the preceding and following contexts were individually coded for each word, in the same way that is reported for the Derby corpus in section 3.5.4.1. Coding three different preceding environments and three different following environments meant there were 9 possible levels of environment. Figure 6.2 below shows how these individual environments differed in F2 for non-final schwa, ordered from the environment with the lowest F2 to highest F2. The labels refer to the preceding and following environments, such that ‘b.f’ means a vowel preceded by a back context and followed by a front context. Although there is variation in F2 level overall, between most adjacent environments there are only small differences. For the purpose of the research questions it would not have been particularly enlightening or necessary to examine phonetic context at this level of detail. As will be explained in more detail in section 6.5.4, the main interest in the phonetic environment is its relative overall level of frontness. The specific combination of preceding and following environment is not the focus. It was therefore decided to collapse the phonetic contexts into three broad groups. This allows for a larger number of tokens in each phonetic context group, meaning more statistical power in the models. Comparing just 3, rather than 9, environments also produced model results that are both easier to interpret and more revealing than examining 9 different phonetic contexts.

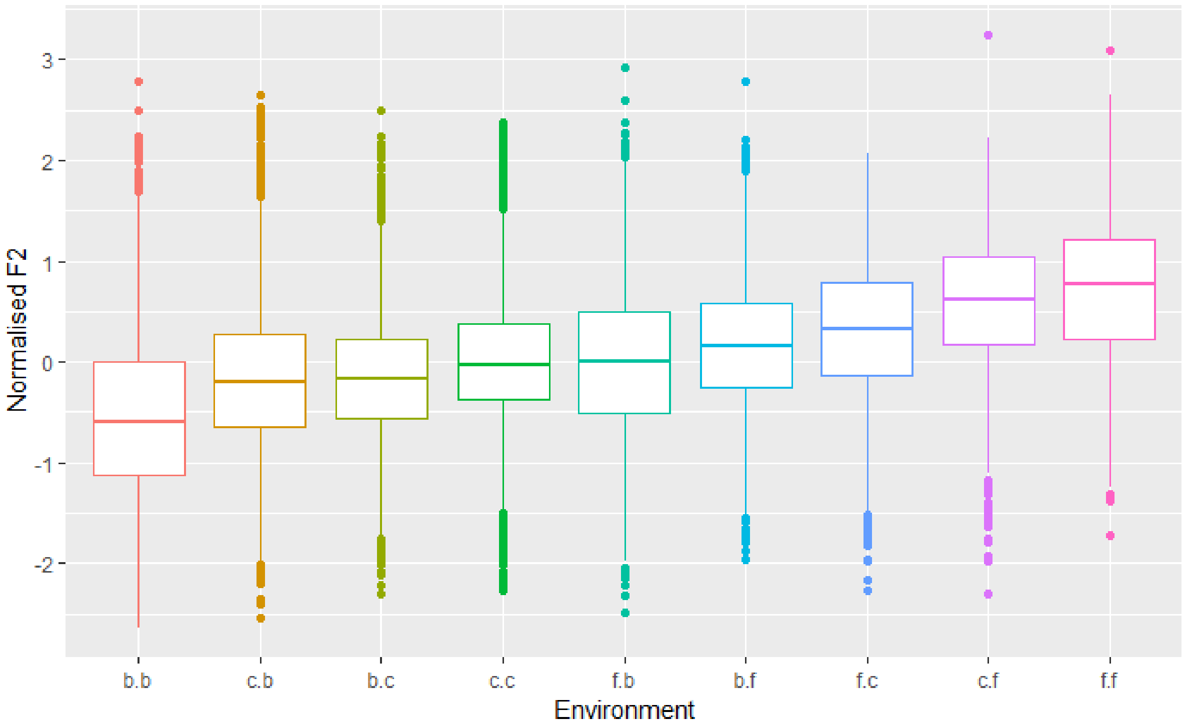


Figure 6.2 *Normalised F2 of non-final schwa by individual environment*

These 9 contexts were therefore combined into three groups depending on the average general F2 of the vowels within each group. Table 6.1 shows the initial classification of the preceding and following environments, and the letter in the cell is the overall category it was placed into for modelling purposes. They were classified in this particular way as there was some variation in the exact ordering of the 9 contexts across the three sub-corpora and KIT and schwa. However, classifying them in this way was consistent with their F2 ordering for both KIT and schwa across all corpora. All environments placed in the front category always had a higher average F2 than those placed in the central category, and all those in the central category always had a higher F2 than those placed in the back category.

***Table 6.1**: Grouping of phonetic contexts*

		Following context		
		b	**c**	**f**
Preceding context	**b**	b	c	f
	c	c	c	f
	f	c	f	f

Figure 6.3 below shows the F2 values for these three groups for both KIT and non-final schwa. Although this is a somewhat crude way of categorising the tokens, Figure 6.3 shows that it creates three groups which are distinct. The central contexts are fairly equidistant between the back and front contexts, and the overall spread of the data is similar between the three groups. These observations argue in favour of this particular grouping of contexts. Additionally, as will be shown later, there are some meaningful differences between the behaviour of these groups in terms of how their F2 values change by speaker birth year and vowel duration.

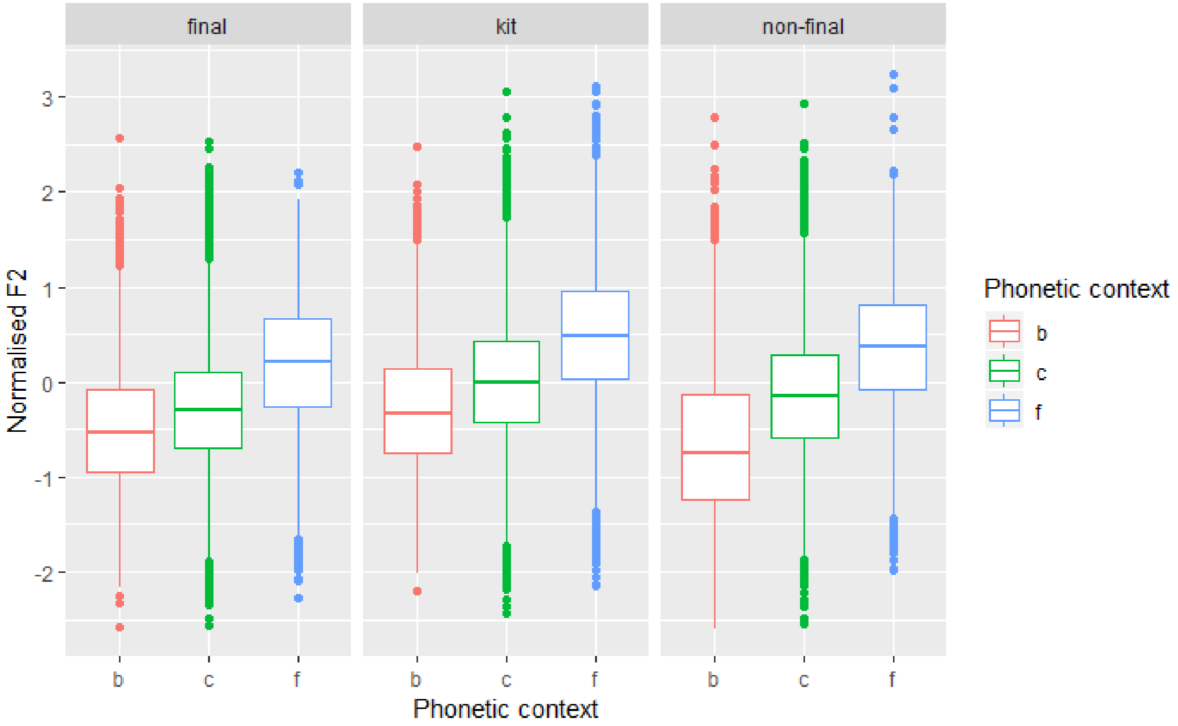


***Figure 6.3** Boxplot of phonetic contexts by vowel on normalised F2*

6.5 Results

6.5.1 Introduction

In this section the analysis of the data from the ONZE corpus is described. The section starts with a brief description of the durations of the vowels, since the variable of duration is key to the models that are presented later on in the chapter. The findings from the F1 analysis are then presented (section 6.5.3), followed by an analysis of the F2 findings (section 6.5.4). Both the F1 and F2 analyses start by outlining the empirical questions and predictions for the data. The data is then analysed with reference to these predictions. At the end of both the F1 and F2 sections there is a summary of the main findings for that particular formant. There is then a short section which examines how the F1 and F2 values play out over the whole vowel space. The results section ends by bringing together the findings from both the F1 and F2 analysis with a more general discussion of what the results mean for the main research questions set out at the beginning of this chapter.

6.5.2 Duration

As was described in section 6.4.5, duration is a key variable within the modelling of the formant data. The analysis presented in the rest of this chapter considers the effect of duration on formant values, and also how this influence changes over time. It is therefore important to also consider what the overall durations are for the different vowels.

Figure 6.4 compares the duration values of the three vowels that are compared in the analysis. Pre-pausal schwa (mean=78.6 ms) is clearly much longer than non-final schwa (mean=50.9 ms), as would be expected being in phrase final position, and this is also what was found in the Derby analysis (see section 4.4.4.1). KIT (mean=54.1 ms) lies in between the two types of schwa, although it is much closer in duration to non-final schwa. Although KIT is slightly longer than non-final schwa they are fairly similar in duration and are both very short. The short duration of KIT here partly reflects the fact that in order to make KIT more comparable with schwa only vowels in multisyllabic words were included. If monosyllabic words were also included, it is likely that KIT would have been longer.

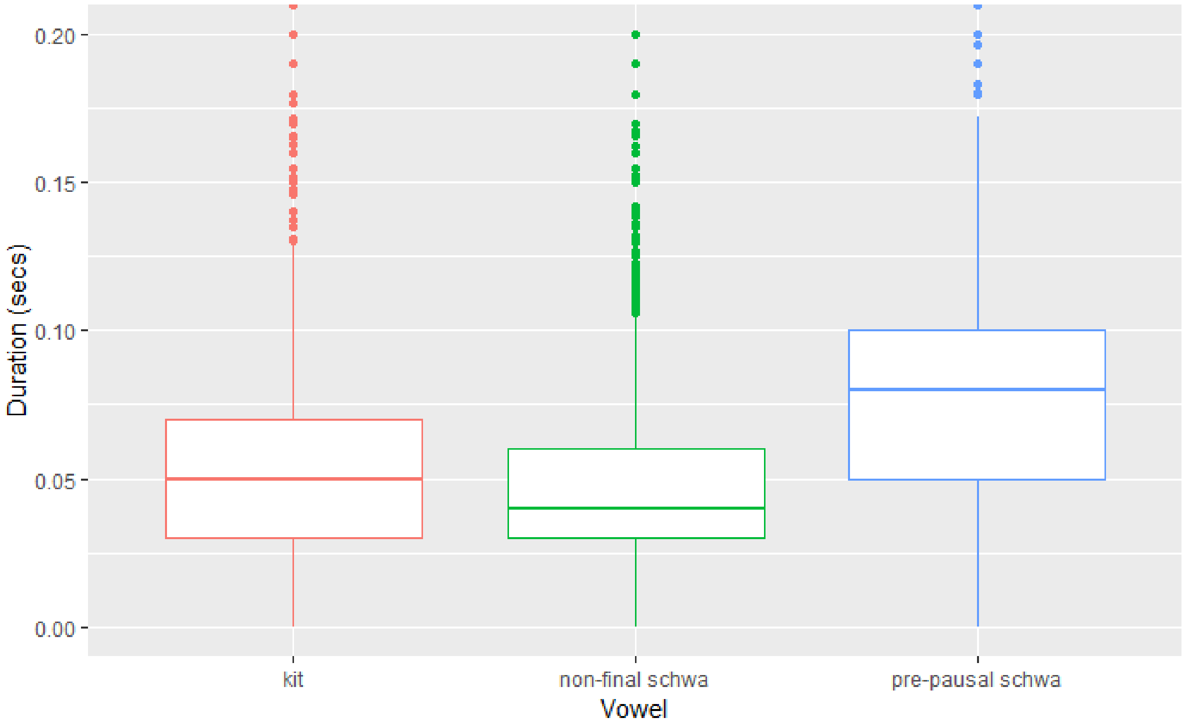


Figure 6.4 *Boxplot of duration by vowel*

6.5.3 F1 findings

This part of the analysis focuses on how F1 varies across time and duration for the different vowels. I first set out the expectations and predictions for the data. An analysis of the data is then presented, broken down into sections for each of these predictions. I then summarise the main findings and discuss how they relate to the broader research questions outlined in 6.3.

6.5.3.1 Predictions

Prediction 1: the influence of speaker year of birth

It is predicted that we will see a rise in the F1 of KIT over time, as KIT has previously been shown to lower in NZE. If schwa changes over time in line with KIT, then we would also expect to see an effect of birth year on the F1 of schwa, with schwa lowering over time, which would be evidenced by an increase in F1. It will be of interest here to see whether KIT becomes more similar to the height of schwa over time. This if course depends on a) whether the heights of KIT and schwa are different to begin with, and b) whether schwa changes in the same way that KIT does.

Prediction 2: the influence of duration

Although pre-pausal schwa is on average much longer than non-final schwa, these predictions still apply to both schwa types, as it is expected that the type of relationship between duration and vowel height is the same for both schwas.

a) Because adjacent consonants have on average a lowering effect on F1, it is expected that vowels with a higher F1 target will change more according to duration, whereas vowels with a low F1 target should change less according to duration. Therefore if schwa has a target which is lower than a high vowel, we would expect there to be a positive relationship between duration and F1 for schwa. This would mean that longer schwas are lower vowels. This would be expected for all years within the corpus. Similarly where KIT is a lower vowel, we would also expect KIT vowels of longer duration to be lower vowels. As will be seen in section 6.5.3.3, we see that in the early years when KIT is a high vowel, there is no such relationship between its duration and vowel height. This therefore suggests that if schwa also targets this higher position, or indeed has no height target at all, then this would be evidenced by a lack of a duration effect.

b) The lower the vowel, the further the articulators have to move between consonants, on average, to reach their target, and therefore lower vowels should be more affected by

> duration. Therefore it is predicted that if schwa becomes lower over time, it will also show an increased effect of duration over time. It would also be expected that this relationship would be true of KIT, as it becomes lower over time. Crucially the steepness of the relationship between duration and vowel height will also provide evidence as to whether KIT and schwa have different targets throughout the period. Where they have different targets we would expect them to have different relationships with duration. This would mean that even if the vowels are not different in height at short durations they could be different at longer duration, evidencing a different target. If schwa has a height target which remains different from that of KIT throughout this period, then we would expect schwa and KIT to be affected differently by duration throughout the period.

Due to the size of the models, it was not possible to model KIT, non-final schwa and pre-pausal schwa all within the same model. The findings for F1 are thus based on two separate models. Model 6.1 compares KIT and non-final schwa, and model 6.2 compares KIT and pre-pausal schwa.

6.5.3.2 Overall findings by birth year

Figure 6.5 shows how the average F1 values of KIT, non-final schwa, and pre-pausal schwa change over time. The KIT and non-final schwa values plotted are from model 6.1, and the pre-pausal values are from model 6.2. Figure 6.5 shows model predictions at the median duration of each individual vowel, in order to give a sense of how typical examples of these vowels are likely to be realised. Confirming previous descriptions of NZE, we see that KIT lowers over time. It is clear that both types of schwa also undergo changes in F1 overtime. All three vowels have an increasing F1 through the period, meaning that the average vowel height lowers. This change was predicted in prediction 1.

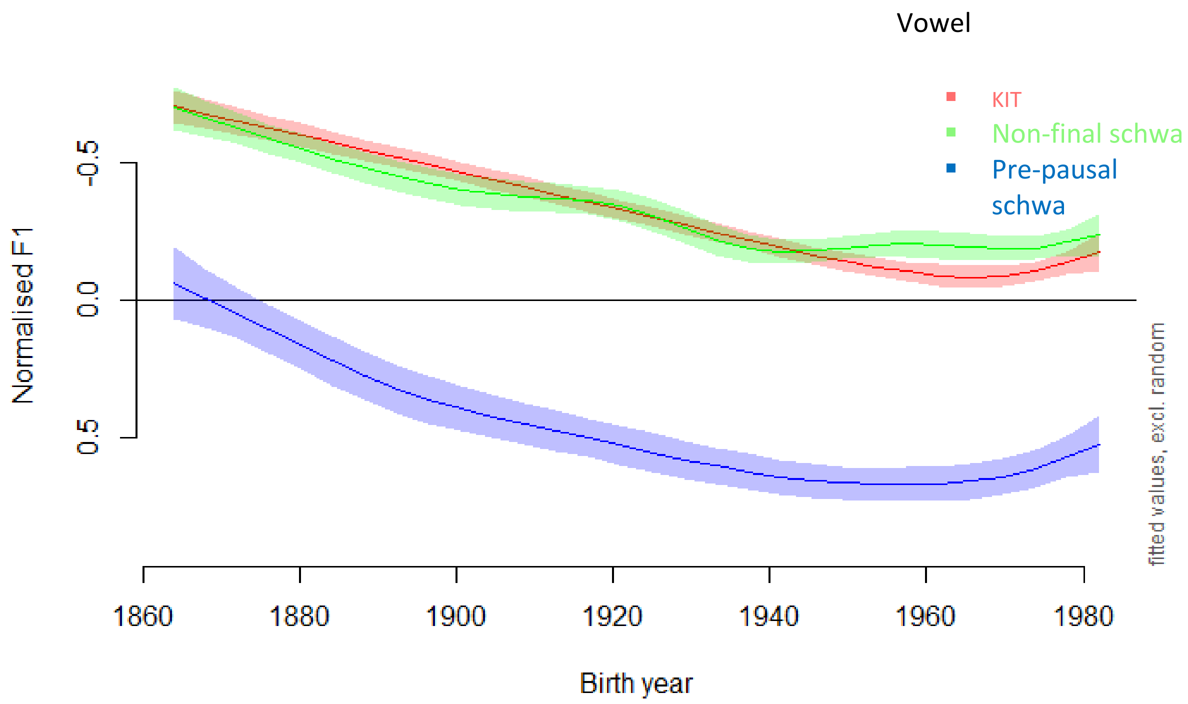


Figure 6.5 *Effect of birth year on normalised F1 of across vowels at median durations*

If we look in more detail at Figure 6.5 we can see that pre-pausal schwa is much lower than KIT or non- final schwa throughout the period. This was the same as was found in the Derby analysis in Chapter 4. Note also that, as was shown in 6.5.2, pre-pausal schwa is much longer than the other two vowels. As we will see in the following section, duration has an effect on F1 values, so the increased duration of pre-pausal schwa helps to further increase its F1 distance from KIT and non-final schwa. With regard to prediction 1, the difference between pre-pausal schwa and KIT remains fairly constant throughout the period, as KIT and pre-pausal schwa change at a similar rate throughout the period.

With regard to the difference between KIT and non-final schwa the situation is different. They are actually very similar in height throughout the period. The similarity between KIT and non-final schwa is somewhat expected given that they occur in the same word medial context, and are of similar durations. In the early part of the period preceding the change in KIT, KIT is a fairly high vowel. That KIT and non-final schwa are so similar here shows that non-final schwa too, is most often realised as a fairly high vowel. This, again, was also found in Derby. What is interesting here, though, is that non-final schwa does not stay at this height throughout the period, but moves lower in the vowel space as KIT does. KIT shows slightly more change over the period than non-final schwa. This is reflected in the fact that the difference smooth between non-final schwa and KIT is significant (F=4.239, edf=5.655, p<0.001). Towards the end of the period, KIT is on average very slightly lower than non-final schwa. As will be shown in the following section, duration values have an effect on F1 so this aspect of the data is related to the fact that non-final schwa is a slightly shorter vowel than KIT.

6.5.3.3 Influence of duration

In this section the influence of duration on vowel height is examined. This relates to the second prediction set out in 6.5.3.1.

Below are three separate figures for each vowel. Figures 6.6 and 6.7 are based on model 6.1 and Figure 6.8 is based on model 6.2. These graphs again show how F1 values differ according to speaker year of birth. The four different colours represent the model estimates for the vowels at different durations, with the red line showing the changes for vowels of particularly short durations (0.03 seconds), and with the blue line representing vowels of relatively long durations (0.12).

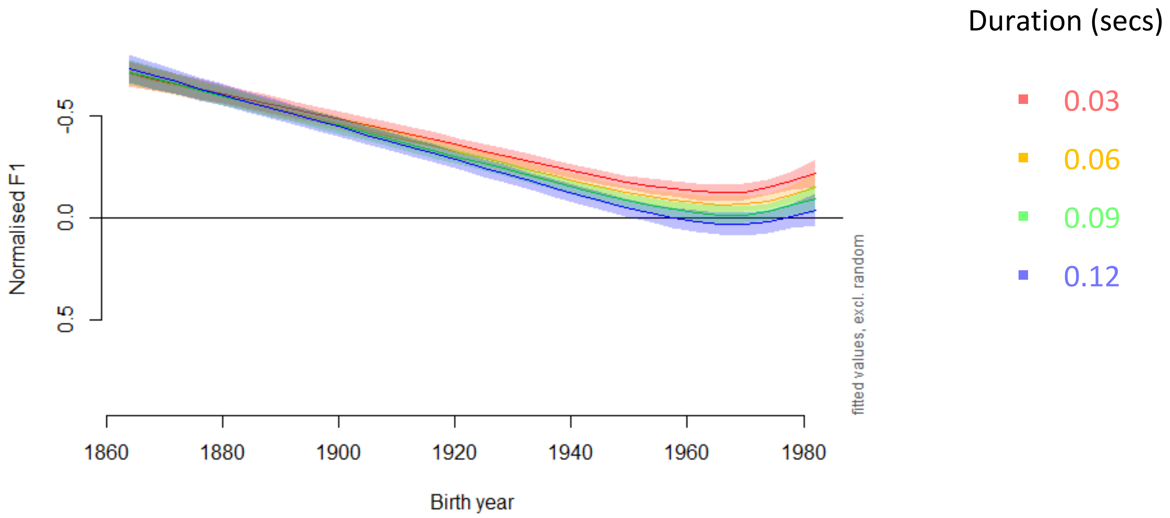


Figure 6.6 *Effect of birth year on normalised F1 of KIT by duration*

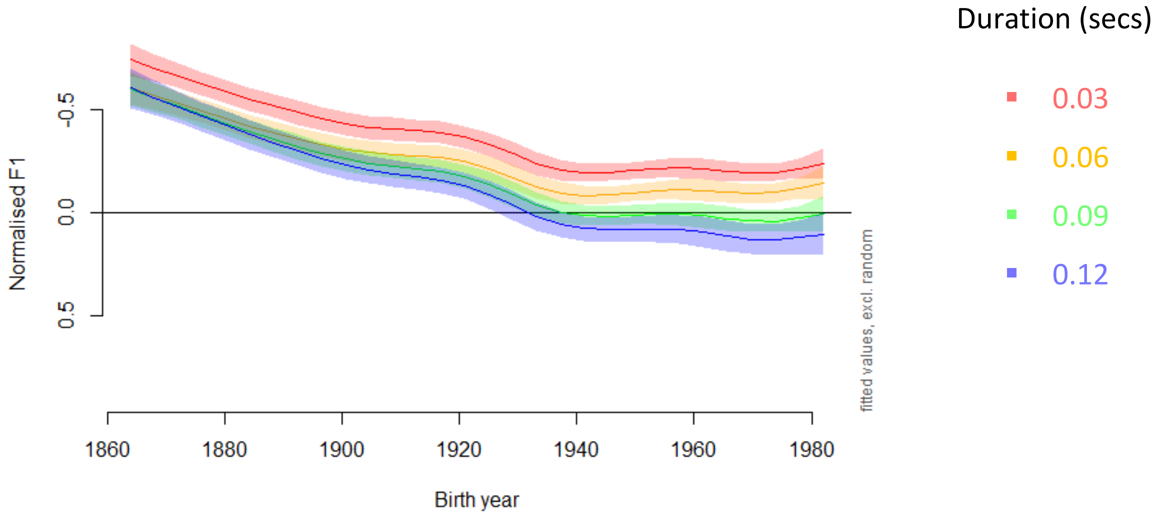


Figure 6.7 *Effect of birth year on normalised F1 of non-final schwa by duration*

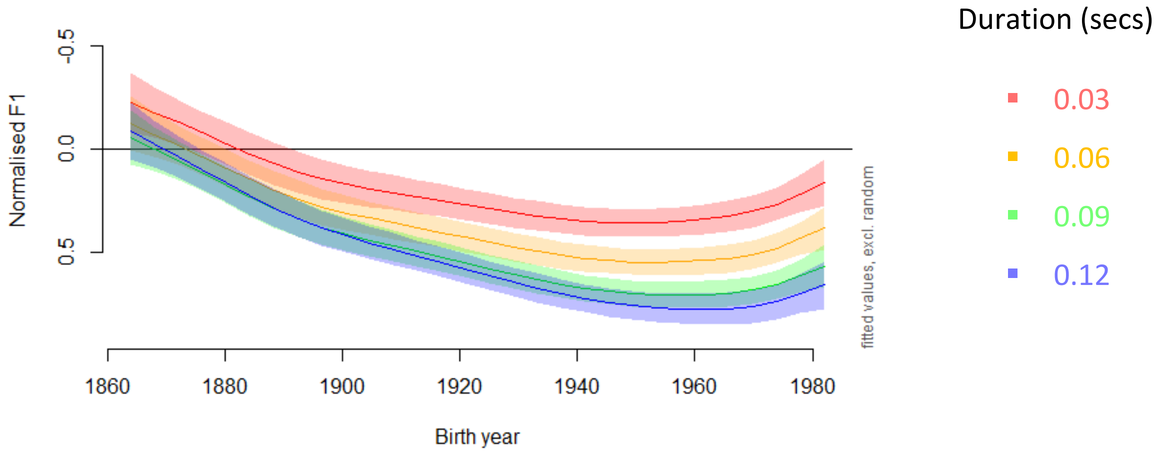


Figure 6.8 *Effect of birth year on normalised F1 of pre-pausal schwa by duration*

If we look at the values for KIT for speakers born before about 1890 there is no clear separation of the different lines; they all overlap. This shows that in the early years of the period, when KIT was a fairly high vowel, duration has no effect on the height of the vowel. This suggests that the height target for KIT here is one which is easy to produce at short durations. Note that, across all birth years, for both non-final and pre-pausal schwa there is at least some separation between the different duration lines. This shows that, in contrast to KIT in the early years, there is a relationship between F1 and duration for schwa across the entirety of the time period. Longer schwas are lower vowels, as indicated in their higher F1 values throughout the period. This is as was predicted in prediction 2a. This shows that even though KIT and non-final schwa are very similar at median durations throughout the period, when we consider the effect of duration their behaviour is different, indicating a slightly lower target for non-final schwa than KIT. This difference between the effect of duration on schwa and KIT is indicated in the significance of the difference term for the duration smooth. For non-final schwa this is indicated in model 6.1 (F=13.298, edf=4.128, p<0.0001), and for pre-pausal schwa this is in model 6.2 (F=17.335, edf=3.614, p<0.0001).

There is, however, a clear interaction between duration and birth year for KIT, as indicated in the significance of the interaction term between birth year and duration in model 6.1 (F=16.820, edf=1.001, p<0.001). As KIT becomes a lower vowel on average, this also coincides with a greater effect of duration on its height. This is apparent if we look at Figure 6.6 and compare the difference between the coloured lines at later and earlier birth years. As already noted, there is no difference between them in the early years, indicating a lack of an effect of duration on F1. However, the difference between the coloured lines gradually increases as speaker birth year increases. This means that later in the period KIT is a lower vowel at longer durations.

Both non-final schwa and pre-pausal schwa show the same type of interaction, whereby duration has an increasingly larger effect on vowel height as the period progresses. Whilst for schwa longer vowels are lower vowels throughout the period, this relationship gets stronger during the period. As for KIT, we can see that for both schwas there is a bigger difference between the coloured lines at later birth years.

This type of interaction effect was predicted in prediction 2b. It was predicted that the KIT vowel would exhibit lowering at longer durations in later years. It was also predicted that if schwa also lowers over time, then an increasing effect of duration on vowel height would be seen. This pattern can be explained in terms of an undershoot effect (Lindblom, 1963), and was also found in the Derby corpus. The undershoot effect is such that at shorter durations it may not always be possible for the articulators to reach a non-high height, because of the effects of adjacent consonants. However, as durations increase it is possible for vowels to be produced with a lower vowel height. Lindblom (1963) shows that high vowels are generally less affected by undershoot as the articulators do not need to move as much between high vowels and adjacent consonants as they do for lower vowels. This means that, in general, vowels with a lower height target would be expected to lower more as durations increase than vowels with a high target. Vowels with a higher target should be able to reach this target more easily, even at very short durations. Therefore their vowel height should change less according to duration.

This can explain the lack of effect of duration on the height of KIT at the start of the period when it was a higher vowel. It can also explain why, as all three vowels get lower, their height is also increasingly affected by duration. This interaction thus further lends support for the fact that the lowering of schwa is a genuine change.

The plots below show the differences between the different vowels at different durations more clearly. Figure 6.9 shows how the vowels compare at very short (0.03 seconds) and long durations (0.12 seconds). Pre-pausal schwa is clearly always much lower than KIT, no matter the birth year or duration. However, the difference is larger at longer durations, which is reflected in the difference smooth for duration between KIT and pre-pausal schwa in model 6.2 being significant (F=17.335, edf=3.614, p<0.001).This shows that the influence of duration on the height of pre-pausal schwa is greater than on KIT, further exemplifying the fact that pre-pausal schwa has a lower target than KIT.

Although KIT and non-final schwa are of a very similar height throughout the period, duration has a stronger effect on non-final schwa than it does on KIT. This can be seen in the fact that in Figure 6.9, where both vowels are of long durations, non-final schwa is actually lower than KIT regardless of spear birth year. This suggests that despite the vowels being of a very similar height overall, non-final schwa actually has a different height target from KIT throughout the period. Indeed the difference smooth for duration between KIT and non-final schwa in model 1 is significant (F=13.298, edf=4.128, p<0.001).

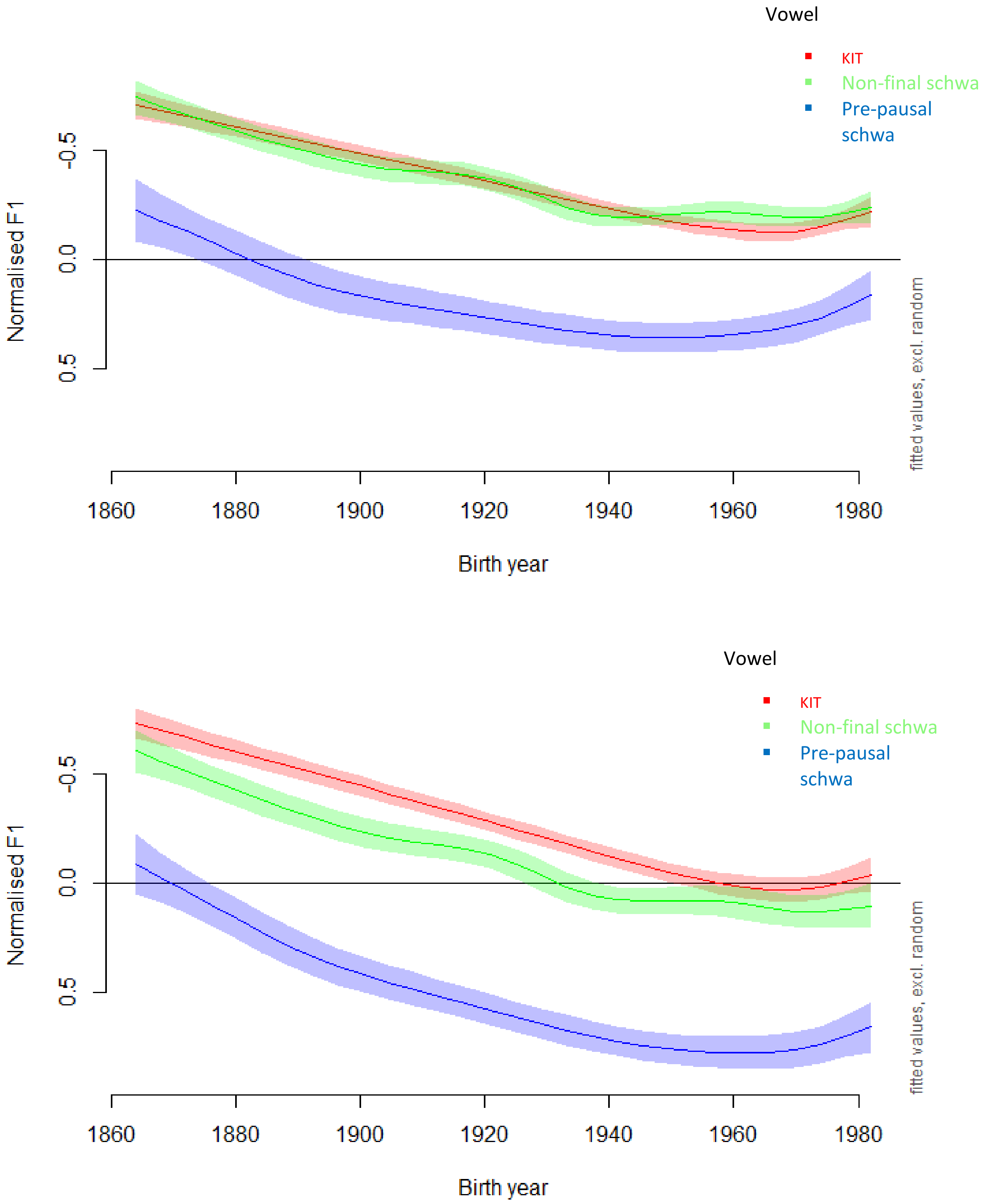


Figure 6.9 *Effect of birth year on normalised F1 of vowels at 0.03 (top) and 0.12(bottom) secs*

It is very clear from Figure 6.9 that even where non-final and pre-pausal schwa are of exactly the same duration, pre-pausal duration is still far lower. This shows that the overall lower height of pre-pausal schwa seen in Figure 6.5 is not simply due to it being a longer vowel overall. This large difference between these vowels can therefore be at least partially attributed to the fact that, by their very nature, pre-pausal schwas do not have a following adjacent consonant. Therefore whilst the height that non-final schwa can reach is affected by the need for the articulators to move in between both a preceding and a following consonant, pre-pausal schwa is only affected by the preceding consonant in this way.

6.5.3.4 Summary of F1 findings

Prediction 1 was that schwa would change over time in the same direction as KIT. This prediction was borne out for both non-final and pre-pausal schwa, with both vowels clearly becoming lower over time. There was also a focus on whether changes in KIT would lead to KIT and schwa becoming more similar over time. Due to the lowering of pre-pausal schwa over time, it remains clearly and significantly much lower than KIT throughout the period. For non-final schwa, even in the earliest years in the period, preceding the change in KIT, non-final schwa and KIT are actually of a very similar height, and due to the lowering of non-final schwa they remain very similar in height throughout the period. The fact that KIT and non-final schwa begin the period at a similar height is not that surprising. It follows the findings in Chapter 4 about non-final schwa's relatively high position in Derby, and the similarly high position that others (e.g. Flemming, 2009; Lilley, 2012; Bekker, 2014) have found non-final schwa to occupy. What is more interesting is the fact that it becomes lower over time, just as KIT does.

Prediction 2 was concerned with the effects of duration on KIT and schwa and how this effect may change over time. Prediction 2a was that schwa vowels would be lower at longer durations, as found in the Derby corpus. This was indeed the case for both non-final schwa and pre-pausal schwa throughout the period. Given that longer vowels are lower vowels, this means that the actual durations of the vowels, as seen in 6.5.2, have an effect on the height of the vowels. Thus the fact that pre-pausal schwas are overall much longer than KIT and non-final schwa accentuates the fact that it is at lower height. The relationship between duration and vowel height is particularly important for non-final schwa. It shows that, even where it is a fairly high vowel at the beginning of the period, it seems to be targeting a lower height.

Prediction 2b regarded the interaction between duration and birth year. It was predicted that as any of the vowels lowered over time, duration would have increasing influence on their height. This was again what was found. All three vowels undergo this change, which means that the different effects of duration on KIT and schwa continues throughout the period. That both schwas undergo this change as well as KIT means that duration continues to have a stronger effect on the height of schwa than KIT.

Overall then, if we ask whether KIT and schwa remain at different vowel heights throughout the period, the answer varies depending on which schwa we look at. Pre-pausal schwa remains different from KIT across the entire span of the corpus, regardless of duration. With regard to non-final schwa, though, the vowels are generally very similar in height. In addition, the fact that they are also very similar in overall duration means that they will indeed tend to surface with similar vowel heights. However, if we ask whether KIT and schwa converge on the same target vowel height, the answer is a no for both types of schwa. Despite changes in the overall height of KIT, and even in the increased positive effect of duration on the F1 of KIT, KIT and schwa still remain different in this regard. Throughout the period, duration still has a stronger effect on both schwas than it does on KIT. That both schwa types still have a stronger effect of duration on F1 than KIT suggests that they still have a lower height target than KIT.

The fact that schwa undergoes similar changes to KIT throughout the period suggests that the two vowels may be linked in some way. The data here alone is not sufficient to prove any causal relationship between KIT and schwa. However, it certainly suggests both that there is a relationship between the two vowels, and that schwa is capable of participating in such sound changes.

6.5.4 F2 findings

In this section the results of the F2 analyses are presented. Unlike for the F1 results, the models that were run, and consequently the figures that are shown, illustrate the patterns separately for different phonetic contexts. These contexts are split up as explained in section in 6.4.5. This means that due to the lack of a following consonant for pre-pausal schwa, the labelling of phonetic contexts for pre-pausal schwa is slightly different, and just refers to the preceding context. However, the labels of back, central and front still refer to the relative effects of each of the phonetic contexts for each vowel.

As for F1, the section is split up into two main parts. First, the overall effects of birth year are considered, and then the effects of duration, including how the effects of duration may change over time.

6.5.4.1 Predictions

Prediction 1: influence of speaker year of birth

a) As per descriptions of the retraction change in KIT in NZE, it is predicted that there will be a clear lowering in the F2 of KIT regardless of phonetic context. If schwa is affected by this change in backness at all, we would also expect to see lowering of F2 in schwa, in at least some environments. It is not necessarily expected that an F2 change would be present in schwa for all three environments, as it is predicted that the different environments will differ in how back they are to begin with. If there is any change in schwa it would be reasonable to expect the three phonetic contexts to be affected differently. If schwa undergoes backing we may expect to see less backing in back phonetic contexts since these schwas are expected to already have a fairly back realisation. It will be of interest to see whether the backing of KIT over time leads to it having a comparable F2 value to either non-final or pre-pausal schwa.

b) As we examine the ways in which different phonetic contexts change over time, we will also be able to observe whether certain phonetic contexts become more similar or further apart. This could result in either an increase or decrease in contextual variability. The effect of birth year on different phonetic contexts is of particular interest for KIT. As discussed in section 2.4.2, there is an assumption that schwa is more contextually variable than other vowels. By analysing KIT in comparison to schwa we can investigate this assumption. Firstly, we will see whether the contextual variability of KIT and schwa changes over time. This will show whether, as KIT becomes more central, it also becomes more contextually variable. If KIT also becomes more contextually dependent as it becomes more central than the differences between KIT in back and front contexts will become greater. Secondly, we will be able to compare the variance of KIT and schwa by phonetic context at the end of the period, when they are more similar in vowel quality, in order to see if schwa is exceptionally variable according to phonetic context.

Prediction 2: influence of duration

a) It is predicted that if duration has an effect on the F2 of KIT in the early years, then it will be the case that KIT will be fronter at longer durations. This is because it is expected that KIT has a fairly front target at the beginning of the period. We would expect any contextual effects to lessen at higher durations and consequently the F2 of KIT to get higher as duration increases. As its overall value becomes more central it is expected that these duration relationships may change in at least some phonetic contexts to reflect the fact that its target is becoming backer.

b) If there is evidence of targetedness in schwa it is expected that duration would also have an effect on the F2 values of schwa. For non-final schwas in particular, this is likely to be evidenced in phonetic contexts having different relationships with duration, since it is flanked by two adjacent contexts. For example, if schwa has a central target it would be expected that duration would have a relatively positive effect on F2 for back contexts and a relatively negative effect for front contexts. This would mean that at longer durations different phonetic contexts will become closer together. For this to be the case, at least one of the examined phonetic contexts would have to change according to duration. We would expect duration to have a relatively negative effect upon F2 in front contexts, and a relatively positive effect on F2 in back contexts. This means that, for example, F2 values should increase for schwas in back contexts in longer schwas i.e. they should become less back.

6.5.4.2 Overall findings by birth year

Because of the complexity of having different phonetic environments to compare, the effects were modelled separately for each vowel. The values and graphical representation given are from model 6.3 for KIT, model 6.4 for non-final schwa, and model 6.5 for pre-pausal schwa. Figure 6.10 shows the changes in each vowel over speaker year of birth at the median duration for each vowel. Each panel plots the changes separately for each phonetic environment.

It is clear that the phonetic context has a strong effect on the F2 value for all three vowels. This is indicated by the fact that there are significant differences between the overall F2 of the different phonetic contexts with each vowel, as indicated in the significance of the parametric terms for phonetic context for each vowel. Central contexts are significantly different from back and front context for all three vowels ($p<0.0001$ in all cases). Table 6.2 below provides the estimates of the difference between the most extreme phonetic contexts, front and back, at the average birth year and duration. We can see that the size of the difference between back and front environments is very similar for KIT and non-final schwa. The fact that overall the sizes of the differences between phonetic contexts are similar for KIT and non-final schwa goes against the idea that non-final schwa is a vowel which is exceptionally context-dependent. If we look at the estimated difference for pre-pausal schwa we can see that the difference between pre-pausal schwas in different phonetic contexts is smaller. This is also apparent from Figure 6.10. The different phonetic contexts are considerably closer together for pre-pausal schwa than they are for other vowels, and the confidence intervals for each context are much larger. This is, of course, not particularly surprising. Whilst the three compared environments for non-final schwa and KIT reflect differences in both the preceding and following context, for pre-pausal schwa only the preceding context differs.

Table 6.2*: Estimated differences between back and front contexts for all vowels (Normalised F2)*

Vowel	Estimated difference between front and back environments
KIT	0.743
Non-final schwa	0.725
Pre-pausal schwa	0.51

If we look at the plot for KIT in Figure 6.10 it is clear that, as predicted in 1a, the F2 clearly lowers over time for all three phonetic contexts. The vowel is much backer at the end than at the beginning of the period, regardless of phonetic context.

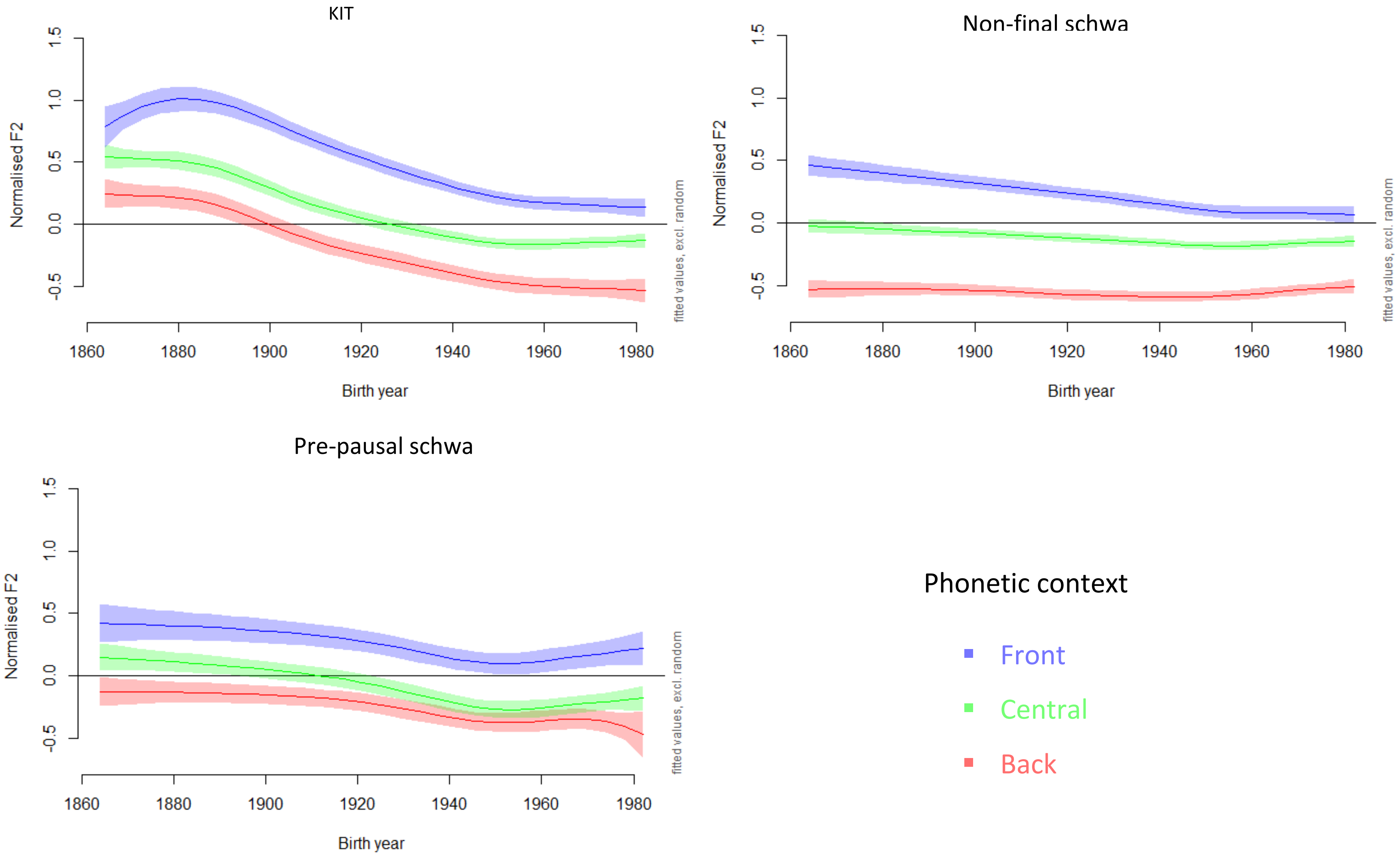


Figure 6.10 *Normalised F2 for vowels by phonetic context and birth year*

If we look at the smooths for the schwas, there is also some change over time, although it is more subtle. In non-final schwa, the way F2 varies over birth year differs according to phonetic context (model 6.4). This was as was predicted in prediction 1a, as it was expected that if non-final schwa does change at all then different environments would be affected differently. The largest change we see over time is in front contexts, where there is clear backing over time. The difference smooth for birth year for front environments shows that the change in front environments is significantly different from central contexts by birth year (F=35.1164, edf=1, p<0.0001). Indeed, it was predicted that if schwa undergoes any backing at all that it should be largest in front contexts. This makes sense, as if schwa is backing, then schwas in front contexts would show the most movement. We also see a small amount of backing in central contexts, but to a lesser degree, which is again in keeping with prediction 1a. This is seen in the significance of the birth smooth (F=9.416, edf=2.711, p<0.0001). In back contexts, however, there is no backing over time at all. This difference between back and central contexts is shown in the significance of the birth difference smooth for back contexts (F=19.192 edf=1.001, p<0.0001). This is, of course, not unexpected, given that schwa already had quite a back realisation in back contexts at the beginning of the period, and so there appears to be a ceiling effect here. There is a very slight increase in the F2 of back contexts in later years, which was not predicted, although at the median duration shown this is a very small effect.

Pre-pausal schwas also show backing over time, as seen in the significance of the birth smooth in model 6.5 (F=18.075, edf=3.674, p<0.0001). However, unlike non- final schwa this effect is the same across all environments. There is no difference in the way that the three environments change over time. This is reflected in the fact that the difference smooths for birth year between environments are not significant in the pre-pausal model. As has already been shown, pre-pausal schwa varies less according to phonetic context than non-final schwa. It is therefore not surprising that there should also be less variation in how different phonetic contexts change over time.

The focus of prediction 1b was both on whether KIT would become more contextually variable as it becomes more central, and on whether KIT would be more contextually variable than non-final schwa at the end of the period. It was predicted that if high contextual variability of non-final schwa is simply to do with the fact it is a short vowel that is overall central, then KIT should become more contextually variable as it becomes more central on average. There is no clear change over time in the variation of KIT according to phonetic context. The difference between phonetic contexts is fairly similar at the beginning and end of the period. Therefore, this prediction is not borne out. There is no indication that high contextual predictability is a unique attribute to short central vowels, and no indication that KIT becomes more 'schwa-like' in its contextual variability as it becomes more central. The other prediction was that if contextual variability is something unique to unstressed vowels then KIT should be less contextually variable than non-final schwa even as it becomes more central. That would mean that even if KIT should change to have the same average F2 values as schwa, that it would differ less according to environment. This prediction was also not borne out. If we compare Figure 6.10 at the end of the period we can see that there is no real difference between the equivalent environments for KIT and non-final schwa. That is to say, KIT and non-final schwa are no different when both are in front contexts, when both are in central contexts, or when both are in back contexts. This clearly offers no indication that contextual variability is something that is unique to having a lack of stress. The other side of this is that KIT and non-final schwa are indistinguishable in backness in the later years of the period, at least at the median durations shown here.

6.5.4.3 Influence of duration

In this section I examine the effect of duration on F2, and how the effects differ across phonetic contexts and also change over time. As was the case with F1, the four different colours represent the model estimates for the vowels at different durations, with the red line showing the changes for vowels of particularly short durations (0.03 seconds), and with the blue line representing vowels of relatively long durations (0.12 seconds).

Figure 6.11 shows the effects of birth year by duration on each phonetic context for KIT. For KIT it is clear that duration has an effect on F2 for all three phonetic contexts. For all three phonetic contexts we find that in the early years longer vowels are fronter. This is as was predicted in prediction 2a. This makes sense as in the earlier years the target is fairly front, so it would therefore be expected that vowels would become fronter as they become longer and have more time for the articulators to reach their target. In back and central contexts this relationship with duration remains throughout the period. However, in front contexts in later years, the vowels do not become fronter as they get longer. To explain this we might propose that the target has changed by this point to be more central, so it would not be expected for vowels to get fronter still at longer durations, when they are already in an environment which exerts a fronting effect.

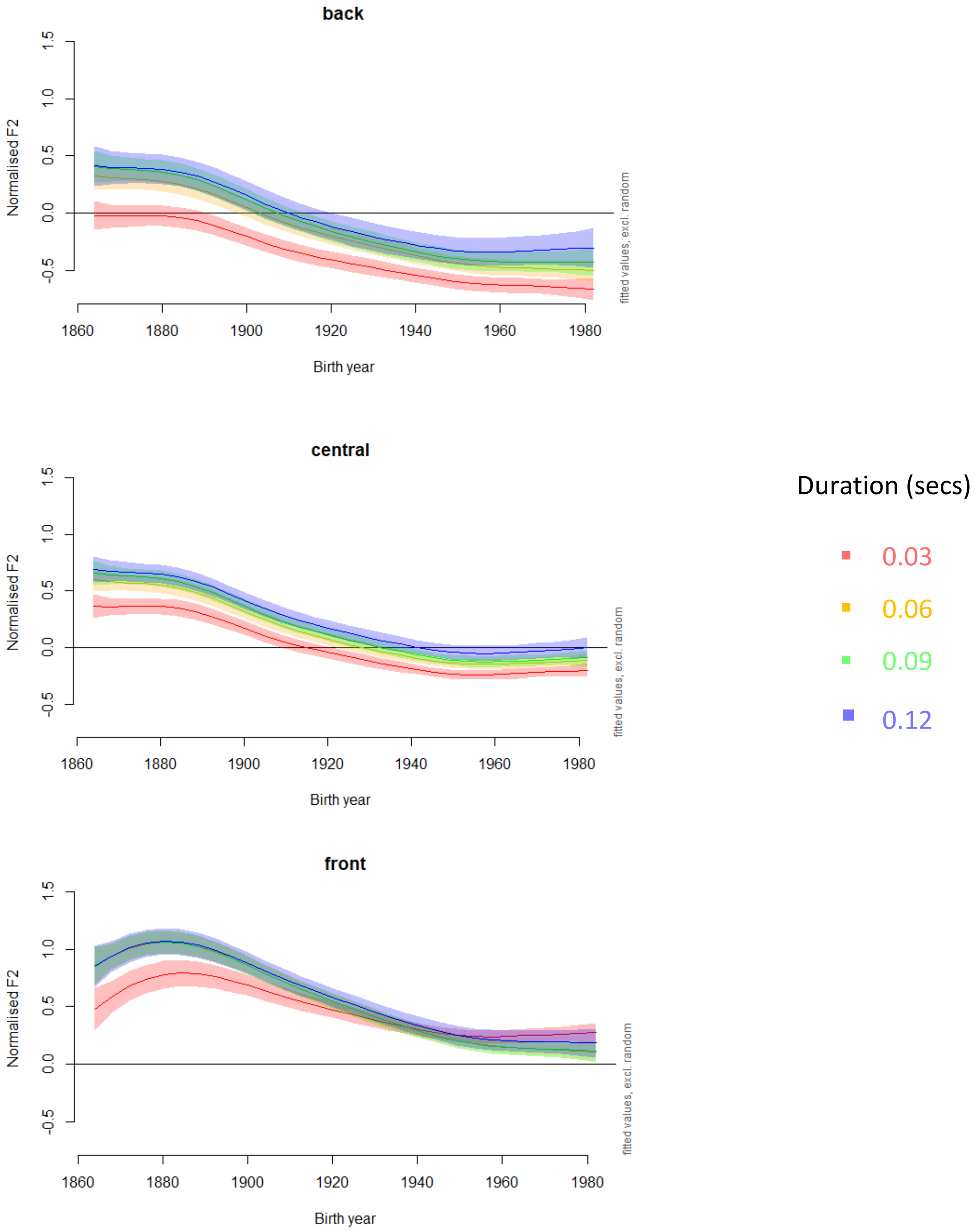


Figure 6.11 *Effect of birth year on normalised F2 of KIT across phonetic contexts by duration*

For non-final schwa, as described in prediction 2b, it was expected that if non-final schwa showed evidence of targetlessness, this would be seen in an effect of duration on F2. It was predicted that there would be less difference between the phonetic contexts at longer durations. For this to be true, either back contexts would have to become fronter at longer durations, or front contexts would have to become backer at longer durations. As shown in Figure 6.12, in the early years there are no real duration effects at all. However, in the later years there is a very clear effect of duration in back contexts. As predicted, they are fronter at longer durations. This effect of duration in back contexts is shown in the fact that the difference smooth for duration between back and central contexts is significant (F=50.341, edf=1.001, p<0.0001) meaning that the effect of duration is significantly different between back and central contexts. The difference between back and central contexts for the interaction between birth year and duration was also significant (F=4.071, edf=1.002, p<0.0001). This means that there is a significantly different interaction between birth year and duration in back contexts than there is in central contexts. As was seen in the previous sections, back contexts become more front over time, although at the median duration this is quite a subtle effect. As can be seen in Figure 6.12 this effect is much clearer at longer durations. This means that at later years we do find some evidence of targeted F2 behaviour in non-final schwa as at longer durations as there is indeed less difference between the contexts than at shorter durations. In fact, it is striking that by the end of the time period, long realisations of schwa across all contexts seem to be almost identical in terms of F2.

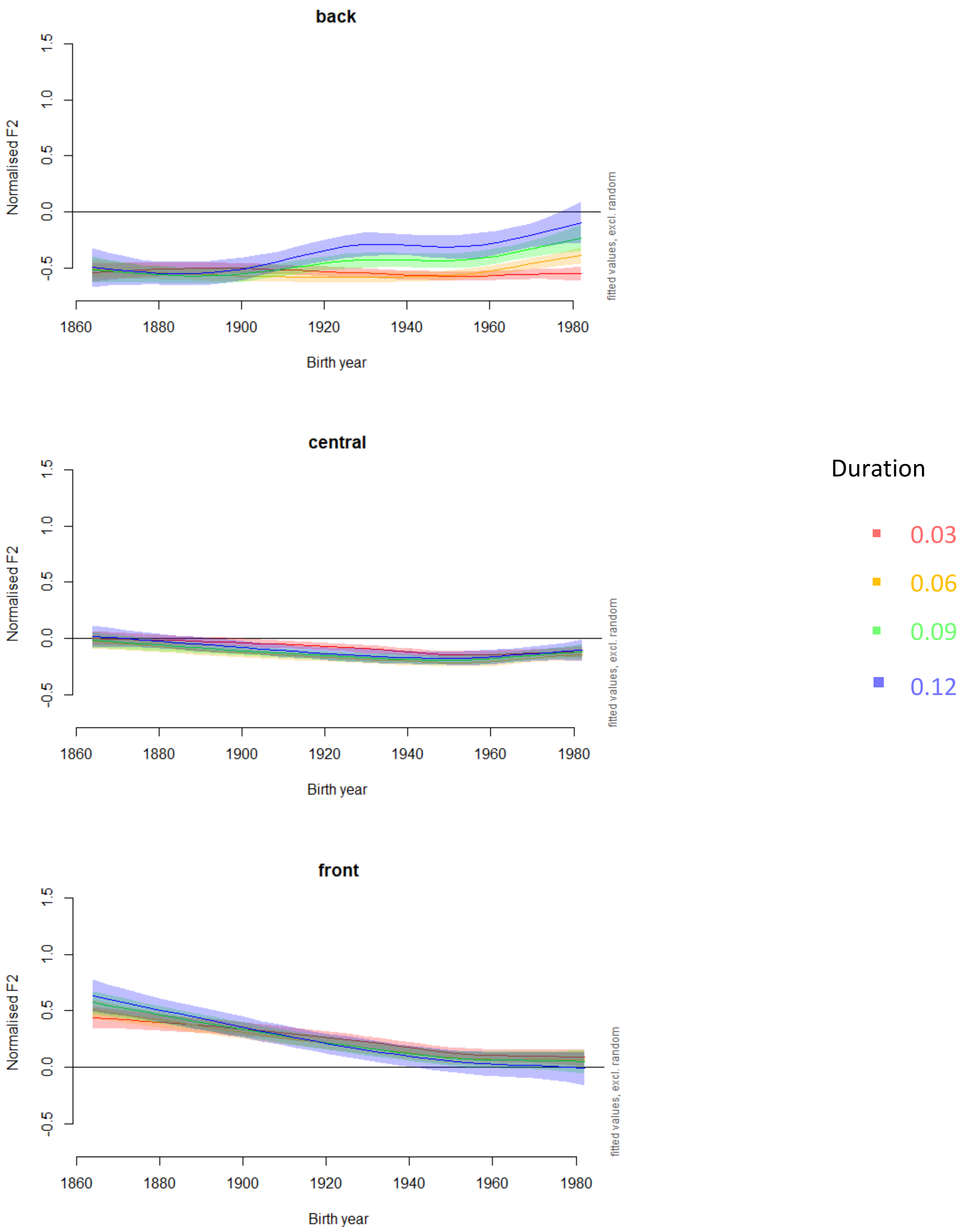


__Figure 6.12__ Effect of birth year on normalised F2 of non-final schwa across phonetic contexts by duration

This is particularly clear if we look at Figure 6.13 below. The top graph is the same as shown Figure 6.10, and it shows the changes over time in the different phonetic contexts at the median duration. This is compared with schwa at longer durations of 0.09 seconds. Although this is quite a long duration for non-final schwa, as seen in Figure 6.5 in section 6.5.2 it is still within its normal range. We can clearly see that at 0.09 seconds the differences between phonetic contexts decrease a lot over time. This is both due to the backing of non-final schwa in front contexts over time, and the fronting in back contexts. Although the lessening of contextual variation at longer durations was predicted, it was not predicted that this effect would increase over time, nor was it predicted that there would be less contextual variation in non-final schwa in later years. This was a slightly surprising finding which will be discussed in more detail later.

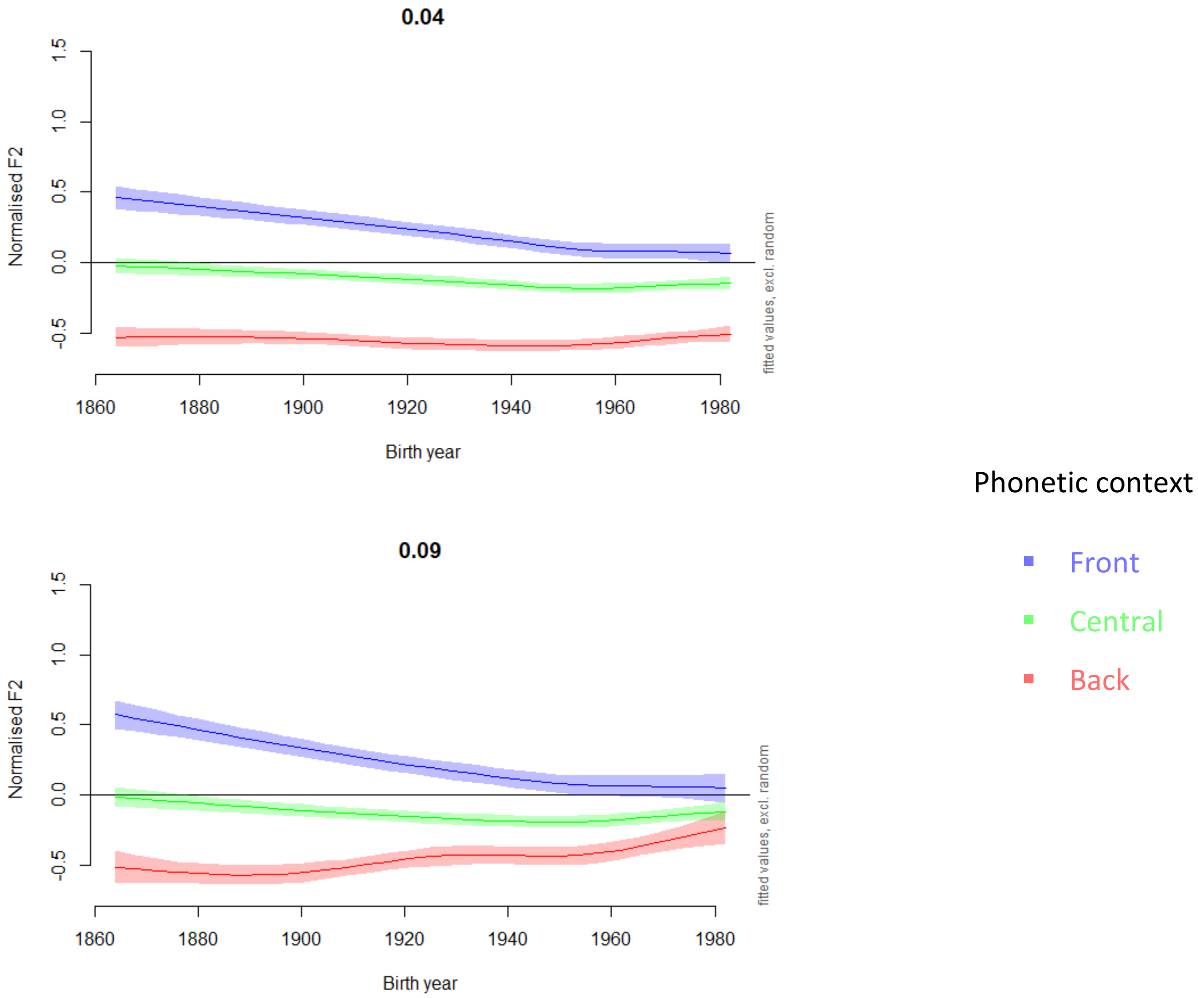


Figure 6.13 *Effect of birth year on normalised F2 of non-final schwa across duration (secs) by phonetic contexts*

For the pre-pausal schwas there was no difference in the effect of duration on different phonetic contexts, so just the effects on central context are shown in Figure 6.14 below. Again, given that there is less difference overall between the different phonetic contexts it makes sense that they would differ less according to duration. Here it can be seen that longer pre-pausal schwas are backer, as indicated in the significance of the duration smooth in model 6.5 (F=15.454, edf=1, p<0.0001). This is an effect which continues throughout the period.

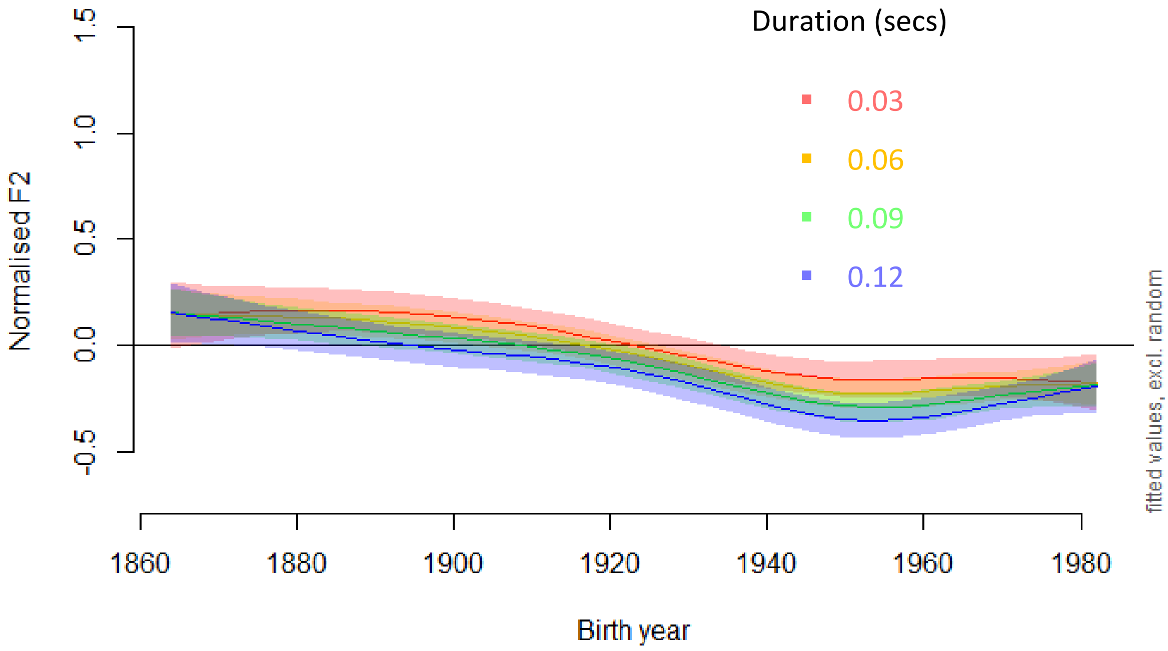


Figure 6.14 *Effect of birth year on normalised F2 of pre-pausal schwa by duration*

The fact that this duration relationship continues throughout the period means that even later in the period, when KIT and pre-pausal schwa are much more similar overall, they still have a different relationship with duration. We can see this in Figure 6.15 below, which directly compares the F2 of KIT and pre-pausal schwa. This figure is based on model 6.6, which compares F2 values for KIT and pre-pausal schwa. Because the phonetic contexts used for pre-pausal schwa and KIT are not equivalent, here just the overall values for KIT and pre-pausal schwa are compared. We can see that, due to the differing effects of duration on KIT and pre-pausal schwa, they are more different at longer durations. This is shown in the significance of the difference smooth for duration (F=80.070, edf=1, p<0.0001). The difference between the effect of duration on KIT and pre-pausal schwa does not change over time, so this suggests the possibility that KIT and pre-pausal schwa may still have different F2 targets. KIT has not backed sufficiently to be indistinguishable from pre-pausal schwa across all durations. The fact that, by the end of the period, KIT remains more distinct in F2 from pre-pausal schwa than it does from non-final schwa is consistent with what was found for F1. Notably though, the difference between KIT and pre-pausal schwa is quite a lot less here than what was seen for F1.

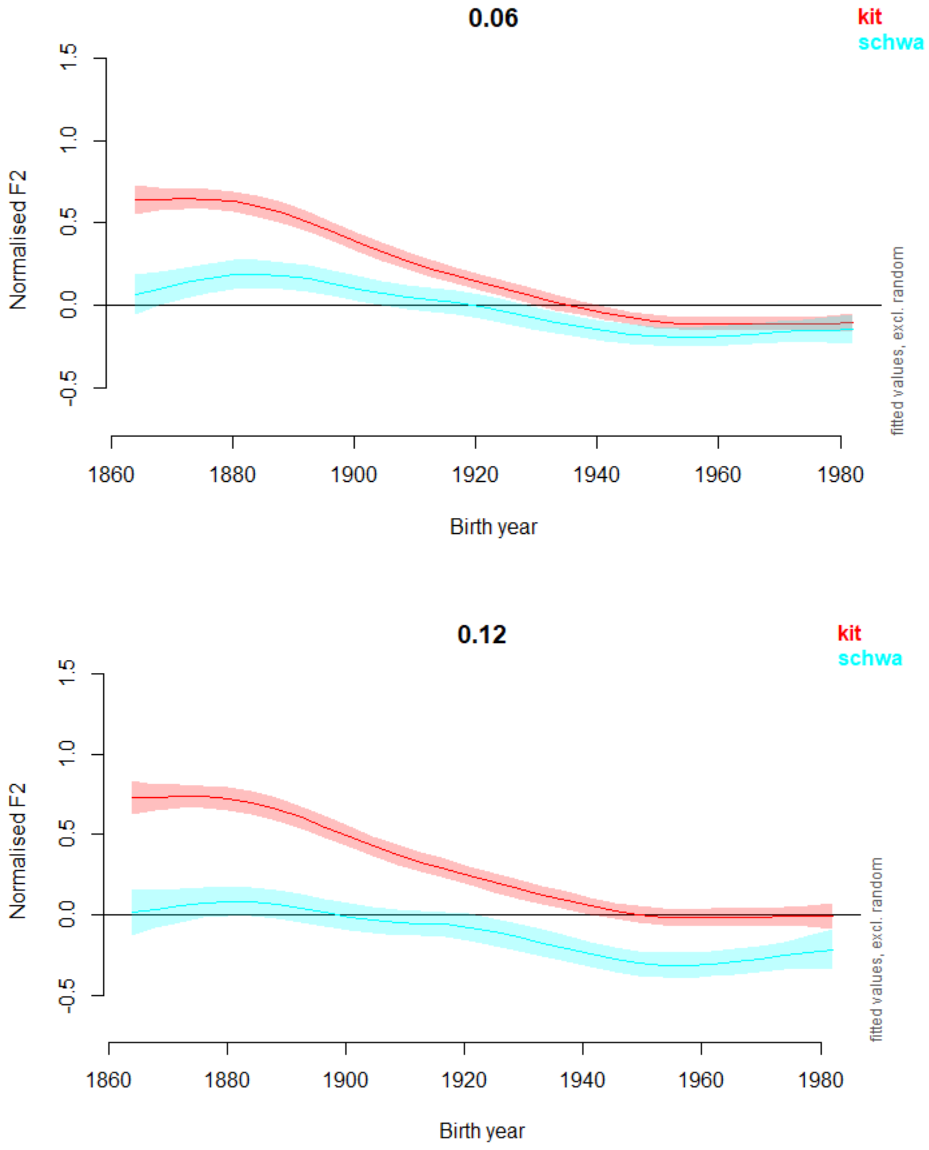


Figure 6.15 *Effect of birth year on normalised F2 of KIT and pre-pausal schwa by duration (secs)*

6.5.4.4 Summary of F2 findings

Prediction 1a was that if schwa is affected by or involved in the change that KIT undergoes over the period, then, like KIT, it should show evidence of backing. This was indeed what was found. Pre-pausal schwa showed backing over time in all environments, whereas non-final schwa only showed it strongly in front environments, and, to a lesser extent in central contexts. This variation in the effects in different phonetic contexts was again also predicted for non-final schwa. That this change occurs in both types of schwa again suggests a link between schwa and KIT.

Despite the changes in non-final schwa it is not the case that these changes lead to it staying separate from KIT in F2. By the end of the period non-final schwa and KIT are very similar to each other across all phonetic contexts. Rather, the change in non-final schwa simply means that KIT does not surpass it in backness. For example, if KIT had undergone backing over time and non-final schwa had not undergone any change then KIT would overall have become backer, but this is not what happens. Given that the change in KIT simply means that the two vowels are very similar in F2 by the end of the period, it is hard to be sure what the exact nature of the relationship between KIT and schwa is. However, it should be noted that pre-pausal schwa does remain backer than KIT at longer durations. Whilst even at longer durations this is not a huge difference, the fact that pre-pausal schwa is more different from KIT than non-final schwa is, is consistent with what was found for F1, where pre-pausal schwa was much lower than KIT, and non-final schwa was more comparable to KIT in height. Of course, it makes sense that pre-pausal schwa would be more different given that it occurs in a different environment to non-final schwa and KIT.

Prediction 2a was that if there was evidence of targetlessness in non-final schwa, this would be evidenced in different phonetic contexts having different relationships with duration. Importantly, clear evidence of this was found, although it was only in the later years of the corpus. That the effect of duration on non-final schwa would change over time was not predicted. There was no particular reason to expect that schwa would become more or less targeted over time, so this is a somewhat curious finding, which will be discussed in more detail in the following overall discussion of the results. However, the fact that evidence is found of schwas from different phonetic contexts becoming closer together at longer durations is important. It means that, at least for the later years in the period, there is evidence of non-final schwa showing a target in F2. In fact there are actually only quite small differences between the different phonetic contexts for non-final schwa in later years at longer durations. These findings certainly do not suggest that non-final schwa is completely targetless or predictable from phonetic context.

Another claim that is often made about schwa is that it varies more according to context than other vowels. Prediction 1b investigated this claim. Because of the changing nature of KIT, and the fact that it becomes more similar to schwa over time, it was possible to investigate both whether schwa was exceptionally variable according to context compared to other vowels, and if so, whether that was due to its central location within the vowel space or its lack of stress. It was found there was no change in the contextual variance of KIT over time, as it became more like schwa in overall F2 value. It was also found that towards the end of the period KIT and non-final schwa were extremely similar at all phonetic contexts. Therefore there was also no evidence found for the idea that schwa is exceptionally variable according to context.

6.5.5 Overall vowel space

Thus far, we have seen how the height and backness of KIT and schwa compare. Before moving on to the overall discussion of the results, it is important to take a brief look at how the changes we have seen separately in F1 and F2 map onto the overall vowel space. Figure 6.16 below compares the 95% confidence intervals for each vowel in the first and last 20 years of the period. Clearly these overall ellipses do not capture any of the duration based variation within the models, but they give an overall picture of the changes each vowel undergoes. It is apparent that schwa undergoes more lowering than backing over time. This is especially true for non-final schwa. As we saw in sections 6.5.4.2 and 6.5.4.3, although non-final schwa undergoes some backing, the degree varies according to phonetic environment. Therefore, when looking at the data overall we wouldn’t necessarily expect a clear backing effect.

In the F2 analysis it was observed that, for non-final schwa, the difference between the phonetic contexts reduced over time. This is reflected here, as the F2 range is reduced over time. In later years there are both less back schwas, and less front schwas. Therefore, for non-final schwa, as its overall height lowers, its F2 variability decreases. This pattern is also seen in KIT. This reflects an important fact which may help explain the unexpected effect of differences between phonetic contexts reducing over time in non-final schwa. The effect of adjacent consonants on F1 should generally be to uniformly lower it. This is why we see a strong relationship between duration and F1 in schwa, as longer schwas are lower schwas. However, the effect of adjacent consonants on F2 varies, depending on the consonants. Although we have looked at the F1 and F2 dimensions separately so far, of course they do not really operate independently. Over the whole vowel space therefore a reduced effect of phonetic context should cause both vowel lowering in schwa and reduced variability in F2 (see sections 2.4.3 p.37-8 and 4.2, p.98). As non-final schwa becomes a lower vowel over time it thus makes sense that its variability by phonetic context would reduce, since these differences are both related to a reduced effect of adjacent consonants. This link is also seen in the fact that it is at longer durations where non-final schwa both lowers the most over time, and where the influence of phonetic context reduces the most over time. This link between F1, F2, and adjacent consonants can also help reconcile the large difference we see between non-final and pre-pausal schwa. Pre-pausal schwa is longer, and has no following phonetic context so it is always less influenced by adjacent consonants than non-final schwa. This allows for it both to be a lower vowel, and to vary less along the backness dimension.

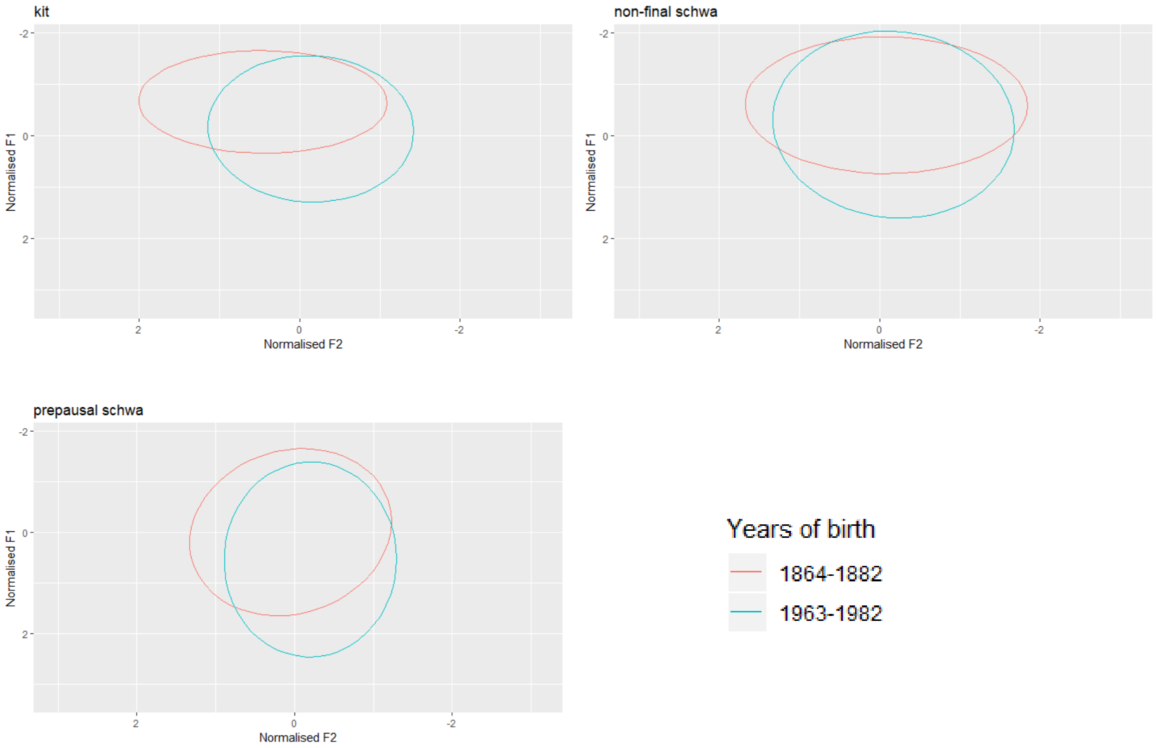


Figure 6.16 *Normalised vowel spaces of all vowels compared across early and later birth years*

6.6 Discussion

Here the overall findings are discussed, with reference to the broader questions posed in section 6.3 of this chapter.

6.6.1 Does schwa change over the period?

The first main goal of the analysis in this chapter was to examine whether schwa was in any way involved in or affected by the front vowel shift in NZE. The particular motivation for examining schwa in this variety was because of the change in the KIT vowels that has led to it becoming lower and more central, and so moving into the part of the vowel space generally occupied by schwa. The rationale was that if schwa is a vowel without an independent target, than it should be unaffected by any change in other vowels.

It was found that schwa, along with KIT, undergoes lowering and backing during the period. The lowering effect was seen more clearly than backing, with both types of schwa showing clear lowering effects over time. However, there was some more subtle backing in pre-pausal schwa, and for non-final schwa in certain environments. This suggests that schwa participates in the same sound change as KIT. Crucially the fact that schwa is changing over time at all, suggests that it cannot be a vowel without an independent target. It appears to have a target, and this target may well change over time.

This therefore suggest that, at least in NZE, schwa is not maximally reliant on adjacent consonants for its articulation, as has been suggested by some (e.g. Kondo, 1994, Bates, 1995). Moreover, the changes in pre-pausal schwa also do not suggest a vowel that simply moves to a phonetically neutral rest position since this too changes. Despite the clear differences in height between pre-pausal and non-final schwa, the sharp division that has often been made between the two types of schwa in targetedness by those such as Flemming (2009) also does not stand up here. Whilst their actual vowel qualities are quite different, both undergo change over time, and both show a clear change in the influence of duration over time.

From the data here it is not possible to be sure of the exact motivation for the changes we see in schwa. The changes in pre-pausal schwa mean that it stays clearly distinct from KIT, more so in the height dimension. This could suggest a motivation for KIT and schwa to remain distinct. However, with regards to non-final schwa the argument for this motivation is slightly weaker, given that non-final schwa and KIT are extremely similar by the end of the period. There are, nevertheless, still some slight differences between KIT and non-final schwa overall; KIT is slightly longer, and non-final schwa is lower at long durations. If non-final schwa did not undergo any change then KIT would have become lower and slightly backer than non-final schwa. The fact that non-final schwa changes means that it does not become lower than non-final schwa, and does not become backer in equivalent environments. For much of the period the height of KIT and non-final schwa is changing in parallel, which does indeed suggest a link between the changes. Despite the overall similarity of KIT and non-final schwa, the fact that we see their height continue to diverge throughout the period does suggest that change in non-final schwa may be caused by an aim to maintain a distinct target from KIT. That these two vowels are in fact not realised very differently may just be reflective of the short durations of both vowels.

Another possibility is that the 'target' of non-final schwa may not be changing as such but more effort may be being made to reach the target, possibly due to the change in KIT. Flemming's (2009) proposal that non-final schwa is articulated with minimal effort is relevant here. Flemming's proposal is that the high value of non-final schwa could be caused by minimised effort because of a lack of need to maintain contrasts with other vowels. It is possible that the changing value of KIT provides more of a motivation to attempt to maintain a contrast with KIT rather than minimise effort. The way schwa changes over time is generally similar to the way that it changes when longer, and this provides some support for the idea that over the course of the period it could be produced with an increasing amount of effort. For example, schwas get lower over time, and when longer; pre-pausal schwa gets backer overall over time, and when longer. The effect of phonetic context on non-final schwa reduces over time, particularly when it's longer.

Regardless of the exact motivation for the changes which we see in schwa, the key finding is that it changes at all. If the articulation of schwa was purely determined by context, and it had no independent target of its own, then we would not expect to see this type of change over time. Therefore, the fact that we do see schwa change over time strongly suggests a vowels with a target, and a vowel that is affected by changes in other vowels.

6.6.2 How similar are KIT and schwa throughout the period?

The second main goal of this chapter was to compare the contextual variance of KIT to schwa. As KIT changes to become more similar to schwa over the period, this provided an opportunity to compare schwa with a stressed vowel that was both similar in duration and vowel quality, in order to see whether schwa had an exceptional level of variability according to context. The aim was twofold: to see whether KIT became more dependent on context as it became more like schwa in vowel quality, and whether schwa was indeed more contextually dependent than KIT by the end of the period, when they were both of similar vowel qualities. Firstly, pre-pausal schwa was a lot less variable according to phonetic context, which was expected as it only has a preceding context. With regard to KIT there was no clear change in its variability by phonetic context; over time the influence of phonetic context on KIT remained fairly static. Therefore it was not the case that as its vowel quality became more ‘schwa like’ it was also becoming more variable. Secondly, and more importantly, at the end of the period KIT and non-final schwa are extremely similar in all phonetic contexts. Therefore there was no evidence found that high variability according to phonetic context is an attribute unique to schwa caused by its lack of stress, or that contextual variation is how schwa distinguishes itself from other vowels (e.g. Kondo, 1994 Flemming, 2009).

Although KIT and non-final schwa differed more at longer durations, they were overall very similar. Therefore, the anecdotal perceptions that schwa and KIT are indistinct and very similar (see section 6.2.2) seem fairly accurate in terms of production as well, at least in terms of non-final schwa. With pre-pausal schwa, however, it is still clearly distinct from KIT. This similarity between KIT and non-final schwa could be a reason for the perception reported by Nokes and Hay (2012), that unstressed syllables are more fully articulated, as if non-final schwa is now more similar to stressed KIT then unstressed and stressed syllables may be perceived as less similar.

6.6.3 Effects of duration on schwa

Although the main reasons for examining this dataset were due to the changes in the KIT vowel over time, it was also possible to examine synchronic duration effects at different points over the period. The effect of duration again pointed to schwa being a vowel with an independent target. As shown in the Derby corpus in section 4.4.4.2 there was a clear relationship with between F1 and duration, with longer vowels being lower vowels. Whilst even this relationship has previously been argued to be schwa simply moving to a physically neutral position (e.g. Bates, 1995), the fact that this relationship becomes stronger over time refutes this idea. In the later years, where non-final schwa was lower, there was also evidence of the effect of phonetic context on F2 reducing at longer durations. Again this was evidence against a targetless schwa, as if schwa was truly a vowel without an independent target, we would not expect it to vary in this way according to duration. This variability according to duration suggests that much of the variation which could be seen as it being targetless is more likely due to non-final schwa often being of a short duration.

6.7 Conclusion

This chapter addressed two of the main research questions set out in Chapter 1. The primary aim of the chapter was to address the question of whether schwa has a phonetic target. In addition, this chapter compared results from both pre-pausal and non-final schwa in order to address the question of whether there is a phonological difference between word final and non-final schwa. Similarly to the findings in Chapter 4, there were large surface differences in their vowel quality. Despite this, the overall behaviour of pre-pausal and non-final schwa was similar. Both types of schwa undergo change in the same way over time, and the effect of duration on the height of both types of schwa becomes stronger over time. This suggests that word final and non-final schwa may not be phonologically different.

With regards to the issue of targetlessness, altogether there are three aspects of the data which all provide important evidence against the idea of targetless schwa. Firstly, schwa changes over the period. As KIT lowers, schwa also lowers, and the increased positive effect of duration on the height of schwa suggests an increasingly lower target over time. Along with KIT there is also somewhat of a backing of schwa over time, although this change is much more slight and is only clearly seen in pre-pausal schwa. These changes in schwa over time are clearly a possible response to the change in the KIT vowel. Although the difference in the backness of KIT and schwa substantially reduces over the period the difference between them in height actually remains fairly static, thus suggesting that the changes in height may be linked. Throughout the period, duration also has a different effect on KIT and schwa. This is clearer in the effect on vowel height than backness, and suggests a continuing different target, at least in height, for KIT and schwa. That schwa changes clearly suggests it has a target independent of phonetic context, and can indeed be affected by changes in other vowels. Secondly, schwa is no more contextually predictable than KIT at the end of the period, which suggests that the level of variability in schwa is not unique to it. Overall the contextual predictability is comparable for KIT and non-final schwa. Thirdly the effect of duration on schwa throughout suggests that it is a vowel with a target, and that it moves closer to that target at long durations. These three facts all provide important evidence that schwa is a vowel with a clear target.

7 Overall discussion of findings

In this chapter the results from the various analysis chapters are brought together in order to answer the broader questions posed in the thesis. The section starts by discussing the results from the comparison of schwa with other unstressed vowels and how these findings relate to the rest of the thesis (7.1). The different measures that were used to examine the question of whether schwa is targetless are then discussed in 7.2, and the results from the Derby and NZE analyses are compared. The last three sections focus on the broader points of whether word final and non-final schwas are fundamentally different (7.3), how differences between Derby and NZE can be explained (7.4), and why schwa may often fall short of its phonetic target (7.5).

7.1 Schwa versus /ɪ/

One of the key issues in this thesis was situating schwa within the context of other unstressed vowels. There is somewhat of a contradiction between the impressionistic literature which transcribes and suggests a clear and categorical difference between /ɪ/ and /ə/ (Gimson, 1962; Roach, 1983), and the instrumental literature on non-final schwa which suggests that schwa is a very variable, if not targetless, vowel (Kondo, 1994; Bates, 1995; Flemming, 2009). Extreme variability in schwa would of course not be conducive to making a clear distinction with other unstressed vowels, and there is a lack of research which actually looks at the phonetic details of the difference between /ɪ/ and /ə/. Chapter 3 therefore examined to what degree there is actually a difference between /ɪ/ and /ə/ in unstressed syllables, and in what way that difference manifests itself.

Whether unstressed vowels are said to be produced with /ɪ/ or /ə/ is linked with spelling. It is generally implied that <i> represents /ɪ/, <a>represents /ə/, and <e> is variable between /ɪ/ and /ə/ (Jones, 1956; Roach, 1983; Wells, 1990). In addition, spelling is also linked to the historical pronunciation of such vowels. Because spelling correlates consistently with /ɪ/ or

/ə/ according to native speaker intuition, I therefore used the variable of spelling to examine the distinction between /ɪ/ and /ə/ in unstressed syllables. <a> was used as a proxy for /ə/, <i> for /ɪ/, and as a supposedly more variable vowel <e> was also investigated. This allowed me to avoid the influence of any preconceptions about how /ɪ/ or /ə/ would be distinguished or which words they would occur in.

The difference between /ɪ/ and /ə/ was primarily in backness, with the <a> vowels being backer than the <i> vowels. This difference in backness corresponds to the difference that the transcriptions /ɪ/ and /ə/ suggest. However, it was found that, contrary to the commonly used transcriptions of [ɪ] and [ə], that overall these vowels were produced at a very similar height in Derby. The lack of difference between these vowels in height also parallels the general lack of height difference which was found between non-final schwa and stressed KIT in the earlier speakers in NZE.

These findings showed, as is suggested in the traditional literature on /ɪ/ and /ə/, that they are indeed two distinct qualities in unstressed syllables, produced in different areas of the vowel space. However, they were not quite as clearly separated as has been suggested, as the difference was only one of backness, not of height. This means that where studies have made undue assumptions about the vowel quality of schwa, their conclusions could be misleading. For example, Fabricius (2002) examined how comparatively similar the *–ed* and *–es suffix* vowels were to /ɪ/ and /ə/. This was quantified as whether the suffix vowels were produced more similarly to /ɪ/ or to the mid central area of the vowel space. Clearly, if the difference between /ɪ/ and /ə/ is primarily not realised as a height difference this is an unsuitable measure.

Of course, as was shown in both Chapters 4 and 6, there is a strong link between the length of schwa and its height, with it being lower at longer durations. It is therefore expected that at longer durations the difference between /ɪ/ and /ə/ could be larger, and more similar to the difference implied in the transcriptions [ɪ] and [ə]. However, the focus of this analysis was on how these vowels are actually produced by speakers in spontaneous speech, and as they were often produced at very short durations the height difference between /ɪ/ and /ə/ was limited.

The second main finding was that vowels spelt with <e> were intermediate between <a> and <i> in backness. This was taken as evidence against the standard view that there is a categorical difference between /ɪ/ and /ə/. Thus, although the difference between <a> and <i> exemplified a degree of contrast between /ɪ/ or /ə/, the intermediate position of <e> shows that unstressed vowel productions may not always fall neatly into being either a clear /ɪ/ or /ə/ token. It was argued that this three way difference arises from residual subphonemic differences in unstressed vowels that likely arise as a result of incomplete neutralisation of vowel qualities in unstressed syllables. Thus, whilst in unstressed syllables there is no longer a three way phonological distinction between these vowels in the way that there would be between /ɪ/, /ɛ/ and /æ/ in stressed syllables, there are still some residual differences.

These findings showed that there are distinct differences between different unstressed vowel qualities, but that words do not always fall neatly into an /ɪ/ or /ə/ category. Importantly, this led unstressed vowels spelt with <i> and also <e> to be excluded from the analyses in Chapters 4 and 6, in order to focus more explicitly on the issue of variation and targetlessness in schwa.

7.2 Targetlessness and variability

Chapters 4 and 6 examined the issue of targetlessness and variability in schwa. On the surface, the findings from schwas that are not before a pause are similar to previous findings on the phonetics of schwa. In both Derby and in the earlier born speakers in NZE, pre-consonantal schwa is found to have quite a high realisation on average. This is seen in the fact that its F1 values are similar to stressed KIT. In non-final schwa, low average F1 values, and clear variation along F2 according to context was found in both Derby and the earlier born speakers in NZE. These findings are similar to what has been found in non-final schwa in other studies (Kondo, 1994; Flemming and Johnson, 2007; Flemming, 2009; Bekker, 2014). However, the aim was not simply to look at average realisations of schwa, but to use a range of variables in order to infer if evidence of a phonetic target could be found. The goal was to examine whether schwa is simply a maximally assimilatory vowel with no target of its own, or whether it has an independent phonetic target. Three key variables were used to examine this question: formant trajectories, vowel duration, and change over time (speaker year of birth). The findings from each of these parts of the thesis will here be discussed in turn.

7.2.1 Formant trajectories

Formant trajectories were examined in Chapter 4 in the analysis of speakers from Derby. In pre-consonantal schwa, formant trajectories were found to be linear, meaning that there was no deviation to a particular F2 value between adjacent consonants. This is the same as was found in both Kondo (1994) and Bates (1995). By contrast, there was a small deviation in the F1 trajectory of schwa, which showed F1 raising towards the middle of the vowel (i.e. articulatory lowering). This is the same as was found in Flemming and Johnson (2009) and Kondo (1994), although it is different to Bates (1995), who finds linear F1 trajectories. This showed that, in F1, schwa did not move linearly between consonants. In addition, the deviation in the schwa trajectory was slightly more than was found in the KIT vowel, suggesting weak evidence for schwa targeting a lower vowel quality than KIT and therefore not being maximally assimilatory to its context. This difference, however, was fairly small.

Despite the overall similarity of these findings to Kondo (1994) and Flemming (2009), I do not argue that this constitutes evidence for targetlessness in schwa, or that it is an exceptionally variable or predictable vowel. This is for two main reasons. Firstly, pre-pausal schwas behaved differently from non-final schwas, and also pre-consonantal word final schwas. Pre-pausal schwas showed a deviation away from their preceding context, meaning that pre-pausal schwas in different phonetic contexts were far more similar at their end than at their beginning. Like pre-consonantal schwas, they also showed a deviation in F1 away from the influence of the preceding context, with F1 values raising through the schwa. This meant that schwas in word final position behaved differently depending on whether they were before another word or a pause. This suggested that the lack of clear F2 trajectories in pre-consonantal schwa is likely to have been a phonetic effect of being surrounded by two consonants, rather than because it is targetless. The issue of the differences between non-final and word final schwa will be further discussed in section 7.3.

In addition, it was not found that all other stressed vowels that were measured showed clear deviations in formant trajectories towards a target. Other vowels that did not have very extreme average F2 values (e.g. the NURSE vowel) also did not show a deviation in their F2 trajectory, meaning that this was not a characteristic unique to schwa. It may therefore be that more central vowels like schwa are not displaced enough by their phonetic context that they would show a deviation in their trajectory. This shows the importance of comparing schwa to other vowels when looking for signs of targetlessness. Because of the lack of unanimous deviations in the F2 trajectories of other vowels it was considered that this was not the best method to use to examine this issue.

7.2.2 Vowel duration

The variable of vowel duration was used to examine the behaviour of schwa in both Derby and NZE. The rationale here was that vowels show more influence from their phonetic context at shorter durations, and thus move away from their targets (Lindblom, 1963). If schwa has a target it was hypothesised that it therefore move closer to this target at longer durations. If it has no independent phonetic target then its realisation should not change according to its duration.

It was found that, in both Derby and NZE, longer schwas were lower. This relationship between duration and vowel height was also found for other vowels such as TRAP, which were also much lower at longer durations. In both Derby and in the earlier born speakers of NZE, however, there was no such relationship for KIT. The lack of such a relationship shows that even at short durations KIT is not caused to deviate much from its target height, and suggests that its target height is one that is easy to produce between two consonants, regardless of duration. If schwa is maximally assimilatory to context it is expected to be produced in the way that it is easiest to produce between two contexts. If schwa is targetless it would therefore be expected to occupy a similar height to KIT. Indeed, at shorter durations, where the influence of phonetic context is greater, the height of pre-consonantal schwa is very similar to KIT, in both Derby and NZE speakers. In pre-pausal schwas, where there is no influence of a following context, pre-pausal schwas are much lower than KIT at all durations. However, at longer durations both pre-consonantal and pre-pausal schwa are lower than KIT. These effects were all found in both Derby and NZE. This clearly shows that schwa has a vowel quality target, and that this height is lower than the one that is maximally easiest to produce.

That these effects were found in two different varieties of English using two different corpora, provides very good evidence for the relationship between F1 and duration in schwa, and consequently for schwa having a phonetic target. Similar evidence for a phonetic target in schwa has also been found for American English. Cohen Priva and Strand (2020) use a very similar method for investigating the existence of a phonetic target in schwa using the Buckeye Corpus (Pitt et al, 2007), and also find that longer schwas have a higher F1. They also examine the relationship between duration and F1 in vowels with a range of heights and find the same relationship for most of them. This again provides evidence that schwa is by no means exceptional in being produced with a lower F1 at shorter durations.

The relationship between F2 and schwa was more varied. As in the formant trajectory analysis, duration has an effect on the F2 of pre-pausal schwas in Derby, with the influence of phonetic context diminishing as it gets longer. For pre-consonantal schwas, however, no effect of duration was found in either Derby or in the earlier born NZE speakers. Thus, this does not show clear movement towards an F2 target. However, this situation changes during the period of NZE which was examined. For the later born speakers there is a clear relationship between duration and F2 in non-final schwa, whereby schwas in different phonetic contexts are more similar at longer durations. In the later born speakers, there is therefore clear evidence of an F2 target. This means that evidence of movement to a phonetic target in F2 was found in NZE but not in Derby. This difference between Derby and NZE will be further discussed in section 7.4.

7.2.3 Overall variability of schwa

Another related dimension of schwa that was investigated was its variability and predictability according to context. It was expected that if schwa was a targetless vowel, that it should show extreme variability in F2, as its F2 value would only be determined by its phonetic context. In Derby, compared to most of the other vowels that were examined, pre-consonantal schwa has a high level of F2 variability. However, KIT and FOOT both have a comparable level of F2 variability to schwa. This is perhaps to be expected, given that KIT and FOOT were the two next shortest vowels. This shows that, although variable, schwa is not exceptionally variable when compared to other similar vowels. Likewise, in NZE, KIT and non-final schwa were found to be comparably variable according to context. This again shows the importance of comparing schwa with other vowels.

On the other hand, the F1 variability of schwa was found to be larger than that of KIT in NZE, and larger than both KIT and FOOT in Derby. As argued in section 2.4.3, a large degree of F1 variability in schwa is not actually suggestive of targetlessness, as maximal assimilation to context should on average cause low F1 values, when in between two consonants. The large F1 variability of schwa shows that, although on average its F1 values may be similar to KIT and FOOT, when it is less influenced by coarticulatory pressures (e.g. at longer durations) it reaches higher F1 values. Therefore, although this large F1 variability of schwa was also found in Bates (1995), I argue that this suggests schwa has a target, not that it is targetless.

7.2.4 Change in schwa over time

In addition to looking at patterns of synchronic variation in schwa, a key part of this thesis was also looking at the behaviour of schwa diachronically. This was in order to see whether the production of schwa was affected by the position of other vowels within the vowel space. If schwa shows overall change over time (i.e. in a way that cannot be explained by changes in its phonetic context) then this points to it having a phonetic target. If schwa is maximally assimilatory and has no independent phonetic target then its positon in the vowel space should not change over time, and it certainly should not show any relationship with other vowels. Schwa has previously been suggested to be a neutral vowel (e.g. Browman and Goldstein, 1992) and has also been suggested not to be subject to the same dispersion constraints as other vowels, meaning that it does not influence and is not influenced by other vowels (Schwartz et al, 1997). NZE was chosen to test these ideas about schwa, as changes in the KIT vowel in NZE towards a lower and more central area of the vowel space (towards the position of schwa) made it a suitable variety to test these claims.

It was found that schwa did indeed undergo change over time. Pre-pausal schwa became slightly backer over the course of the period, although this change was fairly subtle. It was also found that the variation according to phonetic context in non-final schwa reduced over time. In terms of vowel height, however, both pre-pausal and non-final schwa clearly became much lower, following the same trajectory as KIT. We would not expect the realisation of a phonetically targetless vowel to vary in any way which could not be predicted from phonetic context. These results suggest that schwa can interact with the positon of other vowels, and as such is not phonetically targetless. In addition, as discussed in section 2.4.3, we would expect that a maximally assimilatory vowel that is coarticulated with adjacent consonants should be produced as a fairly high vowel, so the lowering of schwa over time clearly suggests that it is not a maximally assimilatory vowel.

The fact that schwa appears to be implicated in the ‘front vowel shift’ in NZE also clearly goes against the idea that it does not interact with other vowels, and that it is phonetically targetless. Importantly, it contradicts the idea espoused by Flemming (2009) that the vowel quality of schwa would only be affected by other contrasting unstressed vowel qualities. More generally, it goes against the idea that schwa can be discounted in theories of sound change, or that unstressed vowels should be systematically ignored when analysing vowel change over time or chain shifts. For example, Labov’s (1994) theory of vowel change describes vowels as operating within two subsystems, tense vowels and lax vowels, which are said to change together. However, schwa is not mentioned so it is not clear how it would fit into this theory, or whether it is simply not expected to undergo change at all.

Labov’s (1994) theory is that, in vowel chain shifts, tense vowels rise along a peripheral track, and that lax vowels fall along a non-peripheral track. Langstrof (2006) argues that this theory does not work for the front vowel shift in NZE, since the phonologically short vowels

TRAP and DRESS have been shown to raise. Langstrof therefore argues against the need for concepts such as ‘subsystems’ and tracks in explaining vowel shifts. The fact that an unstressed vowel also appears to also be implicated in this vowel shift is further evidence from NZE against the division of vowels into subsystems, since it suggests that stressed and unstressed vowels can be involved in the same vowel shift. Instead, Langstrof argues that the likelihood of vowels interacting together in a change can simply be explained by their similarity along various phonetic dimensions, including vowel height and backness, but also duration and diphthongisation. Despite schwa being unstressed and KIT being stressed, they are fairly similar on a number of dimensions. They are both of short durations compared to other vowels, and schwa, like KIT, can often be produced with quite a high realisation, and both also tend to be produced with a fairly low intensity (Watson et al, 2000). In this way then, despite schwa being unstressed and KIT being stressed it is not that surprising that schwa should change with KIT.

The other important aspect of change in schwa was in how the effect of duration changes over time. One particular conception of targetlessness in schwa (e.g. Bates, 1995) is that when the coarticulatory effects of phonetic context are less strong, schwa may show a different type of behaviour. Bates argues that lower schwa realisations in schwas of longer durations such as in pre-pausal position are still consistent with it being a vowel produced with minimum effort. Rather than being maximally assimilated to its phonetic environment, in such contexts the tongue will move towards its natural rest position. There is, however, no reason to expect the natural rest position of schwa to change over time. The results from Chapter 6 therefore also go against this conception of targetlessness. Pre-pausal schwas become lower over time, which is unexpected if the tongue simply moves to a rest positon. In addition, the effect of duration on the height of both pre-pausal and non-final schwa actually gets stronger over time. This means that schwas of the same durations get lower over the period. Again, this does not suggest that schwa is simply moving towards a rest position at long durations. If this was the case, when duration is held constant, schwas would not be expected to show any clear change.

The effect of duration on schwa is also particularly interesting in NZE, as on average by the end of the period KIT and non-final schwa are very similar. However, the differing effect of duration on KIT and schwa mean that they are more different at longer durations. This therefore exemplifies the fact that vowels that appear to be very similar on the surface may show quite different patterns of variation.

Overall, the changes found in schwa both provide evidence that it is not phonetically targetless, and also that it can be involved in vowel changes, so it should not be automatically neglected when looking at vowel shifts.

7.3 Word position differences

It has long been noted that word final schwa tends to have a lower vowel quality than non-final schwa (Jones, 1914; Gimson, 1962; Lass, 1986), and more recently this difference has been verified in phonetic studies of schwa (Flemming and Johnson, 2007; Lilley, 2012, Bekker, 2014). Flemming and Johnson (2007) also note a division in terms of backness with non-final vowels being more variable in backness and word final schwa being more central. These findings somewhat accord with what was found in this thesis. In both the data from Derby and NZE, when before a pause, word final schwa was found to be much lower than non-final schwa, and also less variable in backness. Of course, pre-pausal schwas are not in the same phonetic environment as non-final schwas as they are followed by silence rather than a consonant. In the Derby analysis, when word final schwas that occurred in the same phonetic environment as non-final schwa were examined, that is before another consonant (e.g. *number five, formula one*), this difference was much less stark. These word final schwas showed a very similar level of variability and predictability in F2 to non-final schwa, although they were slightly lower. The height difference between them was, however, fairly minimal and pre-consonantal word final schwas were actually much more similar to non-final schwa than pre-pausal schwa, both in their overall vowel quality, and their behaviour. Whilst the F2 values of pre-pausal schwa varied across its trajectory and according to duration, neither non-final schwa or pre-consonantal word final schwa showed such effects. In addition, pre-pausal schwa had a much stronger deviation in its F1 trajectory than either non-final or pre-consonantal word final schwa.

It is argued, therefore, that although there is an overall difference between schwa word finally and non-finally, at least in the varieties examined here, this difference is primarily caused by phonetic differences in the environments of the two vowels. When word final and non-final schwa occur in the same environment they are actually very similar. The differences between pre-pausal schwas and pre-consonantal schwa is likely partly due to the fact that it is longer than non-final schwa due to phrase final lengthening. This, however, does not completely explain the differences between pre-pausal and non-final schwa, as in both the Derby and the NZE data pre-pausal schwas and non-final schwas of the same duration were still different in vowel height. The difference can also be attributed to the fact that pre-pausal schwa does not have a following consonant so is less liable to coarticulatory influence. This means that the difference may be exaggerated in studies such as Bekker (2014), as Bekker only examines words in citation form, meaning all word final schwas were pre-pausal.

Although word final schwa was much more similar overall to non-final schwa, there were still small differences between pre-consonantal word final schwa and non-final schwa. However, I do not believe that the small differences found here justify the idea that word final and non-final schwa are two fundamentally distinct categories. These small differences may be due to word specific effects (as found for example in Sóskuthy et al, 2018). This means that as word final schwas (but not non-final schwas) are sometimes produced pre-pausally, even when they are not produced pre-pausally they will still be somewhat affected by the fact that they sometimes occur in pre-pausal position. Similarly, Silverman (2011) suggests that differences between non-final and word final schwa could be attributable to the fact that within word motor routines are better practised than those that occur across word boundaries.

The suggestion that word final schwa and non-final schwa are two distinct categories (Flemming, 2009; Lass, 2009) may therefore be inaccurate. Although there are word position differences these are mostly attributable to phonetic context. These differences are also similar to what has been found in the happY vowel, so this difference is not unique to schwa. Ramsammy and Turton (2012) find that in speakers from Manchester the happY vowel is also much lower in pre-pausal contexts. The vast majority of commentary on dialectal differences in schwa has tended to refer to word final rather than non-final schwa.

This is likely because they can occur in pre-pausal context, where schwa is more noticeable due to its longer duration, and phrase final position. However, the chapter on NZE showed that although pre-pausal and non-final schwa are on average quite different in height they still change in the same way over time and according to duration. This therefore has the implication that such future studies should consider including non-final schwa in their analyses as well as word final schwa.

7.4 Differences between Derby and NZE

As discussed in section 7.2.2, the findings for F1 were similar across the Derby and NZE corpora. Both NZE and Derby have a clear difference in vowel height between pre-pausal and non-final schwa. Both also have a clear relationship between duration and F1, whereby F1 is higher at longer durations. However, the findings for F2 were slightly different. For non-final schwa an effect of duration on F2 was only found in NZE and not in Derby. In the later born speakers in NZE there existed a relationship whereby at longer durations schwas of different phonetic contexts were closer in F2 value than they were at shorter durations.

Flemming (2009) suggests that schwa could possibly have a very weakly specified target, but surface as very variable because of the lack of vowels it needs to contrast with. This line of reasoning suggests that the reason we see direct evidence of an F2 phonetic target in the later speakers in NZE but not in the earlier born speakers or in the Derby speakers could be because the changed positon of KIT causes schwa to have more of a strongly specified target. However, in section 7.2.2 it was argued that despite the fact that there is no relationship between duration and F2 in early NZE speakers and Derby speakers, schwa still shows clear evidence of a phonetic target. Therefore, it is unlikely that schwa somehow developed a more strongly specified target over time in NZE.

Instead, these dialectal differences can be explained in terms of the relationship between height and backness, which schwa was predicted to show if it has a target (see section 2.4.3 p.37-8 and section 4.2, p.98). To briefly recall this argument, it is expected that in a targeted schwa higher schwas would also be more variable in backness, as increased coarticulation causes both a smaller F1 and more variable F2. This fact can help explain the differences between the Derby and NZE data. As schwa gets lower over time in NZE, it makes sense that its contextual predictability and variability in F2 would decrease. This relationship between F1 and F2 is also found elsewhere in the data. For example, in both NZE and Derby, pre-pausal schwa is both lower than non-final schwa and is less variable in backness. In Derby, non-final schwas that are lower are also less variable in F2. Therefore, the difference in the relationship between F2 and duration across the corpora can be explained in terms of the more general relationship between vowel height and backness in schwa.

7.5 Summary of evidence of targeted schwa

The main question of this thesis was whether schwa has phonetic target. Throughout the thesis, this issue has been investigated in a number of different ways across two different varieties. In terms of vowel height, it is clear that schwa has target. In Derby and NZE, both pre-pausal schwa and pre-consonantal schwa show a movement towards a lower vowel quality at longer durations. Importantly, all types of schwa are shown to target a height which is lower than the minimum requirement for a vowel.

We also see clear evidence of a target on the F2 dimension. For pre-pausal schwa in Derby, the effect of phonetic context reduces both through the trajectory of schwa, and also when it is at longer durations. In later born speakers in NZE, there is also evidence of an F2 target in pre-consonantal schwa, as longer schwas are less affected by phonetic context.

Change in schwa over time also clearly points to it having a phonetic target, as a targetless vowel is not expected to vary according to anything other than phonetic context. Changes in NZE show that schwa clearly becomes lower over time. This change happens in all environments, including environments where there is less influence from consonantal context. This change happen in both pre-pausal schwa and pre-consonantal schwa, and is shown even when duration is held constant. This, in particular, shows that schwa has an independent phonetic target, and does not simply move towards a rest position when it is less influenced by phonetic context.

Finally, F1 and F2 do not vary independently of each other, and the relationship between them here is exactly as was predicted if schwa has a target. We find that where schwa is lower it also varies less according to phonetic context. This is shown in the changes over time in NZE, as schwa gets lower and less variable according to phonetic context over time. It is also shown in the relationship between pre-consonantal schwa and pre-pausal schwa overall, as pre-pausal schwa is both lower and less variable according to phonetic context than pre-consonantal schwa.

Altogether, it is clear that multiple lines of evidence from two different varieties all point towards schwa having a phonetic target.

7.6 Why is schwa so short?

In this thesis, it has been argued that schwa has a phonetic target, despite the fact that pre-consonantal schwa is often relatively variable in F2 and has a fairly high average realisation. In particular, I have argued that its variability in height according to duration and the way that this relationship changes over time in NZE shows evidence of a phonetic target.

Clearly, variation according to duration is normal, and is something that has been shown in other vowels both within this thesis and in many other studies (e.g. Moon and Lindblom, 1994; Cohen Priva and Strand, 2020). There has been a clear distinction made by some between phonetic and phonological reduction (Fourakis, 1991; Van Bergem, 1993; referred to as acoustic and lexical reduction by Van Bergem). Phonetic reduction is the process whereby vowels move further away from their phonetic target according to factors such as duration, context, effort and stress, and is a gradient process which affects all vowels. Phonological vowel reduction, on the other hand, is where the vowel target is a reduced vowel quality which only occurs in unstressed syllables i.e. schwa. However, this does not mean that schwa itself is a vowel of maximum phonetic reduction or targetlessness, but it can still undergo phonetic reduction in the way that other vowels also do. It has been argued in this thesis that schwa has a phonetic target, but that its often low F1 values and relatively high variability in F2 are caused by phonetic reduction due to factors such as short durations and the effects of its phonetic context. It is therefore unsurprising that schwa would have a low F1 and variable F2 at shorter durations, and compared to stressed vowels of similarly short durations it was not found to be exceptionally variable in F2. Where schwa is of a longer duration, including in pre-pausal position, it is both a lower vowel and is less variable in F2. The variation that is seen in schwa is therefore unexceptional and can be explained by the same phonetic reduction processes that affect all vowels. The often short duration of non-final schwa can explain why it is often produced in a way that has caused others to analyse it as a targetless vowel.

However, this analysis cannot completely explain the phonetic variation that we find within schwa, as it does not tell us why it is so often produced with short durations. If schwa has a vowel quality target which is lower than the vowels /ɪ/ and /ʊ/, then the fact it so often has a short duration which results in higher realisations needs explanation. This is especially the case in NZE, as the short duration of schwa causes it to be realised more similarly to the encroaching KIT vowel than it would be if it was produced at longer durations. Flemming's (2009) analysis is that non-final schwa is very variable in backness and is high due to its short duration and lack of contrast with other unstressed vowel qualities. As non-final schwa does not minimally contrast with other vowel qualities it is seen as being produced with less effort. As discussed in Chapter 2, however, even if schwa does not have to contrast with other unstressed vowels, it still has to contrast with stressed vowels. Indeed, it may be that the short duration of schwa itself is an important cue used in order to distinguish it from other vowels.

Given that schwa appears to have a phonetic target, but that its often short duration mean that it often falls short of this target, this suggests that short durations are themselves likely to be an important cue which is used to distinguish schwa from stressed vowels. Studies such as Fry (1955) have shown that modulations in duration alone can cause listeners perception of stress placement to change, so it is clear that duration is used to help mark stress within a word.

The focus of this thesis was primarily to examine variation in the vowel quality of schwa and to see whether that showed evidence of a phonetic target, which it did. It is, however, important to remember that vowel quality is not the only cue that can be used to differentiate vowels from one another, and it is known that vowels are produced with a combination of cues (Holt and Lotto, 2010). Kirby (2010) discusses the issue of how different cues may be useful in aiding contrasts in different contexts. For example, Nokes and Hay (2012) find that in NZE the pitch variability of females has increased over time, and for males intensity variability has increased. It is possible that this increased variability could have been as a result of KIT encroaching on schwa. However, as they only provide variability statistics over all vowels this is only speculation.

This thesis has only provided production data so we cannot know the degree to which cues such as vowel quality and duration are used to help speakers distinguish schwa from other vowels. However, the fact that schwa often has a short duration when not pre-pausal suggests that short duration could be an important cue for schwa. This short duration can therefore indeed be explained without recourse to the idea that schwa is produced with a lack of effort (e.g. Flemming, 2009). Regardless of the role of duration, this does not change the fact that schwa has been shown to have a vowel quality target. However, the effects of its short duration on its vowel quality can explain why it is often seen as a targetless vowel.

7.7 Future questions

Throughout the chapters in these thesis I have looked at schwa from a number of angles, using findings from data from two different dialects of English. A number of specific avenues for future research from these findings will be discussed below. However, the key argument at the heart of the thesis is the fact that schwa has a phonetic target. In terms of the potential for future research this is particularly important. That schwa is not completely predictable from phonetic context and has a phonetic target means that it is a potential source of interesting variation, in much the same way that other vowels are. This therefore has the broad implication that there should be more consideration of unstressed vowels in studies of language variation and change. In particular, in studies that look at a range of vowels, schwa should not necessarily be systematically excluded in the way that has often been seen in the past.

In Chapter 3 it was shown that unstressed vowels can be a source of interesting individual variation, in terms of the way that different speakers produced unstressed vowels with different spellings. Findings elsewhere in the thesis could also be elucidated with examination of the behaviour of individual speakers. For example, in Chapter 6 it was shown that schwa can change over time, and that it appears to be implicated in the NZE front vowel shift. Comparing the realisation of KIT and schwa in individual speakers would provide further insight into the relationship between these two vowels in NZE. If it could be shown that there is a relationship between the degree of lowering in KIT and schwa across individual speakers than this would provide further evidence that schwa is implicated in the NZE front vowel shift.

As the thesis solely used production data, additional research could make use of perception data. For example, Chapter 3 showed that there are small differences in the production of unstressed vowels spelt with <a>, <e> and <i>. It would be of interest to discover to what extent speakers are aware of these differences and whether these small differences are used to distinguish words. Additionally, the findings regarding NZE in Chapter 6 lead to further questions about how KIT and schwa are distinguished in NZE. It was shown that both KIT and schwa change in similar ways over time, suggesting a relationship between the two vowels. Despite this, the two vowels are very similar by the end of the period examined, only being more different to each other at longer durations. This thus poses the question of how KIT is distinguished from unstressed syllables in NZE, and to what extent cues other than vowel quality may be used in this regard.

It would also be valuable to see whether similar findings to those found in Derby and NZE could also be found in other dialects. For example, it would be of interest to see whether the spelling differences found in Derby also exist in other dialects, and in particular, whether these subtle differences could still be found in varieties that are not said to distinguish between unstressed KIT and schwa e.g. New Zealand English (Gordon et al, 2004). Additionally, since it has been argued in this thesis that non-final and word final schwa are fundamentally the same vowel occurring in different environments, it would be interesting to see further studies that examine non-final schwa as well as word final schwa. In particular, where particular dialectal variants of word final schwa have been reported e.g. in

Manchester (Beal, 2008, Ramsammy and Turton, 2012), it would be interesting to see if these variants also exist in non-final position, when duration is considered.

More generally, the findings regarding NZE lead to the question of whether schwa is also implicated in other types of vowel changes. Schwa is of course a very common vowel in English so it is not hard to find tokens of it in natural speech. Moreover, despite its short duration, the analysis in Chapter 5 showed that it can be analysed using automated methods with comparable results to using more time consuming manual segmentation and measurement. There is therefore no reason why it should not be given more attention in research on language variation and change.

7.8 Summary

In this chapter the findings from the analysis chapters were brought together in order to answer the broader research questions in the thesis. It was shown that, in Derby at least, schwa is not the only vowel that can occur in unstressed syllables, and that it is distinct from /ɪ/. It was also argued that it may be that not all vowel quality differences are fully neutralised in unstressed syllables. Word final schwa and non-final schwa are argued to not be phonologically distinct, and instead most of the differences between them can be explained in terms of phonetic environment. The main argument presented is that schwa has a phonetic target, as evidenced in the fact it changes over time, and that its vowel quality is affected by duration.

8 Conclusion

This thesis has examined phonetic variation and change in schwa in English. The effect of a wide range of variables on unstressed vowels have been examined in order to answer the central questions posed in Chapter 1.

The central question of this thesis was whether evidence of a phonetic target in schwa could be found. Although this question has previously been investigated, these studies have used elicited speech, with schwa in a narrow range of words and contexts (Browman and Goldstein, 1992; Kondo, 1994; Flemming, 2009). This thesis thus filled a gap by examining these issues in spontaneous speech. This allowed for the examination of the realisation of schwa in a much wider range of contexts than these previous studies. The use of two large corpora meant that this investigation included far more speakers and far more data than these previous studies of targetlessness in schwa. The aforementioned studies were focussed quite heavily on the predictability of schwa, which, as discussed in section 2.4.3 and 2.4.4, is not alone an adequate determiner of targetlessness. In addition to looking at predictability, this thesis has examined the issue of targetlessness from a wider range of angles. This included examining variation in vowels according to the naturally varying durations in spontaneous speech, and examining how schwa changes over time. This thesis also made use of comparisons with other vowels in order to assess the targetlessness of schwa.

Like previous studies which have investigated the issue of targetlessness, I analysed the variability of schwa according to phonetic context. Like in other studies, schwa was found to be highly variable in both F1 and F2. However, as explained in section 2.4.3, high F1 variability is not actually indicative of targetlessness. The high F1 variability that was found in schwa throughout (shown in Chapters 3, 4 and 6) was argued to indicate a target in schwa, as it shows that schwa is very often not produced with the low F1 that is expected in a maximally assimilatory vowel. The variability in the F2 of schwa and its predictability according to phonetic context was also found to be great. However, it was often not notably higher than other short phonetically high vowels (Chapters 4 and 6), so this variability was not found to be suggestive of targetlessness.

The way in which schwa varied according across its trajectory was also examined (Chapter 4). Similarly to other studies (e.g. Kondo, 1994, Flemming, 2009) deviations were found in the F1 trajectories, but the F2 trajectories of pre-consonantal schwas were linear. The results found in this part of the analysis provided no clear evidence of either a target or targetlessness in pre-consonantal schwa. Whilst no clear evidence of a target was found, the methods used also did not show clear deviations in the formant trajectories of all stressed vowels. This showed the importance of comparing schwa with other vowels. Previously, where this has not been the case, claims of targetlessness in schwa may have been premature. It was therefore argued that, at least in spontaneous speech, this method was not suitable for assessing targetlessness.

In both the Derby and NZE analyses (Chapters 4 and 6) the way in which schwa varied according to duration was investigated. Both of these analyses strongly suggested that schwa has a target. The average F1 value of pre-consonantal schwa was found to be representative of a fairly high vowel in both varieties. However, in both varieties, schwas that were longer were found to be lower, suggesting that at longer durations schwas move towards a lower height target. This, again, suggested that schwa is not a maximally assimilatory vowel. Crucially, comparisons with KIT showed that schwa has a target which is lower than /ɪ/, and therefore it has a height target which is lower than the minimum required to be perceived as a vowel.

On the other hand, the relationship between F2 and duration was less consistent between Derby and NZE. In Derby and the earlier born NZE speakers there was no relationship between F2 and duration in pre-consonantal schwa. However, in the later born speakers of NZE it was found that longer schwas were less variable in F2, and also differed less according to phonetic context. This therefore showed evidence of movement to a phonetic target at longer durations. This discrepancy between the F2 results across the corpora was explained in terms of a general relationship between F1 and F2 (see sections 4.2, p.98 and 7.4). To summarise this argument, in a targeted schwa higher schwas are also expected to be more variable in backness, due to increased coarticulation causing both a lower F1 and more variable F2. As schwas were found to become lower over time in NZE, it therefore makes sense that they also become less variable in F2. The effect of duration on schwa strongly suggested that it has a phonetic target. Since the average duration of pre-consonantal schwa is, however, generally very short, this explains why it is often produced in a way that has caused others (e.g. Kondo, 1994; Bates, 1995) to analyse it as a targetless vowel.

The final way in which targetlessness was assessed was through examining changes in schwa over time in NZE (Chapter 6). It was found that, as KIT moves towards the position of canonical schwa, schwa itself also changes over time in NZE. These results strongly suggest a target in schwa in NZE, as a targetless schwa would not be expected to show change over time. The changes found in schwa over time both suggest that it is not fully dependent on phonetic context for its realisation, and also that it can interact with the position of other vowels. Therefore, through the combination of various different methods of analysis it has been shown that schwa has a phonetic target, at least in Derby and NZE.

The second related question that was addressed was whether there is a phonological difference between word final and non-final schwa. It was found that there was a large phonetic difference overall between word final and non-final schwa (Chapters 4 and 6). In general, word final schwa was a much lower vowel, and was less variable in F2. It was also a longer vowel. When in pre-pausal position, word final schwa also generally showed clearer evidence of movement towards a phonetic target according to its formant trajectories and duration than non-final schwa. However, analysis of word final schwas in both pre-consonantal and pre-pausal environments indicated that the primary division between schwas was actually based on phonetic context (i.e. whether they were pre-pausal or pre-consonantal), rather than word position. Word final pre-consonantal schwa actually patterned with non-final schwa, rather than pre-pausal schwa. Therefore, it was argued that word final and non-final schwa are not phonologically different, but are simply produced differently on average because of the effects of phonetic environment.

The final question which was addressed was whether schwa is distinct from other unstressed vowel qualities. In Chapter 3, it was shown that there are clear differences between different unstressed vowel qualities, as could be seen through the production of unstressed vowel represented with different spellings. However, there was a high level of variability between speakers in the way that they produced unstressed vowels spelt with <e>, which suggested that the difference between /ə/ and /ɪ/ was not necessarily straightforward, and that there can be a great deal of interspeaker variation in the way that these vowels pattern.

In addition to providing answers to the central research questions posed in Chapter 1, this thesis has contributed more generally to the understanding of how unstressed vowels in English are produced in natural speech. This thesis has shown that, like other vowels, unstressed vowels can be an interesting source of variation across individual speakers. Although only one aspect of variation was explored at the individual level (spelling), these findings suggested that this could be an area for further exploration. In addition, despite the overwhelming focus on stressed vowels in variation and change research, I have shown that unstressed vowels can also change over time (Chapter 6), and therefore should not be so readily discarded when investigating vowel changes. Indeed, Chapter 5 showed that, with appropriate data filtering and outlier removal, unstressed vowels can be successfully analysed using automated methods.

It is important to stress that the findings in this thesis are only from two varieties of English, and therefore not all of these findings will necessarily apply to other varieties of English. However, certain findings were similar across both NZE and Derby. This included the relationship between vowel height and duration, and the overall phonetic differences between pre-pausal schwa and pre-consonantal schwa. This suggests that these findings, at least, may be expected to also be found more widely. It would be of interest to see the extent to which the patterns found in this thesis are true in other varieties of English. For example it would be of interest to see whether, in varieties with a more salient production of word final schwa, (e.g. Manchester; see Beal, 2008; Ramsammy and Turton, 2012) the primary distinction between schwa is also between pre-pausal and pre-consonantal schwas, or whether such varieties may have a more robust distinction between word final and non-final schwa. It would also be of interest to see whether schwa can be shown to undergo any other changes over time in varieties other than NZE.

In summary, this thesis has explored phonetic variation and change in unstressed in spontaneous speech in English. In doing so, I have shown that unstressed vowels vowel vary and change in similar ways to other vowels. Whilst the production of schwa can on the surface look extremely variable, when other variables are considered it is clear that schwa is not targetless. Using a wide range of analyses and variables, and across two varieties of English, I have provided clear evidence that English schwa has a phonetic target.

Appendix-Model summaries

This appendix contains the model summaries for Chapters 4 and 6. Before each summary a short descriptor of the model is listed, followed by the levels for the intercepts, and the model formula.

Model 4.1

Effect of phonetic context on F2 of non-final schwa

Reference levels: preceding context=central, following context=central

F2~Preceding context+Following context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	10377	101.9
speaker	(Intercept)	14534	120.6
Residual		39532	198.8

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1787.74	26.69	66.98
Preceding.back	-168.59	13.95	-12.083
Preceding.front	157.43	19.76	7.966
Following.back	-195.74	18.67	-10.484
Following.front	182.03	29.67	6.135

Model 4.2

Effect of phonetic context on F2 of pre-consonantal word final schwa

Reference levels: preceding context=central, following context=central

F2~Preceding context+Following context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	12688	112.6
speaker	(Intercept)	13551	116.4
Residual		34260	185.1

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1729.39	27.84	62.12
Preceding.back	-134.24	24.11	-5.569
Preceding.front	155.51	36.66	4.242
Following.back	-138.96	12.98	-10.709
Following.front	116.63	21.57	5.406

Model 4.3

Effect of phonetic context on F2 of pre-pausal schwa

Reference levels: preceding context=central, following context=central

F2~Preceding context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	5539	74.42
speaker	(Intercept)	14373	119.89
Residual		14694	121.22

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1603.73	25.66	62.498
Preceding.back	-119.73	21.31	-5.618
Preceding.front	117.89	29.02	4.063

Model 4.4

Effect of phonetic context on F2 of KIT

Reference levels: preceding context=central, following context=central

F2~Preceding context+Following context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	20541	143.3
speaker	(Intercept)	52009	228.1
Residual		25248	158.9

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1957.8	90.28	21.685
Preceding.back	-58.84	41.91	-1.404
Preceding.front	79.25	73.75	1.075
Following.back	-156.91	37.58	-4.176
Following.front	130.94	53.36	2.454

Model 4.5

Effect of phonetic context on F2 of FOOT

Reference levels: preceding context=central, following context=central

F2~Preceding context+Following context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	0	0
speaker	(Intercept)	11008	104.9
Residual		39401	198.5

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1336.02	61.32	21.788
Preceding.back	-149.34	43.56	-3.428
Preceding.front	232.39	56.34	4.125
Following.back	-128.5	39.54	-3.25
Following.front	16.57	49.72	0.333

Model 4.6

Effect of phonetic context on F2 of NURSE

Reference levels: preceding context=central, following context=central

F2~Preceding context+Following context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	2143	46.29
speaker	(Intercept)	35532	188.5
Residual		5552	74.51

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1774.52	68.97	25.729
Preceding.back	-83.92	22.1	-3.797
Preceding.front	25.31	38.68	0.654
Following.back	-50.12	24.64	-2.034
Following.front	-13.19	29.78	-0.443

Model 4.7

Effect of phonetic context on F2 of THOUGHT

Reference levels: preceding context=central, following context=central

F2~Preceding context+Following context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	6352	79.7
speaker	(Intercept)	7603	87.2
Residual		15441	124.3

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1077.37	44.12	24.417
Preceding.back	-133.39	32.66	-4.084
Preceding.front	-38.89	41.65	-0.934
Following.back	32.93	28.6	1.152
Following.front	37.64	37.59	1.001

Model 4.8

Effect of phonetic context on F2 of TRAP

Reference levels: preceding context=central, following context=central

F2~Preceding context+Following context+(1|speaker)+(1|word)

Random effects

Groups	Name	Variance	Std.Dev.
word	(Intercept)	1045	32.32
speaker	(Intercept)	17025	130.48
Residual		12654	112.49

Fixed effects

	Estimate	Std. Error	*t*-value
(Intercept)	1500.738	48.082	31.212
Preceding.back	-81.282	20.504	-3.964
Preceding.front	79.589	25.833	3.081
Following.back	7.858	19.489	0.403
Following.front	127.575	32.777	3.892

Model 4.9

F1 formant trajectory of non-final schwa

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	452.653	9.186	49.27	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	6.525	7.592	11.676	<0.0001
s(speaker)	27.466	30	15.633	<0.0001
s(word)	104.672	427	0.577	<0.0001
s(measurement number,token)	5264.362	6054	822.537	<0.0001

Model 4.10

F1 formant trajectory of pre-consonantal word final schwa

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	472.58	10.36	45.62	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	5.557	6.645	5.725	<0.0001
s(speaker)	27.619	30	15.817	<0.0001
s(word)	67.807	237	0.661	<0.0001
s(measurement number,token)	3944.325	4518	567.195	<0.0001

Model 4.11

F1 formant trajectory of pre-pausal schwa

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	563.2	17.5	32.19	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	5.88	6.979	21.038	<0.0001
s(speaker)	27.1	30	9.876	<0.0001
s(word)	71	169	0	<0.0001
s(measurement number,token)	114	1386	161.755	0.00124

Model 4.12

F1 formant trajectory of KIT

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	428.31	14.32	29.92	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	5.589	6.679	6.207	<0.0001
s(speaker)	6.605	7	22.746	<0.0001
s(word)	20.139	125	0.24	<0.0001

s(measurement number,token)	1117.286	1302	154.711	<0.0001

Model 4.13

F1 formant trajectory of FOOT

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	455.23	24.43	18.63	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	6.043	7.137	5.365	<0.0001
s(speaker)	6.493	7	21.202	<0.0001
s(word)	19.739	55	0.793	<0.0001
s(measurement number ,token)	565.342	714	91.075	<0.0001

Model 4.14

F1 formant trajectory of NURSE

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	488.8	21.5	22.73	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	6.211	7.299	8.67	<0.0001
s(speaker)	6.726	7	26.39	<0.0001

s(word)	6.437	52	0.164	<0.0001
s(measurement number, token)	384.969	474	52.611	<0.0001

Model 4.15

F1 formant trajectory of THOUGHT

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	490.97	23.82	20.61	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	5.056	6.111	8.451	<0.0001
s(speaker)	6.755	7	41.773	<0.0001
s(word)	20.673	75	0.485	<0.0001
s(measurement number,token)	577.975	770	23.514	<0.0001

Model 4.16

F1 formant trajectory of TRAP

F1 ~ s(measurement number, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(measurement number, token, bs = "fs", k = 4, xt = "cr”,m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	693.13	37.37	18.55	<2e-16

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	6.216	7.305	17.152	<0.0001

s(speaker)	7.011	8	12.527	<0.0001
s(word)	6.369	111	0.068	<0.0001
s(measurement number, token)	696.797	847	361.505	0.0115

Model 4.17

F2 trajectories across vowels in back contexts

Reference level: vowel=non-final schwa

F2 ~ Vowel+s(measurement number, bs="cr")+s(measurement number, by=Vowel, bs="cr") + s(speaker, bs="re") + s(word, bs="re")+s(speaker,vowel, bs="re")+s(measurement number, token, bs="fs", xt="cr",k=4, m=1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1390	30.95	44.917	<0.0001
Vowel.Pre-consonantal final schwa	61.45	37.11	1.656	0.0978
Vowel.KIT	327.41	53.84	6.081	<0.0001
Vowel. FOOT	-296.57	65.48	-4.529	<0.0001
Vowel. NURSE	242.51	96.76	2.506	0.0122
Vowel. THOUGHT	-386.8	66.67	-5.801	<0.0001
Vowel. TRAP	15.09	56.75	0.266	0.79

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	1	1	0.531	0.466
s(measurement number):Vowel.Pre-consonantal final schwa	2.115	2.57	0.47	0.673
s(measurement number):Vowel. KIT	6.103	7.194	9.517	<0.0001
s(measurement number):Vowel. FOOT	6.276	7.36	8.546	<0.0001
s(measurement number):Vowel. NURSE	1.001	1.002	0.022	0.882
s(measurement number):Vowel. THOUGHT	8.378	8.872	45.618	<0.0001
s(measurement number):Vowel. TRAP	4.787	5.82	2.945	0.00806
s(speaker)	18.057	30	6.539	<0.0001
s(word)	94.845	248	0.848	<0.0001
s(Vowel,speaker)	41.841	91	1.434	<0.0001
s(measurement number,token)	277.827	2718	3469.308	<0.0001

Model 4.18

F2 trajectories across vowels in central contexts

Reference level: vowel=non-final schwa

F2 ~ Vowel+s(measurement number, bs="cr")+s(measurement number,by=Vowel, bs="cr") + s(speaker, bs="re") + s(word, bs="re")+s(speaker,vowel, bs="re")+s(measurement number, token, bs="fs", xt="cr",k=4, m=1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1742.68	29.62	58.827	<0.0001
Vowel. Pre-consonantal word final schwa	-47.05	25.71	-1.83	<0.0001
Vowel.KIT	172.81	44.52	3.882	<0.0001
Vowel. FOOT	-427.41	75.9	-5.631	<0.0001
Vowel.NURSE	44.97	45.42	0.99	0.322
Vowel. THOUGHT	-567.58	51.89	-10.937	<0.0001
Vowel.TRAP	-225.11	49.38	-4.559	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	1	1	2.441	0.118
s(measurement number):Vowel.final schwa	4.098	5.037	2.06	0.0678
s(measurement number):Vowel.KIT	1.001	1.002	4.271	0.0388
s(measurement number):Vowel.FOOT	4.392	5.375	4.947	<0.001
s(measurement number):Vowel.NURSE	2.36	2.888	1.831	0.134
s(measurement number):Vowel. THOUGHT	7.279	8.226	21.673	<0.0001
s(measurement number):Vowel.TRAP	2.615	3.221	1.566	0.205
s(speaker)	26.605	30	19.914	<0.0001
s(token)	71.087	177	1.326	<0.0001
s(Vowel,speaker)	11.523	89	0.194	<0.0001
s(measurement number,token)	1749.506	2122	1736.022	<0.0001

Model 4.19

F2 trajectories by phonetic context in pre-pausal schwa

Reference level: preceding context=central

F2 ~ preceding context + s(measurement number, bs = "cr")
+ s(measurement number, by = preceding context, bs = "cr") + s(speaker, bs = "re")
+ s(word, bs = "re")
+ s(speaker, preceding context, bs = "re")
+ s(measurement number, token, bs = "fs", xt = "cr", k = 4, m = 1)

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1610.18	24.85	64.798	<0.0001
Preceding context.back	-128.75	22.43	-5.739	<0.0001
Preceding context.front	137.46	29.49	4.66	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(measurement number)	1	1	0.251	0.617
s(measurement number):preceding context.back	1	1.001	21.512	<0.0001
s(measurement number):preceding context.front	4.108	5.045	7.858	<0.0001
s(speaker)	25.448	30	14.03	<0.0001
s(word)	74.449	167	1.249	<0.0001
s(preceding context, speaker)	15.713	78	0.383	<0.0001
s(measurement number,token)	1068.515	1382	408.801	<0.0001

Model 4.20

The effect of duration on F1 compared across all vowels

Reference level: vowel=non-final schwa

F1 ~ Vowel + s(duration, bs = "cr") + s(duration, by = Vowel,
bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") +
s(speaker, Vowel, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	464.229	11.342	40.93	<0.0001
Vowel.Pre-consonantal final schwa	25.012	8.923	2.803	0.00509
Vowel.Pre-pausal schwa	89.082	10.668	8.35	<0.0001
Vowel. KIT	-33.344	14	-2.382	0.0173

Vowel. FOOT	-5.579	16.541	-0.337	0.737
Vowel. NURSE	24.51	27.896	0.879	0.3797
Vowel. THOUGHT	36.598	18.194	2.012	0.0443
Vowel. TRAP	176.894	18.27	9.682	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1	1	50.854	<0.0001
s(duration):Vowel. Pre-consonantal final schwa	2.763	3.426	1.776	0.123
s(duration):Vowel.Pre-pausal schwa	2.315	2.707	3.011	0.0233
s (duration):Vowel. KIT	1	1	5.239	0.0221
s(duration):Vowel. FOOT	1	1.001	1.565	0.21
s(duration):Vowel. THOUGHT	1.096	1.176	2.028	0.1271
s(duration):Vowel. NURSE	1	1	25.9	<0.0001
s(duration):Vowel. TRAP	3.395	3.895	4.813	<0.001
s(speaker)	27.028	30	199.637	<0.0001
s(word)	26.843	1147	0.513	<0.0001
s(speaker,Vowel)	59.829	126	4.464	0.0353

Model 4.21

The effect of duration on the F2 of non-final schwa by phonetic context

Reference level: preceding context=preceding context, following context=central

F2 ~ preceding context + following context + s(duration, bs = "cr") + s(duration,
by = preceding context, bs = "cr") + s(duration, by = following context, bs = "cr") +
s(speaker, bs = "re") + s(word, bs = "re") + s(speaker,
preceding context, bs = "re") + s(speaker, following context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1784.53	28.41	62.81	<0.0001
Preceding context.back	-164.64	15	-10.98	<0.0001
Preceding context.front	162.72	20.73	7.848	<0.0001
Following context.back	-194.58	23.18	-8.395	<0.0001

Following context.front	185.09	33.1	5.591	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1	1	0	0.997
s(duration):preceding.back	2.357	2.855	1.898	0.213
s(duration):preceding.front	2.671	3.202	2.05	0.123
s(duration):following.back	1	1.001	2.321	0.128
s(duration):following.front	1	1.001	1.496	0.221
s(speaker)	25.667	30	54.47	<0.0001
s(word)	131.399	427	0.893	<0.0001
S (speaker, preceding context	7.628	89	0.16	0.206
S(speaker, following context	26.067	89	1.983	0.043

Model 4.22

The effect of duration on the F2 of pre-consonantal word final schwa by phonetic context

Reference level: preceding context=preceding context, following context=central

F2 ~ preceding context + following context + s(duration, bs = "cr") + s(duration, by = preceding context, bs = "cr") + s(duration, by = following context, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, preceding context, bs = "re") + s(speaker, following context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1726.78	28.5	60.6	<0.0001
Preceding context.back	-129.5	24.99	-5.182	<0.0001
Preceding context.front	163.97	37.44	4.38	<0.0001
Following context.back	-143.83	17.46	-8.235	<0.0001
Following context.front	120.27	25.53	4.71	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	3.732	4.376	4.141	0.00211
s(duration):preceding.back	1.001	1.002	0.905	0.341
s(duration):preceding.front	1	1	1.043	0.307

s(duration):following.back	1	1	1.261	0.262
s(duration):following.front	1	1	1.507	0.22
s(speaker)	24.966	30	35.05	<0.0001
s(word)	98.439	235	2.665	<0.0001
s(speaker, preceding context)	8.005	84	0.274	0.106
s(speaker, following context)	20.514	89	1.233	0.0114

Model 4.23

The effect of duration on the F2 of pre-pausal schwa by phonetic context

Reference level: preceding context=preceding context

F2 ~ preceding context + s(duration, bs = "cr") + s(duration, by = preceding context, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, preceding context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1605.25	27.15	59.13	<0.0001
Preceding context.back	-122	24.04	-5.074	<0.0001
Preceding context.front	107.53	32.02	3.358	<0.001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1	1	1.622	0.204
s(duration):preceding.back	1	1	4.706	0.031
s(duration):preceding.front	1	1	0.848	0.358
s(speaker)	25.32	30	24.44	<0.0001
s(word)	59.61	167	1.063	0.00114
s(speaker,preceding.context)	14.08	78	0.72	0.15

Model 4.24

The effect of duration on the F2 of KIT *by phonetic context*

Reference level: preceding context=preceding context, following context=central

F2 ~ preceding context + following context + s(duration, bs = "cr") + s(duration, by = preceding context, bs = "cr") + s(duration, by = following context, bs = "cr") +

s(speaker, bs = "re") + s(word, bs = "re") + s(speaker,
preceding context, bs = "re") + s(speaker, following context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1972.53	94.81	20.81	<0.0001
Preceding context.back	-62.42	41.48	-1.505	0.134
Preceding context.front	32.56	96.51	0.337	0.736
Following context.back	-175.63	36.67	-4.789	<0.0001
Following context.front	84.27	54.58	1.544	0.124

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1	1	3.267	0.0719
s(duration):preceding.back	2.666	3.236	1.285	0.294
s(duration):preceding.front	3.979	4.672	4.24	0.00511
s(duration):following.back	1	1	1.603	0.207
s(duration):following.front	1	1.001	0.175	0.676
s(speaker)	6.859	7	185.8	<0.0001
s(word)	66.42	122	2.432	<0.0001
s(speaker, preceding context)	0.937	20	0.151	0.414
s(speaker, following context)	0.0002215	21	0	0.537

Model 4.25

The effect of duration on the F2 of FOOT *by phonetic context*

Reference level: preceding context=preceding context, following context=central

F2 ~ preceding context + following context + s(duration, bs = "cr") + s(duration,
by = preceding context, bs = "cr") + s(duration, by = following context, bs = "cr") +
s(speaker, bs = "re") + s(word, bs = "re") + s(speaker,
preceding context, bs = "re") + s(speaker, following context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	1341.66	67.35	19.92	<0.0001
Preceding context.back	-162.69	44.11	-3.688	<0.001
Preceding context.front	222.69	61.25	3.636	<0.001
Following context.back	-121.24	40.22	-3.014	0.00299
Following context.front	24.62	92.35	0.267	0.79

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1.969637	2.459	3.719	0.0159
s(duration):preceding.back	1.057268	1.101	0.837	0.3923
s(duration):preceding.front	1.000281	1.001	1.31	0.2541
s(duration):following.back	1.000068	1	1.502	0.2221
s(duration):following.front	1.31818	1.556	0.22	0.8125
s(speaker)	5.627394	7	7.413	<0.0001
s(word)	2.426583	53	0.051	0.3427
s(speaker, preceding context)	0.0008326	18	0	0.6307
s(speaker, following context)	0.0001463	17	0	0.947

Model 4.26

The effect of duration on the F2 of NURSE *by phonetic context*

Reference level: preceding context=preceding context, following context=central

F2 ~ preceding context + following context + s(duration, bs = "cr") + s(duration, by = preceding context, bs = "cr") + s(duration, by = following context, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	1779.62	73.27	24.29	<0.0001
Preceding context.back	-88.39	23.39	-3.778	<0.001
Preceding context.front	17.24	41.85	0.412	0.681

Following context.back	-56.48	26.59	-2.124	0.0366
Following context.front	-28.3	33.94	-0.834	0.407

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1	1	0.667	0.4165
s(duration):preceding.back	1	1	0.629	0.4298
s(duration):preceding.front	1	1	0.652	0.4217
s(duration):following.back	1	1	1.153	0.2859
s(duration):following.front	1	1	0.454	0.5025
s(speaker)	6.89E+00	7	100.6	<0.0001
s(word)	1.85E+01	49	1.897	0.0132

Model 4.27

The effect of duration on the F2 of THOUGHT *by phonetic context*

Reference level: preceding context=preceding context, following context=central

F2 ~ preceding context + following context + s(duration, bs = "cr") + s(duration, by = preceding context, bs = "cr") + s(duration, by = following context, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, preceding context, bs = "re") + s(speaker, following context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t value*	Pr(>|t|)
(Intercept)	1086.848	42.938	25.312	<0.0001
Preceding.back	-112.89	32.314	-3.494	<0.001
Preceding.front	-63.655	40.924	-1.555	0.122
Following.back	3.864	24.484	0.158	0.875
Following.front	21.399	31.137	0.687	0.493

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1.550967	1.88	4.753	0.0236
s(duration):preceding.back	1.00003	1	1.508	0.2212

s(duration):preceding.front	1.000035	1	0.556	0.4572
s(duration):following.back	2.834371	3.439	2.397	0.0866
s(duration):following.front	2.077353	2.567	1.524	0.1698
s(speaker)	5.898398	7	17.516	<0.0001
s(word)	15.0001	75	0.376	0.0155
s(speaker,preceding context)	3.728966	19	0.467	0.467
s(speaker,following context)	0.001407	21	0	0.4411

Model 4.28

The effect of duration on the F2 of TRAP by phonetic context

Reference level: preceding context=preceding context, following context=central

F2 ~ preceding context + following context + s(duration, bs = "cr") + s(duration, by = preceding context, bs = "cr") + s(duration, by = following context, bs = "cr") + s(speaker, bs = "re") + s(word, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|t|)
(Intercept)	1504.708	51.514	29.21	<0.0001
Preceding context.back	-82.146	20.866	-3.937	<0.001
Preceding context.front	68.861	26.253	2.623	0.00948
Following context.back	2.648	19.88	0.133	0.894
Following context.front	120.51	33.214	3.628	<0.001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(duration)	1.915	2.383	0.859	0.4052
s(duration):preceding.back	1	1	0.114	0.7358
s(duration):preceding.front	3.507	4.202	3.162	0.0163
s(duration):following.back	1	1	0.777	0.3794
s(duration):following.front	1.383	1.634	0.248	0.7689
s(speaker)	7.528	8	37.4	<0.0001
s(word)	13.613	108	0.168	0.1487

Model 6.1

F1 Interaction between birth year and duration compared between KIT and non-final schwa

Reference level: vowel=KIT

Normalised F1~vowel + s (birth year, bs = "ad") + s(birth year, by = vowel, bs = "ad") + s(duration, bs = "cr") + s(duration, by = vowel, bs = "cr") + ti(birth year, duration) + ti(birth year, duration, by = vowel) + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, vowel, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	-0.28498	0.01261	-22.61	<0.0001
Vowel.schwa	0.03509	0.01606	2.185	0.0289

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(birth year)	3.639	3.892	122.7	<0.0001
s(birth year):vowel.schwa	5.655	6.141	4.239	<0.001
s(duration)	1.843	2.309	16.49	<0.0001
s(duration):vowel.schwa	4.128	4.96	13.3	<0.0001
ti(birth year,duration)	1.001	1.002	16.82	<0.0001
ti(birth year,duration):vowel.schwa	2.79	3.191	3.023	0.0306
s(speaker)	207.146	555	4.986	<0.001
s(word)	1693.878	4773	1.584	<0.0001
s(speaker,vowel)	419.444	1089	2.435	<0.001

Model 6.2

F1 Interaction between birth year and duration compared between KIT and non-final schwa

Reference level: vowel=KIT

Normalised F1~vowel + s(birth year, bs = "ad") + s(birth year, by = vowel, bs = "ad") +s(duration, bs = "cr") + s(duration, by = vowel, bs = "cr") + ti(birth, duration) + ti(birth, duration, by = vowel) + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, vowel, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	-0.2835	0.01338	-21.19	<0.0001
Vowel.schwa	0.64577	0.02471	26.14	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(birth year)	3.365	3.558	90.6	<0.0001
s(birth year):vowel.schwa	2.989	3.295	5.472	<0.001
s(duration)	1.902	2.385	16.57	<0.0001
s(duration):vowel.schwa	3.614	4.227	17.34	<0.0001
ti(birth year,duration)	1.007	1.014	17.14	<0.0001
ti(birth year,duration):vowel.schwa	1.001	1.002	2.192	0.139
s(speaker)	116.344	554	3.89	0.0627
s(word)	808.106	2628	1.51	<0.0001
s(speaker,vowel)	475.66	1038	4.608	<0.0001

Model 6.3

F2 Interaction between birth year, duration and phonetic context for KIT

Reference level: preceding context=central, following context=central

Normalised F2 ~ context + s(birth year, bs = "ad") + s(birth, by = context, bs = "ad") +

s(duration, bs = "cr") + s(duration, by = context, bs = "cr") + ti(birth year, duration) + ti(birth, context, by = context) + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, context, bs = "r e")

Parametric coefficients

	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	0.04242	0.01438	2.951	0.00317
Context.back	-0.33149	0.02624	-12.63	<0.0001
Context.front	0.41138	0.02035	20.22	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(birth year)	5.523	5.856	99.572	<0.0001
s(birth year):Context.back	1.927	2.173	2.322	0.0762
s(birth year)Context.front	3.53	4.019	8.219	<0.0001
s(duration)	6.691	7.493	24.505	<0.0001
s(duration):Context.back	2.901	3.52	3.941	0.0059
s(duration):front	2.32	2.876	6.455	<0.001
ti(birth year,duration)	4.641	5.945	4.125	<0.001
ti(birth year,duration):Context.back	2.984	4.086	0.617	0.6735
ti(birth year,duration):Context.front	2.892	3.248	5.989	<0.001
s(speaker)	259.202	554	5.145	<0.0001
s(word)	778.438	1993	3.294	<0.0001
s(speaker,context)	375.978	1603	1.084	<0.0001

Model 6.4

F2 Interaction between birth year, duration and phonetic context for non-final schwa

Reference level: preceding context=central, following context=central

Normalised F2 ~ context + s(birth year, bs = "ad") +

s(birth year, by = context, bs = "ad") + s(duration, bs = "cr") + s(duration, by = context, bs = " cr") + ti(birth year, duration) + ti(birth year, context, by = context) + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	-0.12172	0.01224	-9.946	<0.0001
Context.back	-0.40161	0.01441	-27.86	<0.0001
Context.front	0.32309	0.0217	14.89	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(birth year)	2.711	2.865	9.416	<0.0001
s(birth year):Context.back	1.001	1.001	19.193	<0.0001
s(birth year):Context.front	1	1	35.164	<0.0001
s(duration)	4.245	5.059	11.549	<0.0001

s(duration):Context.back	1.001	1.002	50.341	<0.0001
s(duration):Context.front	2.315	2.875	1.752	0.2003
ti(birth year,duration)	5.413	6.88	2.01	0.0758
ti(birth year,duration):Context.back	5.363	6.872	4.071	0.0003
ti(birth year,duration):Context.front	1.499	1.883	2.189	0.0897
s(speaker)	260.835	537	7.41	<0.0001
s(word)	1721.37	4301	2.57	<0.0001
s(speaker,context)	464.065	1577	2.179	<0.0001

Model 6.5

F2 interaction between birth year, duration and phonetic context for pre-pausal schwa

Reference level: preceding context=central

Normalised F2 ~ context + s(birth year, bs = "ad") + s(birth year, by = context, bs = "ad") + s(duration, bs = "cr") + s(duration, by = context, bs = "cr") + ti(birth year, duration) + ti(birth year, context, by = context) + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, context, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	-0.08308	0.02324	-3.574	0.000357
Context.back	-0.17174	0.03155	-5.444	<0.0001
Context.front	0.33818	0.03907	8.655	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(birth year)	3.674	4.08	18.075	<0.0001
s(birth year):context.back	2.328	2.79	2.387	0.0528
s(birth year):front	1	1	1.347	0.2459
s(duration)	1	1	15.454	<0.0001
s(duration):context.back	1	1	1.872	0.1714
s(duration):front	1	1	0.163	0.6866
ti(birth year,duration)	3.533	4.404	1.452	0.1966
ti(birth year,duration):context.back	1	1	0.01	0.9195
ti(birth year,duration):context.front	1	1	0.228	0.6327
s(speaker)	162.983	484	0.964	<0.0001

s(word)	113.57	633	0.442	<0.0001
s(speaker,context)	71.71	1010	0.099	0.0242

Model 6.6

F2 interaction of birth year and duration compared between KIT and pre-pausal schwa

Reference level: vowel=KIT

Normalised F2 ~ vowel + s(birth year, bs = "ad") + s(birth year, by = vowel, bs = "ad") + s(duration, bs = "cr") + s(duration, by = vowel, bs = "cr") + ti(birth year, duration) + ti(birth year, context, by = vowel) + s(speaker, bs = "re") + s(word, bs = "re") + s(speaker, vowel, bs = "re")

Parametric coefficients

	Estimate	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	0.10056	0.01321	7.61	<0.0001
Vowel.schwa	-0.17158	0.02551	-6.727	<0.0001

Approximate significance of smooth terms

	edf	Ref.df	*F*	*p*-value
s(birth year)	5.62	5.978	139.681	<0.0001
s(birth year):vowel.schwa	2.087	2.192	37.601	<0.0001
s(duration)	6.506	7.338	36.058	<0.0001
s(duration):vowel.schwa	1	1	80.87	<0.0001
ti(birth year,duration)	4.66	5.794	10.211	<0.0001
ti(birth year,duration):vowel.schwa	3.739	4.894	0.789	0.543

References

Bailey, G. 2016. Automatic detection of sociolinguistic variation using forced alignment. In H. Jeoung (Ed*.), Penn working papers in linguistics 22.2: Selected papers from NWAV 44* (Pp 11–20). Philadelphia: Penn Graduate Linguistics Society.

Bates, S.A.R (1995). *Towards a Definition of Schwa: An Acoustic Investigation of Vowel Reduction in English.* Ph.D. this. University of Edinburgh: United Kingdom.

Barry, W. J. (1998). Time as a factor in the acoustic variation of schwa. In *Fifth International Conference on Spoken Language Processing.* Sydney, Australia.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (Release 2): Linguistic data consortium*. University of Pennsylvania, Philadelphia, PA, USA.

Bauer, L., & Warren, P. (2008). New Zealand English: phonology. In Burridge and B. Kortmann (eds.) *Varieties of English*, *3*, (Pp 39-63).Berlin: Walter de Gruyter.

Beal, J. (1999). *English Pronunciation in the Eighteenth Century: Thomas Spence's Grand Repository of The English Language*. New York, Oxford University Press.

Beal, J. (2008). English dialects in the North of England: morphology and syntax. In B. Kortmann and C. Upton (Eds.), *Varieties of English, 1* (Pp 373-403). Berlin: Walter de Gruyter.

Bekker, I. (2014). The weak vowels of South African English: A critical review and comparative acoustic analysis. *Southern African Linguistics and Applied Language Studies*, 32(1), 133-147.

Bermúdez-Otero, R. and Trousdale, G. (2012). Cycles and Continua: On Unidirectionality and Gradualness in language change. In T. Nevalainen and E. Closs (Eds.). *The Oxford Handbook of the History of English* (Pp 691-720). New York: Oxford University Press.

Boersma, P. & Weenink, D. 2011. *Praat: doing phonetics by computer* [Software]. Ver. 5.2.38.

Brand, J., Hay, J., Clark, L., Watson, K., & Sóskuthy, M (2019). Systematic covariation of monophthongs across speakers of New Zealand English. In S. Calhoun, P. Escudero, M. Tabain and P. Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (Pp 1879-1882). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Browman, C.P. & Goldstein, L. (1990). “Targetless” schwa: an articulatory analysis. In G. J. Docherty and R.D Ladd (Eds.). *Papers in laboratory phonology ll: Gesture, segment, prosody* (Pp 26-56). Cambridge: Cambridge University Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Cohen Priva, U., & Strand, E. (2020). Schwa has an articulatory target in American English. Under review, https://psyarxiv.com/gwkvh/.

Dalby, J. M. 1986. *Phonetic structure of fast speech in American Engl*ish. Doctoral dissertation. Indiana University, USA.

Davidson, L. (2006). Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica*, 63(2-3), 79-112.

Dinkin, A.J & Dodsworth, R. (2017). Gradience, Allophony and Chain Shifts. *Language Variation and Change*, 29, 101-127.

Easton, A., & Bauer, L. (2000). An acoustic study of the vowels of New Zealand English. *Australian Journal of Linguistics*, 20(2), 93-117.

Evanini, K., Isard, S., & Liberman, M. (2009). Automatic formant extraction for sociolinguistic analysis of large corpora. In *Tenth Annual Conference of the International Speech Communication Association* (Pp 1655-1658).Baixas, International Speech Communication Association.

Fabricius, A. (2002). Weak vowels in modern RP: an acoustic study of happY-tensing and KIT/schwa shift. *Language Variation and Change*, 14(2), 211-237.

Fabricius, A., Watt, D. & Johnson, D.E. (2009). A Comparison of Three Speaker- Intrinsic Vowel Formant Frequency Normalization Algorithms for Sociophonetics. *Language Variation and Change*, 21, 413-435.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner and D. Steriade (Eds.) *Phonetically Based Phonology* (Pp 232-276). Cambridge: Cambridge University Press.

Flemming, E. (2009). The phonetics of the schwa vowel. In D. Minkova (Ed.). *Phonological Weakness in English: From old to present-day English* (Pp 78-95). Basingstoke: Palgrave MacMillan.

Flemming, E. & Johnson, S. (2007). Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association*, 11, 83-96.

Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York papers in linguistics*, *2*(11), 1-28.

Foulkes, P., & Docherty, G. (Eds.). (2014). *Urban voices: Accent studies in the British Isles*. Routledge. London: Edward Arnold.

Foulkes, P., Docherty, G., Hufnagel, S. S., & Hughes, V. (2018). Three steps forward for predictability. Consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard*, 4(2).

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical society of America*, 90(4), 1816-1827.

Fromont, R., & Hay, J. (2012). LaBB-CAT: An annotation store. In *Proceedings of the Australasian Language Technology Association Workshop 2012* (Pp 113-117). University of Otago

Fromont, R., & Watson, K. (2016). Factors influencing automatic segmental alignment of sociophonetic corpora. *Corpora*, *11*(3), 401-431.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27(4), 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech*, 1(2), 126-152.

Geng, C., Mooshammer, C., Nam, H., & Hoole, P. (2010). Schwa deletion under varying prosodic conditions: Results of a pilot study. In S. Fuchs, P. Hoole, C. Mooshammer and M. Zygis (Eds.), *Between the regular and the particular in speech and language* (Pp 145-169). Frankfurt am Main: Peter Lang.

Gick, B. (2002). An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica*, 59(1), 38-48.

Gick, B., & Wilson, I. (2006). Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. *Laboratory phonology*, *8*, 635-659.

Gimson, A.C. (1962). *An Introduction to the Pronunciation of English*. London: Edward Arnold.

Gimson, A.C. (1984). The RP accent. In P. Trudgill (Ed.), *Language in the British Isles* (Pp 45-54.) Cambridge: Cambridge University Press.

Glowacka. (2001). Unstressed vowel deletion and new consonant clusters in English. *Poznan Studies in Contemporary Linguistics*, *37*, 71-94.

Gonzalez, S., Grama, J. and Travis, C.E. (2020). Comparing the performance of forced aligners used in sociophonetic research. *Linguistics Vanguard*, 6 (1).

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: its origins and evolution*. Cambridge: Cambridge University Press.

Gordon, E., Maclagan, M., & Hay, J. (2007). The ONZE corpus. In J. Beal, K. Corrigan, and H. L. Moisl (Eds.) *Creating and digitizing language corpora* (Pp 82-104). London: Palgrave Macmillan.

Jones, D. (1956). *The Pronunciation of English, 4th Edition.* London: J.M. Dent.

Halle, M., & Mohanan, K. P. (1985). Segmental phonology of modern English. *Linguistic inquiry*, *16*(1), 57-116.

Heselwood, B. (2007). Schwa and the phonotactics of RP English. *Transactions of the Philological Society*, 105 (2), 148-87.

Hay, J., Maclagan, M. & Gordon, E. (2008). *New Zealand English*. Edinburgh: Edinburgh University Press.

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., and LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83-91.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, 97(5), 3099-3111.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization*. Attention, Perception, & Psychophysics*, 72(5), 1218-1227.

Jones, D. (1914). *The Pronunciation of English*, 2nd Edition. London: J. M. Dent.

Kendall, T. and E. R. Thomas. 2018. Vowels: Vowel manipulation, normalisation and plotting. R-package version 1.2. https://cran.r-project.org/web/packages/vowels/index.html

Kiesling, S. F. (2005). Variation, stance and style: Word-final-er, high rising tone, and ethnicity in Australian English. *English world-wide*, 26(1), 1-42.

Kirby, J. P. (2010). *Cue selection and category restructuring in sound change*. Doctoral dissertation. The University of Chicago, USA.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3), 129-140.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5), 1208-1221.

Kondo, Y. (1994). Targetless Schwa: Is that how we get the impression of stress timing in English? *Proceedings of the Edinburgh Linguistics Department Conference*, 63-76.

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Mason, Nelson Education.

Langstrof, C. (2006*). Vowel Change in New Zealand English-Patterns and Implications*. PhD thesis. University of Canterbury, New Zealand.

Labov, W. (1994*). Principles of linguistic change. Vol. 1: Linguistic factors*. Oxford: Blackwell publishing.

Lass, R. (1986). On Schwa. *Stellenbosch Papers in Linguistics*, 15, 1-30.

Lass, R. (1999). Phonology and morphology. In R. Lass (Ed.) *The Cambridge history of the English language volume III: 1476-1776* (Pp 56-186). Cambridge: Cambridge University Press.

Lass, R. (2009). On schwa: Synchronic prelude and historical fugue. In D. Minkova (Ed.). *Phonological Weakness in English: From old to present-day English* (Pp 47-77). Palgrave MacMillan.

Leach, H. (2018). *Sociophonetic variation in Stoke-on-Trent's pottery industry*. PhD thesis. University of Sheffield, UK.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press, Cambridge, MA.

Lilley, J. (2012). *The characterization of variation American English schwa using hidden Markov Models*. Ph.D. dissertation. University of Delaware, USA.

Lindblom, B. (1963). Spectrographic Study of Vowel Reduction. *The 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 W.J. Hardcastle and A. Marchal (Eds.) *Speech production and speech modelling* (Pp 403-439). Springer, Dordrecht.

Lindblom, B., Sussman, H. M., Modarresi, G., & Burlingame, E. (2002). The trough effect: implications for speech motor programming. *Phonetica*, 59(4), 245-262.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, 49(2B), 606-608.

LoCasto, P. C., & Connine, C. M. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics*, 64(2), 208-219.

MacKenzie, L., & Turton, D. (2020). Assessing the accuracy of existing forced alignment software on varieties of British English. *Linguistics Vanguard*, 6(1).

Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English" short" front vowel shift. *Language variation and change*, *19*(1), 1.

McMahon, M. (1998). Phonology. In S. Romaine (Ed.), *The Cambridge history of the English language volume IV 1776-1997* (Pp 375-535). Cambridge: Cambridge University Press.

Mesthrie, R. (2017). Class, gender, and substrate erasure in sociolinguistic change: A sociophonetic study of schwa in deracializing South African English. *Language*, 93(2), 314-346.

Milroy, L., Milroy, J. & Docherty, G.J. (1997) *Phonological Variation and Change in Contemporary Spoken British English*. ESRC End of Award Report, R000234892.

Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical society of America*, 96(1), 40-55.

Nokes, J., & Hay, J. (2012). Acoustic correlates of rhythm in New Zealand English: A diachronic study. *Language Variation and Change*, 24(1), 1-31.

Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, *60*(1), 45-69.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech (2nd release)*. Columbus, OH: Department of Psychology, Ohio State University.

Polgardi, K. (2015). Syncope, syllabic consonant formation, and the distribution of stressed vowels in English 1. *Journal of Linguistics*, 51(2), 383-423.

Proctor, M., Lo, C. Y., & Narayanan, S. (2015). Articulation of English vowels in running speech: A real-time MRI study. In The Scottish Consortium for ICPhS 2015 (Eds.) Proceedings *of the 2015 International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.

Rakerd, B., Sennett, W., & Fowler, C. A. (1987). Domain-final lengthening and FOOT-level shortening in spoken English. *Phonetica*, 44(3), 147-155.

Ramsammy, M. & Turton, D. (2012). High [əɹ] or low[ʌ]? An investigation of the unstressed vowel system in Mancunian English. Paper presented at the *Manchester Salford New Researchers Forum in Linguistics*, 2-3rd November 2012.

R Core Team (2012). R: A language & environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Roach, P. (1983). *English Phonetics and Phonology*. Cambridge: Cambridge University Press.

Rosenfelder, I. Fruehwald ,J. Evanini, K. Seyfarth, S. Gorman, K. Prichard, H. and Yuan, J. 2014. *FAVE (Forced Alignment and Vowel Extraction) Program Suite* v1.2.2.

Schiel, Florian. 1999. Automatic phonetic transcription of non-prompted speech. In Ohala, J., Hasegawa, Y. Ohala, M., Granville, D.M and Bailey, A.C. *14th Proceedings of the International Phonetic Association* (Pp 607-610). San Francisco, University of California

Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of phonetics*, 25(3), 255-286.

Silverman, D. (2011). Schwa. In M. van Oostendorp, C.J. Ewen, E. Hume and K. Rice (Eds.) *The Blackwell companion to phonology* (Pp 1-628-42). New York, Blackwell.

Sóskuthy, M. (2014). *Formant Editor: Software for editing dynamic formant measurements* (version 0.8.1).

Sóskuthy, M., Foulkes, P., Hughes, V., & Haddican, B. (2018). Changing words and sounds: The roles of different cognitive units in sound change. *Topics in cognitive science*, 10(4), 787-802.

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, *24*(3), 320-335.

Thomas, E.R. & Kendall, T. (2007). NORM: *The Vowel Normalization and Plotting Suite*. Retrieved November 1 2017 from http://lingtools.uoregon.edu/norm/index.php

Toft, Z. (2002). The phonetics and phonology of some syllabic consonants in Southern British English. *ZAS Papers in Linguistics*, *28*, 111-144.

Tollfree, L. (1999). South East London English: discrete versus continuous modelling of consonantal reduction. In P. Foulkes & G. Docherty (Eds.) Urban Voices, Accent Studies in the British Isles (Pp. 163-184). London: Arnold.

Trudgill, P. Gordon, E., & Lewis, G. (1998). New-dialect formation and Southern Hemisphere English: The New Zealand short front vowels. *Journal of Sociolinguistics*, 2(1), 35-51.

Turton, D. (2014). *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes*. PhD thesis. University of Manchester, UK.

Van Bergem, D. R. (1989). Phonetic and linguistic aspects of vowel reduction. In *Proceedings of the Institute of Phonetic Sciences Amsterdam Vol.13* (Pp 97-105). University of Amsterdam.

Van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech communication*, *12*(1), 1-23.

Van Oostendorp, M. (1998). *Vowel quality and phonological projection*. Doctoral dissertation. Tilburg University, Netherlands.

Warren, P. (2018). Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association*, *48*(3), 305-330.

Watson, C. I., Harrington, J., & Evans, Z. (1998a). An acoustic comparison between New Zealand and Australian English vowels. *Australian Journal of linguistics*, 18(2), 185-207.

Watson, C., Harrington, J., & Palethorpe, S. (1998b). A kinematic analysis of New Zealand and Australian English vowel spaces. In *Fifth International Conference on Spoken Language Processing (*Pp 2363-2366), Sydney, Australia.

Watson, C. I., Maclagan, M., & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change*, 12(1), 51-68.

Watt, D., & Allen, W. (2003). Tyneside English. *Journal of the International Phonetic Association*, 33(2), 267-271.

Wells, J. C. (1982). *Accents of English (Vol. 1)*. Cambridge: Cambridge University Press.

Wells, J.C. (1990). *Longman Pronunciation Dictionary*. Harlow: Longman.

Young, S., Evermann, G. Hain, T., Dan Kershaw, Liu, X. Odell, J .Ollason, D. Povey, D. Valtchev, V. and Woodland, P. 2006. *The HTK Book (For Version 3.4)*. Cambridge: Cambridge University Engineering Department

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, 38(2), 260-271.

Zwicky, A. M. 1972. Note on a phonological hierarchy in English. In P. S. Robert, and S. M. Ronald (eds.), *Linguistic Change and Generative Theory* (Pp 275-301). Bloomington, IN: Indiana University Press.