Perception of social-indexical information in gender-ambiguous voices

Anna Katarzyna Kubisz

Doctor of Philosophy

University of York
Language and Linguistic Science

January 2016

Abstract

This thesis presents innovative research which uses gender-ambiguous speech to investigate perceptions of speaker-indexical information.

In a series of three perceptual experiments perceptions of speaker age, gender and social class are researched. In Experiment 1 listeners heard audio samples, on the basis of which they were asked to evaluate speaker age, gender and social class using a Visual Analogue Scale.

Experiment 2 was performed in the interests of investigating how perceptions of the same speaker-indexical information as in Experiment 1 might be shifted when providing the listener with visual information about the supposed speaker. For example, upon seeing a young female face when hearing a phonetic variant, the listener might rate the variant differently from the answer s/he gave in response to the same stimulus in Experiment 1.

In Experiment 3, a new social factor, ethnicity, was introduced. The aim of this experiment was to investigate perceptions of speaker-indexical information when listeners were exposed to visual cues to the ethnicity of the supposed speaker. As Experiment 2, Experiment 3 tested whether speaker-indexical information could be shifted as a result of the manipulation.

Furthermore, this research offers a multivariate investigation of perception of speaker-indexical information based on Tyneside English. Perceptions of the variants of the FACE, GOAT and NURSE vowels, T-to-R and variants of /p t k/ are tested.

Finally, the findings for groups of listeners with high and lower exposure to Tyneside English are compared and contrasted.

List of contents

List of tables

List of figures

Acknowledgements

My sincere thanks go to Dr Carmen Llamas and Dr Dominic Watt for their insight and expertise, as well as guidance through the PhD process.

My heartfelt thanks go to Daniel Lawrence, Dr Andrew Euan MacFarlane and Dr Márton Sóskuthy for the help with R.

I would also like to thank Huw Llewelyn-Jones for the technical support including help with the recordings and the choice of software for manipulating the stimuli.

Big thanks to:

My Parents who supported me emotionally and financially, and were there for me whenever I needed them.

My Brother and Sister for being who they are.

Daniel for being amazingly patient and putting up with me.

Everyone I've met during my PhD and to those who helped me with data collection or any other aspect of the PhD work.

Thank you!

Author’s declaration

This thesis has not previously been submitted for any degree other than Doctor of Philosophy of the University of York. This thesis comprises my original work, except where otherwise stated. Other sources are acknowledged by explicit references.

1 Introduction

The research presented in this thesis investigates perceptions of speaker social-indexical information from gender-ambiguous-sounding speech. The investigation was carried out with the aim of exploring how information about speaker age, gender and social class is evaluated by the listener. The perceptions are explored at the segmental level via the use of localised as well as non-localised phonetic variants identified in Tyneside English. This chapter discusses the meaning of non-linguistic information in speech processing and specifies the main research questions. It lays out the contribution of the thesis within the body of sociolinguistic research and provides an overview of the following chapters.

1.1 Objectives and research questions

Linguistic variation occurring in language is a natural consequence of language use. The question which motivated this research is a broader one, of how the large amount of variation in spoken language is managed by the listener. This question has been narrowed down to one, to which the answer seems quite obvious on the surface, of how speaker-indexical information is accessed from speech by the listener.

The first main aim of this thesis was to add to the body of research relating to the interaction between linguistic and non-linguistic information in the process of speech perception with the objective of helping to move forward our understanding of the process. In light of findings reported by Foulkes et al. (2010), who explored perception of gender from child speech, this research takes a step toward a multivariate investigation of perception of speaker-indexical information. One of the explored aspects includes social-indexical information which is embedded in a linguistic form produced by the speaker.

In a traditional and somewhat simplified view, in the process of language communication there need to be at least two participants: a speaker who sends a message, and a listener who receives and decodes the message. Whilst some of the information carried by the message includes semantic information, it is only part of what the receiver can learn from the acoustic signal. A single message has a capacity to provide the listener with an abundance of information apart from the semantic content itself. For instance, the listener may learn a variety of things about the speaker. The speaker's voice carries information about biological differences between individuals. One of the most readily accessible types of such information is speaker sex. It is usually accessed from the pitch, which depends on the length of the vocal tract among other things. In addition to providing information about speaker sex, pitch can also tell us how tall the speaker is. Another type of information resulting from anatomical differences and carried by the voice is speaker age. We can easily identify the speaker's age and distinguish a child, teenager, adult or older speaker by relying on pitch or other voice characteristics. However, these are not the only types of information about the speaker which are accessed from his or her voice. Other types of information tell us, for instance, about the speaker's personality or mood. We can also learn how friendly someone sounds or what his/her attitude toward his/her interlocutor is, for example, whether he/she respects the interlocutor. Furthermore, information about health of the speaker is present in voice as well. We may be able to hear whether someone is suffering from a cold, is a smoker or had his/her larynx removed.

As has been pointed out so far, listeners attend to acoustic information at different levels. While biological information or personality traits carried by the voice are two types of information, there is also the socially constructed aspect of speech, which provides us with cues to the speaker's gender, social class but also age. Socially constructed information is created as an outcome of an "agreement" between the speaker and listener. Social information is context- and region-specific in the sense that a linguistic feature recognised in one geographical region or community as, for example, widely used by females may not be recognised as such in a different community or region. It is typically argued that gender is one of the socially constructed features. Nevertheless, gender differences are de facto grounded in biology. It is the child's sex which primarily decides whether the caregiver will teach him/her features of linguistic repertoire used for example more by men than by women (Foulkes et al., 2005). The speaker's social class is another type of information which is socially constructed. Linguistic forms used by the speaker provide cues to his/her social standing because different forms are used by working-, middle- or upper-class speakers, although to different degrees.

Although the speaker's age can be identified from biological information present in voice, it can also be socially constructed as gender and social class are. As a result, older and younger speakers differ in terms of linguistic forms they use. The differences may be lexical. For example, young speakers often use youth slang, which older speakers may not be familiar with (Agha, 2004: 24). However, the differences may also be phonological. For instance, speakers may choose standard pronunciations as they enter working age and try to secure a job and then return again to using less standard forms as they approach retirement. These examples are cases of age-grading, which can be reflected in the use of linguistic forms but also in other aspects of life (Labov, 2001; Sankoff, 2005; Wagner, 2012).

From the above account it can be seen that speaker age, gender and social class may be socially constructed. In turn, linguistic forms used to carry this information become indexical (for further discussion see Section 2.2.2). A question one may ask is: what is indexicality manifesting, exactly? Why is there as much variation in speech? Perhaps it would be easier if everyone spoke in the same way irrespective of their sex, age or social class. In fact, we already have unavoidable variation resulting from biological differences between speakers. The listener needs to be able to deal with large amounts of variability in spoken language. Since increased speaker variability may result in longer processing time (Mullennix & Pisoni, 1990), it could be argued that less variation should minimise processing costs on the part of the speaker in terms of production, as well as on the part of the listener in terms of perception. Nevertheless, despite high levels of speaker variability, listeners successfully process speech (Kraljic & Samuel, 2006; Kraljic et al., 2008; Magnuson & Nusbaum, 2007). Moreover, as numerous production and perception studies show, it seems that during speech processing language users apply more fine variation in the acoustic signal in the form of indexical information (Drager, 2005; Eckert, 2000; Hay & Drager, 2010; Johnstone & Kiesling, 2008). Speakers and listeners are willing to make the effort of using even more differences than the ones resulting from biology. One such example is geographical variation. Geographical dialects are not necessary for successful communication and yet they exist. Social variation in speech is another example. As it turns out, social variation is very systematic and, at the same time socially constrained (Milroy & Milroy, 1997).

Language users encode social information about themselves in the utterances they produce. The listener, on the other hand, uses this information as a proxy for the speaker's age, gender or social standing. He/she is willing to deal with extra variation in order to help him/her understand the social context of the message. As a result, the message becomes rich in information about the speaker, complementing biological and semantic information. It should be expected that this, more fine-grained variation facilitates communication.

The other main aim of this thesis was to investigate the role of language variation in speech perception. First and foremost, the obvious should be stated: language is not stable. Language is in a constant state of flux. One such example would be the social values attributed to linguistic forms (Agha, 2003). For instance, to be a young Geordie female, that is a Tyneside English speaker, sixty years ago meant something different from what it means nowadays. The dress style was different, the attitude towards older members of society was more formal, and the language forms used by a young female in the 1960's were different from today's forms (Corrigan, 2012; Watt, 2002).

Nevertheless, it is language variation which leads to indexicality being ascribed to linguistic forms. Milroy and Milroy (1985: 19) argue that accounting for social values attributed to linguistic forms is the main focus of sociolinguistics. The present research attempts to investigate the relation between variation and indexicality from the perspective of speech perception. Previous research has established that speakers may use indexical forms to construct their identities and to signal belonging to a social group (see Eckert, 2008 for an overview). Furthermore, speakers are aware of listeners responding to indexicality in speech. If this is the case, it seems that indexing social information on the part of the speaker may be a partially conscious act, which in essence, would be a socially conditioned performance. As will be shown in the following chapter, a number of studies of the perception of speaker indexicality have been carried out. These studies, as well as the experiments presented in the thesis, explore how use of socially-indexed variation indexes the speaker's characteristics in the ears of the listener.

One of the more specific questions this thesis is attempting to address is that of the cues the listener relies upon to access the types of information described above when visual information about the speaker is not present. For example, how is speaker-indexical information accessed when listening to a stranger on the phone? The other, more detailed question is whether the listener is able to access indexical information when he/she cannot rely on fundamental frequency information typically found in the male or female speaker's voice, for instance, when the speaker speaks with a gender-ambiguous-sounding voice or an artificial larynx. By focusing on different types of social-indexical information: age, gender and social class, this thesis investigates whether these three types of indexical information are decoded by the listener with the same consistency.

Another aspect of the perception of indexical information investigated in this research is whether and how perceptions of acoustic information may be shifted by visual cues representing the supposed speaker. Would cues to the social class or ethnicity of the speaker skew perception of indexical information present in speech? Finally, the research attempts to answer the question of whether relevant indexical information is more readily accessible to listeners who have had high exposure to the socially-indexed linguistic form than to the listener who has experienced low exposure to the form.

1.2 Relevance and broader implications

Socioperceptual studies have researched perception of such speaker-indexical information as age (Drager, 2005; Walker, 2007), ethnicity (Purnell et al., 1999) and geographic origin (Clopper et al., 2005). It has also been studied how the listener's expectations about the speaker influences perceptions of speech. For instance, it has been shown that expectations about speaker gender (Johnson et al., 1999; Strand, 1999), speaker-geographical information (Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999) or speaker age (Hay et al., 2006b) shifted perceptions of speech (cf. Section 2.6).

The present research offers a systematic, as well as a multivariate investigation of perception of speaker-indexical information. To begin with, three types of information: age, gender and social class are investigated simultaneously testing a number of variables. Furthermore, this research attempts to improve our understanding of how expectations about the speaker - for example, his/her age, gender, social class or ethnicity presented in the visual form - may shift perceptions of speaker-indexical information in speech. In this sense, the present research takes a somewhat similar route of investigation by comparison with the studies quoted above, since it is primarily interested in perception and evaluation of speaker-indexical information from speech. Nevertheless, the innovation introduced by this research is the use of gender-ambiguous voice in stimuli. By applying this line of investigation this thesis contributes to the body of sociolinguistic research.

1.3 Thesis outline

Following the Introduction, Chapter 2 presents an overview of the literature describing standard language ideology, which provides the theoretical framework for interpreting findings of this thesis. It then proceeds to reviewing a body of research on parameters responsible for successful identification of voice gender, such as fundamental frequency and formant frequencies. Subsequent to that, an overview of a body of research investigating perception of speech and speaker-indexical information from the sociolinguistic perspective is presented. Next, supraliminal priming and multimodal speech processing are briefly discussed and exemplar theory is mentioned. Finally, a description of socially-indexed phonetic variants in Tyneside English is provided.

In Chapter 3 the first of a series of perceptual experiments carried out for the purpose of this thesis is presented. This study was conducted to investigate whether listeners respond to speaker social-indexical information - the speaker's age, gender and social class - in gender-ambiguous speech. The listeners were provided only with acoustic information and were asked to evaluate perceived speaker-indexical information.

Chapter 4 presents the second of the experiments. This study investigates how perception of speaker-indexical features: age, gender and social class on the basis of short speech samples may be affected by providing the listener with visual information about the supposed speaker, his or her age, gender and social class.

Finally, in Chapter 5, the last experiment of the series is presented. This experiment investigates perceptions of speaker age and gender from a different perspective. This time, visual information about the ethnicity of the supposed speaker is introduced to explore whether it may affect perception of speaker-indexical information.

The results presented in Chapters 3-5 for each individual experiment are brought together for discussion in Chapter 6.

Finally, Chapter 7 summarises the findings of the thesis and present conclusions. It mentions some of the limitations of the methodology and highlights opportunities for future research to build on the outcomes of the present experiments.

2 Literature Review

2.1 Introduction

This chapter outlines the theoretical and contextual background relevant to this thesis. Standard language ideology as the framework within which findings are interpreted is discussed in section 2.2. Section 2.3 introduces multimodal speech processing. A brief account of exemplar theory is presented in section 2.4 followed by an overview of the body of literature on the phonetic characteristics of male and female voices are presented in section 2.5. This section focuses on the fundamental frequency (F0) in voice and its importance in speaker gender identification. Section 2.6 presents an overview of sociolinguistic and sociophonetic research on speech perception with particular consideration given to recovery of speaker-indexical information from the acoustic signal. Since two of the experiments presented in the thesis are priming experiments, the notion of priming along with a brief overview of previous studies on priming are presented in section 2.7. Finally, section 2.8 discusses phonetic variation in Tyneside English, which has been chosen as a basis for constructing stimuli for this research.

2.2 Standard language ideology

Language ideologies are political and economic in nature. They serve specific social groups, and as such divide society by means of language (see Woolard & Schieffelin, 1994 for a review). According to Silverstein

> "(…) ideologies about language, or linguistic ideologies, are any sets of beliefs about language articulated by the users as a rationalization or justification of perceived language structure and use" (Silverstein, 1979: 193).

One such language ideology is standard language ideology, which will serve as a theoretical background for the research in this thesis.

In general, using a standard variety of a language is something common and obvious to language speakers in Western culture. Students are used to writing and speaking in the standard when at school, papers are written in the standard and radio and television are also broadcast in the standard. The main feature of a standard language is its uniformity (Milroy, 2007: 133). What this means in practice is that phonetic, morphological or syntactic variation is diminished (Cheshire, 1999: 132) and speakers are encouraged to use exclusively one variant, which is at the same time claimed to be the correct one (Milroy, 2007: 133-134). Standard language ideology is prescriptivist in nature and stigmatises any variability. However, as Milroy (2007: 134) points out, it is almost impossible to maintain a standard which would have no variation, which is especially the case on the level of spoken language. It goes without saying that we notice variation in everyday situations: speakers style-shift, use vernacular forms, etc. Therefore, Milroy (2001: 543) argues that a uniform language becomes more of an ambition than a realistic goal.

Standardisation is a process which aims to help to consolidate the society and the state. As such, it is a political process. In a number of European countries standardisation processes were part of national unification processes. Written Standard English started emerging in the sixteenth century (Milroy, 2001: 542), though this did not cause other English dialects to perish. Another such example is Germany, where even today geographical dialects are mutually unintelligible.

How is the standard decided upon and where does the standard language come from? Milroy (2007: 137) points out that standard forms are typically based on the language varieties used by social groups associated with prestige. Such groups may include, for example, journalists, lawyers or businesspeople. What these people have in common, argues Milroy, is that they are very capable language users whose language skills help them make a living. Because the forms come from powerful social groups who are respected in society, also the forms are perceived as prestigious (Milroy, 2001: 532).

Standardisation results in two main outcomes. Alongside developing the most important and prestigious language variety, it also results in diminishing the importance and meaning of other varieties (Milroy, 2007: 138). While the varieties which lose their significance are geographical dialects, the standard variety, gaining in importance, is a social construct (Trudgill, 1999: 124). In the case of the UK, such stigmatised varieties became urban non-standard varieties (Milroy, 2001: 548). Unlike rural non-standard varieties, in the past scholars did not view urban non-standard varieties as contributors to the present state of the standard language, and so they were severely stigmatised (Milroy, 2001: 548). An example of such an urban non-standard dialect is Tyneside English, the variety which has been used for the purpose of this thesis.

Even though Standard English is only one of many language varieties, to an average language user it is equivalent to the language itself, and other varieties are compared to it (Agha, 2004: 24; Milroy, 2001: 539; Milroy, 2007: 136). A number of bodies known as language academies have been also established to guard purity and correctness of the standard language in various countries. Perhaps the best known of them is the Académie française. There are also other guardians of language purity, who can be often found in papers and magazines complaining about speech, grammar and vocabulary of other speakers, often younger ones.

As has been mentioned, standard language is believed by language users to be the correct variety (Milroy, 2007: 134). The educational system and public institutions spread and support standardisation, ensuring that at the same time non-standard varieties do not receive any such support (Agha, 2004: 24; Milroy & Milroy, 1985: 36).

Milroy (2007: 133) argues that in a number of countries where standard languages are spoken, speakers also live in the culture of standard language, which affects their attitudes towards language. It is usually the case that speakers are unaware of this correlation between standard language and language attitudes. One of the common beliefs is that standard language is the correct variety. Another such popular belief is that change and variation in language lead to "decay and corruption" (Milroy, 2007: 139; Milroy, 1999: 175) by jeopardising the only correct form sanctioned by standardisation.

2.2.1 Attitudes

Attitudes and beliefs such as those mentioned above lead to discrimination against speakers who refrain from or cannot use standard forms (Milroy, 2007: 135). They are also responsible for the formation of beliefs and stereotypes about certain social groups on the basis of the language the speakers in these groups use. This would not be the case if variation ceased to exist. Nevertheless, despite the efforts of standardisation and centralisation processes, variation is present in the language. While most linguists view all languages and dialects including terms of correctness as equal (Ager, 2003: 51), speakers living in standard language cultures tend to perceive the standard variety as more prestigious than any non-standard variety (Milroy, 2007: 137; Milroy, 2001: 532). It seems that, in general, language users hold prescriptivist attitudes towards language, even though they also use non-standard forms. At the same time, however, speakers from working class backgrounds are sometimes harsh judges of localised variants. It is lower-class speakers, but not only, who view localised variants as stigmatised. They sometimes also attempt to compensate for using them by hypercorrecting in more formal contexts (Gal, 1989: 350).

Perceptual dialectology and folk linguistics, as well as sociolinguistics, have investigated speakers' attitudes towards language (Montgomery, 2015; Pearce, 2012; Preston, 1993) and provided us with evidence supporting the above claims.

Milroy (2007: 135) and Milroy (1999: 178) point out that because it is not "politically correct" for speakers to show prejudice against others on the basis of their ethnicity, gender or social class, instead people comment on their language. By doing so, they in fact disclose their prejudice against social groups which are already disadvantaged, for example, working-class people or ethnic minorities (Milroy, 1999: 178). This way, users of a typically more prestigious or mainstream language variety manage to discriminate against speakers of a less prestigious variety, effectively discriminating against their ethnicity, gender or social background. As a result, language forms used by these groups of speakers become stigmatised by other community members (Milroy, 2007: 137). Gal (2006) provides the example of a Hungarian language minority in Austria, whose speakers are discriminated against the mainstream Hungarian language users for the use of obsolete forms. Another example constitutes German minority speakers who go to Germany from Hungary to improve their economic conditions. However, on their return, they are discriminated against by fellow minority speakers for using mainstream language forms.

Which forms are stigmatised in the UK? Milroy (2004: 167) claims that while in the US the society is divided according to race and ethnicity, in the UK it is divided according to social class. Nevertheless, the two divides are true in both the US and the UK. These divides are reflected in the language ideologies in the UK and US. Milroy (1999: 178, 183) further points out that the most disadvantaged social groups in the UK are working-class speakers, who de facto are users of non-standard urban varieties. It is these varieties that are stigmatised by society the most.

Due to attitudes shaped by standard language ideology, language speakers discriminate against other speakers on the basis of non-standard forms the other speakers use. However, information about the social standing of the speaker is not the only type of information accessible from speech. In their studies Labov (1966; 1972) and Trudgill

(1974) have shown that the use of linguistic variables reflects social attributes of the speaker which may include age, gender, social class or ethnicity. Effectively, linguistic forms index social information (gender, social class or ethnicity) (see Section 2.2.2). It could be argued that these linguistic forms become symbols of certain people and activities. Nevertheless, it should be pointed out that what is being indexed by a linguistic feature is not necessarily universal but specific to a place and community. The same phonetic forms may have indexical meaning in one place but none whatsoever in another place.

It seems that it is not only standard language ideology which is responsible for the formation of language attitudes. In general, attitudes are formed as a result of the interaction between a language form and its social meaning, that is, indexicality. It has been mentioned that while standard language ideology gives certain forms a privileged status, others become stigmatised as a result of the same ideology. However, there are also other processes, in which social groups of speakers or individual speakers themselves give linguistic forms specific social meanings not only in reference to the social class of the speaker but also their other social dimensions such as ethnicity (Purnell et al., 1999), gender (Foulkes et al., 2010), sexual orientation (Podesva, 2011), localness (Johnstone et al., 2006) or activities that the speakers engage in (Irvine & Gal, 2009: 403). Such forms are used in normal, everyday situations. Linguistic anthropologists and sociolinguists provide similar explanations for the existence of indexical forms. Irvine and Gal (2009: 403) argue that once language users become aware of these linguistic forms and their social meaning, they form language ideologies which help them justify the existence of these forms. Bucholtz and Hall (2010: 21) also claim that indexicality grows out of ideology. Milroy argues that ideologies supporting the use of non-standard variants may be created as a reaction to standard language ideology (Milroy, 2004: 170). Later, on the basis of these ideologies, listeners evaluate and judge speakers in social terms. It should be also pointed out at this point that when there are two variants, of which one is standard and the other one stigmatised, it is the stigmatised variant that becomes indexical (Trudgill, 1986). As a result, speaker evaluation will be in negative terms.

In my thesis I will try to show that these attitudes are still present in society. Across a series of three experiments listeners with high- and low-exposure to Tyneside English, a localised urban dialect used for the purpose of this thesis, will be asked to evaluate localised as well as non-localised variants in terms of perceived social-indexical information of the speaker.

2.2.2 Indexicality

There are correlations between linguistic forms and their social meaning. Specific language forms are associated with certain groups of speakers (Woolard, 2008: 437). Using vocal fry used to be specific to young women in California in the same way as using *tej* [teɪ] is a dialectal feature typical of speakers in the Poznań (Poland) area. Nevertheless, indexicality is time-, context- and place-specific. Indexical relationships change and do not refer to all speakers at the same time.

In order for an indexical form to fulfil its role, it needs to be understood by the speaker and listener in a similar or identical way. In other words, it needs to be perceived by the listener and speaker as indicating the same set of social information. This makes perfect sense, since the linguistic exchange between the speaker and listener takes place in a social context and if the speaker uses a phonetic variant indexing his or her social class or locality, it needs to be decoded by the listeners as indexing these types of social information. If this condition is not met, the whole *purpose* of an indexical form is lost. The question we can further ask is whether using indexical forms is a conscious attempt on the part of the speaker to signal indexical information. Given that an indexical form may be used by a speaker if he or she wishes to identify with a certain social group or behaviour, it could be claimed that this is at least partially a conscious process. The fact that speakers are aware of the fact that listeners respond to indexicality also supports the claim of the conscious use of indexical forms.

There are different levels of indexicality, in the sense that indexical meaning can be attributed to entire languages and dialects but also smaller linguistic units, such as phonetic variants. This thesis, for example, investigates perception of indexical information from phonetic variants.

Johnstone (2010) explains how linguistic forms become indexicalised: that is, whom listeners begin to associate certain forms with. Johnstone claims that in order for a certain form to obtain social-indexical meaning, listeners need to be "told" that a particular form has such a social meaning (Johnstone, 2010: 32). For example, there may be two pronunciations of a word. While one is a standard realisation and the other a non-standard, each of them is used by a distinct social group. As such, the pronunciation indexes belonging to the group. Johnstone et al. (2006) provide monophthongisation of /aʊ/ (i.e. the

MOUTH vowel) as an example of a linguistic feature indexing localness, working class and gender in Pittsburgh, US. Using Silverstein's (2003) classification, the authors distinguish three steps in the development of indexicality: first-order, second-order and third-order indexicality (Johnstone et al., 2006: 82-84). Johnstone et al. described the process with reference to the example of a Pittsburgh realisation of the diphthong /aʊ/. In first-order indexicality, the socially mobile speakers are aware of the fact that monophthongal /aʊ/, for example in *house* realised as [ha:s], indexes a working-class, usually male speaker. Johnstone et al. argue that this type of information is not accessible to socially non-mobile speakers who are not exposed to, nor aware of, any other realisations of /aʊ/. In the second-order, the awareness of the interaction between the linguistic feature and its social meaning has spread in the community. In consequence, the speakers, as well as other members of society who do not use the feature, become aware of what it indexes. Furthermore, speakers start to notice the two forms and consciously monitor their use, depending on whether they wish to index local identity or distance themselves from it. In the final step, when the third-order indexicality is achieved, the linguistic feature becomes representative of the local identity. Speakers as well as listeners are aware of the relation between the feature and its social connotations. It will be also noticed that a shift has occurred in terms of social information being indexed. The linguistic feature does not signal working-class information any more instead, it indicates the localness of the speaker (Johnstone et al., 2006: 93, 99).

Johnstone et al. (2006) also compare Silverstein's classification with Labov's (1972) and Agha's (2003). Labov distinguishes three steps of development of a linguistic feature into indexicality, which are *indicator, marker* and *stereotype*. Agha, on the other hand, talks about enregisterment of language forms in terms of social meaning they acquire.

2.2.3 Speaker identity

The presence of indexical forms in language is a result of the existing variation. It is by now a well-known fact that indexicality is closely linked with speaker identity. Indexical forms are used by speakers to construct their identity and indicate their belonging to a specific social group (Bucholtz & Hall, 2010: 21; Eckert, 2000; Eckert 2012: 92; Eckert & Podesva, 2011). Joseph (2010: 9) provides a further interpretation of the process, claiming that linguistic forms which speakers decide to use create the speakers. It is through the language that speakers “make themselves”.

This and the previous sections attempted to demonstrate that identity construction is supported by language ideology. Just as indexicality can index different social groups, so in the same way can different types of identities be constructed. Speakers choose to identify themselves with a particular group, be it Tynesiders, Mancunians, a church group or a book club, and they want to sound this way. This also explains why people insist on using non-standard stigmatised forms, even though the awareness of standard language ideology is present in almost every aspect of human life, and speakers of different social strata are able to differentiate between standard and non-standard forms. It seems that the rationale behind using non-standard forms is a result of alternative ideologies speakers develop in response to standard language ideology (Milroy, 2004: 170).

Speakers seem to be aware that the listeners are able to retrieve information about, for example, which social group the speaker belongs to from his/her speech. The aim of this section was to show that standard language ideology (and stemming from it), language ideologies justifying the use of non-standard forms, indexicality and speaker identity ideology are very much interconnected. A number of studies have investigated language production in the light of these theories. However, this thesis focuses on standard language ideology from the perspective of language perception and interprets the results within this theoretical framework.

2.3 Multimodal speech processing

In the widely understood process of perception the human brain uses a number of senses, that is vision, hearing, touch, taste and smell (O'Callaghan, 2012: 95). In fact, multiple senses are used in any process of perception. One such example is speech perception. It may seem that speech perception predominantly relies on audio information, after all most of us would not be able to understand the message without hearing it and we experience no perceptual difficulty when listening to the radio. Nevertheless, listening to the radio is actually not a natural way of communicating, also we usually see our interlocutor when we talk to him or her. Rosenblum argues that "speech is inherently multimodal" (Rosenblum, 2005: 53). Indeed, speech perception is a complex process during which the listener uses information coming from at least two different channels. These two channels provide audio and visual information which is put together into a unique perceptual event. As a matter of fact, audio and visual information becomes inseparable. The *McGurk effect* (McGurk & MacDonald, 1976) is one of the first and well-known examples proving evidence in support of the above claim.

Nevertheless, for more than sixty years speech perception was investigated as a unimodal process focusing only on acoustic information in natural and synthetic speech (Bernstein, 2012: 22; Meng et al., 2009: 401; Remez, 2012: 4). As a result, other modalities such as for example the visual channel were neglected. In fact, the unimodal approach is a simplified way of investigating the phenomenon and it is far from natural speech perception where the listener can see the speaker and thus perceive speech multimodally. Whilst unimodal acoustic speech perception has been investigated quite extensively, it is multimodal speech perception that we still need to learn more about. After the McGurk and MacDonald discovery the research interests shifted towards audio-visual and visual-only perception (Bernstein, 2012: 21) and even though vision's impact on speech perception has been recognised we are yet to fully understand how seeing the speaker while listening can benefit perception (Remez, 2012: 4).

Why the visual channel? Vision is the dominant sense in the process of perception (O'Callaghan, 2012: 98). Therefore, it is not surprising that visual information seems to be the most important channel in addition to the audio channel in speech processing (Rosenblum, 2005: 52). There are multiple examples of studies which show that looking at the speaker improves speech perception in any type of conditions (Remez, 2012: 4; Rosenblum, 2005: 52). There is evidence that seeing the speaker's face or lips increases intelligibility (Sakamoto et al., 2014). There is also evidence that seeing the upper part of the speaker's face and head enhances perception in normal hearing conditions (Davis & Kim, 2006). Moreover, word recognition in noise is higher from audio-visual speech than from audio only (Buchwald et al., 2009). Finally, audio-visual speech enhances phonetic learning in infants (Teinonen et al., 2008).

At the same time we can notice from the *McGurk effect* that visual cues can shift our perception of audio information. Nevertheless, audio information can also impact what we see. Thus it is rather clear that speech perception relies on information from the combined channels. So how does it work? When may visual cues shift perception of the audio signal?

Therefore, experiments presented in this thesis attempt to address the shortcomings of traditional automatic speech processing which focused on the audio channel when investigating the phenomenon. Furthermore, this thesis attempts to improve our understanding of multimodal speech perception by investigating the role of visual information on perception of the acoustic signal. Specifically, it investigates if and how visual information about the speaker may affect perception of speaker social-indexical information from the acoustic signal.

2.4 Exemplar theory as a model of speech perception

Exemplar theory is a model of speech perception based on exposure. In this model memories of words (exemplars) are stored in the long-term memory along with the contextual information about the utterance and information about the speaker (Pierrehumbert, 2001; Johnson, 2006). While speaker characteristics include speaker-indexical information: age, gender, social class, etc., contextual information may include background noise, gestures performed by the speaker, or visual information. These types of linguistic and non-linguistic information are, in turn, accessed during speech recognition.

A theory similar to exemplar theory was already proposed in 1909/1923 by Semon (Schacter, Eich & Tulving, 1978 as cited in Goldinger 1998). In this model every phonetic representation of the word "left a unique memory trace" (Goldinger, 1998: 251). When a new word was heard, memories of words stored in the mind were activated. The activation process depended on how similar the existing exemplars were to the new word.

As was the case in the older model, in the present model exemplars stored in the long-term memory are activated during speech recognition (Johnson, 2006: 494; Pierrehumbert, 2001). Listeners hear thousands of words and store them in the long-term memory. Pierrehumbert (2001: 140) argues that these exemplars are stored as clouds of categories which, in turn, are made of individual memories of words. Exemplars of words which are phonetically similar are stored together while exemplars which are phonetically different are separated and stored under different categories (Pierrehumbert, 2001: 140). For example, the category of female speaker is constructed from exemplars of female speech experienced by the listener (Johnson, 2006: 494). However, De Schryver et al. (2009), argue that not all exemplars need to build a category. Instead, only a subset of exemplars is enough to build a category.

Johnson (2006: 493) and Pierrehumbert (2001: 141) also claim that during speech processing not all exemplars are activated but only the most recent ones. The older exemplars, on the other hand, fade away and are not activated to the same degree. At the same time, categories which are frequent are easily activated and do not need many exemplars to be stored in them (Pierrehumbert, 2001). As the language user encounters new exemplars, they are compared with the existing ones on the basis of how similar they are in terms of phonetic properties and classified under appropriate category (Pierrehumbert, 2001).

A number of studies in psychology and linguistics have been carried out within the framework of the exemplar theory. Lacerda (1998) investigated language acquisition in infants. Luce and Lyons (1998) examined whether representations of words were abstract or stored as exemplars. If word representations were abstract, a change of speaker and no change of speaker would not impact the recognition process of words. If, on the other hand, words were stored as exemplars, a change of speaker would result in slower processing and lower accuracy of word recognition. Luce and Lyons compared reaction times and word recognition accuracy in three conditions (same word, different words and new words) and reported that the reaction times were lower in the same words condition.

Later, Johnson (2006) investigated perceptions of speaker gender. He argued that when the listener heard a stereotypically female voice, female exemplars were activated during speech recognition (2006: 494-496).

In a recent study, Pufahl and Samuel (2014) investigated indexical effects of environmental sounds heard during speech perception. They argued that apart from speaker-indexical information other indexical information included also background noise or environmental sounds, such as an alarm clock. The results showed that words were identified with lower accuracy when there was a change in the voice or the environmental sound. Therefore, Pufahl and Samuel argued that the indexical effects observed for environmental sounds were similar to the ones for words (Pufahl & Samuel, 2014: 23).

Indexical effects in speech perception have been investigated also in sociolinguistics, for example by Hay and Drager (2010), Hay et al. (2006b) (see Section 2.6).

I will be taking on board insights from exemplar theory, although I will not be embracing it completely.

The following section provides an overview of research on F0 in perception of speaker gender

2.5 Voice gender identification

Previous studies on voice gender identification from phonemes (Lass et al., 1976, Hillenbrand & Clark, 2009) reported high accuracy of identification. This fact seems quite obvious on the face of it, since people in general have no difficulty recognising whether they are talking to a man or a woman in everyday situations such as, for example, talking on the phone. People are also usually able to identify the gender of a radio news announcer. Listeners decode this type of information easily even if not infallibly. However, recognition and identification of voice gender is more complex than it appears to be. Every now and then we may come across a speaker on the radio whose gender is not easy to recognise right away and we actually need to stop and think. Nevertheless, even in such situations we reach a decision fairly quickly. How do we arrive at it? What phonetic cues do we tune into?

While some of the differences in voice features between speakers are socially conditioned and, as such, are learned and acquired by speakers, others are a result of anatomy, for example differences in the lengths of the adult male and female vocal tracts. There is a correlation between mean fundamental frequency and the length of the vocal tract (Mackenzie Beck, 1997: 280). Voice fundamental frequency is one of the features correlating with the length of the speaker's vocal tract. The shorter the vocal tract, the higher the fundamental frequency of the voice tends to be, and vice versa. While the average length of an adult male vocal tract is 17 cm, a female vocal tract is around 14 cm long (Ladefoged, 1996). The average fundamental frequency in European languages for male speakers is 120 Hz and 220 Hz for female speakers (Laver, 2002: 451). In conversational speech, the maximum pitch values are between 50 to 250 Hz for male speakers and between 120 and 480 Hz for female speakers (Laver, 2002: 451). These numbers show an overlap in pitch frequency ranges between the two genders. As a result, a male voice with higher pitch could be mistaken for a female voice of lower pitch, and the other way round (Biemans, 2000; Foulkes et al., 1999). Furthermore, individual pitch ranges fluctuate during the speaker's lifetime. While fundamental frequency can exceed 500 Hz (from about 390 Hz) in babies, it drops to between around 450 and 350 Hz in young children and then further to about 200 and 230 Hz in teenagers between 15 and 17 years of age (Mackenzie Beck, 1997: 281). The ranges stabilise for adult speakers, but then they shift again in advanced old age. In females, a drop in pitch can be observed, while in males the opposite process takes place.

Ranges of fundamental frequency can vary also depending on the culture a speaker lives in. For example, it has been reported that in New Zealand the mean fundamental frequency of Maori speakers, was significantly higher than that of Pakeha speakers (New Zealanders of European origin) (Szakay, 2006; 2008). A correlation between age and ethnicity was observed, as Maori mean pitch increased with age (Szakay, 2006). This finding was of significant interest in light of the previous research on pitch differences between genders, which usually indicated that the differences are due to anatomical differences in pharynx and vocal cord structure. Szakay (2006) argued that these differences could be a result of social and cultural influence.

As far as the audio stimuli used in this thesis are concerned, when shifting fundamental frequency (F0) in Adobe Audition (Adobe, 2007) to obtain gender-ambiguous-sounding samples, formant frequencies (FF) were automatically adjusted to changes in pitch (cf. Section 3.2.1). However, the user manual was not very explicit in explaining the process[1]. Whilst it could be argued that manipulations applied to the stimuli could have been more controlled, it was decided to shift FF along with F0 in the interest of the auditory naturalness of the stimuli. For example, Assmann et al. (2006) and Assmann and Nearey (2007) argue in favour of a correlation between FF and F0 in natural speech. Furthermore, manipulating F0 alone seems not to result in a shift of perceived speaker gender. Hillenbrand and Clark (2009) reported that male voices with raised F0 were not reported by listeners to sound female. Finally, Klatt and Klatt (1990: 851) established that raising F0 with no additional aspiration noise results in a more nasal sounding voice.

[1] The author attempted contacting Adobe Audition to find out what algorithm was applied by the software (https://forums.adobe.com/message/8263211). However, no response was obtained. It is more than likely that the company wanted to protect the code. Therefore, measurements of F1 and F2 for open and close vowels for both speakers were performed. The results showed that in manipulated stimuli there was about 10% shift of F1 and F2 with reference to original recordings.

Previous research suggests that as far as investigations of the role of F0 on speaker gender discrimination are concerned there are two opposing views. While part of the body of research points to F0 as the main cue used by listeners to identify the gender of the speaker, the other part of the research provides counter evidence, arguing that listeners are, in fact, able to identify the gender of the speaker in the absence of acoustic information typically provided by the speaker's F0. The studies presented below attempt to establish which of the phonetic cues, F0 or FF, cue the gender of the speaker to a more consistent degree.

One of the earliest studies which found evidence in favour of fundamental frequency in identification of speaker gender was Coleman (1976). Coleman combined fundamental frequency and FF of opposing sexes to determine whether listeners relied on F0 or FF when determining gender of the speaker. Isolated vowels were produced with a laryngeal vibrator at 120 and 240 Hz, which represented typically male and female frequencies. The results differed for male and female voices, showing that listeners did not rely on F0 consistently and that in some cases fundamental frequency alone did cue speaker gender information but in others it did not. While male F0 was a salient cue to speaker gender identification, it was not the case for female F0. Identification of female voices, whether based on female F0 or female vocal tract resonance characteristics, was not as successful as identification of male voices. Thus, even though the role of F0 in speaker gender identification was inconsistent, the study reported a high accuracy of identification for male voices on the basis of F0 alone.

Another early study, by Lass et al. (1976), also argued in favour of F0 rather than FF being the most prominent cue to speaker gender identification. As in Coleman's study, speaker gender was identified from isolated vowels. Male and female speakers produced vowels in three conditions: modal, whispered and low-pass filtered speech at 255 Hz. Lass et al. found high accuracy of voice gender identification in normal speech and low-pass filtered vowels where information about FF was inaccessible to the listeners. The results showed that in the modal voice condition the accuracy of speaker gender categorisation was 96%, while it was 75% in the whispered condition and as high as 91% in the low-pass filtered condition. Because the results in low-pass filtered condition were more accurate than in the whispered condition, the authors concluded that it was fundamental frequency that is the main carrier of talker gender information. Yet it is worth drawing attention to the fact that, overall, identification of female voices was less accurate than that of male voices.

Murry and Singh (1980) reported that when identifying the gender of the speaker, listeners relied on different phonetic cues depending on the gender of the voice and on the type of stimuli, such as a phrase or a vowel. They reported F0 to be the main cue when listeners differentiated between speakers of the same gender, for example between individual speakers in a male speaker group, irrespective of whether the stimulus was a phrase or an isolated vowel. In the female condition, especially in the case of phrase stimuli, voice quality was a more informative cue than fundamental frequency.

A later study, which reported results supporting the role of fundamental frequency in gender identification, was Gelfer and Mikos' (2005) study, which also investigated identification of speaker gender from isolated vowels. Vowels were produced by male, female, as well as transgender speakers who were in the process of changing gender from male to female. The fundamental frequencies of the transgender speakers had to be at least 165 Hz when producing vowels. The results showed that even when there was a mismatch between F0 and FF in terms of gender, listeners still relied on information carried by F0. When male FF were matched with female F0, vowels were judged to be produced by a male speaker only 19.3% of the time. Stimuli with female F0 and FF were identified correctly as female 73.8% of the time. When F0 was substituted with male values but FF remained female, vowels were rated as produced by a female speaker just 20.2% of the time. Finally, when the FF of the transgender stimuli were matched with typically female F0, these stimuli were rated as female-sounding 84.0% of the time. On the other hand, when transgender FF were matched with typically male F0, only 16.7% of the stimuli were rated as female-sounding.

Hillenbrand and Clark (2009) also investigated the influence of F0 and FF on perception of speaker gender. Hillenbrand and Clark re-synthesised vowel stimuli which were originally produced by male and female speakers. Both F0 and FF were manipulated in such a way that they had values for the opposite genders. Hillenbrand and Clark investigated vowels. They argued that using more than one vowel in fact made using FF in speaker gender identification challenging, since listeners were unable to rely on formant frequencies alone if they did not know what vowel was being produced (Hillenbrand & Clark, 2009: 1150). While F1 and F2 frequencies are tightly linked with vowel quality so that it is quite difficult to discriminate speaker gender on the basis of these values alone, F3 is not as strongly linked with vowel quality, and so it serves the purpose much better (Hillenbrand & Clark, 2009: 1153). Hillenbrand and Clark reported that information carried by F1, F2 and F3 frequencies together were only slightly worse predictors of speaker gender than fundamental frequency. While F0 yielded about 96% accuracy in speaker gender identification, FF gave a slightly worse result, at about 92%. Yet Hillenbrand and Clark pointed out that to obtain this result, listeners needed to have a point of reference and to know which frequencies were low and which were high in the token which was pronounced (Hillenbrand & Clark, 2009: 1154). Knowing this helped in vowel identification. Additionally, Hillenbrand and Clark discovered that manipulating only F0 often produced the effect of the opposite sex. This effect was especially strong in syllables produced by females, whereby upon shifting F0 these syllables were often perceived by listeners to be male-sounding (2009: 1161). Overall, in terms of gender identification F0 and FF together gave better results than F0 or FF alone. Thus, Hillenbrand and Clark also argued that there is a mutual interdependence of F0 and FF (2009: 1162) and that in fact, recognising the vowel produced depends on prior gender identification (2009: 1150). On the other hand, there are no such obstacles when it comes to gender recognition from F0, since F0 depends on the sex of the speaker. Nevertheless, if vowel identity is known to the listener, FF work better than F0 in cueing speaker gender (Hillenbrand & Clark, 2009: 1150).

Barreda and Nearey (2012) also investigated the correlations between F0, FF and non-linguistic features such as voice gender. The findings revealed that in fact F0 influenced perception of speaker gender. The authors created a seven-step continuum between /ʌ/ and /æ/. Vowel FF were manipulated in such a way that while F1 and F2 of /ʌ/ had male values, /æ/ had female values. In addition, three F0 levels were applied: one at 120 Hz which reflected typically male F0 values, one at 240 Hz which reflected typically female values, and one which was the mean of the two extreme values. F3 values, along with other higher formants, were also set to three different conditions. The lowest F3 value, was typical for male speakers, the highest typical for female speakers, and the middle F3 value was a mean of the highest and lowest values. Furthermore, in order to ensure that the stimuli sounded more natural, breathiness was added to the female stimuli such that not just high F0 was characteristic of female voices. Listeners were asked to categorise vowels as /ʌ/ (*hud*) or /æ/ (*had*), but also to judge the gender of the speaker. This enabled gaining an insight into how F0-related changes in vowel quality influence perceived changes in speaker characteristics. The results of the study revealed that F0 values corresponded strongly to listeners' judgements about speaker gender and as such clearly affected perceived vowel quality. For example, upon hearing high F0, listeners rated more vowels as /ʌ/. By contrast, listeners rated fewer vowels as /ʌ/ when the stimuli had the lowest F0. However, when information about speaker gender was present, the effect of F0 became less significant. Therefore, Barreda and Nearey concluded that F0 influenced perceptions of vowel quality on the condition that it influenced listeners' expectations about the speaker. In other words, the speaker's expected identity had an influence on perception of vowel quality. As a result of their findings, Barreda and Nearey argued in favour of there being an indirect relationship between F0 and vowel quality. This relationship influenced listener perceptions of speaker characteristics, such as perceived speaker gender indirectly changing perception of vowel quality.

Coleman (1971), attempted to establish whether it was possible for listeners to distinguish voice gender in the absence of acoustic information typically present in the speaker's fundamental frequency. Stimuli used in this study consisted of two isolated vowels /i/ and /u/. Male and female speakers produced the vowels using an electrolarynx operating at a constant fundamental frequency of 85 Hz. This technique resulted in producing the effect of oesophageal speech, a type of speech usually developed in patients who have had their larynx and vocal cords removed. Oesophageal speech is characterised by low and flat F0 between 50 and 100 Hz. Despite applying flat fundamental frequency at 85 Hz listeners were still able accurately to identify the gender of the speaker. Overall, the gender of 88% of the speakers was identified correctly (Coleman, 1971: 570). Nevertheless, listeners were better at identifying male speakers than female speakers. As will be recalled, other studies described in this chapter also reported higher identification rates for male stimuli in comparison to female stimuli. Coleman's results support the hypothesis that gender-specific acoustic information does not rely heavily on fundamental frequency. When equal values of F0 were applied to male and female speech, listeners relied on the vocal tract resonances of individual speakers. Coleman argued that while F0 is self-evidently a parameter differentiating male and female acoustic performance, FF may be the second most important parameter (Coleman, 1971: 576).

Later research addressed the issue of the higher accuracy of male voice identification over female voice identification which was signalled in some of the reported studies. A seminal study by Klatt and Klatt (1990) reported aspiration to be the key element to identification a voice as female. This finding was supported by Johnson et al. (1999), who found out that voices judged as the most stereotypically female were more breathy than other female voices, even though the fundamental frequency was lower than in the case of the voices judged as less stereotypically female. Biemans (2000) also reported breathiness as a typically female feature in a number of languages including American and British English.

While in the past male voices have been synthesised successfully, this cannot be said about female voices. In their attempts to synthesise a natural-sounding female voice, Klatt and Klatt established that it was breathiness rather than F0 which was chiefly responsible for perceived femaleness of the voice. It should be also kept in mind that while typically female voices are on average more breathy than male voices, there is variation within genders (Klatt & Klatt, 1990: 853). Thus, some males can sound more breathy than average male voices whereas some females can sound less breathy than average male voices (Klatt & Klatt, 1990: 852, 853). It could be concluded from Klatt and Klatt's study that in fact, perception of voice gender depends on an array of acoustic features which include fundamental frequency and FF as well as breathiness (Klatt & Klatt, 1990).

A study by Assmann et al. (2006) also investigated the correlation between F0 and vowel formant frequencies (F1, F2, F3) in the perception of speaker gender and naturalness of speech. Their study was different from the previous studies discussed in this section in the sense that it investigated perceptions of whole sentences rather than isolated vowels. Most importantly, however, Assmann et al. (2006) claimed that in natural speech there is a correlation between FF and F0. They also claimed that listeners are subconsciously aware of this relation. Two male and two female speakers produced two sentences each. Ten steps of manipulated F0 were matched with ten steps of manipulated FF in such a way that high F0 was matched with high FF, low F0 was matched with low FF, and finally values of F0 were cross-matched with values of FF.

Listeners evaluated the perceived gender of the speakers of the stimuli and the naturalness of the stimuli using a graphical slider scale. While gender was identified accurately in sentences matched for gender, sentences with F0 and FF cross-matched for gender were rated near the midpoint of the scale, which implied that they were perceived as gender-ambiguous. Furthermore, Assmann et al. (2006) drew attention to the fact that sentences originally produced by male speakers were perceived as overall more male-sounding, while sentences originally produced by females were rated as more female-sounding. This was also true in conditions where F0 and FF had been manipulated. Thus, Assmann et al. (2006) concluded that factors other than F0 and FF influence the perception of speaker gender. As far as evaluations of the naturalness of the manipulated sentences were concerned, F0 and FF matched for gender were rated as more natural-sounding than if F0 and FF were mismatched for gender (Assmann et al., 2006: 891).

A following study by Assmann and Nearey (2007) also investigated how F0 and FF contributed to the perceived naturalness of the speaker's voice. As per Assmann et al. (2006), Assmann and Nearey argued in favour of a relation between F0 and fundamental frequency that listeners were aware of. Mean female F0 and FF values, as well as mean male F0 and FF values, were used to synthesise vowels. F0 and FF were mismatched for gender. Listeners were asked to identify the most natural-sounding voice in a continuum. The choices provided in 25-step continua included only F0 and FF values naturally found in human voices. Also, the middle point of each continuum had values occurring originally in the natural voice (Assmann & Nearey, 2007: 36). The results showed that, for example, male voices with female F0 were matched with female FF, and so on. Again, these results also provided compelling evidence of listeners being aware of the link between F0 and FF in natural speech.

In a follow-up experiment, F0 and FF values were shifted to make the voices sound gender-ambiguous. That is, they were set to the middle values on the continua. In these conditions, listeners selected FF which were also in the middle ranges between male and female averages, supporting Assmann and Nearey's hypothesis. The results showed that in each experiment the combinations of F0 and FF provided by listeners reflected similar natural voice characteristics.

Hubbard and Assmann (2013) took their investigation of the role of F0 in a slightly different direction. Their study investigated the influence of F0 on speaker identification and listener adaptation to voice gender. Their findings showed that F0 was not a main cue in a gender identification task, yet it was a main cue in a gender adaptation task (see below). Adaptation to speaker F0 resulted in a shift in gender identification of test stimuli.

Stimuli were VCV syllables produced by male and female speakers and resynthesized for the experiment. There were two conditions: in the first condition, F0 was present in the male and female voices, whereas in the second condition, it was substituted with broadband noise, giving the effect of whispered speech. After hearing the test stimuli, listeners were prompted to identify the speaker as female or male. The results showed that in the F0-removed condition, there were no after-effects of the adaptors on speaker gender identification. However, the results for the F0-present condition showed that adaptation occurred and female adaptors resulted in a lower number of female categorisations, while male adaptors produced lower numbers of male identifications. When gender-neutral adaptors in the F0-present condition were used, test tokens were categorised in between male and female values. Comparing the findings in the two conditions allowed the authors to argue in favour of the role of F0 on voice gender after-effects.

This section presented research providing evidence in favour of F0 or F0 and FF as parameters used in perception of speaker gender. The following section presents an overview of sociophonetic research conducted to investigate perceptions of speaker indexical-information.

2.6 Perception of speaker-indexical information

This part of the chapter provides an overview of the body of literature on perception of sociophonetic variation. The first part addresses studies investigating recognition and identification of speaker-indexical information from speech. For example, previous socioperceptual studies focused on identifying speaker-indexical information such as ethnicity (Purnell et al., 1999), geographical origin (Clopper et al., 2005) or age (Drager, 2005; Walker, 2007). In the second part of the section, the focus shifts to the issue of the influence of exemplar priming on perception of speaker-indexical information. Research has shown that speech perception is influenced by the social characteristics of the speaker such as age, gender or regional origin (Drager, 2011; Hay & Drager, 2010; Niedzielski, 1999; Strand, 1999).

2.6.1 Identifying speaker identity from acoustic information

One of the studies which showed that listeners are able to determine the individual identity of the speaker from phonetic information alone was a study by Remez et al. (1997). Stimuli had intonation and voice quality filtered out. Thus, the samples did not have properties of natural speech but were comprehensible. Male and female speakers provided the sentences used in the experiment. Three different sentences were played, two manipulated and one natural. A natural sentence and one of the manipulated sentences were produced by the same speaker. Listeners were asked to match sentences produced by the same speaker. The results showed that even though listeners did not know the speaker, they had no problems identifying him or her.

A later study by Purnell et al. (1999) showed that in the absence of visual information listeners were able to make correct judgements about the ethnic and dialectal background of unknown speakers on the basis of a single word *hello*. Purnell et al. showed that speakers were discriminated against based on the dialect they spoke, which meant that their ethnicity was accurately recognised from speech. The same speaker, who was of African American origin and spoke African American Vernacular English (AAVE), Chicano (Mexican American English) and Standard American English, produced the tokens. Listeners were instructed to identify the dialect of the each of the tokens. The results showed that on hearing a single word, listeners were able to identify a dialect correctly more than 70% of the time.

In a later study, Foulkes et al. (2010) investigated the perception of gender identification from child speech. One of the obvious characteristics of child speech is that the F0 does not carry gender-specific information to the same degree as in the case of adults. In this study, indexical information was recognised at the segmental level from phonetic variants embedded in single words. Foulkes et al. (2010) point out that sociolinguistic variation is learned by children at early stages of their lives and so even young children vary in their use of phonetic variants. Three boys and three girls, aged 3;0-4;1, provided natural samples for the experiment.

Three groups of listeners participated in the study. The first group was formed of Tyneside listeners who had high exposure to the variants used in the study, since they were speakers of the investigated dialect. The second group was formed of native British English speakers from different parts of the country, and the last group consisted mostly of residents of Tucson, Arizona. While the first group of listeners was aware of phonological variation in their own area, this knowledge decreased with every following group of listeners. Listeners were asked to categorise the speaker as a girl or a boy. They were also told the children were from Newcastle, UK.

The results revealed that the accuracy of identification was very similar for all three groups of listeners: 48.7% for the Tyneside listeners, 49.4% for the non-Tyneside UK listeners, and 46.5% for the American listeners. However, word-medial laryngealised tokens, which were most often produced by male speakers, were less often categorised by Tyneside listeners as produced by girls. Other groups of respondents did not follow this pattern. [^{h}t], which is most often produced by female speakers, was overall less often categorised as a variant produced by girls than boys by all three groups of respondents. However, only in the Tyneside group of respondents, was girl identification not significantly lower than boy identification.

A study by Drager (2005) investigated the influence of speaker age on perception of speech. Specifically, she tested whether voice age had an effect on perception of vowel boundaries and found evidence to support this claim. For the purpose of Drager's study a DRESS-TRAP continuum was created by shifting values of F1 and F2. A ten-step continuum from *bad* to *bed* for two male and two female speakers was created. Listeners were instructed to categorise the tokens they heard as either *bad* or *bed* and to assess the age group of the speaker upon hearing the word *bad*.

In New Zealand English younger male speakers lead the raising of the TRAP vowel. Therefore, younger male speakers use more raised variants of the TRAP vowel, whereas variants used by older speakers are not as raised (Drager, 2005: 61).

The results showed that when listeners found a speaker to sound older, they categorised fewer of his tokens as TRAP and more tokens as DRESS. On the other hand, if listeners found the speaker's voice to be younger-sounding, they categorised more of his tokens as TRAP.

In a later study also using New Zealand phonetic variants, Walker (2007) investigated the perception of speaker social class and age. She tested whether different realizations of a phonetic variable could influence listeners' perceptions of the speaker's social features if the variable was embedded in a socially-marked grammatical construction. Two sociophonetic variables were used in the study, phrase-final /t/ and intrusive /r/ preceded by the MOUTH vowel. Intrusive /r/ preceded by the MOUTH vowel is more common in the speech of young, lower-class speakers who pronounce it with a lower F3 than speakers of upper classes. While middle-class females often produce released phrase-final /t/, other groups of speakers pronounce it without release. Phonetic variants were embedded in preterit sentences with *come* and *done*, as well as sentences indicating possession with *have-got*. These constructions are associated by language users with younger speakers. In addition, *come* and *done* constructions are associated with working-class speakers, and as such are stigmatised.

The results showed that even in the grammatical context of the sentences that the phonetic variants were embedded in, listeners were sensitive to the information about speaker age and class carried by the phonetic variants. The perception of class changed with the variants in such a way that more conservative variants were perceived as having been produced by a speaker of a higher class. As far as age ratings were concerned, conservative variants were again rated as older-sounding.

2.6.2 Exemplar priming

Another strand of research on speaker-indexical information focused on the influence of visual information on processing of the acoustic signal (Drager, 2011; Hay et al., 2006b; Niedzielski, 1999). It investigated whether visual cues contributed to auditory signal processing in any way. While visual stimuli themselves can have different forms they can also affect different types of speaker-indexical information.

2.6.2.1 Priming with information about speaker gender

An early study which investigated perception of speaker gender was the study by Johnson et al. (1999). This study was an extension of work on the *McGurk effect* (McGurk & MacDonald, 1976). In Johnson et al. (1999), listeners' expectations about the gender of the American English speaker and how the speaker should sound overshadowed the acoustic information present in the signal. A continuum from [ʊ] (*hood*) to [ʌ] (*hud*) was created. It was tested whether the gender of the speaker's face would change the perception of the vowel. The results of the study confirmed the hypothesis and revealed that priming listeners with gender information did shift listeners' perceptions of the vowels they heard.

The second priming experiment reported by Johnson et al. (1999) provided interesting results. In this experiment, participants were instructed to imagine the speaker. While one group of participants was asked to imagine a female speaker, the other was asked to imagine a male speaker. Furthermore, the F0 in the stimuli was manipulated to achieve gender-ambiguous-sounding speech. In a forced-choice experiment, listeners were asked to identify the vowel they heard as [ʊ] *hood* or [ʌ] *hud.* The results showed that even imagining the speaker shifted listeners' perceptions of the vowels they heard.

The results of the experiments discussed above indicate that to determine speaker gender, listeners relied on visual information or merely visualisation of the speaker over acoustic information. Attention should be drawn to the fact that even when the most stereotypical female voice had a lower fundamental frequency than the non-stereotypical female voice, the gender of the speaker in the video shifted vowel perception. Overall, these studies demonstrated that if listeners' attention is paid to speaker-indexical information, it has a direct effect on speech perception.

2.6.2.2 Priming with information about speaker age and social class

Researchers have also investigated the influence of speaker age and social class on speech processing. For example, Hay et al. (2006b) provided evidence that cues to speaker social-indexical information may skew listeners' perception of the speaker. Hay et al. (2006b) used perceptions of NEAR/SQUARE merger in progress occurring in New Zealand English, whereby the majority of younger speakers used merged forms (Hay et al., 2006b: 459), with NEAR being the primary pronunciation. Merged forms were also more often used by lower-class speakers.

Listeners were asked to categorise NEAR/SQUARE words when visual cues to speaker age and social class were provided in the form of pictures showing a younger or older speaker and a middle- or working-class speaker. It was reported that upon hearing the same audio stimulus, participants who believed that they themselves did not use merged NEAR/SQUARE variants performed more accurately in distinguishing NEAR/SQUARE words when seeing a picture of an older speaker. However, the results were the opposite when the speaker in the picture was younger.

In contrast, participants who themselves claimed to have the NEAR/SQUARE merger were not influenced by the information about the age of the speaker in the picture. The results showed they had misidentified the stimuli at the same rate irrespective of the age of the supposed speaker presented in the picture (Hay et al., 2006b: 479).

As far as the social class of the speaker is concerned, a picture of a middle-class-looking person biased listeners to perceive the stimuli as less merged, whereas a picture of a working-class speaker had the opposite result.

Drager (2011) also investigated how information about speaker age may influence perceptions of a vowel within a word. Listeners were primed with visual cues to information about speaker age. The TRAP and DRESS vowels, which are in the process of a chain shift in New Zealand English, were investigated (cf. Drager, 2005). A vowel continuum between DRESS and TRAP was synthesised from the natural vowels.

Manipulation involved only the values of F1 and F2. Male and female FF were used in the male and female continua respectively. Male voices were matched with pictures of men (younger or older) and female voices were matched with pictures of women (younger or older). During the experiment, listeners saw a picture of the speaker and after a short delay they heard a word from the continuum, which they were asked to identify as *head* vs. *had* or *bed* vs. *bad*.

The results revealed once more that speaker social information may impact perception of the acoustic signal. In addition, Drager reported that listener social attributes such as age also influenced perception. In her study the age of the supposed speaker as well as the age of the listener influenced categorisation of the vowels.

2.6.2.3 Priming with information about speaker geographical origin

In her seminal study, Niedzielski (1999) investigated the influence of speaker geographical information on speech perception at the segmental level by means of looking at awareness of Canadian Raising among Michigan listeners. In her study of the perception of dialectal variation, Detroit speakers were exposed to sentences produced by a female speaker from Detroit and asked to categorise her vowels. Both Michigan and Canadian speakers produce raised variants of /aʊ/, which is known as Canadian Raising. Niedzielski primed her participants with information about the speakers' geographical origin. While for half of the participants the top of the response sheet read *Michigan*, for the other half it read *Canada*. In addition, listeners were instructed that they were listening to a Detroit speaker or to a Canadian speaker from across the border. This study resulted in a number of findings. Firstly, the results showed that listeners rely on speaker-indexical information when categorising vowels. In addition, providing listeners with social information about the speaker shifted listeners' perceptions of vowels they heard. Listeners were more likely to categorise vowels as raised if *Canada* was written at the top of their answer sheet, even though /aʊ/ is produced in the same way on both sides of the Canadian and US border. The results revealed that listeners in the Detroit condition were more likely to identify the diphthong as lower than it was in fact produced (Niedzielski, 1999: 67).

Furthermore, Niedzielski's study showed that listeners' perceptions of their own dialect also influenced their judgements of the speech they heard. While listeners were aware of the fact that Canadian speakers stereotypically produce raised variants of the /aʊ/ diphthong, they perceived their own realisations as different from those of Canadian speakers, rather than identical to them. Thus, Michigan speakers associated raised variants with Canadian English only. In addition, Canadian variants were perceived as stigmatised by the Detroiters and Canadians as well. As the same time, Detroit listeners perceived the variants they produced as standard. Therefore, it could be further concluded that listeners from Michigan were not yet aware of the fact that the Detroit dialect has been affected by Northern Cities Chain Shift.

Later, Hay et al. (2006a) replicated Niedzielski's experiment using New Zealand conditions. They investigated perceptions of the /ɪ/, /æ/ and /ɛ/ vowels, which are pronounced differently in Australian and New Zealand English. While speakers of the two varieties were aware of the different realisations of /ɪ/, this was not the case for /æ/ and /ɛ/. Therefore, raised variants of /æ/ and /ɛ/ found in Australian English were not as stigmatised in New Zealand. Six-point synthetic continua of the natural vowels were prepared and listeners were asked to match the synthetic vowels with the vowels produced by a New Zealand speaker.

Hay et al.'s experiment also had two conditions. In the Australian condition, listeners had *Australian* written at the top of their answer sheet, whereas in the New Zealand condition they had *New Zealander* written on the answer sheet. Participants' attention was not attracted to these tags on the answer sheets in any way (Hay et al., 2006a: 357).

The findings reported in their study were consistent with findings reported by Niedzielski. The regional tag alone appeared to shift listeners' perceptions of the vowels they heard. Therefore, the participants in the Australian condition reported hearing raised and fronted /ɪ/ more often. On the other hand, participants in the New Zealand condition more often reported hearing the lowered and centralised variant. The results reported for the remaining two vowels were consistent with the results for /ɪ/.

Another priming study in which speaker social information biased listeners' perceptions of the speaker was Hay and Drager (2010). One group of participants was told the truth about the speaker, and the other half were told the opposite. While attention in one group of participants was drawn to a kiwi plush toy, which was associated with New Zealand, the other group's attention was attracted by kangaroo and koala mascots associated with Australia. During the experiment, New Zealand listeners were asked to match synthetic vowels with the vowels produced by a male New Zealander. Synthetic vowels comprised a six-step continuum, from the most Australian- to the most New Zealand-sounding tokens. The results showed that the symbols of geographical places were enough to shift listeners' perceptions of the vowels they heard. Listeners in the Australian condition judged more vowels as Australian-sounding, while listeners in the New Zealand condition judged more vowels as New Zealand-sounding.

The findings of the studies presented in this section are important to our understanding of speech perception, as well as to our understanding of how linguistic information connects with non-linguistic information. They show that non-linguistic information about the speaker, such as his or her age, gender or regional label, can shift perceptions of the acoustic information present in speech.

The sociophonetic studies cited in the previous sections have established that speaker-indexical information can be identified by the listener at the segmental level. In their investigations the studies used synthesised speech which had male or female acoustic features. Therefore, the question arises whether listeners would be able to identify speaker-indexical information if they would not know whether the speaker was male or female. This research attempts to fill in the gap in knowledge by addressing this question. Furthermore, this research is innovative in its approach since it uses natural-sounding gender-ambiguous speech to investigate perceptions of speaker-indexical information. Because one of the aspects of speech perception investigated in the thesis is the influence of visual cues on perception of indexical information, the following section provides a short account of research on priming.

2.7 Priming

While the previous section discussed examples of sociolinguistic studies which used priming, the present section discusses the concept of priming itself since two of the experiments presented and discussed in this thesis use visual priming. Thus, an explanation of what priming is and how it works along with a brief review of priming studies showing the automaticity of category activation and its influence on human behaviour is presented below.

Bargh (2006: 54) and Bargh et al. (1996b: 231) claim that socially complex behaviour can occur automatically without a person's conscious knowledge, intention or awareness of it happening. It takes place as a result of activation of a concept or category which can be triggered by as little as being randomly exposed to a stimulus. A stimulus can be for example, a word; it can be visual or auditory. In turn, activation of a concept or category results in activation of a trait or stereotype (Dijksterhuis et al., 2000: 532) such as for example, the elderly stereotype or the elderly trait. Activation of traits and stereotypes further influences individuals' social perceptions as well as attitudes, responses and behaviour (Bargh et al., 1996a: 233; Blair & Banaji, 1996: 1142; Dijksterhuis et al., 2000: 532; Wheeler & Berger, 2007: 357).

As has been already mentioned, trait or stereotype activation is an automatic process. Bargh et al. (1996: 125) argue that even on brief exposure to stimuli, automatic attitude activation in priming is a default state. Yet Dijksterhuis (2013: 244) points out that the link between "consciousness" and behaviour is not as strong as has been thought. Attention and consciousness are in fact separate (Dijksterhuis, 2013: 246). This disconnection can be observed in a number of actions which are performed by individuals unconsciously, even if some attention is needed to perform them.

There are two types of priming which lead to two types of perception of the stimulus. One is subliminal priming and the other supraliminal priming, the latter of which is used in two of the experiments described in the present thesis. While subliminal perception occurs below the level of consciousness, supraliminal perception is above the level of consciousness. This means that duration of exposure to the stimulus is different in subliminal and supraliminal priming. In subliminal perception the individual sees the stimulus for a fraction of a second, for example 35 ms (Oyserman & Wing-Sing Lee, 2010: 267), although the duration may differ (Bargh & Chartrand, 2014: 319). After the prime is shown, it is followed by another stimulus which masks the prime to ensure that the priming is below the level of consciousness (Bargh & Chartrand, 2014: 319). As a result of this manipulation, people are not consciously aware of the stimulus, and so even though the senses acknowledge the presence of the prime, the individual is not able to report it (Dijksterhuis et al., 2006: 80-81). In supraliminal perception, on the other hand, the individual sees the stimulus long enough to be able to acknowledge it and is able to report it. Furthermore, Dijksterhuis et al. (2006: 87) argue that the effects of subliminal and supraliminal priming give identical results. Balconi and Ferrari (2012) also found similarities between the results of subliminal and supraliminal priming. Balconi and Ferrari investigated perception of emotions from faces in subliminal and supraliminal priming. Black and white faces showing the emotions happy, sad, angry, surprised, or neutral were stimuli presented to participants for the duration of 30 ms in the subliminal condition and 200 ms in the supraliminal condition. Participants were instructed to pay attention to the stimulus while their brain activity was recorded by EEG. The analysis showed similar results in both priming conditions when perception of different emotional states was concerned. Similar brain areas for negative and positive emotions were activated in subliminal, as well as supraliminal priming.

The reasons for choosing supraliminal priming for the experiments in this thesis were multiple. Supraliminal priming is more accurate (Balconi & Ferrari, 2012), it results in a stronger manipulation than subliminal priming, and its impact is also stronger (Bargh & Chartrand, 2014: 318, 323). In addition, supraliminal priming is a well-established method (Bargh et al., 1996; Dijksterhuis & van Knippenberg, 1998), whereas subliminal priming is often used to support the results of supraliminal priming in the sense that it provides additional evidence in support of how unconscious and automatic the perception process was (Bargh & Chartrand, 2014: 318).

Even though there is a substantial body of literature on subliminal priming, applying it in Experiments 2 and 3 in this thesis was not necessary. When applying subliminal priming we want to be absolutely certain that it affects the participant. Therefore, individual recognition thresholds need to be discriminated, a task which requires, for example, a lengthy adaptation task (Bargh & Chartrand, 2014: 319). Thus, Bargh and Chartrand (2014: 319) advocate using stimulus exposure of a brief duration. However, there can be a possible issue with this approach. For example, a stimulus that is flashed for a short time may be overlooked by the participant if they happen to blink. In the experiments in this thesis, it was important to ensure that the participant was exposed to the visual stimuli rather than leaving things to chance. In addition, the visual stimuli used in Experiment 2 were detailed and showed gender, age and social class. Therefore, in order to ensure that participants were able to absorb all the information, it was decided to use supraliminal rather than subliminal priming. If participants were unable to report the images I could not have the certainty that the visual stimuli were perceived as expected.

A later experiment, by Dijksterhuis and van Knippenberg (1998) investigated the influence of supraliminal priming on how participants performed intellectually. Two experimental groups of participants were instructed to think about a typical professor and to write down their associations with the concept of a professor. One of the groups had two minutes to do it whilst the other had nine minutes. The control group, on the other hand, was not primed. Next, participants in both groups were asked questions from the game “Trivial Pursuit” testing them on their general knowledge. The results showed that participants primed with the stereotype of the professor had better results on the test than participants in the control group. In addition, participants who were primed for nine minutes performed better than participants who were primed for two minutes.

Glaser (1992: 99) argued that pictures in comparison to words have a stronger priming effect. In a different experiment, Bargh et al. (1996b) investigated the impact of priming on behaviour in social context. Participants were subliminally primed with a black and white image of the face of a young African American man. Subliminal priming was expected to automatically activate a negative stereotype of a Black male. The control group were subliminally primed with a face of a young White male. Both groups of participants were asked to perform a mundane visual task. Throughout the duration of the task participants were primed with an African American or White male face. After 130 trials an error message appeared on the screen saying the data were not saved and the participants were told they would need to start the task from the beginning. Participants’ reaction to the “computer error” and their facial expressions upon hearing the news were recorded by a hidden camera. The results showed that participants primed with an African American male face showed more hostility than participants primed with a White male face.

Another example of an automatic supraliminal priming experiment was an experiment in which gender stereotypes were activated (Blair & Banaji, 1996). A third of the primes were male, a third female and a third gender-neutral. Primes were personality traits or non-traits, for example a profession, an activity or object. Targets were male or female names which needed to be categorised as male or female. The results showed that if the prime and the target were of the same gender, participants responded faster than if the prime and target were of the opposite genders.

Abbate et al. (2013) investigated the effect of supraliminal priming with words associated with helping people. In the first experiment, after performing a scrambled sentence test participants were asked to donate money to a student association helping students in need. The results showed that in comparison with the neutral condition, in the priming condition participants were more likely to donate money even though they were students themselves.

Ferguson's (2008) study is an example of a study where subliminal priming was used. Participants were primed with words *eat, taste*, and *hungry*. The results revealed that participants' attitudes towards food were more positive as a result of priming.

Stenner et al. (2014) showed that motor processing as well as auditory perception can be manipulated with the use of subliminal priming. During an experiment comprised of different tasks participants were primed subliminally. Primes and targets were arrows, where primes pointed towards or away from the targets. Participants heard two tones and needed to judge which was the louder of the two. The results showed that the compatibility of prime and target did not influence perceived loudness of the tones. However, it affected participants' efficiency, and so when primes and targets were incompatible listeners reaction times were longer and when the primes and targets were compatible listeners reaction times were shorter.

Priming has been also used in disciplines other than psychology, for example, in sociolinguistics, which could be seen in Section 2.6.2.

2.8 Tyneside English

The following section describes the cultural importance of Newcastle and Tyneside in the North-East of England. The section also outlines phonetic variation in Tyneside English with a specific focus on socially-indexed variants which have been used in the perceptual experiments presented in this thesis. The localised variants discussed below are sociolinguistically marked in terms of speaker gender, age and social class. It was decided to use Tyneside English phonetic variants for the following reasons. In Tyneside English, vowels and stops can be realised using a number of variants, as will be shown in the following section. Because Newcastle is considered to be the cultural and economic hub of the North East region of England, its dialect has been extensively researched and described (Beal et al., 2012; Docherty & Foulkes, 1999; Foulkes et al., 2005; Milroy et al., 1994a; 1994b; Watt, 2000; 2002; Watt & Allen, 2003; Watt & Milroy, 1999). Furthermore, Tyneside English is stereotypically perceived as the urban dialect spoken in all of the North East (Beal et al., 2012; Pearce, 2009).

In this thesis perceptions of Tyneside-localised variants were compared and contrasted with perceptions of other localised variants used in the wider North-East region, or non-marked non-localised variants. And so, localised vowel and consonantal variants were contrasted with non-localised variants.

Newcastle is a port city formerly known for its coal and shipbuilding industries. It has also been the cultural centre of the North-East. After the BBC was established in 1922, Newcastle became an important broadcasting station, representing the North-East region (Vall, 2011: 18). In 1929 the BBC introduced broadcasts with regional accents and in the second half of the twentieth century the Newcastle station broadcast some successful programmes such as *Wot Cheor, Geordie!* or *Vox Pop* (Vall, 2011: 17). When independent commercial television in the shape of *Tyne and Tees Television* was introduced in the region in 1959, its studios were also set up in Newcastle (Vall, 2011: 41, 43).

After a period of economic recession following the closing down of collieries and decline of heavy industry in the second half of the twentieth century, the North-East was slowly revived. Tyneside took the lead also this time. The Metro Centre, the largest retail facility in the country, was opened in Gateshead in 1986 and remained the largest shopping centre in the country until the Westfield London centre was opened in 2011.

At present Newcastle's strength is education (Lancaster, 2007: 36). Recently Newcastle and Gateshead are being promoted as leisure and culture destinations trying to appeal to wider audiences from the entire country. Whilst Northumberland Street is the major shopping street in Newcastle, the banks of the River Tyne, known as the Quayside, host a number of restaurants, bars and night clubs. The Baltic Centre for Contemporary Arts can be found in Newcastle Quayside and Sage Gateshead, a music venue, on the opposite bank of the Tyne, in Gateshead Quays. There are museums and cultural centres in all of the North-East, for example Sunderland has the Northern Gallery for Contemporary Art. There is also the Middlesbrough Institute of Modern Art. Nevertheless, Newcastle remains the city with the highest number of cultural venues in the region.

Although the North-East has universities including Durham, Teesside in Middlesbrough and the University of Sunderland, Newcastle with its two universities, Newcastle University and Northumbria University, has the highest student population. Therefore, Tyneside still attracts people from other parts of the region in search of employment.

Great Britain is characterised by an abundance of local dialects. The North East of England, with Tyneside English being one of many spoken in the region, is no different. However, outsiders tend to have a distorted view of the North East. They seem to neglect a number of distinct dialects, such as the Sunderland or Middlesbrough dialects, that are present in the region, and consider Tyneside English to be spoken anywhere up north (Beal et al., 2012; Pearce, 2009). However, each of the dialects in the region is characterised by distinctive phonetic features.

Variation in the use of some vowels and consonants is one of multiple cues revealing social and regional characteristics of the speakers within the North East (Beal et al., 2012: 26). However, there is also considerable variation within Tyneside English in terms of the use of phonetic variants. In fact, Tyneside English is characterised by an array of localised phonetic variants, which are marked sociolinguistically, as they are not only gender- but also age- and class-specific (Watt & Allen, 2003: 269). It is these features that distinguish Tyneside speakers from speakers from south of the River Wear or from Teesside (Beal et al., 2012). It is also these features that distinguish speakers within Tyneside English. Therefore, a set of socially-marked Tyneside English variants have been selected for the purpose of this thesis. The vowel variables are the FACE, GOAT and NURSE vowels. The remaining variables involve realisations of the voiceless plosives /p, t, k/ in word-final and -medial positions. It should be mentioned, however, that the production studies cited below are not recent, as a result of which the patterns of use of some of the variants may not be up to date.

Two perceptually prominent vowels in Tyneside English are the FACE and GOAT vowels (Beal et al., 2012; Watt, 2000). Watt (2000; 2002) lists three types of realisations of the FACE and GOAT vowels and groups them into monophthongs, centring diphthongs and closing diphthongs. The most commonly-occurring and hence least marked variants of FACE and GOAT in Tyneside English are the monophthongal realisations, [eː] and [oː]. These realisations are also found in other varieties of North East English, and as such, are non-localised (Beal at al., 2012: 31).

Monophthongal [eː] and [oː] are found among male and female speakers of different ages and social backgrounds in Tyneside English. The only exceptions are older working-class male speakers who use the centring diphthong [ɪə] as a realisation of the FACE vowel. Nevertheless, older working-class males use [eː] and [oː] quite frequently themselves. The GOAT vowel is realised as monophthongal [oː], the centring diphthongal [ʊə] or the fronted monophthongal [ɵː][2] in this group of speakers (Beal at al., 2012; Watt, 2000, 2002; Watt & Milroy, 1999).

While the diphthongal FACE and GOAT variants [ɪə] and [ʊə] are found in all of the North East, in Tyneside English they are considered to be traditional and old-fashioned, and are most often used by older working-class males (Beal at al., 2012; Watt, 2000; 2002). [ɪə] can be also found in the speech of younger working-class males, although much less frequently than among older working-class males (Watt & Milroy, 1999). [ʊə] is less frequently used by other groups of male speakers than older working-class ones. For example, older middle-class and younger working-class speakers use it less frequently, and younger middle-class speakers use it very rarely (Watt & Milroy, 1999).

The closing diphthongs are [eɪ], for FACE vowel, and [oʊ], for GOAT. Overall, [eɪ] is not a common variant in Tyneside English, yet it is becoming more popular among younger middle-class speakers. It is used most often by young female middle-class speakers, followed by young middle-class male speakers (Watt, 2000). Both diphthongs are also Standard English forms.

The closing diphhtong [oʊ] is also widely used in other parts of the country. In Tyneside English, this realisation is used by young middle-class speakers (Beal et al., 2012; Watt, 2000). The fronted monophthongal [oː], on the other hand, is largely found among male speakers and is used most frequently by younger middle-class males, but also by older and younger working-class males. However, the variant is becoming less common in general and female speakers tend to refrain from using it (Watt, 2000; Watt & Milroy, 1999).

Finally, Watt and Allen (2003: 269) and Viereck (1968: 69, 70) provide more examples of the realisation of the GOAT vowel which make vowel contrast in Tyneside English sill more varied. For example, [ɪə] can be found in words like [stɪən] *stone,* [hɪəm] *home* and [bɪən] *bone,* and [aː] in words like *snow* [snaː]. These pronunciations occur more frequently in the speech of older working-class male speakers and are considered to be old-fashioned even by Viereck (1968).

[2] Although the fronted monophthongal [ɵː] is also used in other varieties of English, for example in Yorkshire English (Haddican et al., 2013), for the purpose of this research it was treated as a localised variant.

Another vowel associated with significant variability in the region is the NURSE vowel, which can be realised as the localised retracted [ɔː], fronted [øː] and centralised [ɜː] (Beal et al., 2012; Maguire, 2007; Watt, 1998; Watt & Milroy, 1999).

While the first variant is now rare and used mostly by older working-class male speakers, the two other variants are more commonly used in Tyneside English than [ɔː]. The centralised [ɜː] is most common, and is also non-localised. Watt (1998) and Watt & Milroy (1999) point out that the fronted variant [øː] is marked for age and gender, as it is found mostly among female speakers, and especially younger middle- and working-class females, who use it more frequently than [ɜː].

In general, localised vowel variants seem to be used by older and usually male speakers. Younger speakers, especially females, tend to prefer non-localised variants, which are widely found across the region and the country (Beal et al., 2012).

Furthermore, overall, a decrease in the use of localised, traditional forms can be observed in Tyneside English (Watt, 2000). In their place, new, non-regional forms are being adopted. The process leads to a reduction of the number of vowel variants in use and implies dialect levelling, which results in the formation of a more uniform repertoire of phonetic variation, one that is closer to other varieties of British English (Watt, 2000; 2002). At the same time, the non-localised forms new to the region seem to be less socially and geographically marked.

The situation is quite the opposite in the case of consonant productions. Some localised consonantal realisations seem to be becoming *more* widely used by speakers of Tyneside English (Beal et al., 2012: 45). One of the characteristic features of the North East dialects is a wide variety of realisations of voiceless stops. For example, /t/ in Tyneside English can be realised as T-to-R variant, a glottalised stop, a pure glottal stop, a fully released /t/, or a pre-aspirated /t/ (Beal et al., 2012; Watt & Milroy, 1999). /p/ and /k/ are also glottalised, pre-aspirated or realised as glottal stops in Tyneside English (Beal et al., 2012; Watt & Milroy, 1999).

Word-final /t/ can be realised as [ɹ]. This feature is commonly referred to as T-to-R and it is found in some restricted contexts: across word boundaries, in intervocalic position. It can occur in a limited number of common verbs and non-lexical words, for example *put on*, *shut up* or *get off* (Beal et al., 2012; Buchstaller et al., 2013; Milroy et al., 1994b; Watt & Milroy, 1999: 30). Given the fact that T-to-R is found most frequently among older working-class female speakers and, to a lesser extent, among younger working-class females, the process is marked sociolinguistically (Watt & Milroy, 1999).

Another marked feature of Tyneside English is glottalisation or glottal reinforcement of voiceless plosives (Beal et al., 2012; Docherty et al., 1997; Docherty & Foulkes, 1999; Milroy et al., 1994a; Milroy et al., 1994b; Watt & Allen, 2003; Watt & Milroy, 1999). Overall, glottalisation is a feature preferred by male speakers and it is found in intervocalic position after or before a stop gap and is realised as a laryngealised voice, in words like for example: *city, bottle, wanted, copy, happy, lucky* or *local* (Watt & Allen, 2003: 268). These variants are associated with a double articulation and as such they are usually transcribed as [ʔ͡p], [ʔ͡t] and [ʔ͡k] (Beal et al., 2012). At the same time, glottalised variants, and particularly glottalised /p/ and /k/, are claimed to be recessive (Docherty et al., 1997: 306).

As far as glottalisation of /p, t, k/ in intervocalic position is concerned, of the three stops, /p/ is the most often realised as a glottalised stop, and while this realisation is almost categorical for Newcastle male speakers, it is used by female speakers 58 per cent of the time (Beal et al., 2012; Docherty et al., 1997; Milroy et al., 1994a).

Glottalised /t/ is also a significantly frequent realisation of /t/ in intervocalic contexts. This feature is characteristic of the speech of older working-class male speakers who use it 82 per cent of the time (Beal et al., 2012: 39; Watt & Milroy, 1999). Female speakers, on the other hand, use [ʔ͡t] 42 per cent of the time (Beal et al., 2012: 39).

Of the three stops, /k/ is glottalised the least often. Male speakers use the feature 82 per cent of the time. Women, by contrast, use it 37 per cent of the time (Beal et al., 2012: 38).

Voiceless plosives in Tyneside English can be also realised as pure glottal stops [ʔ]. In Tyneside, /t/ is realised as [ʔ] more often than are /p/ or /k/ (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy et al., 1994a; Milroy et al., 1994b). While realisations of /t/ as glottal stops are quite commonly used in southern parts of the North East region, in Tyneside English this type of realisation is mainly found word-medially before syllabic /l/ (Beal et al., 2012). Apart from this context, intervocalic word-medial glottal stops are not very common in Tyneside English, where young speakers use it to some extent (Beal et al., 2012: 46). According to Watt and Milroy (1999), Milroy et al. (1994a) and Milroy et al.

(1994b), this realisation most commonly occurs in the speech of young middle-class females, who are also responsible for spreading the variant.

Glottalled /p/, on the other hand, is not very common in Tyneside English and amounts to only 1 per cent of use for male and female speakers. It is slightly more widely used, for example, in Sunderland and especially Middlesbrough, where it has been a growingly popular variant among young female speakers (Beal et al., 2012: 38).

As far as realisations of /k/ as [ʔ] are concerned, they have not been found in Tyneside English (Beal et al., 2012: 39).

Finally, realisations of word-final pre-pausal voiceless stops as pre-aspirated variants, with /t/ being the most frequently pre-aspirated of the three variables, have been reported in the North East (Beal et al., 2012; Jones & Llamas, 2008). These variants have been also found in Tyneside English, where they are most frequent in young females' speech (Docherty & Foulkes, 1999; Foulkes et al., 2005; Watt & Allen, 2003).

Table 2.1 sets out the variables and their phonetic variants including a summary of sociolinguistic distributions of these variants. The table functions as a reference table for the three experiments presented in the thesis.

Variants	Speakers
Variants of the FACE vowel	
[ɪə] – a localised variant	most often used by older working-class (WC) male speakers
[eː] – a non-localised variant	used across all groups of speakers
Variants of the GOAT vowel	
[ɵː] – a fronted monophthongal localised variant	most often used by younger MC males but also older and younger WC males
[iːə] – an archaic localised variant	found in the speech of older WC males
[ʊə] – a centring diphthongal localised variant	most often used by older WC males
[oː] – a non-localised monophthongal variant	used across all groups of speakers
Variants of the NURSE vowel	
[øː] – a fronted localised variant	most often used by young MC and WC females, but also older WC females
[ɔː] – a retracted localised variant	most often used by older WC males
[ɜː] – a non-localised centralised variant	used across all speaker groups
Variants of word-final pre-vocalic /t/	
[ɹ] – T-to-R	most often used by older WC female speakers
[ʔ͡t] – glottalised /t/	most often found in older MC and WC male speech
Variants of intervocalic /t/	
[ʔ͡t] – glottalised /t/ - a localised variant	most often found in older MC and WC male speech
[ʔ] – glottalled /t/ - a localised variant	most often used by younger MC female speakers

[t] – released /t/ – a non-localised variant	used across all speaker groups
Variants of word-medial /p/	
[$\widehat{ʔp}$] – glottalised /p/ - a localised variant	most often used by male speakers
[p] – released /p/ – a non-localised variant	used across all speaker groups
Variants of word-medial /k/	
[$\widehat{ʔk}$] – glottalised /k/, a localised variant	most often used by male speakers
[k] – released /k/, a non-localised variant	used across all speaker groups
Variants of word-final post-vocalic /t/	
[^{h}t] – pre-aspirated /t/, a localised variant	characteristic of younger female speakers
[t] – released /t/, a non-localised variant	used across all speaker groups

Table 2.1 Usage patterns of the FACE, GOAT and NURSE vowel variants in Tyneside English followed by realisations of word-final pre-vocalic /t/, intervocalic /t/, intervocalic /p/, intervocalic /k/ and word-final post-vocalic /t/ (Beal et al., 2012; Docherty et al., 1997: 306; Docherty & Foulkes, 1999; Foulkes et al., 2005; Milroy et al., 1994a; Milroy et al., 1994b; Viereck, 1968; Watt, 2000; Watt, 1998; Watt & Allen, 2003; Watt & Milroy, 1999).

Perceptions of speaker-indexical information from the variants described above are presented in the following chapter in Experiment 1. At the same time, Experiment 1 is the first in the series of three experiments investigating perceptions of speaker-indexical information from gender-ambiguous speech.

3 Experiment 1

3.1 Introduction

The general hypothesis of this thesis is that speaker-indexical information can be recovered by the listener from socially-correlated phonetic variants, which sound gender-ambiguous. Previous research has investigated female and male voice identification (Biemans, 2000; Johnson et al., 1999; Klatt & Klatt, 1990). Even though it has been established that listeners are quite accurate at identifying adult female and male voices, it is still unclear how listeners identify gender in the speech signal (Munson & Babel, 2007). Literature provides evidence that fundamental frequency impacts on femininity and masculinity judgments (Foulkes et al., 2010; Munson & Babel, 2007). However, there is also evidence that voices perceived as stereotypically female are not always characterised by high fundamental frequency (Johnson et al., 1999; Klatt & Klatt, 1990).

It has been also reported that listeners are able to distinguish male from female speakers in the absence of acoustic information normally found in speaker fundamental frequency (Assmann & Nearey, 2007; Coleman, 1971; Hubbard & Assmann, 2013; Lass et al., 1975). These findings imply that parameters of the vocal tract are not the only factors influencing whether a speaker sounds feminine or masculine, which further implies that gender-specific acoustic information does not rely exclusively on fundamental frequency.

Because fundamental frequency is only one of the cues to speakers' gender identification, it is hypothesised that when speaker-social information embedded in fundamental frequency is not accessible to the listener, this type of information can be identified from other cues such as localised gender-correlated phonetic variants.

Therefore, this chapter examines whether speaker social-indexical information can be identified at the segmental level. By investigating and comparing perceptions of speaker-social information provided by Tyneside listeners and listeners from other areas of the North-East region, the present chapter investigates and compares the perception of speaker-social information in listeners with high and low exposure to the dialect.

This experiment builds on earlier research on the perception of speaker-indexical information in child speech (Foulkes et al., 2010). Following Foulkes et al.'s findings, it is hypothesised that listeners with high exposure to the dialect and particular variant realisations should be more sensitive to speaker-indexical information carried by these variants. However, listeners with low exposure to the dialect are not expected to be able to access this information with as much detail as listeners with higher exposure. The present research also goes further in investigating perceptions of speaker-indexical information than did Foulkes et al. (2010). In addition to speaker gender, it investigates perceptions of speaker age and social class. The three types of speaker-indexical information will be discussed in detail in the Results section.

3.2 Method

3.2.1 Stimuli

A total of four voices were used to construct stimuli for this thesis. Two phoneticians, who were not native speakers of Tyneside English, recorded target stimuli using Tyneside English variants and two other speakers recorded fillers used in the experiments. Speakers were in their forties and mid-twenties.

Stimuli selected for this thesis represent specific phonological contexts. Vowels occur in three phonological contexts: word-finally in open syllables, preceding a nasal or a fricative. For example, the words in the GOAT group included: *grow, stone, dough, go, home* and *toe*. Consonantal stimuli were also embedded in one-syllable words, where the variables under investigation occurred in intervocalic word-medial and final positions. All consonants were adjacent to vowels, and so for example in the T-to-R group *get off, put on* and *shut up* were used, whereas in the glottalised /t/ group, *bottle*, *wanted* and *city* were found.

Preliminary tests with Adobe Audition 3.0 (Adobe, 2007) revealed that regarding the range of possible pitch manipulation and the final outcome in terms of voice naturalness, male voices gave better results than female voices. In other words, when working with male voices, it was possible to apply a wider range of pitch manipulations before the voice started to sound unnatural. The results were less optimistic for female voices, which would lose their naturalness before they started to sound gender-ambiguous. Therefore, only male voices were used in the experiments.

The tokens were recorded in a recording studio to .wav sound files at a sampling rate of 44.1 kHz and 16 bit mono resolution. All tokens were manipulated in Adobe Audition 3.0 (Adobe, 2007) using the Pitch Shifter function to raise pitch and obtain the effect of gender-ambiguous-sounding voice. In addition to preserving the tempo of the samples, high precision and default appropriate settings were selected. Pitch Shifter allows changes in fundamental frequency (F0) by semitones and cents, where 1 semitone is equal to 100 cents. Each token was manipulated individually between 1.0 and 4.0 semitones. The average F0 of target stimuli was 135 Hz.

For the purpose of the experiments presented in the thesis, speaker F0 was shifted to obtain the effect of a gender-ambiguous-sounding voice. The algorithm implemented by the Pitch Shifter in Audition allows the speech tempo to be preserved and the formant values to be adjusted to changes in pitch (Adobe, 2007). Because this thesis investigates the perception of localised phonetic variables in the absence of gender-specific F0, the aim was to manipulate only one of the phonetic cues, that is, F0. Preserving tempo and adjusting formant values to changes in F0 sustained other acoustic features of the recordings. Furthermore, this approach allowed controlling for F0 and drawing more specific conclusions about the acoustic cues responsible for perceptions of speaker-indexical information.

Nevertheless, obtaining gender-ambiguous voices which sound natural is a challenging task. Adjusting FF to changes in F0 resulted in overall natural sounding voices which in some cases may have sounded slightly unnatural in comparison to non-manipulated human voices. Having said that, it should be stressed that even if the voices sounded somewhat unnatural by using the approach described above, the potential issue of the voices sounding robotic was avoided.

It should be pointed out that upon raising F0, male voices may seem to sound younger than before manipulation. Female voices, on the other hand, may seem to sound older. This interaction between age and gender will be found in the experimental results presented in the following parts of the thesis.

Another issue resulting from the experimental design is the fact that one of the speakers, speaker B, is sometimes reported in the statistical analysis as more female sounding, as well as more middle-class sounding than speaker A. This difference between speakers A and B is an unintended outcome of having used more than one speaker when recording the stimuli. Furthermore, it should not be expected to have both speakers sound identical, and it might be inevitable that listeners would evaluate their voices differently.

Figure 3.1 below presents the interface of the Pitch Shifter in Audition with exemplary settings. Figures 3.2 and 3.3 present the spectrogram and waveform of the word *stir* before and after manipulation when F0 was shifted by 3.5 semitones.

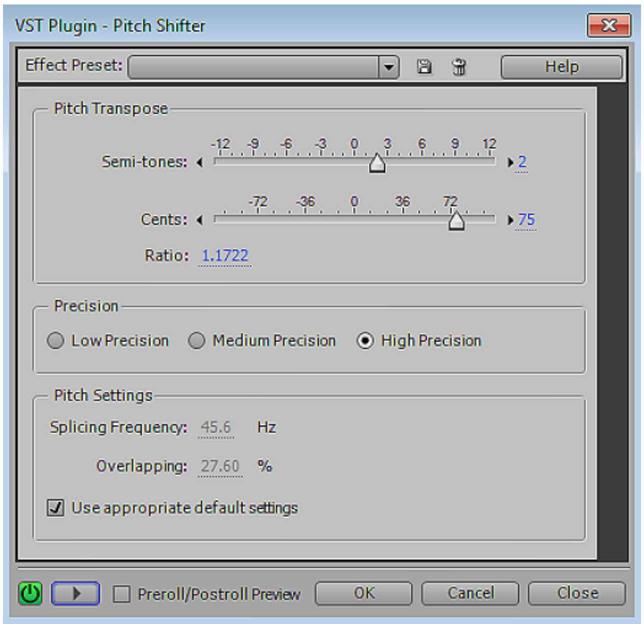


Fig. 3.1 The interface of the Pitch Shifter in Audition (Adobe, 2007).

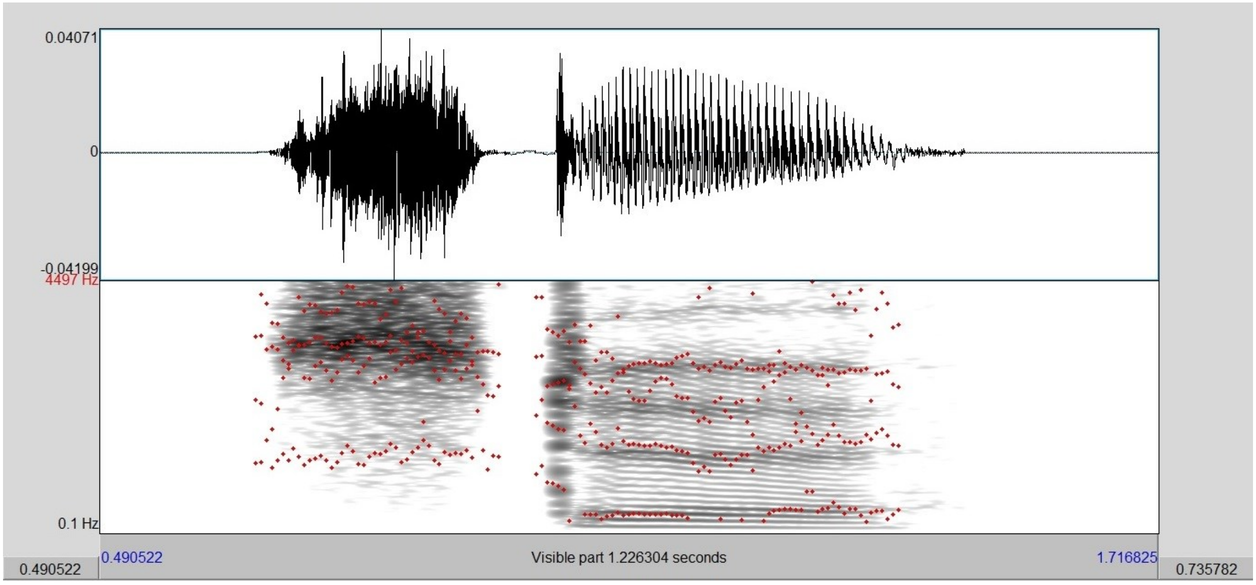


Fig. 3.2 The [øː] variant in *stir*. Before manipulation of F0.

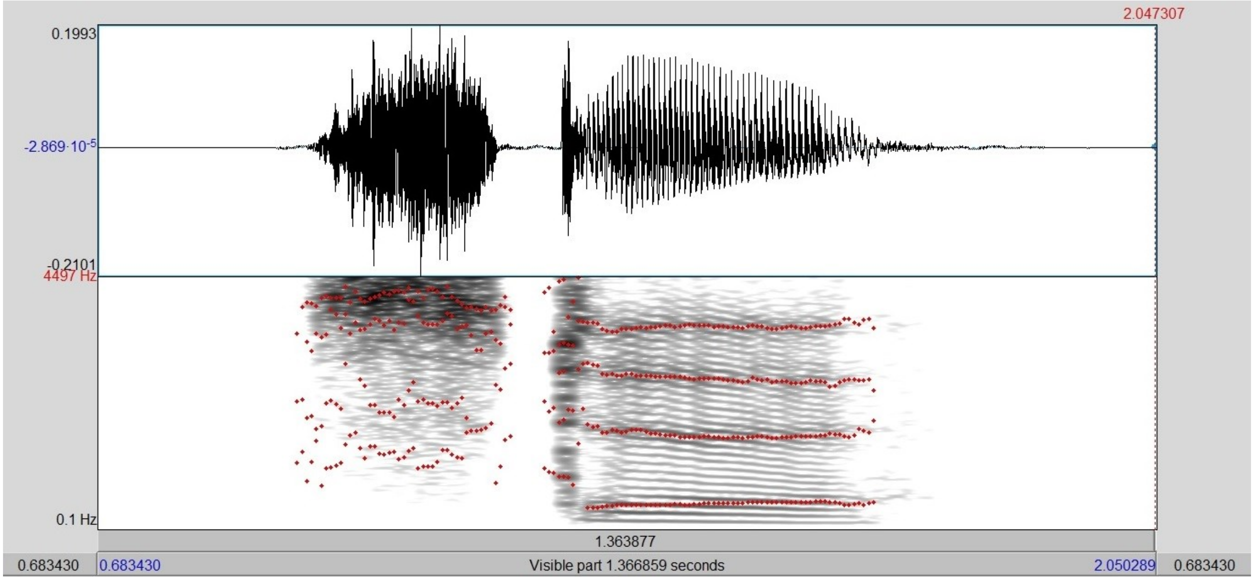


Fig. 3.3 The [øː] variant in *stir*. Shifted by 3.5 semitones.

All tokens were normalised for volume in Adobe Audition CS5.5 (Adobe, 2012) using the Match Volume function. A single token was pre-selected and the remaining tokens were matched in volume to the pre-selected token using the file total RMS (root mean square) power function and limiting settings to ensure that the output files were not clipped or overly loud.

In one instance, token duration required elongation using the Stretch and Pitch function in CS5.5. Previous shifting of F0 resulted in a speeded up and unnatural output of the token. While duration was stretched by 50 per cent using the iZotope Radius algorithm built into Audition, pitch and speech characteristics were preserved.

At the end of the process, the naturalness and gender-ambiguity of the stimuli and fillers were judged by a male and a female sociophonetician with expertise in the dialects of North East England.

The experiments use single-word stimuli. The advantage of using single words over connected speech is that listeners can focus with greater ease on the specific type of information present in the acoustic signal (Munson, 2007). At the same time, this approach allows the researcher to control for more parameters and therefore to draw more reliable conclusions from the data when analysing which phonetic cues listeners rely on.

In constructing each stimulus, it was important to ensure that listeners responded only to the variant under investigation, as opposed to any other segment of a word. This was achieved by realising the variant of interest as localised, with all other segments being as constant as possible.

3.2.2 Measuring subjective judgements – the Visual Analogue Scale

The present experiment investigates subjective evaluations of speaker age, gender and social class. Even though measuring subjective ratings based on attitudes, opinions, feelings, mood or sensory stimuli seems to be quite a difficult task, a number of scales have been developed to deal with it. One of such scales is the Visual Analogue Scale (VAS).

VAS is a continuous scale. It is usually 100 mm long with increments every 1 mm, making it in this way easy to analyse the results. VAS is usually horizontal, although vertical versions can be found as well. Depending on the aim of the research when using VAS, the numeric scale can be shown, or a straight line can be used instead. VAS was first introduced in the 1920's as a tool to assess the performance of employees (Hayes & Patterson, 1921). It was described in more detail by Aitken (1969), and ever since it has been applied in clinical research, where it was used as a self-assessment tool for the patients, at the same time giving the clinicians an idea of how the patients felt (Aitken, 1969: 17; Zealley & Aitken, 1969: 21). VAS has been widely used to assess pain or fatigue, but also mood or quality of life in patients with cancer, etc. Even though VAS has been a popular tool to use in clinical research, it has not been used much in surveys investigating people's opinions or feelings (Couper et al., 2006). However, recently VAS was used by Llamas and Watt in a study of national attitudes in the Scottish-English border region where speakers' attitudes were measured (Llamas & Watt, 2014). A little earlier, Munson (2011) used VAS to measure perceptual boundaries between fricatives arguing that the technique allowed more detailed results to be obtained than using the binary choice technique (Munson, 2011: 2633).

Advocates of VAS argue that using the scale is an efficient way of collecting subjective evaluations and ratings from respondents (Marsh-Richard et al., 2009). Furthermore, the scale is easy to use and requires little effort on the part of the respondent (Ahearn, 1997). VAS is also argued to be more flexible than a number of other known scales which might be crucial in the case of subjective ratings. For example, one of the popular and widely-used scales in surveys is the Likert scale (Likert, 1932), which usually has 5 or 7 pre-defined points from among which respondents can select their choices. Unlike the Likert scale, VAS gives respondents more flexibility and independence in terms of the choices of possible ratings. Rather than being restricted to any pre-defined points on the scale, respondents can use its entire length, which results in more precise measurements (Aitken, 1969; Llamas & Watt, 2014). At the same time VAS is a more sensitive scale than a discrete scale. Given that it increments by 1 on a scale from 0-100 and is an interval scale, VAS has a potential to detect even small changes in ratings or differences between respondents who share broadly the same point of view (Llamas & Watt, 2014: 613). As a result, VAS measurements are fine-grained which enables a detailed analysis (Llamas & Watt, 2014: 612, 613; Redinger & Llamas, 2015: 173).

In their study comparing VAS with a radio button scale and numeric input scale, Couper et al. (2006) reported that VAS took the longest to complete in comparison with other scales (Couper et al., 2006: 243). Furthermore, they also found out that out of the three scales, VAS was the one that was the most often skipped and left unrated by respondents (Couper et al., 2006: 243). This might suggest that VAS is actually a challenging scale to use for respondents. Also Torrance et al. (2001) point out that perhaps VAS might not be very easy for respondents to use. Nevertheless, the idea of VAS itself is simple and can be very easily explained. In fact, VAS is so readily accessible that it was used with children as young as 9.8 years of age (Shields et al., 2003: 227). In their study, where children assessed the intensity of pain they could feel, Shields et al. (2003) noticed that older children, 11 to 14 years of age, with better developed abstract thinking, were better at understanding VAS than were younger children, 5 to 10 years of age. Thus, it seems that another criterion which should be fulfilled for VAS to be a successful scale is the respondents' ability to think in abstract terms irrespective of their age (Shields et al., 2003: 232).

Some fundamental criteria need to be fulfilled in order for a VAS to be a valid scale. The anchors at both ends of the scale need to be clearly defined (Aitken, 1969; Torrance et al., 2001) and they ought to be unipolar rather than bipolar. Bipolarity introduces semantic differential and, as such the scale is more difficult to grasp for the respondents (Wewers & Lowe, 1990: 228, 233). Bipolar scales were used in research to measure mood, where the anchor points were, for example, *most happy* vs. *most depressed* (Zealley & Aitken, 1969) or *happy* vs. *sad* (Stern & Bachman, 1991). Unipolar scales, on the other hand, are often used to measure the strength of pain, for example. The anchor points describe the two extremes of the feeling: *no pain* (no phenomenon) *vs. the worst pain I can imagine* (Wewers & Lowe, 1990: 228).

Another issue which should be mentioned when discussing VAS is the fact that interpretation of the anchor points as well as any point on the scale depends on the participant's experience (Wewers & Lowe, 1990: 234) as well as his/her attitudes or beliefs. Thus, the interpretation of two seemingly similar or identical responses might in fact differ across individuals. Even though VAS has some clear advantages, individual respondents might use the scale differently in terms of whether they use the entire length of the scale, as desired by the researcher, or whether they use only certain points on the scale, or even limit the scale to the two end points, in which case the benefits of VAS are largely lost. Couper et al. (2006) and Torrance et al. (2001), on the other hand, draw attention to the problem of respondents avoiding using the end points of the scale. Rather than use the entire length of the scale, the respondents might exclude these extreme values on the scale from use, or they might tend to place their responses in the middle of the scale.

Because of the differences across users of VAS mentioned above, it is important to normalise data from VAS before performing statistical analysis. The easiest way to do so is to apply random effects which account for variation within the participants (cf. Section 3.3).

This short overview provides some pros and cons of using VAS. Nevertheless, it is believed that the advantages of VAS outweigh its possible limitations. Thus it was decided to implement the scale in the present experiment.

3.2.3 Procedure

The experiment was conducted in laboratory conditions and administered in SurveyGizmo (SurveyGizmo, 2014). At the beginning of the experiment there was a short training session during which participants familiarised themselves with the types of scales used in the experiment. Data from the training session were excluded from analysis. After the training participants were given time to ask questions. A total of 531 single-word stimuli and fillers were presented via headphones at a comfortable hearing level, one at a time. Each stimulus was played once only. The entire session took about 90 minutes and there were three five-minute-long breaks during which participants were asked to complete a Sudoku puzzle.

During the experiment, a visual representation of a stimulus was displayed on screen. The sound was played after an image and scale were displayed. The onset delay for audio was about a second.

Each target word was evaluated four times along four dimensions: perceived speaker gender - maleness and femaleness - and perceived speaker age and social class. These alternatives were presented in a mixed order, in such a way that every stimulus was rated along only one dimension per block and on all four of them in total. Each target word was followed by a filler word.

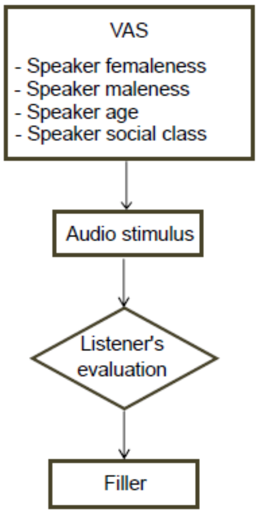


Fig. 3.4 Experimental procedure. The flowchart illustrates the process of speaker femaleness, maleness, age and social class evaluation.

The role of the pictures was to help listeners with low exposure to Tyneside English understand the recordings. The images also served as an additional element in the experiment, which alleviated a possible feeling of boredom. In order to avoid visual priming, to filler words, pictures excluded images of men or women except for two instances. Care was also taken to ensure that the pictures used in the experiment were not associated with males or females. Images included photographs, drawings, cartoons and computer icons. They presented objects, concepts or activities illustrating words played to the listeners. Nevertheless, it could be argued that for some listeners some of the pictures could be associated with men or women. For example, with the word *bat,* a picture of a bat or a picture of a baseball bat was shown. With *blur,* a picture of blurred birds in motion or a picture of a blurred view through a windshield was shown. It could be argued that participants may have stereotypically associated driving fast with men. The same might be true for a baseball bat. A bowl filled with dough, however, might be stereotypically associated by some with women. However, associations of this type may depend on an individual and his/her experience; as a result they may vary from person to person. The words were not tested for gender bias in the sense that it was not tested if any of the words were seen as activities or objects typically associated with men or women.

Listeners were instructed to listen to each stimulus and evaluate it using a Visual Analogue Scale (VAS) slider with a 0 to 100 point scale, incrementing by 1 point and logging participant choices on the *x* axis. Listeners were also asked to go with their first impressions and to not "overthink" their choices.

Fig. 3.5 VAS scale for evaluating perceived speaker femaleness. The target word is *blur*.

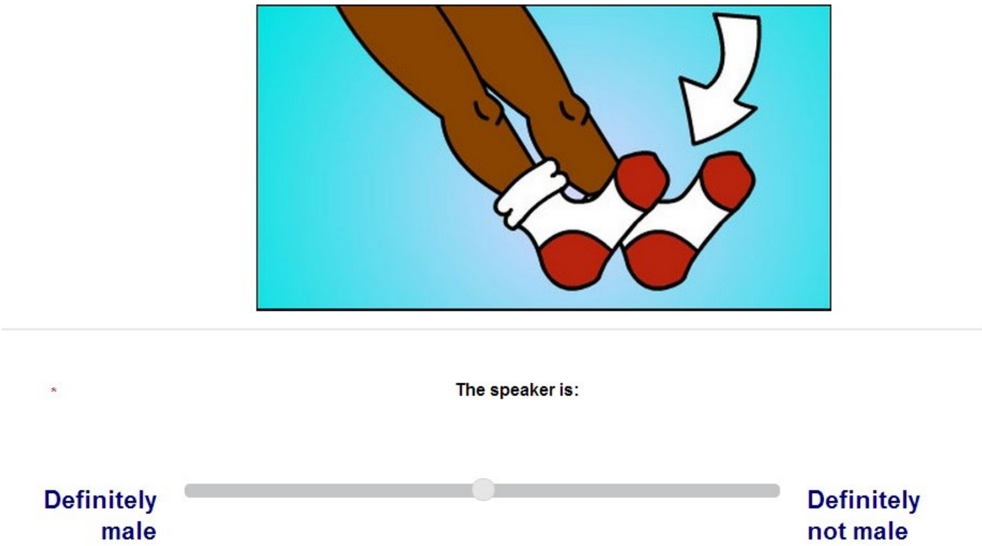


Fig. 3.6 VAS scale for evaluating perceived speaker maleness. The target word is *put on*.

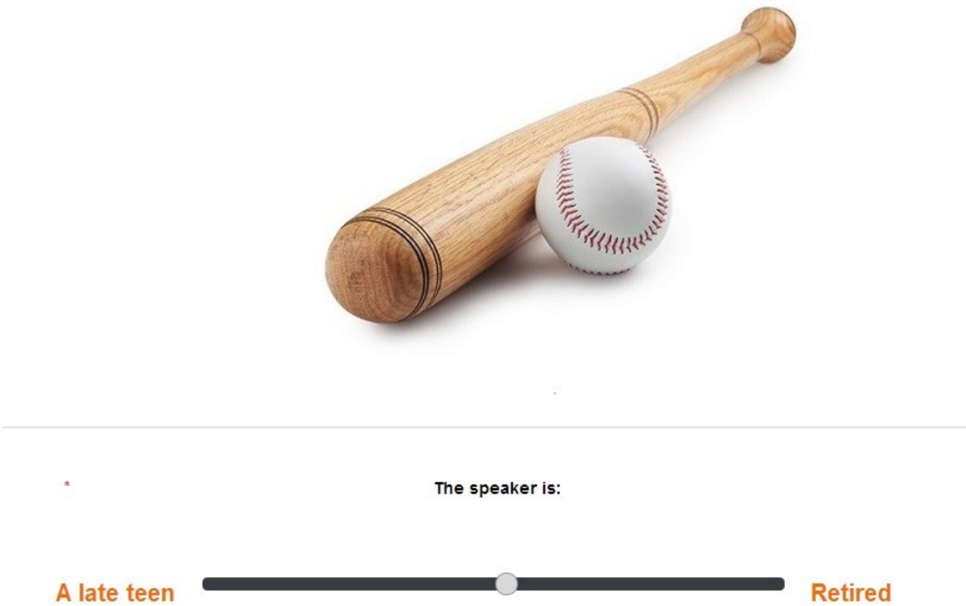


Fig. 3.7 VAS scale for evaluating perceived speaker age. The target word is *bat*.

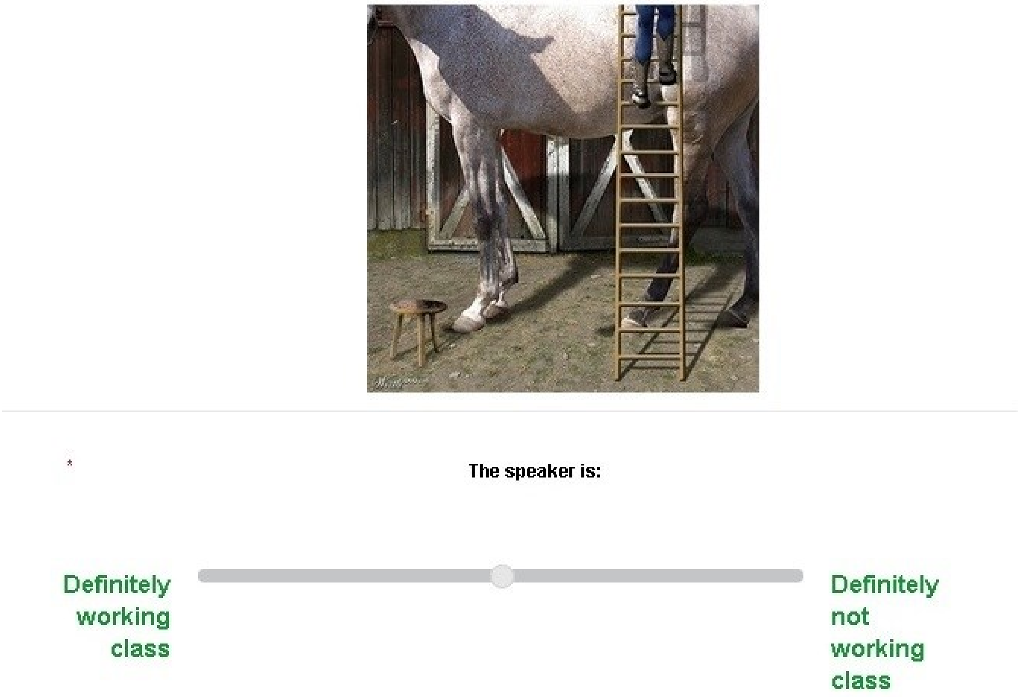


Fig. 3.8 VAS scale for evaluating perceived speaker social class. The target word is *get off*.

As can be noticed, the anchor points of the VAS were clearly defined and the concepts formed the end-points of a continuum, for example *Definitely female - Definitely not female* or *Definitely male - Definitely not male*. This was the reason for developing two separate scales to evaluate speaker gender in the present experiment.

Wording in each of the scales was colour-coded for the benefit of the participant. Distinctive colours aimed to associate a particular colour with a particular scale. Colour-coding was consistent, and so the femaleness scale used red wording, the maleness scale navy blue, the age scale orange, and the social-class scale green.

Stimuli were presented in a fixed order and the slider was reset to a midpoint position on the scale after each evaluation. Additionally, the slider did not allow for stimuli to be left unrated and so, in order to proceed, participants had to move it even if they moved it back to where it was at the start.

Data were saved on an external server owned by the software provider.

3.2.4 Participants

The main group of listeners who participated in the experiment were from Tyneside, and so they were in the group of listeners with high exposure to the dialect under investigation. Listeners were volunteers recruited from the undergraduate and graduate student bodies at the University of York and Newcastle University. The majority of participants recruited at the University of York were in their first year, and thus had lived away from Tyneside for about 6 months. Four of the York students were more advanced in their coursework and had lived in York longer, between 1 and 4 years. However, all of the students had family and friends in Tyneside and visited home often. As far as participants from Newcastle University are concerned, they had lived on Tyneside all their lives.

The comparison group of listeners were from other areas in the North-East of England, excluding Tyneside. These listeners were recruited at the University of York and University of Sunderland. They also had friends and family in the North-East. The reason for choosing participants from the wider North-East was that perceptual differences between the two groups of respondents were expected. Even though most of the phonetic variants chosen for the experiment are almost identical in the accents across the region the frequency of use of these variants varies across the region. For example, the FACE variant [ɪə] is more common in Newcastle, Gateshead, Sunderland and Durham than in Middlesbrough and Darlington (Beal et al., 2012: 30). Out of the three stops, /p/ is glottalisated word-medially most often in Middlesbrough and Newcastle and the least often in Sunderland (Beal et al., 2012: 39). Whilst glottalised /t/ is most common in Newcastle, it is glottalised /k/ that is most frequent in Middlesbrough (Beal et al., 2012: 39). At the same time, for example the GOAT variant [ɵː] is found only in Newcastle and Gateshead whilst other variants are found in Sunderland, Durham, Middlesbrough and Darlington (Beal et al., 2012: 31). It was assumed that these differences in the frequency of use of particular variants would result in perceptual differences of speaker-indexical information carried by the variants.

Listeners in both groups were close in age. With the exception of three members of the Tyneside group, whose age range was 25-34, the rest of the participants were aged 18-24. In the North-East group twenty-six participants were aged 18-24, six were 35-34, and one was 45-54.

In the Tyneside group twenty-four female and seven male listeners participated in the experiment. In the North-East group twenty-one listeners were female and twelve were male. Although the aim was to obtain a balanced sample of male and female participants, this proved difficult in practice.

In terms of social background in the Tyneside group, 15 participants considered themselves to be from a working-class background and 16 described themselves as middle-class. In the North-East group 21 listeners had working-class backgrounds and 12 had middle-class backgrounds.

Only five participants in each of the groups claimed to speak a foreign language, five of whom in total spoke more than one non-native language. Participants came from a wide variety of fields, the most popular being speech and language therapy, history, philosophy and psychology as well as English language and linguistics.

None of the listeners reported a hearing impairment but five suffered from a mild cold.

In the first stage of data collection each participant was paid £12 upon completion of the experiment, while in the second stage participants were paid £10. This difference resulted from funds available to the researcher.

3.3 Statistical Methods

Statistical tests in this thesis were carried out using the lme4 (linear mixed-effects 4) library in the software package R (Bates & Maechler, 2015). Three types of mixed-effects tests were applied in the thesis: linear, logistic and ordinal regression. Mixed-effects have been used in linguistics and psychology for some time now and their understanding has been improving as well. As the name suggests, mixed-effects models incorporate fixed and random effects. They have some obvious advantages over more traditional statistical methods because they account for by-subject and by-item variation, that is, variation involving participants and linguistic material, respectively (Barr et al., 2013: 256). On the one hand, mixed-effects deal with variation across individual subjects while, on the other, they make it possible to generate models reflecting trends in the population (Baayen, 2008: 407). As a result, mixed-effects produce more robust results.

Random effects are elements which cause random variation in the data (Baayen, 2013: 350). An example of random variation is individual differences between participants (Barr et al., 2013). Therefore, applying random effects in the model is a way of dealing with variation in the sampled population. For example, in Experiment 1 individual participants varied in terms of threshold use of the Visual Analogue Scale (VAS) slider. Some participants did not use the entire length of the scale, avoiding the end-points, while others used all of the scale and a few used only its end points. Applying random effects in this case allowed for normalising the data and the differences across participants. Additionally, individual respondents varied in their evaluations of the phonetic variants. This variability stemmed from differences in sensitivity to the linguistic material across the population and was also accounted for by applying random effects.

An example of variation in the linguistic material is multiple observations per item (Barr et al., 2013). In all of the experiments in this thesis, a repeated measures within-items design was implemented, which means that all stimuli were evaluated by all participants and that there were multiple responses per participant, which resulted in clustering of the data. Thus, to deal with clustering of the data as well as random variation, the by-subject and by-item random effects were applied (Baayen et al., 2008: 390; Barr et al., 2013: 263, 266). In fact, all of the models presented in this thesis, that is models analysing social class of the speaker as well as perceived age and gender of the speaker, account for by-subject and by-item random variation, where *listener* was entered as a random slope and *audio* as a random intercept.

Each participant was assigned a different intercept value and slope, which enabled accounting for individual differences between participants in the perception of the phonetic variants under investigation. Similarly, assigning a different intercept value to each audio-sample accounted for audio-specific idiosyncrasies, since audio sample had different phonetic variants in them and were produced by two different speakers. The reason behind choosing *audio sample* rather than *word* to account for by-item random variation was that *word* was excluded from the analysis because it was not of interest whether any of the lexical items themselves were perceived as more working- or middle-class sounding.

Barr et al. (2013: 258, 257) advocate making careful choices of the right random effects for inclusion in the model as well as using a maximum number of them. Overall, adding random effects increases the power of the test (Barr et al., 2013: 261). In order to specify random effects correctly potential areas of clustering within subjects and items should be considered (Barr et al., 2013: 262).

As a rule of thumb, random intercepts, by-subject or by-item depending on the design, should be applied when the experimental design enables multiple responses per participant (Barr et al., 2013: 263, 275). A model incorporating random intercepts shows how individuals differ from the mean. Thus, individual intercepts differ from the mean; however, they do not differ from the mean intercept in their slopes. Instead, all intercepts have the same slope. However, a model which applies random slopes can tell us more. We can find out whether there is large variability among participants in their evaluations of the dependent variable. Unlike a model with random intercepts, a model with random slopes enables variability in slopes across participants. In other words, random slopes tell us how sensitive individual participants are to the measured variable (Barr et al., 2013: 260), and also how they differ in their sensitivity to the variable in question. Therefore, it is possible to observe how a specific word or phonetic variant is processed by an individual and how the effect of the variant differs across participants (Baayen, 2008: 399). For example, a high level of variability in slopes across subjects implies that the effect of the tested variable varies between subjects. However, it should be pointed out that in practice, accounting for variation across participants by applying random slopes often reduces the statistical significance of findings.

When performing a mixed-effects analysis a stepwise backward or forward approach can be applied. In this thesis, to find the best-fit model, a backward stepwise selection of models was performed by removing fixed factors or interactions between predictors which did not yield significant results. The full model was tested first. In such a model all fixed effects as well as interactions between them were entered and modelled as fixed effects. However, if predictors or interactions between them were not statistically significant in the full model or any of the following ones, they were deleted. Next, models were compared by means of ANOVA to determine the best-fit model to the data. The way to compare models to establish which of them is a better fit is to compare Akaike's information criterion (AIC) for each consecutive model. The lower the AIC, the less complex the model and, as such, it is a better-fit to the data (Baayen, 2013: 347). If removing a predictor from a model reduces the AIC noticeably, this means that the predictor was not relevant to the model. If it is the case that a predictor or interactions between predictors were not significant, but the AIC score still indicated that they contributed to the model, such interactions are included in the best-fit model and reported in the results.

Random effects from the full model stayed unchanged throughout the process of fitting the best model to the data.

3.4 Results

This section presents and discusses the results of the perceptual experiment. First, the FACE, GOAT and NURSE vowel variants are discussed in terms of perception of speaker gender, age and social class. Next, results for T-to-R, glottalised voiceless stops and pre-aspirated /t/ are presented. In fact, glottalised /t/ is investigated in two different environments. When it is investigated as a T-to-R variable, it is examined in the contexts in which T-to-R occurs, that is, across word boundaries, in intervocalic position, in words such as *get off* or *shut up* (Beal et al., 2012; Buchstaller et al., 2013; Milroy et al., 1994b; Watt & Milroy, 1999: 30) (cf. Section 2.8). The second environment in which glottalised /t/ is investigated is the same one in which glottalled and released /t/ occur, that is, word internally, in intervocalic position, in words such as *city* or *water* (Beal et al., 2012: 39; Watt & Milroy, 1999) (cf. Section 2.8).

Within evaluations of each type of speaker-indexical information the results from the Tyneside and North-East groups of listeners were investigated separately.

As has been mentioned, in the present experiment a VAS scale was implemented. This was a continuous scale and so linear regression mixed-effects tests were applied to analyse the data.

Depending on what the test was measuring, the dependent factor was *the perceived femaleness of the speaker, the perceived maleness of the speaker, the perceived age of the speaker* and finally *the perceived social class of the speaker* (working-class or non-working-class). Possible fixed effects used in a linear regression included the following:

- The phonetic variant (number of variants depended on the variable)
- Speaker (A or B)
- Listener (participant)
- The gender of the listener (male or female)
- The social class of the listener (working-class or middle-class)

In addition, a series of linear regression mixed-effects tests attempting to find statistical differences between the two groups of respondents were carried out. In these tests, *dialect exposure* (high or low) was entered as a new fixed effect.

3.4.1 Analysis of listener perceptions of vowels

The following section focuses on evaluation of speaker-indexical-social information of localised and non-localised vowel variants occurring in Tyneside English.

3.4.1.1 Perceptions of the variants of the FACE vowel

Table 3.1 presents patterns of use of the FACE vowel variants in Tyneside English.

Variants of the FACE vowel	Speakers
[ɪə] – a localised variant	most often used by older working-class (WC) male speakers
[eː] – a non-localised variant	used across all groups of speakers

Table 3.1 Usage patterns of the FACE vowel variants in Tyneside English (Beal et al., 2012; Watt, 2000; Watt & Milroy, 1999).

Figures 3.9 and 3.10 show evaluations of perceived speaker femaleness and maleness, respectively. As can be seen, both variants were evaluated as overall male-sounding. Also, the best-fit logistic regression mixed effects models of perceived femaleness (Table 3.2) and maleness (Table 3.3) did not show a statistically significant difference between evaluations of the two variants in terms of gender (maleness and femaleness). It can be noticed that Figures 3.9 and 3.10 are mirror images of one another. This would imply that listeners were consistent with their evaluations.

Furthermore, from the results presented in the graphs it can be concluded that evaluations provided by listeners seem to reflect the results of the production studies (Beal et al., 2012; Watt, 2000; Watt & Milroy, 1999).

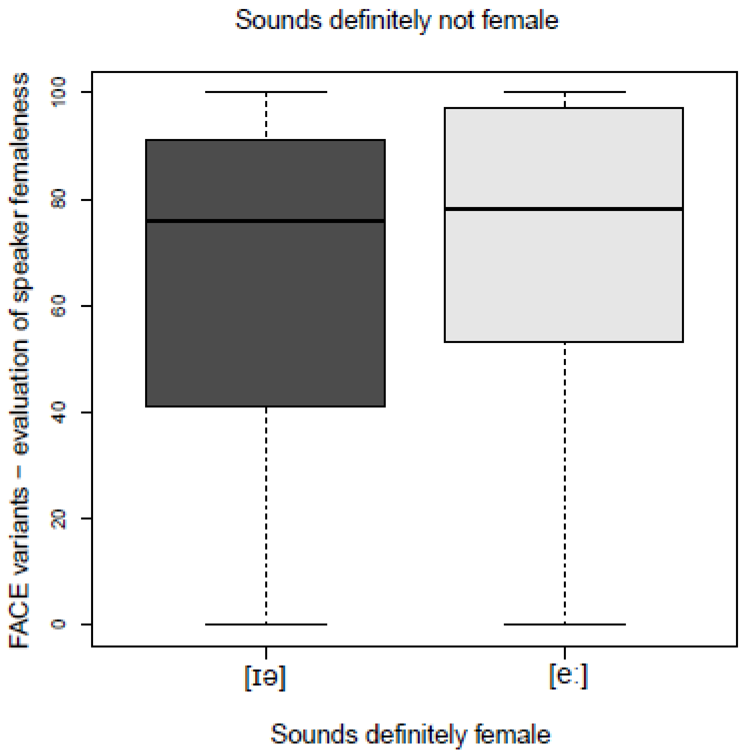


Fig. 3.9 FACE **localised [ɪə] and non-localised [e:] variants -- evaluation of speaker gender. Tyneside group.**

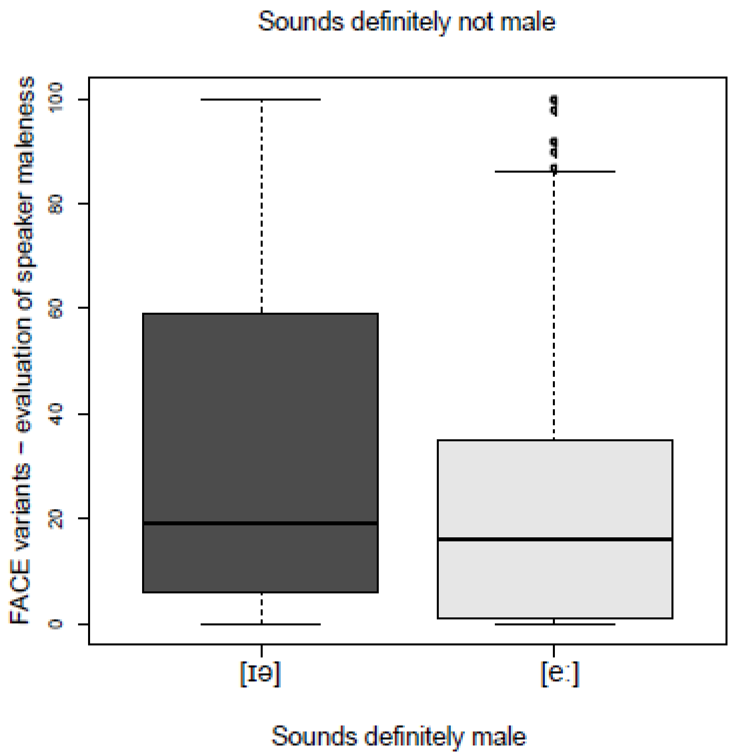


Fig. 3.10 FACE **localised [ɪə] and non-localised [e:] variants -- evaluation of speaker gender. Tyneside group.**

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	73.661	7.057	10.438	<.001	***
Variant [ɪə]	-3.495	8.575	-0.408	0.7043	
Listener gender (Male)	-15.069	7.944	-1.897	<.1	

Table 3.2 Best-fit linear regression mixed effects model of perception of speaker femaleness for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. Random slopes were (1 + variant | listener) and random intercepts were (1 | audio). Number of observations (N)=186; Listener=31. Tyneside group.

Comparison of models showed that the best-fit model to the data for perceived femaleness of the speaker included *variant* as well as *gender of the listeners* as fixed effects (Table 3.2). The same fixed effects were used in the best-fit model to the data for perceived maleness of the speaker (Table 3.3).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	23.070	7.029	3.282	<.05	*
Variant [ɪə]	5.559	8.396	0.662	0.5441	
Listener gender (Male)	11.453	8.527	1.343	<.1	

Table 3.3 Best-fit linear regression mixed effects model of perception of speaker maleness for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. N=186; Listener=31. Tyneside group.

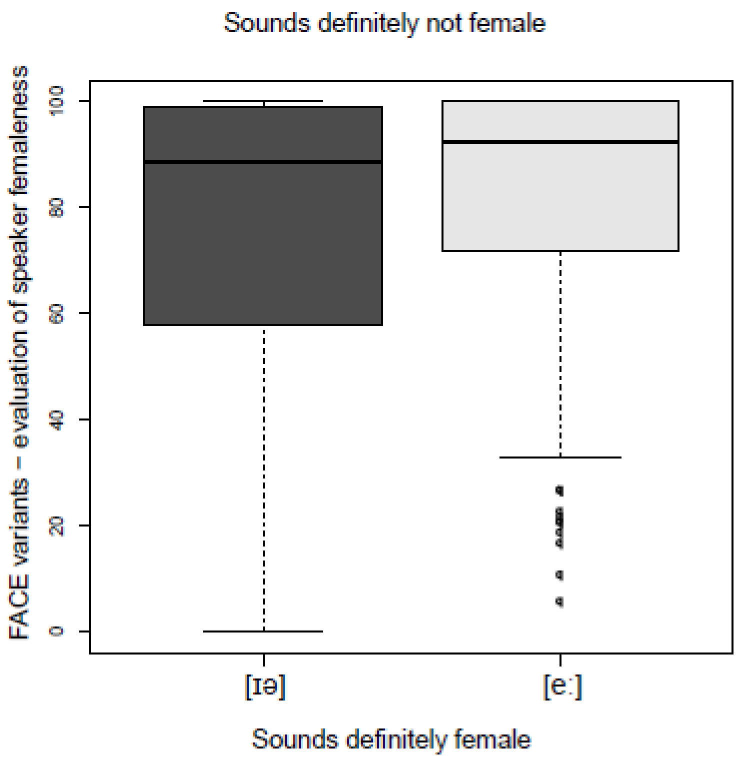


Fig. 3.11 FACE localised [ɪə] and non-localised [eː] variants -- evaluation of speaker gender. North-East group.

Figure 3.11 presents evaluations of perceived femaleness of localised and non-localised FACE variants. This time, evaluations are provided by North-East listeners who comprised the comparison group. As can be noticed, the North-East group evaluated both variants as more definitely not female-sounding (Fig. 3.11) by comparison with Tyneside listeners (Figs. 3.9 & 3.10). On these grounds an argument could be put forward that perhaps North-East listeners were slightly less sensitive to speaker-indexical information carried by the segments than Tyneside listeners. Furthermore, it is often the case that people tend to judge localised variants as male-sounding. Nevertheless, statistical analysis of perceived femaleness for the North-East group of listeners showed no effect of the variant (Table 3.4).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	79.848	5.852	13.644	<.001	***
Variant [ɪə]	-5.481	7.054	-0.777	0.476	

Table 3.4 Best-fit linear regression mixed effects model of perception of speaker femaleness for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. N=198; Listener=33. North-East group.

Mixed-effects linear regression comparing the Tyneside and North-East groups of listeners showed no statistically significant differences in terms of the perceived maleness of the speaker when the FACE variants were investigated. As far as the perceived femaleness of the speaker in the two groups was concerned, the results showed no effect of phonetic variant (Table 3.5). However, there was a tendency to evaluate [ɪə] as more female sounding when listeners belonged to the group with high exposure to the dialect ($p<0.05$).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	83.445	6.526	12.786	<.001	***
Variant [ɪə]	-4.542	7.710	-0.589	0.586	
Listener gender (Male)	-11.431	5.008	-2.282	<.05	*
Dialect exposure	-9.910	4.579	-2.164	<.05	*

Table 3.5 Best-fit linear regression mixed effects model of perception of speaker femaleness for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. N=382; Listener=64. Tyneside and North-East groups.

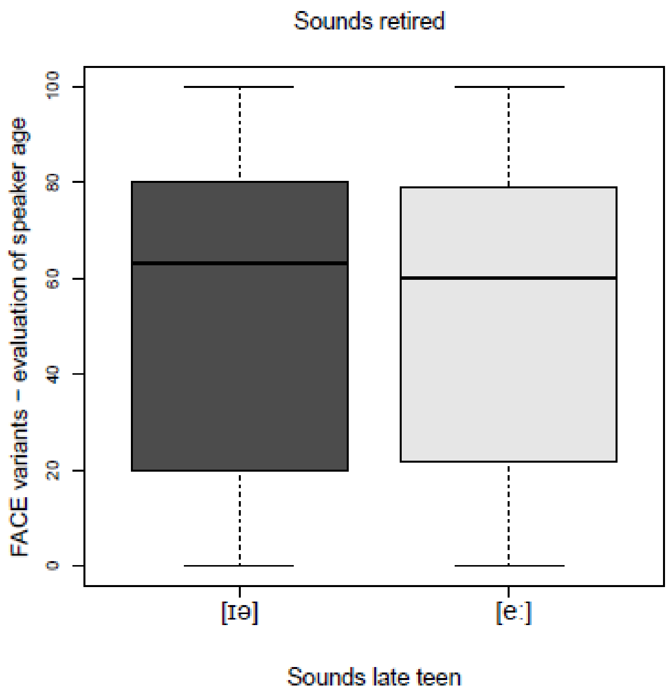


Fig. 3.12 FACE localised [ɪə] and non-localised [eː] variants -- evaluation of speaker age. Tyneside group.

In terms of the perceived age of speakers (Fig. 3.12), the results show that the localised and non-localised variants were judged equally in the middle range of the scale, which might imply that most participants found the voices to be mature-sounding. Furthermore, the spread of evaluations would suggest that listeners varied considerably in their perceptions. This was confirmed by the statistical results presented in Table 3.6. The results show that there was a tendency for the localised variant [ɪə] as well as the non-localised [eː] to be evaluated as older sounding. However, the results were not significant. The best-fit model to the data included variant and speaker as fixed effects.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	38.075	11.496	3.312	<.05	*
Variant [ɪə]	0.397	11.361	0.035	0.974	
Speaker (B)	21.241	11.907	1.784	0.172	

Table 3.6 Best-fit linear regression mixed effects model of perception of speaker age for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. N=186; Listener=31. Tyneside group.

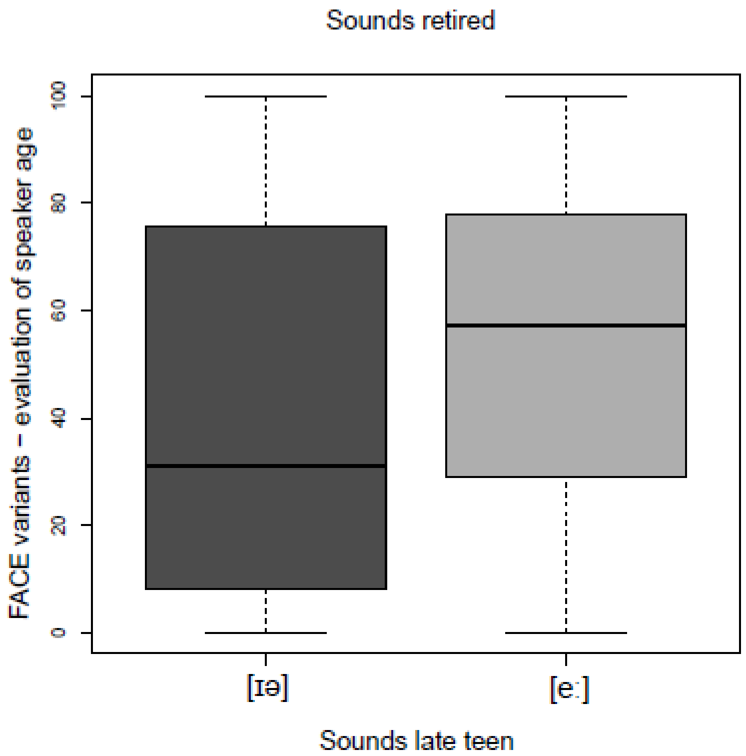


Fig. 3.13 FACE localised [ɪə] and non-localised [eː] variants -- evaluation of speaker age. North-East group.

Evaluations of speaker age show that North-East listeners found the two FACE variants, and especially the localised [ɪə], to be slightly younger-sounding in comparison to Tyneside listeners (Fig. 3.13). While perhaps in the case of [ɪə] we can notice age-gender interaction (cf. Section 3.2.1), the non-localised variant is characterised by a wide spread of evaluations, indicating that speakers of any age could use it. Table 3.7 presents statistical results for evaluations of speaker age in the North-East group of respondents. *Variant* and *speaker* were entered as fixed effects in the best-fit model. The spread of evaluations was so large (Fig. 3.13) that there was no effect of variant on perception of speaker age.

Furthermore, no statistically significant results between the Tyneside and North-East groups were reported in terms of the perceived age of the speaker.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	40.07	12.47	3.212	<.05	*
Variant [ɪə]	-12.11	12.35	-0.981	0.395	
Speaker (B)	19.12	12.91	1.481	0.235	

Table 3.7 Best-fit linear regression mixed effects model of perception of speaker age for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. N=198; Listener=33. North-East group.

As far as perceptions of speaker social class are concerned, it can be noticed that in the Tyneside group of respondents the localised [ɪə] was found to sound more working-class (Fig. 3.14) which corroborates the findings of the production studies. By contrast, the pan-northern variant [eː] was found to be less working-class sounding when compared with the localised variant. Furthermore, when taking a closer look at the graph it can be noticed that the [eː] variant is characterised by a considerably wide spread of evaluations, including working-class judgements as well as non-working-class judgements. What is interesting, especially in comparison with the FACE results obtained for gender and age, is that participants seemed to be quite sensitive to the social-class characteristics of the speakers. As far as statistical analysis is concerned, the best-fit model to the data included *variant*, *speaker* and *gender* of the listeners as fixed effects (Table 3.8). A linear regression mixed effects test showed a statistically significant difference ($p<0.01$) between evaluations of the localised variant [ɪə] and pan-northern variant [eː] in terms of speaker social-class (Table 3.8). The localised [ɪə] was more likely to be evaluated as working-class in comparison to [eː]. In addition, the results show that when the gender of the listener was male, there was a higher probability of variants being evaluated as working-class ($p<0.01$).

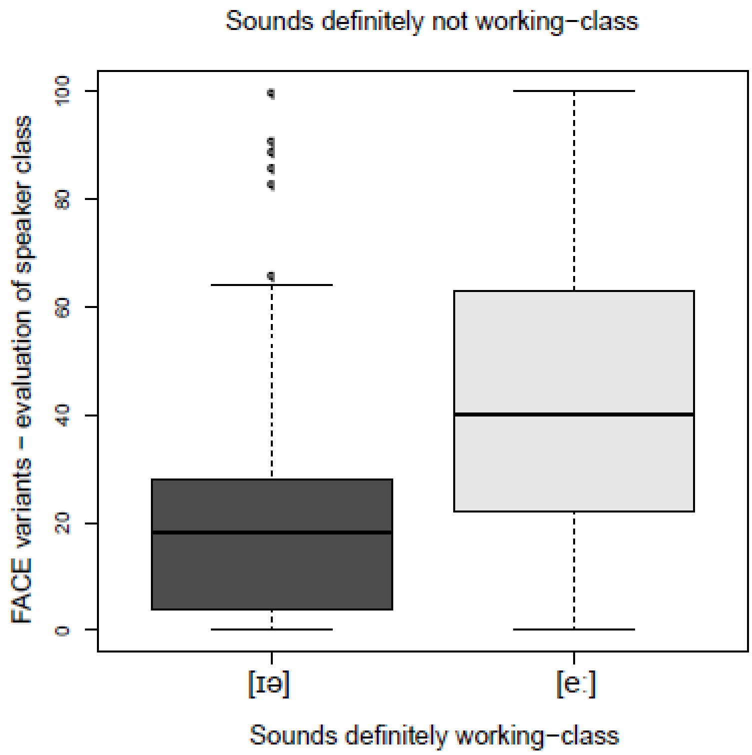


Fig. 3.14 FACE localised [ɪə] and non-localised [eː] variants -- evaluation of speaker class. Tyneside group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	40.375	5.373	7.514	<.001	***
Variant [ɪə]	-21.656	4.684	-4.623	<.01	**
Speaker (B)	8.621	4.273	2.017	0.137	
Listener gender (Male)	-14.637	5.240	-2.794	<.01	**

Table 3.8 Best-fit linear regression mixed effects model of perception of speaker social class for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. N=186; Listener=31. Tyneside group.

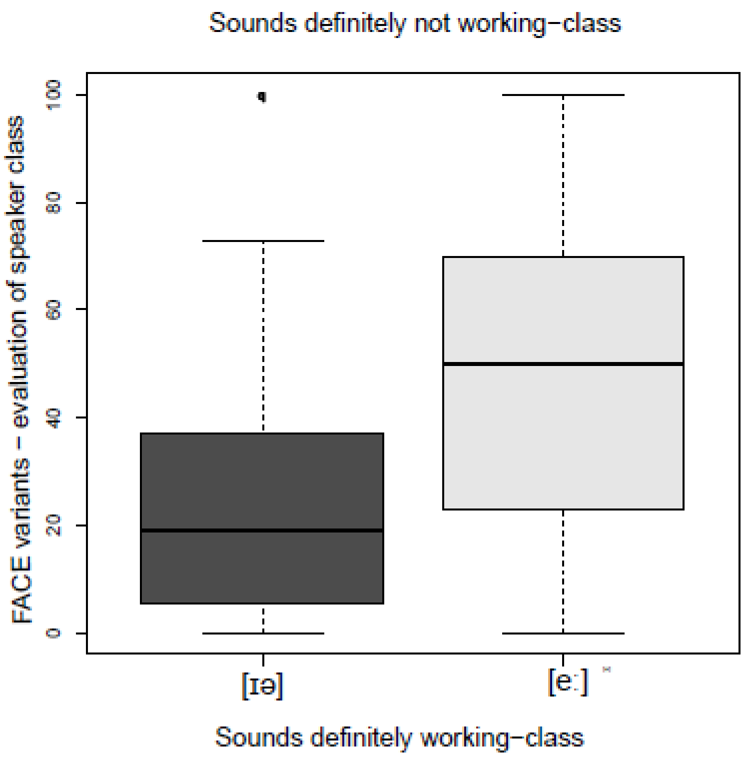


Fig. 3.15 FACE localised [ɪə] and non-localised [eː] variants -- evaluation of speaker class. North-East group.

Evaluations of speaker social-class show that the North-East group also found the two variants distinctly different (Fig. 3.15). However, North-East listeners did not vary in their perceptions from Tyneside listeners. While localised [ɪə] was found to be working-class sounding, the pan-northern [eː] was evaluated as less working-class sounding, with median evaluations around the midpoint on the scale.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	40.609	5.997	6.771	<.001	***
Variant [ɪə]	-24.797	6.245	-3.970	<.01	**
Speaker (B)	10.696	5.344	2.001	0.139	
Listener gender (Male)	0.379	4.717	0.080	0.936	

Table 3.9 Best-fit linear regression mixed effects model of perception of speaker social class for variants of the FACE vowel, where [ɪə] is a localised variant compared against the non-localised [eː]. N=198; Listener=33. North-East group.

In evaluations of speaker social class the best-fit model for the North-East group of respondents included *variant*, *speaker* and *listener gender* as fixed effects (Table 3.9). Even though *speaker* and *listener gender* were not significant factors, they contributed to the model. The results show that among the North-East group of respondents the localised variant was found to be more likely to be rated as having been spoken by a more working-class sounding speaker ($p<0.01$). In this respect, the two groups of respondents did not differ at all. This was further confirmed by a mixed effects linear regression test comparing the two groups of respondents, whereby no statistical differences were found. Nevertheless, it seems that listeners in both groups were quite sensitive to social-class information present in the talkers' speech.

3.4.1.2 Perceptions of the variants of the GOAT vowel

Table 3.10 presents patterns of use of the GOAT vowel variants in Tyneside English.

Variants of the GOAT vowel	Speakers
[ɵː] – a fronted monophthongal localised variant	most often used by younger MC males but also older and younger WC males
[iːə] – an archaic localised variant	found in the speech of older WC males
[ʊə] – a centring diphthongal localised variant	most often used by older WC males
[oː] – a non-localised monophthongal variant	used across all groups of speakers

Table 3.10 Usage patterns of the GOAT vowel variants in Tyneside English (Beal et al., 2012; Watt, 2000; Watt, 1998; Viereck, 1968).

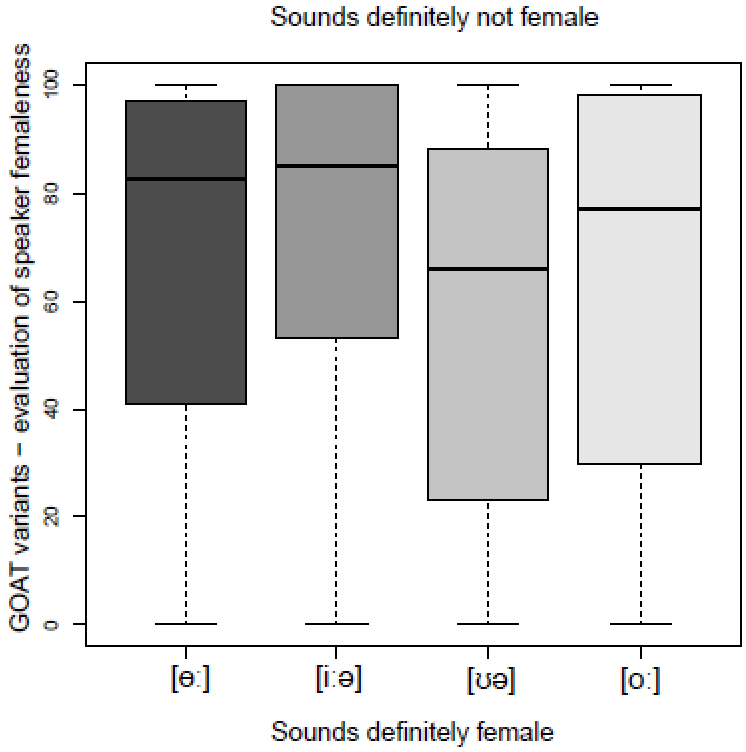


Fig. 3.16 GOAT localised [ɵː], [iːə] and [ʊə] and non-localised [oː] variants – evaluation of speaker gender. Tyneside group.

Figure 3.16 presents evaluations of speaker gender of the three localised GOAT variants. It seems that listeners found the fronted monophthongal and archaic variants to be very similar-sounding in terms of perceived speaker femaleness – that is, as not female sounding. The centring diphthong [ʊə] received the widest spread of evaluations in the second quartile, which might suggest that listeners did not find the variant to be particularly male sounding (Fig. 3.16). Finally, the non-localised variant [oː] was found to be overall not female sounding, albeit less so than the fronted monophthongal [ɵː] and archaic [iːə] variants.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	68.435	7.626	8.974	<.001	***
Variant [ɵː]	4.371	12.981	0.337	0.745	
Variant [iːə]	5.984	13.238	0.452	0.662	
Variant [ʊə]	-7.677	13.261	-0.579	0.577	
Listener gender (Male)	-17.571	7.280	-2.413	<.05	*

Table 3.11 Best-fit linear regression mixed effects model of perception of speaker femaleness for variants of the GOAT vowel, where [ɵː], [iːə] and [ʊə] are localised variants compared against the non-localised [oː]. N=372; Listener=31. Tyneside group.

In the best-fit model investigating perceived femaleness of the speaker, *variant*, and *listener gender* were entered as fixed effects (Table 3.11). While *variant* yielded no statistical effect, some effect of the *listener gender* was reported. The variants were more likely to be evaluated as female when the gender of the listener was male ($p<0.05$).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	34.510	7.167	4.815	<.001	***
Variant [ɵː]	-7.172	13.199	-0.543	0.600	
Variant [iːə]	-6.397	13.324	-0.480	0.642	
Variant [ʊə]	-0.591	13.370	-0.044	0.965	

Table 3.12 Best-fit linear regression mixed effects model of perception of speaker maleness for variants of the GOAT vowel, where [ɵː], [iːə] and [ʊə] are localised variants compared against the non-localised [oː]. N=372; Listener=31. Tyneside group.

In the Tyneside group of respondents, in the best-fit model to the data of perceived speaker maleness only *variant* was entered as a fixed effect (Table 3.12). However, *variant* was not reported to be statistically significant.

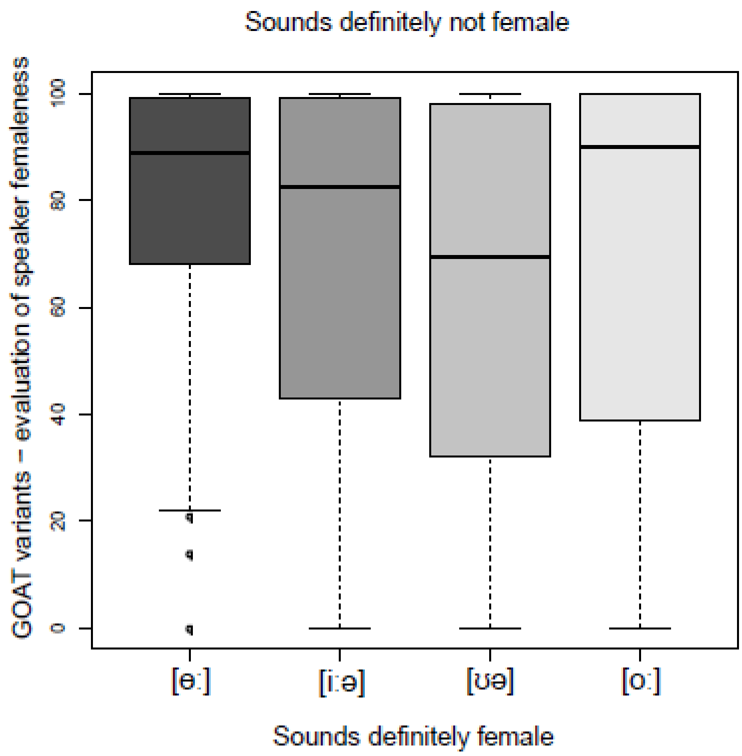


Fig. 3.17 GOAT localised [ɵː], [iːə] and [ʊə] and non-localised [oː] variants – evaluation of speaker gender. North-East group.

The fronted monophthongal [ɵː], which is most often used by younger middle-class males but also older and younger working-class males, was evaluated as more definitely not female-sounding by North-East listeners than it was by Newcastle listeners (Figs. 3.16 & 3.17). The archaic [iːə], on the other hand, was rated as more female sounding among the North-East group of respondents in comparison to the Tyneside ratings. The centring diphthong [ʊə], used most often by older working class males, was evaluated as less female sounding by the North-East group of respondents. Finally, the non-localised [oː] was evaluated as less female sounding (Fig. 3.17) by the North-East group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	70.571	7.596	9.291	<.001	***
Variant [ɵː]	9.172	14.037	0.653	0.531	
Variant [iːə]	-1.362	14.347	-0.095	0.927	
Variant [ʊə]	-7.536	14.242	-0.529	0.610	

Table 3.13 Best-fit linear regression mixed effects model of perception of speaker femaleness for variants of the GOAT vowel, where [ɵː], [iːə] and [ʊə] are localised variants compared against the non-localised [oː]. N=396; Listener=33. North-East group.

As far as the statistical results for perceived femaleness among the North-East group of respondents are concerned, the best-fit model was the simplest model which included *variant* as a fixed effect (Table 3.13). As can be noticed, none of the phonetic variants had a significant effect on evaluations of speaker femaleness. The results for perceived maleness in this group of respondents are reported in Table 3.14 below. As before, no effect for *variant* was found.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	32.818	8.348	3.931	<.01	**
Variant [ɵː]	16.451	19.219	0.856	0.432	
Variant [iːə]	-16.437	16.417	-1.001	0.345	
Variant [ʊə]	-0.545	16.561	-0.033	0.974	

Table 3.14 Best-fit linear regression mixed effects model of perception of speaker maleness for variants of the GOAT vowel, where [ɵː], [iːə] and [ʊə] are localised variants compared against the non-localised [oː]. N=396; Listener=33. North-East group.

Listeners found the localised and archaic variants [ɵː] and [iːə] to be overall male-sounding, which is consistent with the results presented for evaluations of perceived femaleness. However, as can be seen in Figure 3.17, the spread of evaluations shows that some listeners found these variants to be altogether female-sounding. As far as evaluations of the [ʊə] variant are concerned, a spread of evaluations that is noticeable especially in the second and third quartiles (Fig. 3.17) suggests that listeners had quite varied perceptions of this variant and some of the listeners did not have a strong association with the variant as being male sounding.

As far as statistical differences between the Tyneside and North-East groups of listeners are concerned, no such differences were reported for the GOAT variants either in the case of perceived femaleness or with respect to the perceived maleness of the speaker.

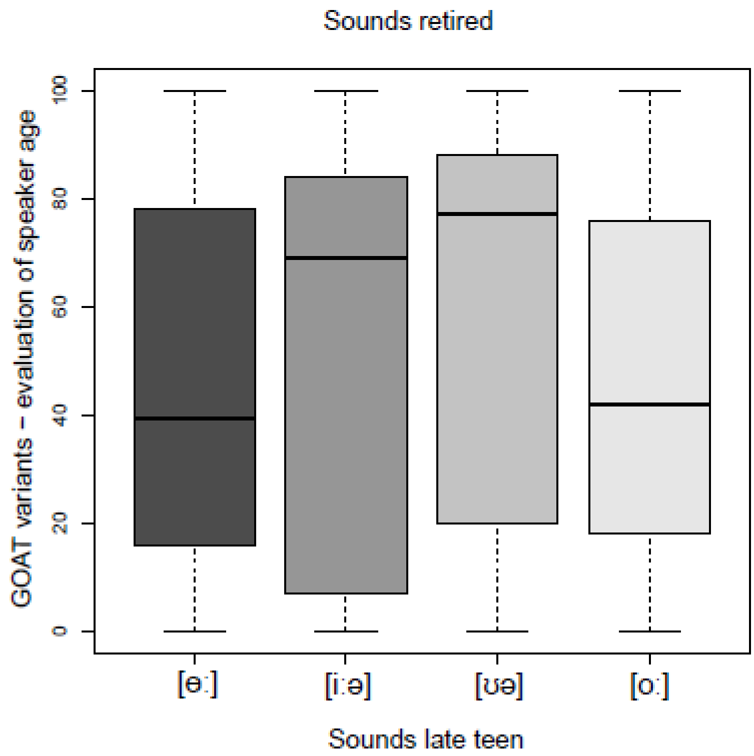


Fig. 3.18 GOAT localised [ɵ:], [i:ə] and [ʊə] and non-localised [o:] variants – evaluation of speaker age. Tyneside group.

The results for perceived speaker age among the Tyneside group of respondents seem to be reflecting the findings of the production studies only to a limited extent. While the non-localised variant [o:] and the localised variant [ɵ:] were found to be in general mature-sounding, the evaluations of the localised variant [ʊə] and the archaic variant [i:ə] were not as clear-cut. Even though the evaluations of [ʊə] and [i:ə] show a significant dispersion of ratings in the second quartile, the median itself is located towards the end of the scale, which means that the variants were generally perceived as older-sounding. This, in turn, correlates with the findings of the production studies. This finding is particularly interesting when compared with the results of the gender evaluation of this variant. Of the four GOAT variants, [ʊə] was judged as somewhat female-sounding by the highest number of respondents. This links with the fact that the same variant was perceived as old-sounding. This is the first instance of a possible interaction between perceived speaker age and gender identified in this experiment (cf. Section 3.2.1). It will be investigated in the following sections whether more examples of such interaction can be reported.

As far as age evaluations of the archaic [iːə] are concerned, a spread of evaluations can be noticed, especially in the second quartile (Fig. 3.18). This might suggest that listeners did not necessarily have any strong associations of this variant with older working-class males simply because of the young age of the listeners and the fact that the variant was already recessive in the mid- to late-1990s. Nevertheless, the median evaluation is towards the end of the scale, which shows that the majority of respondents perceived in a similar way to the [ʊə] variant.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	48.042	4.272	11.245	<.001	***
Variant [ɵː]	-3.074	6.323	-0.486	0.639	
Variant [iːə]	4.265	7.268	0.587	0.567	
Variant [ʊə]	13.781	7.020	1.963	<.1	

Table 3.15 Best-fit linear regression mixed effects model of perception of speaker age for variants of the GOAT vowel, where [ɵː], [iːə] and [ʊə] are localised variants compared against the non-localised [oː]. N=372; Listener=31. Tyneside group.

As far as statistical results for evaluations of speaker age are concerned, the best-fit model to the data in the Tyneside group of respondents was the simplest model including *variant* as a fixed effect (Table 3.15). Except for a slight tendency for the centring diphthong [ʊə] to be heard as having been spoken by an older speaker, none of the variants had a significant effect.

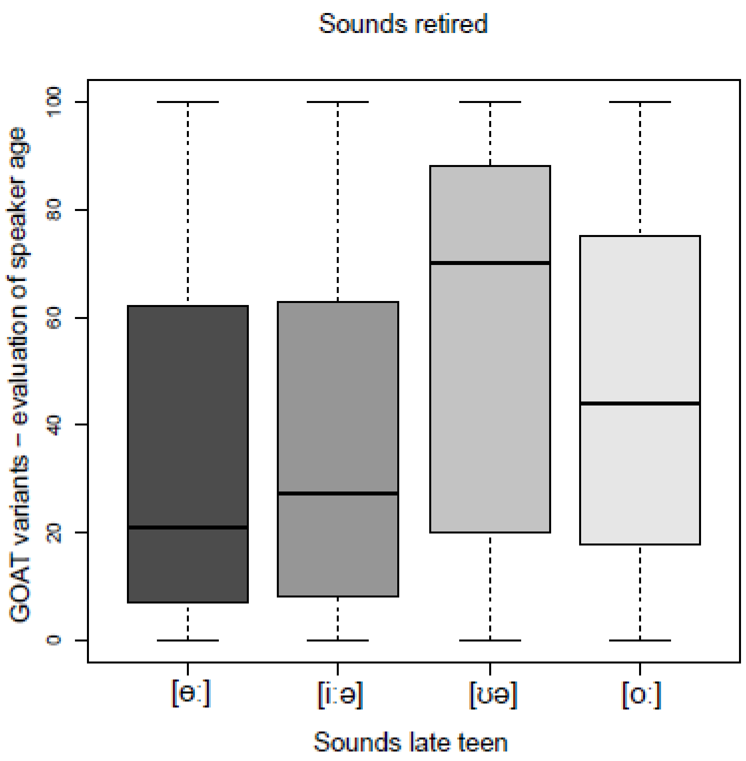


Fig. 3.19 GOAT localised [ɵː], [iːə] and [ʊə] and non-localised [oː] variants – evaluation of speaker age. North-East group.

The evaluations of speaker age among the North-East group of respondents present interesting results (Fig. 3.19). Except for the non-localised variant, North-East listeners judged all the localised variants as younger-sounding than did Tyneside listeners. The largest difference in evaluations can be noticed for the archaic diphthong [iːə], then the localised [ɵː], and finally the centring diphthong [ʊə]. While the non-localised [oː] was evaluated around the mid-range of the scale, which might suggest that the voices were perceived as mature yet young-sounding, the other two variants, localised [ɵː] and [iːə], which were also evaluated as definitely male-sounding, were rated as younger-sounding than [ʊə] and [oː].

These results might suggest that listeners with low exposure to the variants under investigation might be even more prone to the age-gender effect discussed earlier.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	47.667	4.440	10.737	<.001	***
Variant [ɵː]	-13.761	7.824	-1.759	0.110	
Variant [iːə]	-13.580	7.935	-1.711	0.117	
Variant [ʊə]	10.017	7.542	1.328	0.218	

Table 3.16 Best-fit linear regression mixed effects model of perception of speaker age for variants of the GOAT vowel, where [ɵː], [iːə] and [ʊə] are localised variants compared against the non-localised [oː]. N=396; Listener=33. North-East group.

Table 3.16 presents the best-fit model to the age data for the North-East listener group. The model was simple, where the only fixed effect was *variant*. As can be noticed, none of the variants had a statistical effect upon the evaluations of speaker age. As may be recalled, the results in the Tyneside group (Table 3.15) were almost identical. Furthermore, among the Tyneside and North-East groups of listeners no statistically significant differences were reported for the GOAT variants in terms of perceived speaker age. Thus, the two groups of respondents did not differ in their perceptions of speaker age.

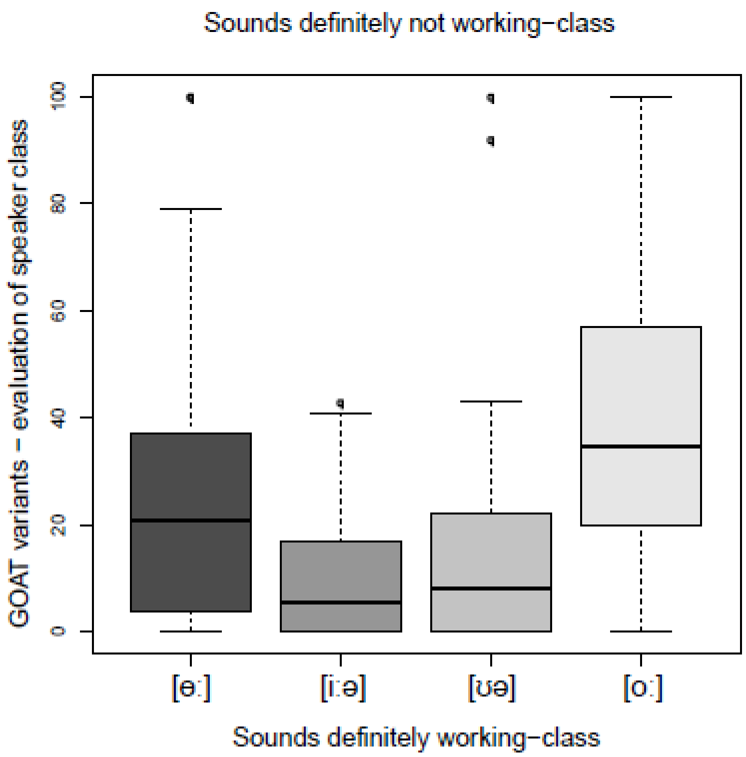


Fig. 3.20 GOAT localised [ɵː], [iːə] and [ʊə] and non-localised [oː] variants – evaluation of speaker social class. Tyneside group.

As far as perceptions of the GOAT variants in terms of speaker social class are concerned, it can be noticed that once again, Tyneside listeners were quite sensitive to speaker social-indexical information encoded by the phonetic variants under investigation, albeit to a varied degree (Fig. 3.20). A closer look at Figure 3.20 reveals that the archaic [iːə] and local [ʊə] variants were, in fact, evaluated as working-class sounding. While the localised [ɵː] was found to be much less working-class sounding than the other two localised variants, it was also perceived to be more working-class sounding in comparison with the non-localised variant [oː]. Once again, it seems that the listeners were quite sensitive to social-class information in the case of the archaic variant.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	34.428	4.634	7.429	<.001	***
Variant [ɵː]	-7.670	6.709	-1.143	0.285	
Variant [iːə]	-23.960	6.851	-3.497	<.01	**
Variant [ʊə]	-26.754	6.388	-4.188	<.01	**
Speaker (B)	15.361	5.562	2.762	<.05	*

Table 3.17 Best-fit linear regression mixed-effects model of perception of speaker class for variants of the GOAT vowel, where [ɵː], [iːə] and [ʊə] are localised variants compared against the non-localised [oː]. N=372; Listener=31. Tyneside group.

As far as statistical results for speaker class in the Tyneside group of respondents are concerned, the best-fit model to the data included *variant* and *speaker* as fixed effects (Table 3.17). It can be noticed that two of the localised variants, the archaic [iːə] and localised [ʊə] were significantly more likely to be perceived as having been produced by a working class-speaker ($p<0.01$). In addition, an effect for *speaker* was reported. That is, when words were produced by speaker B, they were more likely to be seen as having been uttered by a less working class-sounding person ($p<0.05$).

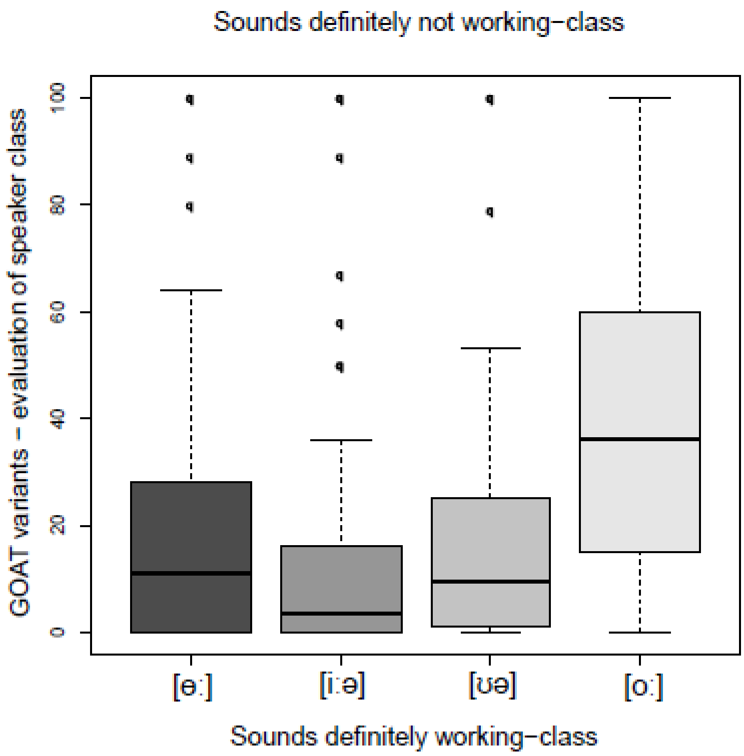


Fig. 3.21 GOAT localised [ɵ:], [i:ə] and [ʊə] and non-localised [o:] variants – evaluation of speaker social class. North-East group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	34.275	3.780	9.067	<.001	***
Variant [ɵ:]	-16.187	5.342	-3.030	<.05	*
Variant [i:ə]	-21.901	5.335	-4.105	<.01	**
Variant [ʊə]	-23.478	5.169	-4.542	<.001	***
Speaker (B)	7.678	4.043	1.899	<.1	

Table 3.18 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of the GOAT vowel, where [ɵ:], [i:ə] and [ʊə] are localised variants compared against the non-localised [o:]. N=396; Listener=33. North-East group.

For the North-East group of participants the best-fit model to the speaker-social class data included *variant* and *speaker* as fixed effects (Table 3.18). It can be noticed that the localised [ʊə] was the most likely to be evaluated as having been spoken by a working-class speaker ($p<0.001$). Right behind it was the archaic variant [i:ə] ($p<0.01$) and finally the last localised variant [ɵː] ($p<0.05$). Thus, all localised variants were more likely to be evaluated as having been uttered by a working-class speaker in comparison to the non-localised [oː]. These results show that although both groups of respondents were sensitive to the social-class information encoded by the variants, it seems that the results obtained in the Tyneside group of listeners (Table 3.17) reflected findings of the production studies more adequately than the results reported for the North-East group of listeners (Table 3.18). Nevertheless, a mixed effects linear regression test showed no statistically significant differences between the two groups of listeners.

3.4.1.3 Perceptions of the variants of the NURSE vowel

Table 3.19 presents patterns of use of the NURSE vowel variants in Tyneside English.

Variants of the NURSE vowel	Speakers
[øː] – a fronted localised variant	most often used by young MC and WC females, but also older WC females
[ɔː] – a retracted localised variant	most often used by older WC males
[ɜː] – a non-localised centralised variant	used across all speaker groups

Table 3.19 Usage patterns of the NURSE vowel variants in Tyneside English (Beal et al., 2012; Watt & Milroy, 1999; Watt, 1998).

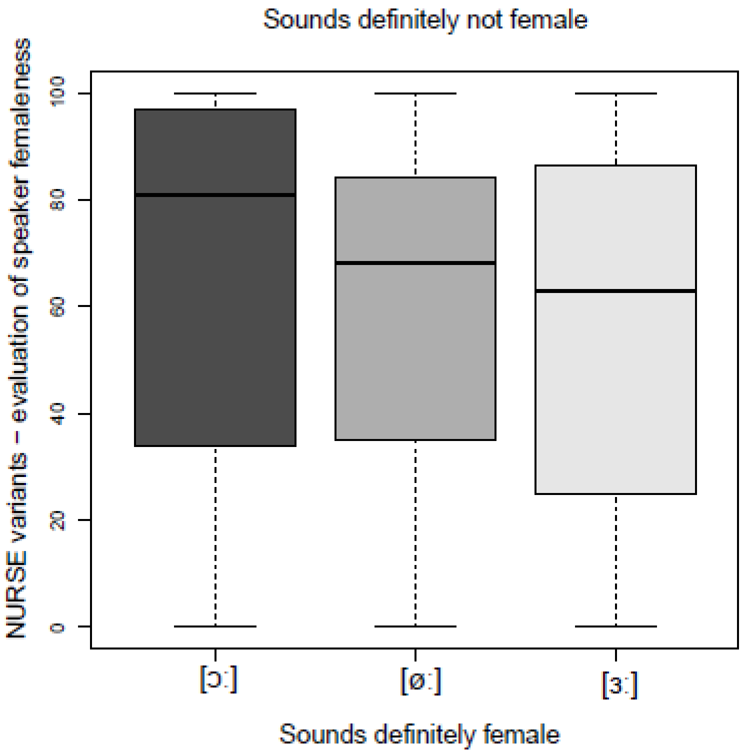


Fig. 3.22 NURSE localised [ɔː], [øː] and non-localised [ɜː] variants -- evaluation of speaker gender. Tyneside group.

The final vowel investigated in the experiment is the NURSE vowel (Fig. 3.22). When looking at Fig. 3.22 it can be noticed that in the Tyneside group of respondents the fronted variant [øː] was evaluated as more female sounding than the retracted variant [ɔː].

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	65.973	7.020	9.398	<.001	***
Variant [ɔː]	12.066	7.423	1.625	0.141	
Variant [øː]	10.716	8.133	1.318	0.230	
Speaker (B)	-14.304	6.944	-2.060	<.1	
Listener gender (Male)	-16.844	9.610	-1.753	<.1	

Table 3.20 Best-fit logistic regression mixed effects model of perception of speaker femaleness for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=372; Listener=31. Tyneside group.

Regarding the statistical results of perceived femaleness of the speaker in the Tyneside group, the best-fit model to the data included *variant*, *speaker* and *listener gender* as fixed effects (Table 3.20). None of the phonetic variants or other predictors had a significant effect on evaluation of speaker femaleness. However, there was a slight tendency towards evaluating the speaker as more female-sounding when the words were produced by speaker B. There was also a tendency among male respondents to rate the speaker as more female-sounding. A closer look at Table 3.21 shows that the statistical results reported for the perceived maleness of the speaker differed from the results for perceived femaleness. This is the first instance of such discrepancy in terms of perception of speaker gender. The best-fit model included *variant* and *speaker* as fixed effects. While the retracted variant [ɔː] was more likely to be rated as having been spoken by a male speaker ($p<0.05$), for the fronted variant [øː] this was only a slight tendency. It is worth pointing out that these findings are in line with the results of the production studies. Furthermore, the effect of speaker voice was reported. When words were produced by speaker B, listeners were more likely to see them as having been produced by a less definitely male-sounding speaker.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	38.458	6.149	6.255	<.001	***
Variant [ɔː]	-15.156	5.968	-2.540	<.05	*
Variant [øː]	-14.703	6.492	-2.265	<.1	
Speaker (B)	17.030	5.161	3.300	<.05	*

Table 3.21 Best-fit logistic regression mixed effects model of perception of speaker maleness for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=372; Listener=31. Tyneside group.

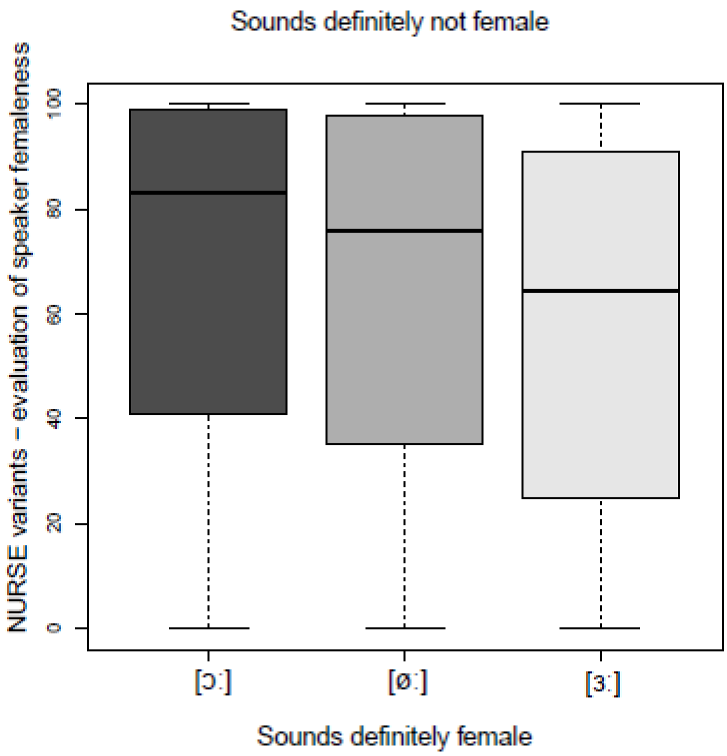


Fig. 3.23 NURSE localised [ɔː], [øː] and non-localised [ɜː] variants -- evaluation of speaker gender. North-East group.

In the North-East group's evaluations of the perceived femaleness of the speaker, the local fronted variant [øː] was found to be less female-sounding in comparison with the Tyneside group of respondents (Figs. 3.22 & 3.23). The other two variants, that is, the retracted [ɔː] and centralised [ɜː], were evaluated almost identically by both groups of listeners. Generally, the NURSE variants were found to be male-sounding, although a considerable spread of evaluations in the second quartile across both groups of listeners can be noticed.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	76.386	8.974	8.512	<.001	***
Variant [ɔː]	15.290	9.807	1.559	0.156	
Variant [øː]	17.232	10.788	1.597	0.161	
Speaker (B)	-17.424	9.228	-1.888	<.1	
Listener gender (Male)	-10.979	5.996	-1.831	<.1	
Listener class (WC)	-10.088	6.233	-1.618	0.114	

Table 3.22 Best-fit logistic regression mixed effects model of perception of speaker femaleness for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=396; Listener=33. North-East group.

The statistical results for the perceptions of speaker femaleness were identical for both groups of respondents (Tables 3.20 & 3.22). For the North-East group, in the best-fit model investigating perceptions of speaker femaleness, *variant*, *speaker*, *listener gender* and *listener social class* were entered as fixed effects.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	32.908	5.032	6.539	<.001	***
Variant [ɔː]	-24.771	5.669	-4.369	<.01	**
Variant [øː]	-24.359	6.208	-3.924	<.01	**
Speaker (B)	19.043	5.369	3.547	<.01	**

Table 3.23 Best-fit logistic regression mixed effects model of perception of speaker maleness for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=396; Listener=33. North-East group.

The best-fit model to the data included *variant* and *speaker* as fixed effects (Table 3.23). Both localised variants ([ɔː] and [øː]) were more likely to be perceived as having been uttered by a male speaker (p<0.01). In addition, when words were produced by speaker B, listeners were more likely to perceive the speaker as less definitely male-sounding.

Even though no statistically significant differences between the Tyneside and North-East groups of listeners in terms of the perceived femaleness of the speaker were reported, such differences were found for the perceived maleness of the speaker (Table 3.24). In the presented model, *variant*, *speaker* and *exposure to the dialect* were entered as fixed effects. As can be observed, both localised variants, [ɔː] and [øː] were more likely to be perceived as male (p<0.01). However, listeners with high exposure to Tyneside English were more likely to perceive these variants as less male-sounding (p<0.05).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	30.246	5.139	5.886	<.001	***
Variant [ɔː]	-20.092	5.102	-3.938	<.01	**
Variant [øː]	-19.650	5.553	-3.539	<.01	**
Speaker (B)	18.050	4.696	3.843	<.01	**
Dialect exposure	11.064	4.969	2.226	<.05	*

Table 3.24 Best-fit logistic regression mixed effects model of perception of speaker maleness for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=768; Listener=64. Tyneside and North-East groups.

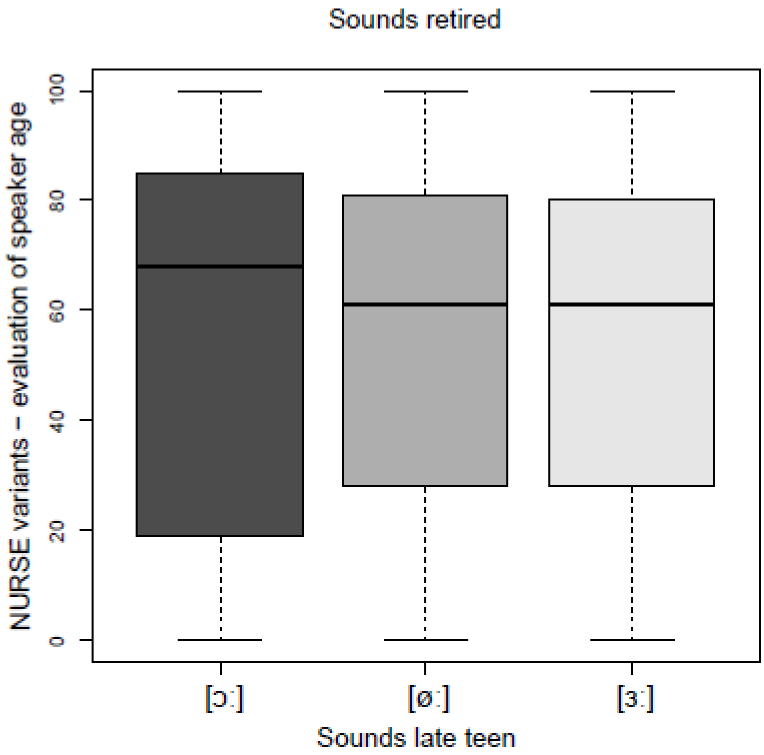


Fig. 3.24 NURSE localised [ɔː], [øː] and non-localised [ɜː] variants -- evaluation of speaker age. Tyneside group.

As far as evaluations of speaker age among the Tyneside group of respondents are concerned (Fig. 3.24), a slight difference in the medians for the retracted [ɔː] and fronted [øː] illustrate a difference in perception. While the variant [øː] was evaluated as slightly less old-sounding, the retracted variant [ɔː] was found to be slightly older-sounding in comparison. The non-localised variant, on the other hand, was evaluated almost identically to the fronted variant. As could be expected from the spread of evaluations, no statistically significant differences were reported (Table 3.25). The best-fit model to the data was the simplest one, and included *variant* as the only fixed effect.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	54.565	4.793	11.384	<.001	***
Variant [ɔː]	2.683	6.206	0.432	0.672	
Variant [øː]	-1.597	5.534	-0.289	0.780	

Table 3.25 Best-fit logistic regression mixed effects model of perception of speaker age for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=372; Listener=31. Tyneside group.

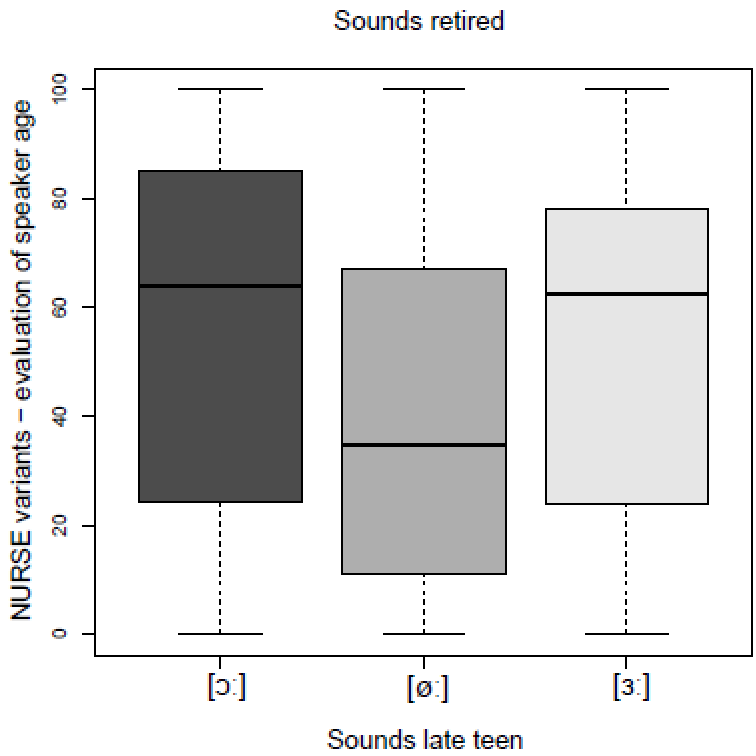


Fig. 3.25 NURSE localised [ɔː], [øː] and non-localised [ɜː] variants -- evaluation of speaker age. North-East group.

As can be observed, among the North-East group of respondents the fronted [øː] was evaluated as younger-sounding in the Tyneside group (Figs. 3.24 & 3.25). The retracted [ɔː] and non-localised [ɜː], on the other hand, were rated as older-sounding than the fronted variant. These evaluations matched evaluations reported by the Tyneside group. Overall, speakers using these two variants sounded mature. The statistical results reported in Table 3.26 show that the fronted variant [øː] was more likely to be perceived as having been produced by a younger speaker ($p<0.05$). The best-fit model included *variant* and *speaker* as fixed effects.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	48.083	5.435	8.846	<.001	***
Variant [ɔː]	0.411	6.275	0.066	0.949	
Variant [øː]	-17.840	7.103	-2.511	<.05	*
Speaker (B)	9.691	5.568	1.741	0.120	

Table 3.26 Best-fit logistic regression mixed-effects model of perception of speaker age for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=396; Listener=33. North-East group.

As far as the perception of age of the speaker from the NURSE variants was concerned, a mixed-effects linear regression test showed that there were no statistically significant differences between the Tyneside and North-East groups of listeners.

The results shown in Figure 3.26 show evaluations of speaker social class by the group of Tyneside respondents. It can be noticed that, as per the previous variables, listeners were again quite sensitive to social class information carried by the three variants. The results corroborate the findings of the production studies. Furthermore, perceptions of these variants in terms of social class were statistically significant (Table 3.27). Both the retracted [ɔː] and fronted [øː] variants were likely to be evaluated as having been spoken by a working-class speaker ($p<0.001$ and $p<0.01$ respectively). As far as the best-fit model is concerned, it included *variant* as the only fixed effect.

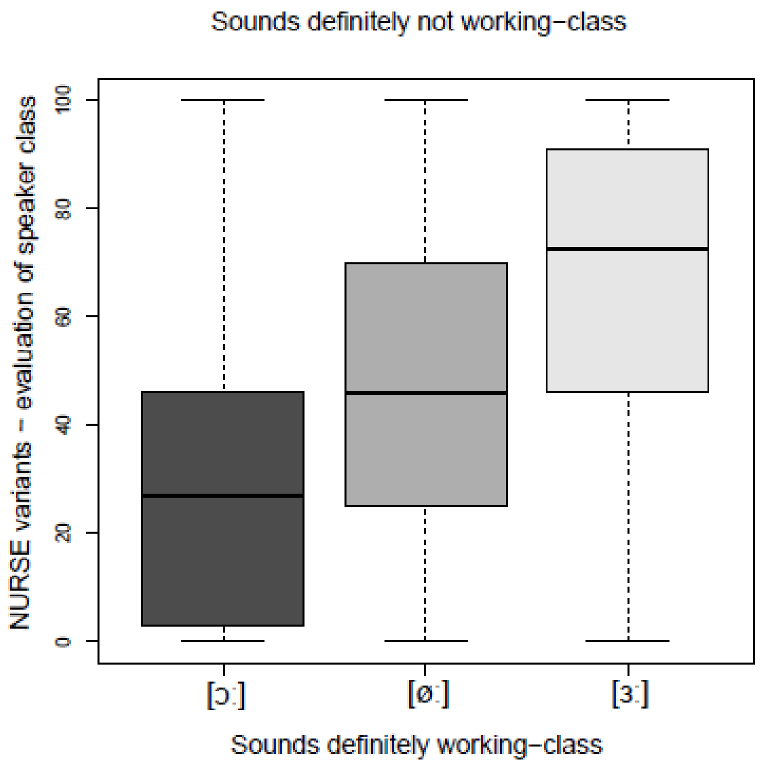


Fig. 3.26 NURSE localised [ɔː], [øː] and non-localised [ɜː] variants -- evaluation of speaker class. Tyneside group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	68.140	4.325	15.755	<.001	***
Variant [ɔː]	-37.645	6.081	-6.191	<.001	***
Variant [øː]	-18.785	5.604	-3.352	<.01	**

Table 3.27 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=372; Listener=31. Tyneside group.

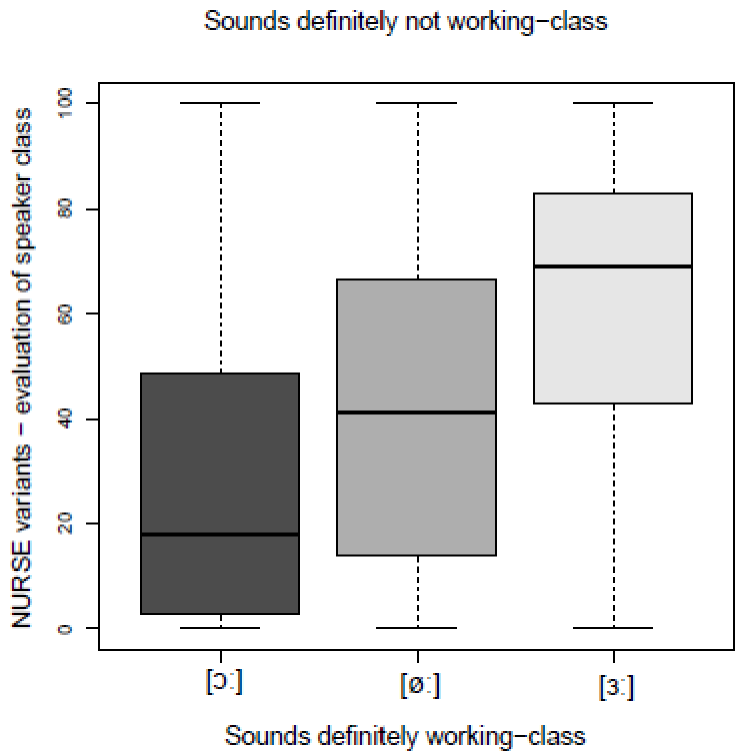


Fig. 3.27 NURSE localised [ɔː], [øː] and non-localised [ɜː] variants -- evaluation of speaker class. North-East group.

Figure 3.27 presents graphical results of evaluations of speaker social class on the basis of the NURSE variants by the North-East group of respondents. As can be observed, the results are virtually identical with the ones reported in the Tyneside group of listeners (Fig. 3.26). As has been already mentioned, the present results reflect the findings of the production studies. Table 3.28, on the other hand, presents the statistical results for the comparison group. Both the fixed effects used in the best-fit model as well as statistical results were identical for the two groups of respondents.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	61.424	3.554	17.283	<.001	***
Variant [ɔː]	-31.648	5.888	-5.375	<.001	***
Variant [øː]	-19.338	4.546	-4.254	<.01	**

Table 3.28 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=396; Listener=33. North-East group.

As might be expected from the results presented above, a comparative linear regression test did not report any significant differences between the Tyneside and North-East groups of listeners as far as the perceived social class of the speaker was concerned.

The results for vowels show that the variants did not differ statistically in terms of the perceived speaker gender. Nevertheless, perceptual differences could be observed in the graphs. The age-gender interaction resulting from the stimuli manipulation was also reported. Finally, localised and non-localised variants were perceived as different in terms of social class. The following section (3.4.2) presents the results for consonants.

3.4.2 Analysis of listener perceptions of plosives

The following section focuses on the evaluation of speaker gender, age and social class of localised and non-localised variants of voiceless plosives occurring in Tyneside English.

3.4.2.1 Perceptions of T-to-R in intervocalic, across word boundaries contexts

Table 3.29 presents patterns of use of the word-final intervocalic /t/ in Tyneside English.

Variants of word-final pre-vocalic /t/	Speakers
[ɹ] – T-to-R	most often used by older WC female speakers
[ʔ͡t] – glottalised /t/	most often found in older MC and WC male speech

Table 3.29 Realisations of word-final intervocalic /t/ in Tyneside English (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy et al., 1994a; Milroy et al., 1994b; Watt & Milroy, 1999).

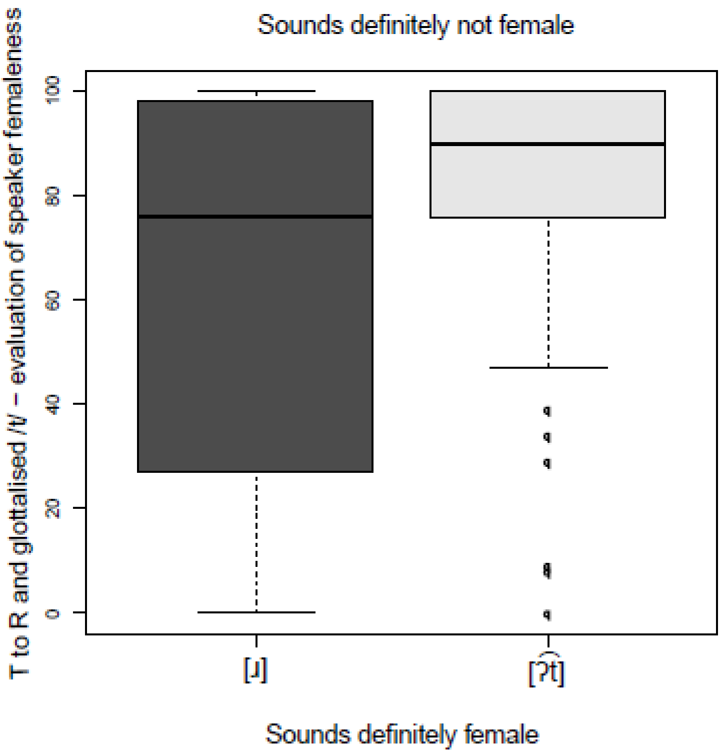


Fig. 3.28 T-to-R [ɹ] and glottalised /t/ [ʔ͡t] -- evaluation of speaker gender. Tyneside group.

Listener perceptions of speaker femaleness and maleness were mirror images of one another. This would imply that listeners had clear and consistent perception of the two variants in terms of the speakers' perceived gender. It is also worth drawing attention to the fact that the results of the perception test correlate with the results of production studies (Docherty & Foulkes, 1999; Milroy et al., 1994a; Milroy et al., 1994b; Watt & Milroy, 1999). While [ʔ͡t] was strongly perceived to be male-sounding (see Fig. 3.28), [ɹ] received a wider spread of evaluations. This would imply that listeners in fact perceived the variant as more female sounding.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	86.593	5.149	16.816	<.001	***
Variant [ɹ]	-21.043	7.478	-2.814	<.05	*
Listener gender (Male)	-14.005	6.069	-2.308	<.05	*

Table 3.30 Best-fit linear regression mixed-effects model of perception of speaker femaleness for variants of the word-final intervocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=186; Listener=31. Tyneside group.

When evaluating speaker femaleness, the best-fit linear regression mixed effects model to the data included *variant* and *listener gender* as fixed effects (Table 3.30). The results show that the [ɹ] variant was significantly more likely to be rated as female-sounding than the [ʔ͡t] variant (p<0.05). At the same time, an effect of *listener gender* was reported. When the listener was male, a variant was more likely to be evaluated as female-sounding (p<0.05).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	10.821	5.042	2.146	<.1	
Variant [ɹ]	24.344	8.079	3.013	<.05	*
Listener gender (Male)	7.365	4.668	1.578	0.125	

Table 3.31 Best-fit linear regression mixed-effects model of perception of speaker maleness for variants of the word-final intervocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=186; Listener=31. Tyneside group.

In the evaluations of the perceived speaker maleness of [ɹ] and [ʔ͡t], the best-fit model to the data included *variant* and *listener gender* as fixed effects (Table 3.31). Even though the last predictor did not have a significant effect, it contributed to the model. The results show that the [ɹ] variant was more likely to be evaluated as less male-sounding than the [ʔ͡t] variant (p<0.05).

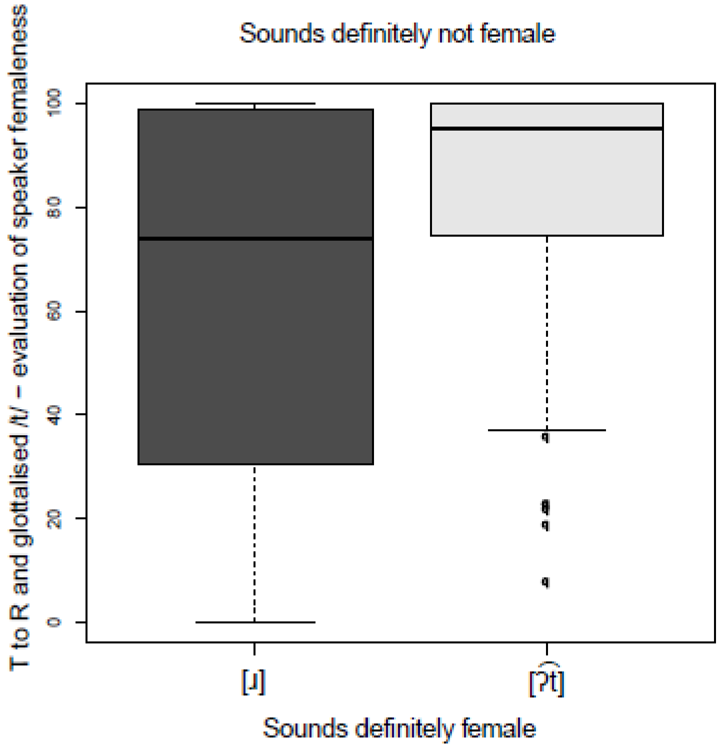


Fig. 3.29 T-to-R [ɹ] and glottalised /t/ [ʔ͡t] variants -- evaluation of speaker gender. North-East group.

Evaluations of the perceived femaleness of [ɹ] and [ʔ͡t] show almost identical results across both groups of listeners (Figs. 3.28 & 3.29). [ʔ͡t] was evaluated as slightly more male-sounding in the North-East group. While [ʔ͡t] was evaluated as definitely not female-sounding, [ɹ] was characterised by a wide spread of evaluations and found to be more female-sounding than [ʔ͡t]. From the results presented for the Tyneside and North-East listeners, it can be concluded that [ɹ] and [ʔ͡t] are quite strongly associated with female and male speakers respectively, in the entire North-East region, not just Tyneside. This, of course, results from the fact that [ɹ] and [ʔ͡t] are found well beyond Tyneside.

The statistical results for the perceived femaleness and maleness of the speaker presented in Tables 3.32 and 3.33 show that [ɹ] was more likely to be evaluated as definitely female-sounding ($p<0.01$) and definitely not male-sounding ($p<0.05$). Both models had a single fixed effect, which was the *variant*.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	83.972	4.352	19.296	<.001	***
Variant [ɹ]	-20.307	6.129	-3.313	<.01	**

Table 3.32 Best-fit linear regression mixed-effects model of perception of speaker femaleness for variants of the word-final pre-vocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=198; Listener=33. North-East group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	13.462	3.742	3.598	<.01	**
Variant [ɹ]	16.002	6.050	2.645	<.05	*

Table 3.33 Best-fit linear regression mixed effects model of perception of speaker maleness for variants of the word-final pre-vocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=198; Listener=33. North-East group.

There were no statistically significant differences between the Tyneside and North-East groups of listeners when either the femaleness or the maleness of the speaker for [ɹ] and [ʔ͡t] were evaluated.

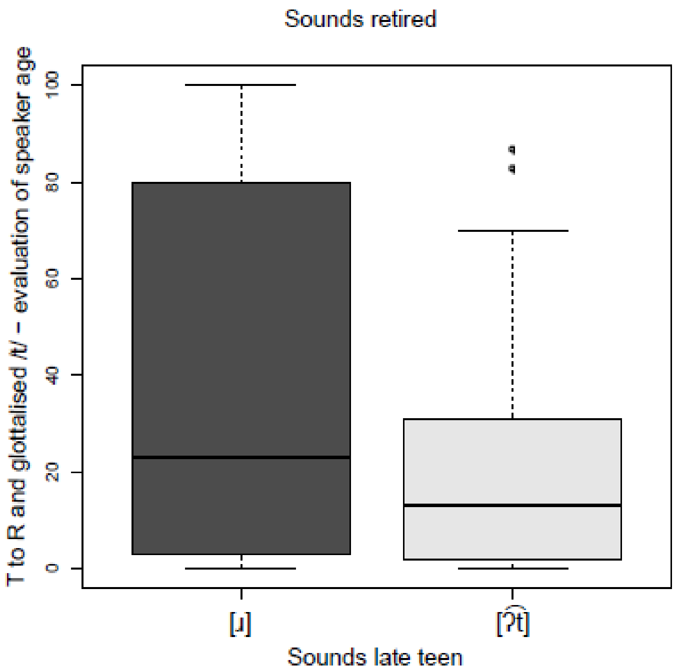


Fig. 3.30 T-to-R [ɹ] and glottalised /t/ [ʔ͡t] variants -- evaluation of speaker age. Tyneside group.

Although a statistically significant difference between perceived speaker age with respect to [ɹ] and [ʔ͡t] was found in the data for the Tyneside group of listeners (p<0.05) (Fig. 3.30 & Table 3.34), it seems that evaluations of speaker age to some degree reflect the findings of the production studies, which report [ɹ] to be used mostly by older speakers and [ʔ͡t] by younger and older speakers (Beal et al., 2012; Docherty et al., 1997; Milroy et al., 1994a; Watt & Milroy, 1999). Nevertheless, a relation between perceived speaker age and gender can be observed in the case of [ʔ͡t] (cf. Section 3.2.1). A similar age-gender interaction has been already reported for one of the GOAT variants, which was found to be somewhat female-sounding and, at the same time, older-sounding (see Figs. 3.18 & 3.19).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	18.548	4.646	3.992	<.01	**
Variant [ɹ]	22.656	7.523	3.012	<.05	*

Table 3.34 Best-fit linear regression mixed-effects model of perception of speaker age for variants of the word-final pre-vocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=186; Listener=31. Tyneside group.

As far as statistical analysis is concerned, the simplest model was the best-fit model to the data. The model included *variant* as a fixed effect. The results show that [ɹ] was significantly older-sounding than [ʔ͡t] (p<0.05) (Table 3.34).

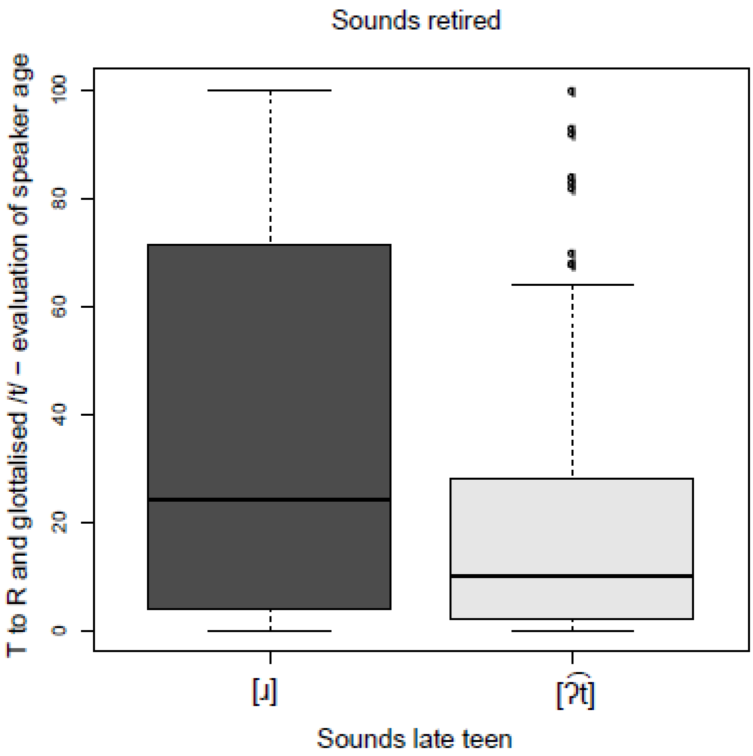


Fig. 3.31 T-to-R [ɹ] and glottalised /t/ [ʔ͡t] variants -- evaluation of speaker age. North-East group.

Evaluations of perceived age present a similar picture to that found for the gender results in terms of the compatibility of the two listener groups (Figs. 3.30 & 3.31). For both groups, the variant most often used by female speakers ([ɹ]) was found to be considerably older-sounding than the variant most often used by male speakers, that is, [ʔ͡t]. The statistical results show, however, that among the North-East group of respondents there was merely a tendency towards evaluating [ɹ] as older-sounding (Table 3.35). These results do not corroborate findings of production studies. In the best-fit model to the data presented in Table 3.35, *variant* and *listener gender* were entered as fixed effects.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	23.130	6.095	3.795	<.01	**
Variant [ɹ]	17.826	8.157	2.185	<.1	
Listener gender (Male)	-9.014	5.590	-1.612	0.114	

Table 3.35 Best-fit linear regression mixed-effects model of perception of speaker age for variants of the word-final pre-vocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=198; Listener=33. North-East group.

It could be concluded from the results presented above that the group of Tyneside listeners was especially sensitive to the age-gender interaction resulting from the design of the experiment (cf. Section 3.2.1). [ɹ], evaluated as considerably more female-sounding than [ʔ͡t], was also found to be older-sounding (Tables 3.30 & 3.34). [ʔ͡t], in contrast, rated as definitely male-sounding, was also found to be much younger-sounding. However, a separate linear regression test demonstrated that there were no statistically significant differences between the two groups of respondents in terms of the perceived age of the speaker.

In terms of the evaluation of speaker social class, a clear difference in the perception of the two variants can be noticed (Fig. 3.32). The [ɹ] variant was found to be definitely working-class sounding. The [ʔ͡t] variant, by contrast, was perceived as less definitely working-class sounding. These results corroborate the findings of the production studies. Furthermore, there was a wider spread of evaluations of the [ʔ͡t] variant, which suggests that listeners varied more in their judgments. In terms of speaker social class, the best-fit model to the data was the simplest model, and included *variant* as a fixed effect (Table 3.36). The results show that [ɹ] was found to be significantly more working-class sounding than [ʔ͡t] ($p<0.001$). Once more, social-indexical information encoded in these variants seems to be quite salient to Tyneside listeners.

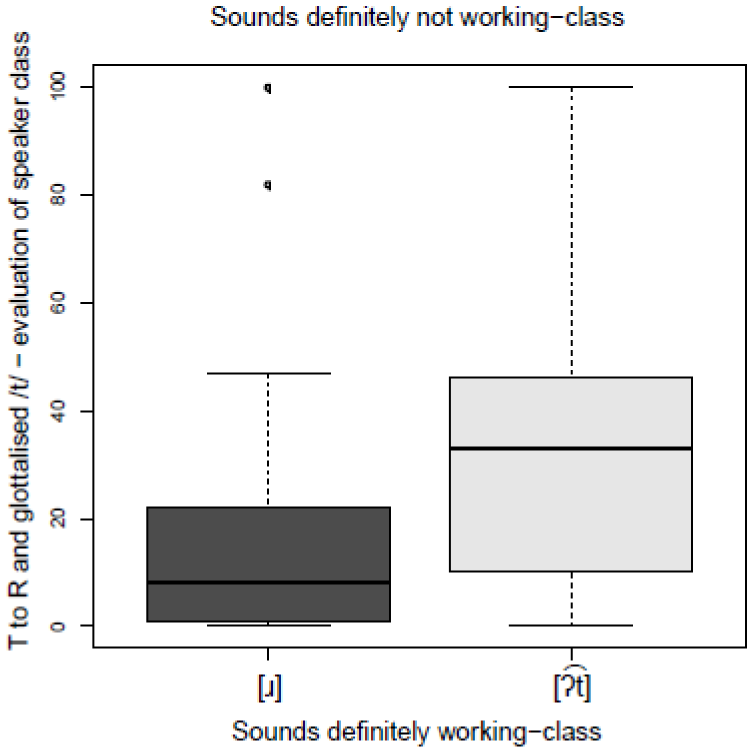


Fig. 3.32 T-to-R [ɹ] and glottalised /t/ [ʔ͡t] variants -- evaluation of speaker class. Tyneside group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	34.140	4.174	8.179	<.001	***
Variant [ɹ]	-21.097	4.509	-4.679	<.001	***

Table 3.36 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of the word-final pre-vocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=186; Listener=31. Tyneside group.

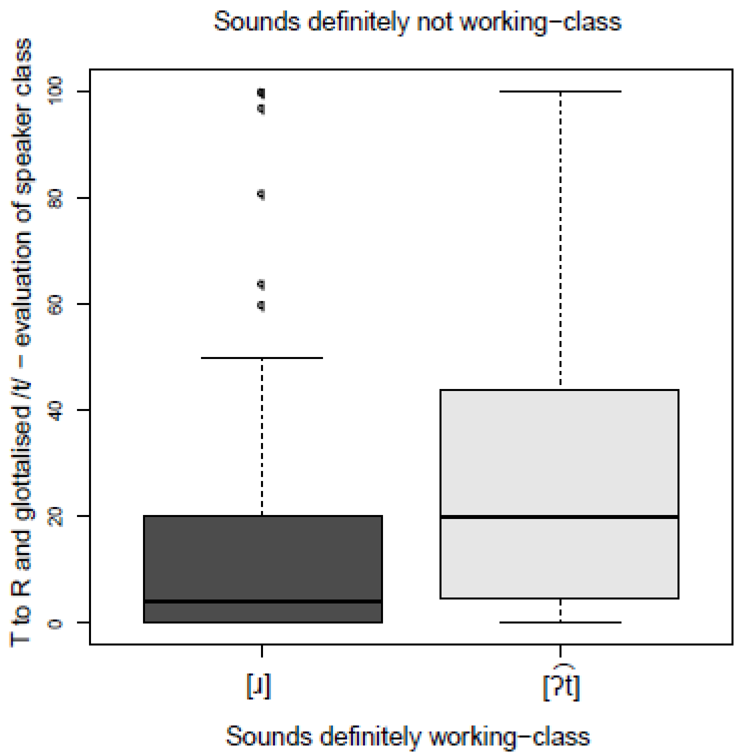


Fig. 3.33 T-to-R [ɹ] and glottalised /t/ [ʔ͡t] variants -- evaluation of speaker class. North-East group.

Evaluations of perceived social class for [ɹ] seem to agree across both groups of listeners. This variant was found to be working class-sounding, which corroborates the findings of production studies.

Even though [ʔ͡t] was also found to be less working-class sounding in comparison to [ɹ] in the North-East group, a difference in the evaluation of this variant across the two groups of listeners can be noticed. North-East listeners evaluated the variant as more working class-sounding than Tyneside listeners (Figs. 3.32 & 3.33). In fact, among the North-East group of respondents, both variants were perceived as slightly more working class-sounding.

Nevertheless, statistically significant results were also reported for the two variants in the North-East group. In fact, the fixed effects as well as the statistical results were identical for both groups of respondents (Table 3.37). Thus, the results indicate a clear difference in the perception of the two variants. In other words, social-indexical information encoded in these variants seems to be quite salient to listeners from the North-East, including Tyneside.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	26.871	3.925	6.847	<.001	***
Variant [r]	-12.700	2.931	-4.333	<.001	***

Table 3.37 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of the word-final pre-vocalic /t/, where [ɹ] and [ʔ͡t] are localised variants. N=198; Listener=33. North-East group.

As might be expected from the results presented above, there were no statistically significant differences between the two groups of respondents in terms of their perceptions of the social class of the speaker.

3.4.2.2 Perceptions of glottalled, glottalised and released /t/ in word-medial, intervocalic contexts

Table 3.38 presents patterns of use of /t/ in word-medial, intervocalic position in Tyneside English.

Variants of intervocalic /t/	Speakers
[ʔ͡t] – glottalised /t/ - a localised variant	most often found in older MC and WC male speech
[ʔ] – glottalled /t/ - a localised variant	most often used by younger MC female speakers
[t] – released /t/ – a non-localised variant	used across all speaker groups

Table 3.38 Realisations of intervocalic /t/ in Tyneside English (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy et al., 1994a; Milroy et al., 1994b).

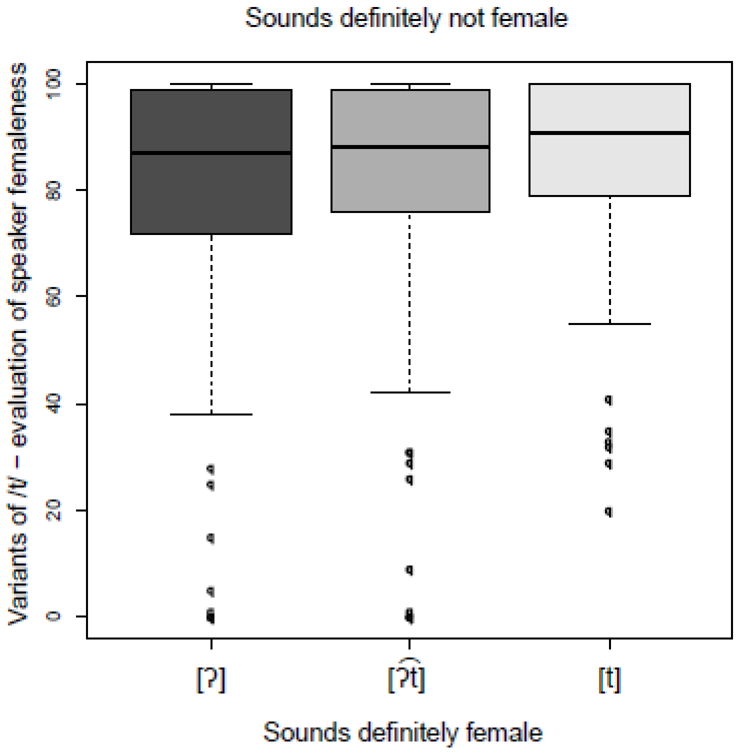


Fig. 3.34 Glottal stop [ʔ], glottalised /t/ [ʔ͡t] and released /t/ [t] variants -- evaluation of speaker gender. Tyneside group.

Figure 3.34 presents evaluations of perceived speaker gender for three realisations of /t/ found in the Tyneside dialect. As can be observed, the spread of ratings for [ʔ͡t] stimuli is very similar in this and the previous group, where [ʔ͡t] was contrasted with [ɹ] (Fig. 3.28). However, as has been mentioned (cf. Section 3.4), [ʔ͡t] stimuli in different phonological contexts were used in opposition to [ɹ] and [ʔ] and [t]. It can be noticed that listeners did not differ in their perceptions of the speaker's gender for the three variants (Fig. 3.34). [ʔ], [ʔ͡t] and [t] were quite consistently evaluated as male-sounding. Furthermore, when taking a closer look at the graphs, it can be seen that even though listeners found these variants to be definitely male-sounding, [t] was found to be the most male-sounding and the most definitely not female-sounding in evaluations of perceived maleness and femaleness respectively. This is interesting since the variant is non-localised and is used by males and females alike. Tables 3.39 and 3.40 reveal no statistically significant results for the variants under investigation.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	75.349	5.357	14.067	<.001	***
Variant [ʔ]	6.849	6.168	1.110	0.304	
Variant [ʔ͡t]	4.398	4.735	0.929	0.387	
Speaker (B)	13.677	4.724	2.895	<.05	*
Listener gender (Male)	-14.738	4.556	-3.235	<.01	**

Table 3.39 Best-fit linear regression mixed-effects model of perception of speaker femaleness for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=279; Listener=31. Tyneside group.

When evaluating femaleness, the best-fit model to the data included *variant*, *speaker* and *listener gender* as fixed effects. As has been mentioned, there was no effect for *variant*. Nevertheless, an effect for *speaker* was reported ($p<0.05$), as well as an effect for *listener gender* ($p<0.01$). Variants were more likely to be evaluated as definitely not female-sounding when produced by speaker (B), but less likely to be found definitely not female-sounding when the gender of the listener was male.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	25.198	6.905	3.649	<.05	*
Variant [ʔ]	-10.854	7.253	-1.497	0.195	
Variant [ʔ͡t]	-8.086	5.854	-1.381	0.222	
Speaker (B)	-12.144	6.041	-2.010	0.100	

Table 3.40 Best-fit linear regression mixed-effects model of perception of speaker maleness for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=279; Listener=31. Tyneside group.

In evaluations of speaker maleness, *variant* and *speaker* were entered as fixed effects (Table 3.43). However, no significant results were reported.

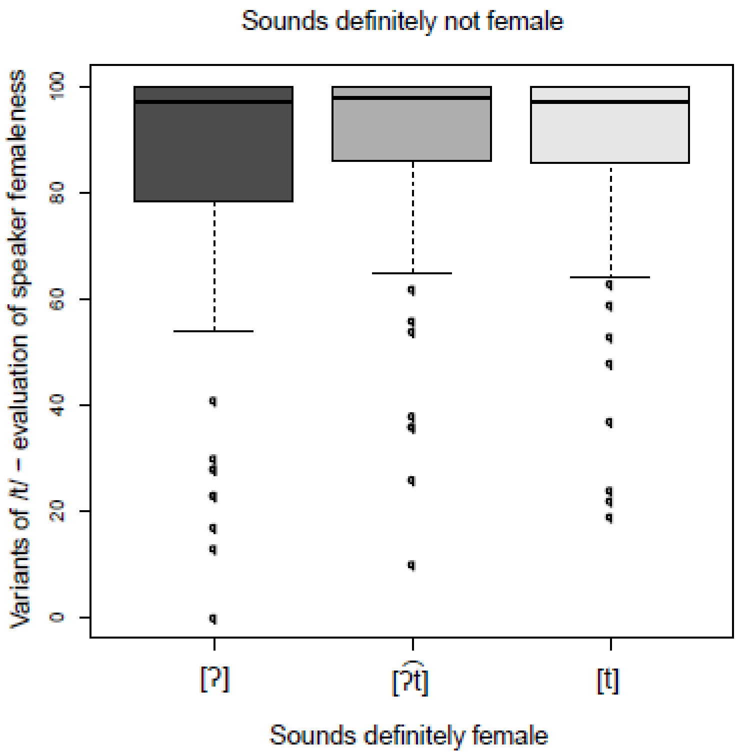


Fig. 3.35 Glottal stop [ʔ], glottalised /t/ [ʔ͡t] and released /t/ [t] variants -- evaluation of speaker gender. North-East group.

As can be seen in Figure 3.35, for the North-East group of respondents all variants were evaluated, similarly to the Tyneside group, as not female-sounding (cf. Fig. 3.34). As with the evaluations of [ɹ] and [ʔ͡t], here also [ʔ͡t], [ʔ] and [t] were found to be overall less female-sounding (in fact almost categorically male-sounding), by the North-East group than by the Tyneside group. Interestingly, these results for [ʔ͡t] and [ʔ] for the North-East group do not reflect the findings of the production studies fully. For example, [ʔ], a variant most often used by females, was found to be definitely male-sounding. Furthermore, the non-localised variant [t] was also perceived as male-sounding, which seems unusual since in the previous cases non-localised variants were evaluated as less definitely male-sounding. This might indicate that listeners heard such variants as being used by speakers of both genders. However, no statistical significance was reported for any of the three variants in the data for the North-East group alone (Tables 3.41 & 3.42). In evaluations of speaker femaleness as well as maleness, the best-fit model included *variant*, *speaker*, and *listener gender*. While none of the phonetic variants had a significant effect, male respondents had a tendency to evaluate them as more female-sounding in evaluations of femaleness (p<0.01) (Table 3.41) and less male sounding in evaluations of maleness (p<0.05) (Table 3.45).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	87.427	4.899	17.847	<.001	***
Variant [ʔ]	1.671	5.231	0.320	0.757	
Variant [ʔ͡t]	4.220	3.917	1.077	0.320	
Speaker (B)	6.280	3.817	1.645	0.160	
Listener gender (Male)	-11.569	3.918	-2.953	<.01	**

Table 3.41 Best-fit linear regression mixed-effects model of perception of speaker femaleness for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=297; Listener=33. North-East group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	22.612	7.382	3.063	<.05	*
Variant [ʔ]	-14.074	7.685	-1.831	0.123	
Variant [ʔ͡t]	-13.040	6.046	-2.157	<.1	
Speaker (B)	-7.644	6.290	-1.215	0.278	
Listener gender (Male)	8.888	3.756	2.366	<.05	*

Table 3.42 Best-fit linear regression mixed-effects model of perception of speaker maleness for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=297; Listener=33. North-East group.

Even though there were no statistically significant differences between the Tyneside and North-East groups of listeners in terms of the perceived maleness of the speaker, some differences were reported for the perceived femaleness of the speaker (Table 3.43). Listeners from the group with higher exposure to Tyneside English were more likely to evaluate the phonetic variants as less male-sounding (p<0.05).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	83.349	4.503	18.509	<.001	***
Variant [ʔ]	4.180	4.591	0.910	0.395	
Variant [ʔ͡t]	4.302	3.532	1.218	0.270	
Speaker (B)	9.805	3.582	2.737	<.05	*
Listener gender (Male)	-8.385	3.285	-2.553	<.05	*
Dialect exposure	-6.317	3.003	-2.104	<.05	*

Table 3.43 Best-fit linear regression mixed-effects model of perception of speaker maleness for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=576; Listener=64. Tyneside and North-East groups.

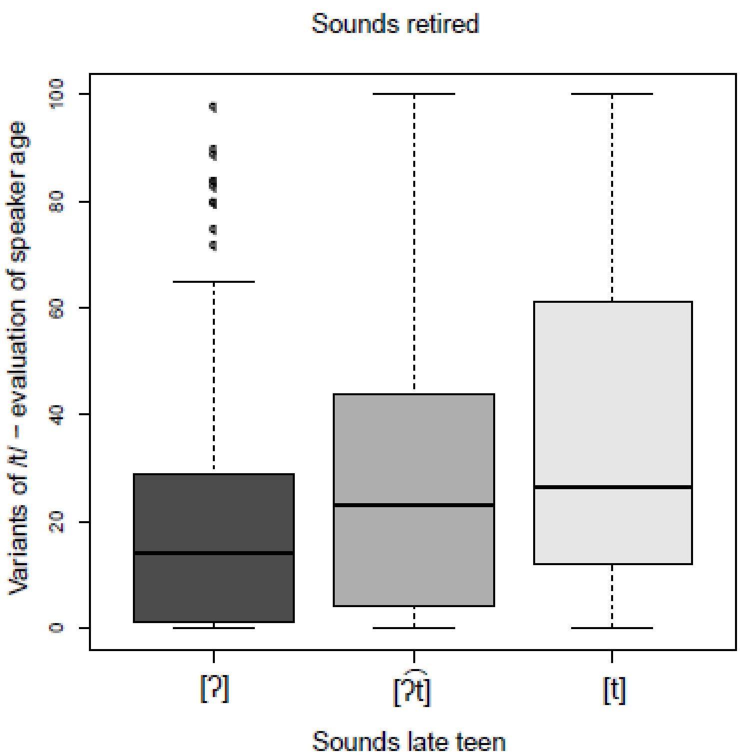


Fig. 3.36 Glottal stop [ʔ], glottalised /t/ [ʔ͡t] and released /t/ [t] variants -- evaluation of speaker age. Tyneside group.

In terms of the perception of speaker age (Fig. 3.36), among the Tyneside group of respondents a difference can be noticed between evaluations of [ʔ͡t] and [ʔ]. As far as the [t] variant is concerned, it was found to be the oldest-sounding of the three variants under investigation. This would mean that listeners were sensitive to information about speaker age encoded by the phonetic variants to some degree. The perceptual differences reflect the pattern of results of the production studies, which have established that [ʔ͡t] is most often used by older and younger speakers, while [ʔ] is used by younger speakers. Interestingly enough, however, from the results presented in the graphs it can be seen that overall, all the variants were found to be young-sounding. This might result from the fact that all three variants were evaluated by listeners to have been produced by a male speaker (cf. Section 3.2.1). As far as statistical results in the evaluations of speaker age are concerned, the best-fit model of age evaluations, which was the simplest model, did not report significant effects for any of the variants (Table 3.44). Instead, a slight tendency towards [ʔ] being rated as produced by a younger speaker was reported.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	36.183	5.148	7.029	<.001	***
Variant [ʔ]	-14.451	6.839	-2.113	<.1	
Variant [ʔ͡t]	-5.989	6.901	-0.868	0.413	

Table 3.44 Best-fit linear regression mixed-effects model of perception of speaker age for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=279; Listener=31. Tyneside group.

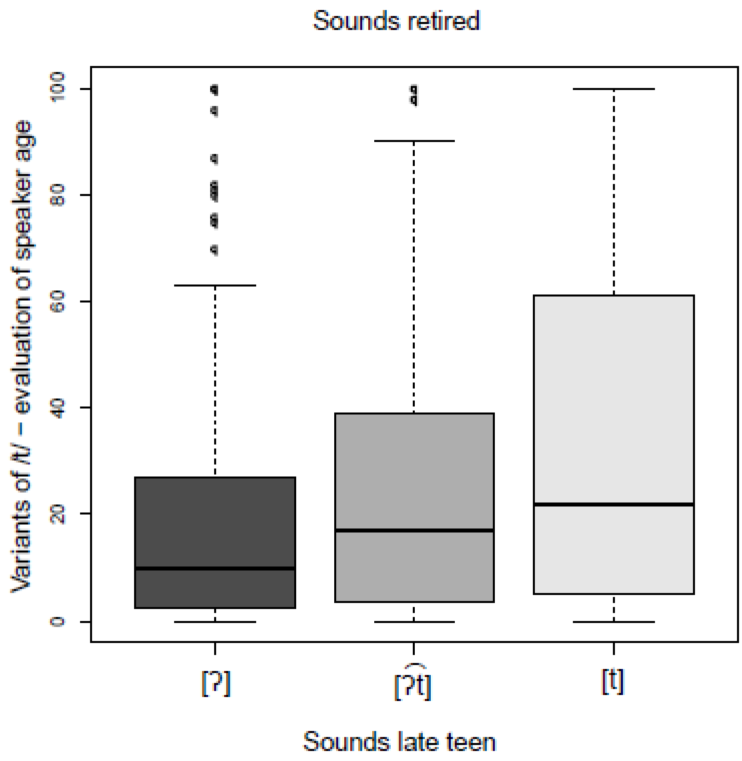


Fig. 3.37 Glottal stop [ʔ], glottalised /t/ [ʔ͡t] and released /t/ [t] variants -- evaluation of speaker age. North-East group.

As can be seen in Figure 3.37, speaker age was rated very similarly by both groups of respondents (cf. Fig. 3.36). As with the Tyneside listeners, the North-East group found the [ʔ] to be older-sounding than [ʔ͡t]. [t], on the other hand, was found to be older-sounding than [ʔ͡t]. However, a slight shift downward in terms of the age evaluations provided by the North-Easteners was observed. In other words, this group found all three variants to sound only slightly younger than did the Tynesiders. This might mean that listeners with lower exposure to the dialect under investigation were more sensitive to the age-gender dependency resulting from the design of this experiment. Once again, it can be seen that the results seem to be reflecting the findings of production studies to some degree. In terms of the statistical results for age rating among the North-East group, the best-fit model reported in Table 3.45 included *variant* as the fixed effect. No significant results for *variant* were found.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	31.666	4.954	6.392	<.001	***
Variant [ʔ]	-10.571	6.424	-1.645	0.131	
Variant [ʔ͡t]	-5.858	6.278	-0.933	0.381	

Table 3.45 Best-fit linear regression mixed-effects model of perception of speaker age for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=297; Listener=33. North-East group.

A separate linear regression test showed that there were no statistically significant differences between the Tyneside and North-East groups of respondents in terms of perceived speaker age for [ʔ], [ʔ͡t] or [t].

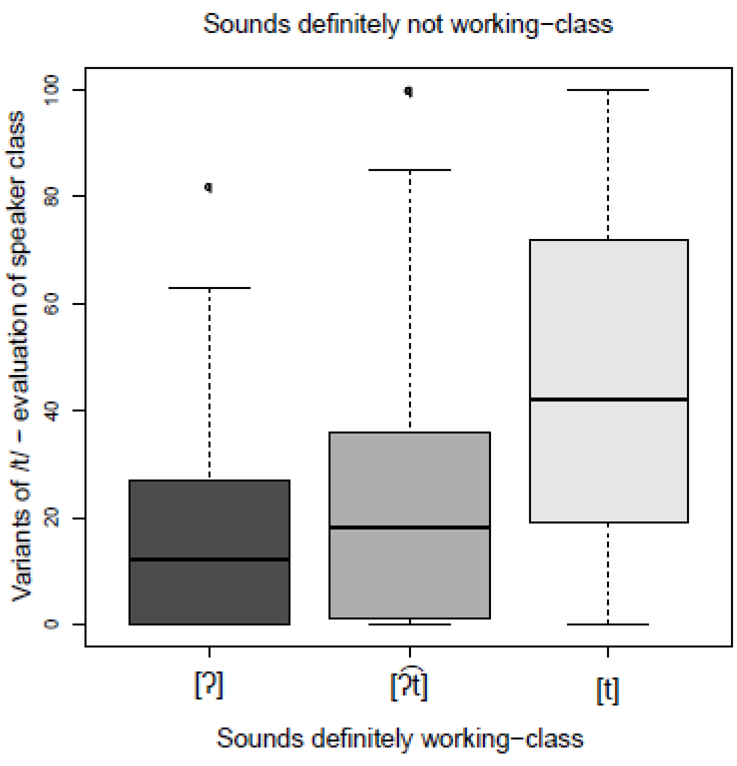


Fig. 3.38 Glottal stop [ʔ], glottalised /t/ [ʔ͡t] and released /t/ [t] variants -- evaluation of speaker class. Tyneside group.

Evaluations of speaker social class for the variants of /t/ by the Tyneside group of respondents are presented in Fig. 3.38. The results presented in the graph do not reflect the results of the production studies, especially when comparing evaluations of the two local variants with the evaluations of the non-local variant. While [ʔ͡t] is used by both working- and middle-class speakers, the listeners evaluated it to be less definitely working-class sounding than [ʔ], which is used by middle-class speakers. The mainstream variant, [t], was found to be perhaps middle-class sounding, with the median evaluation around the mid-point of the scale. This, as well as the fact that the variant received the widest spread of evaluations, might suggest that listeners associated [t] with speakers of all social strata.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	21.153	13.221	1.600	0.160	
Variant [ʔ]	-3.183	14.271	-0.223	0.831	
Variant [ʔ͡t]	-5.376	11.240	-0.478	0.657	
Speaker (B)	27.065	11.390	2.376	<.1	
Listener gender (Male)	-7.869	4.586	-1.716	<.1	

Table 3.46 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=279; Listener=31. Tyneside group.

The best-fit model showing results for the perceived social class of the speaker included *variant*, *speaker* and *listener gender* as fixed effects (Table 3.46). None of the independent variables was significant. Nevertheless, a slight tendency towards reporting the variants as less working-class sounding was observed when the words were produced by speaker B. In addition, there was a tendency to perceive variants as more working-class sounding when the listener was male.

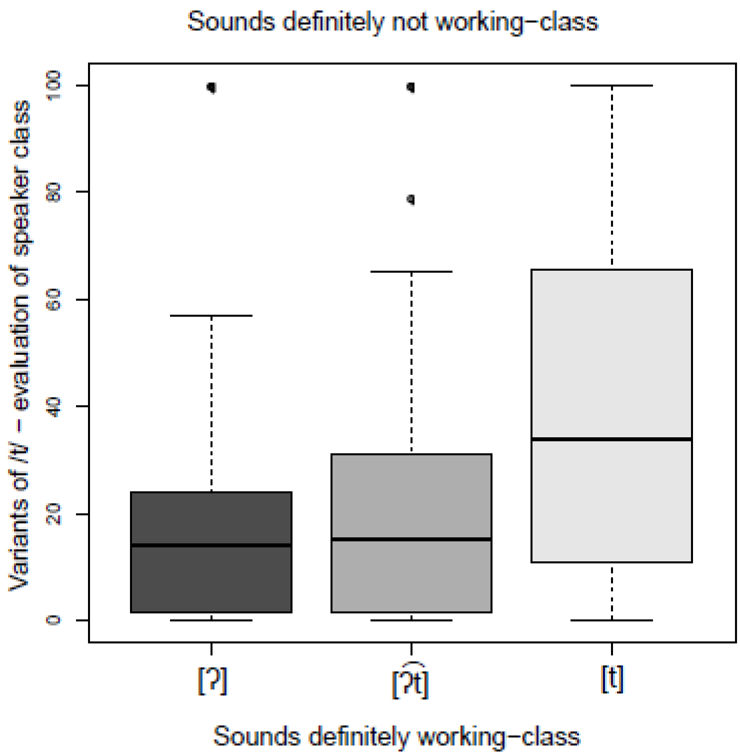


Fig. 3.39 Glottal stop [ʔ], glottalised /t/ [ʔ͡t] and released /t/ [t]variants -- evaluation of speaker class. North-East group.

A closer look at Fig. 3.39 which presents evaluations of speaker social class by the North-East group of respondents, will in fact show no differences between the two groups of listeners. [ʔ] was judged as somewhat more working-class sounding than [ʔ͡t]. The non-local variant, on the other hand, was judged as much less working-class sounding in comparison, which might mean that it was heard to be middle-class sounding. The statistical results presented in Table 3.47 tell us that in fact [ʔ] as well as [ʔ͡t] were more likely to be evaluated as having been uttered by a working-class speaker than was [t] ($p<0.05$). It can be noticed that the North-East group of respondents felt more strongly about the differences between localised variants and the non-localised variant in terms of speaker social class (cf. Table 3.46).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	38.832	5.972	6.502	<.001	***
Variant [ʔ]	-20.757	7.467	-2.780	<.05	*
Variant [ʔ͡t]	-18.485	7.396	-2.499	<.05	*

Table 3.47 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of /t/, where [ʔ] and [ʔ͡t] are localised variants compared against the non-localised released /t/. N=297; Listener=33. North-East group.

To test for any possible differences between the Tyneside and North-East groups of respondents in terms of the perceived social class of the speaker, a separate linear regression test was carried out. However, the test did not report any significant differences for [ʔ], [ʔ͡t] or [t].

3.4.2.3 Perceptions of glottalised and released /p/

Table 3.48 presents patterns of use of /p/ in word-medial intervocalic contexts in Tyneside English.

Variants of word-medial /p/	Speakers
[ʔ͡p] – glottalised /p/ - a localised variant	most often used by male speakers
[p] – released /p/ – a non-localised variant	used across all speaker groups

Table 3.48 Realisations of intervocalic /p/ in Tyneside English (Docherty et al., 1997: 306; Milroy et al., 1994a; Milroy et al., 1994b).

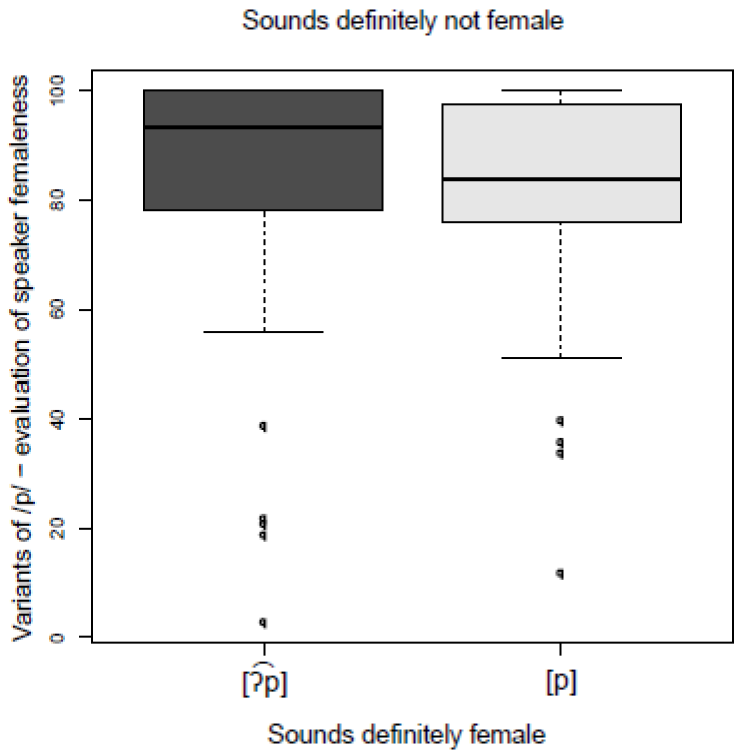


Fig. 3.40 Glottalised /p/ [ʔ͡p] and released /p/ [p] variants -- evaluation of speaker gender. Tyneside group.

As was the case with glottalised and released /t/, glottalised and released /p/ were both evaluated as overall male-sounding (Fig. 3.40). Listeners seemed to be quite consistent when judging the femaleness of the speakers, which is reflected by the narrow spread of evaluations.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	74.031	6.690	11.066	<.01	**
Variant [ʔ͡p]	6.714	6.985	0.961	0.405	
Speaker (B)	12.862	6.932	1.855	0.160	

Table 3.49 Best-fit linear regression mixed-effects model of perception of speaker femaleness for variants of /p/, where [ʔ͡p] is a localised variant compared against the non-localised released /p/. N=186; Listener=31. Tyneside group.

In the best-fit model of evaluations of perceived femaleness of the *speaker*, *variant* and *speaker* were entered as fixed effects (Table 3.49). As can be observed, no significant effects were reported. As far as the perceived maleness of the speaker is concerned, the fixed effects as well as the results were identical to those of the previous model.

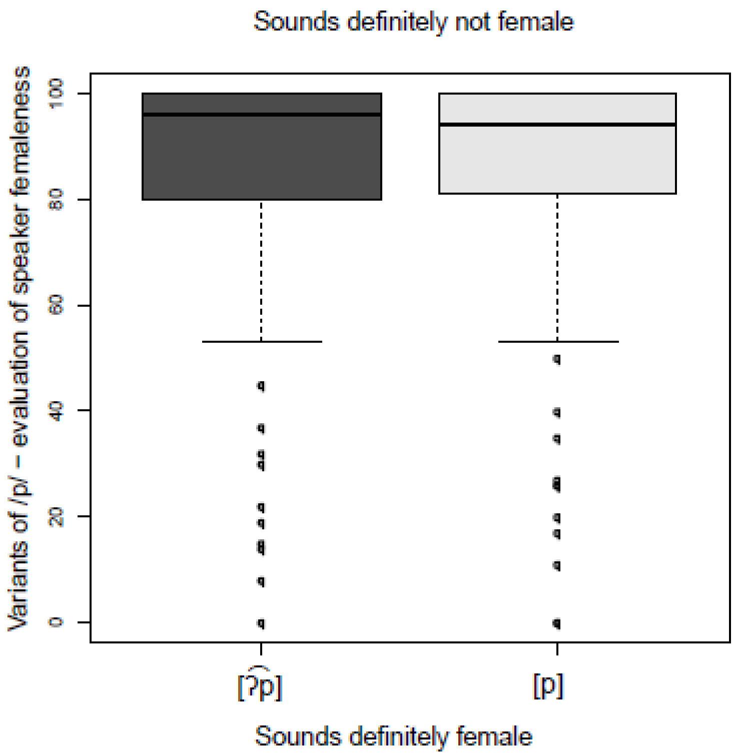


Fig. 3.41 Glottalised /p/ [ʔ͡p] and non-localised released /p/ [p] variants -- evaluation of speaker gender. North-East group.

A comparison of the evaluations of femaleness by the North-East and Tyneside groups (Figs. 3.40 & 3.41) shows that [ʔ͡p] was heard to be definitely not female-sounding, as was released [p]. While overall in both groups the variants were judged as male-sounding, for the North-East group of respondents, both variants were found to be equally male sounding. As expected, the statistical results did not show any significant effect in terms of perception of speaker gender. In the best-fit model of speaker femaleness, *variant*, *speaker* and *listener gender* were entered as fixed effects (Table 3.50). It should be pointed out that a slight tendency towards evaluating the speaker as more female-sounding was observed among male respondents. In the results reporting the perceived maleness of the speaker, no statistically significant results were found.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	77.319	6.078	12.722	<.001	***
Variant [ʔ͡p]	6.237	5.996	1.040	0.374	
Speaker (B)	12.562	5.988	2.082	0.126	
Listener gender (Male)	-7.956	4.324	-1.840		

Table 3.50 Best-fit linear regression mixed-effects model of the perception of speaker femaleness for variants of /p/, where [ʔ͡p] is a localised variant compared against the non-localised released /p/. N=195; Listener=33. North-East group.

As far as perceptions of speaker femaleness and maleness on the basis of [ʔ͡p] and [p] were concerned, no statistically significant differences were found between the Tyneside and North-East groups of listeners.

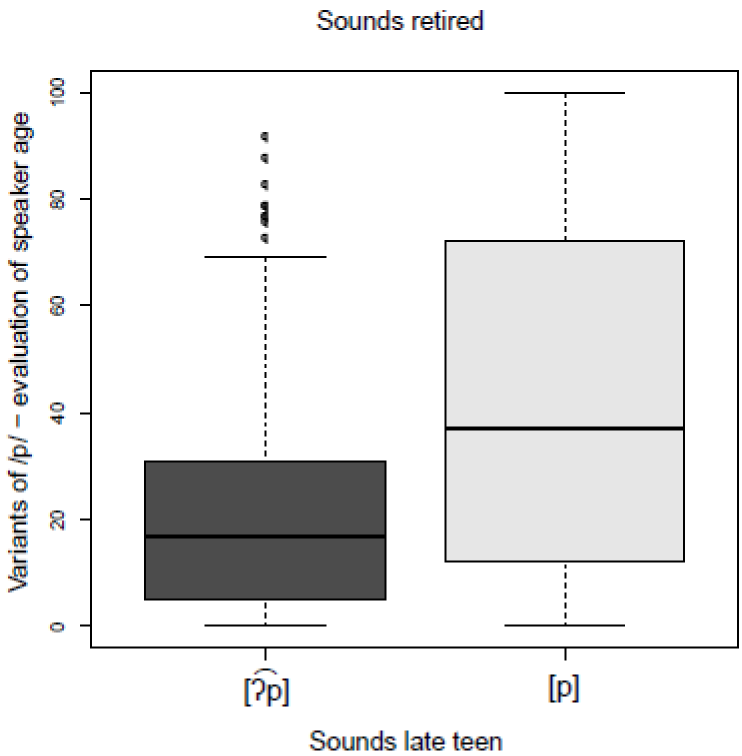


Fig. 3.42 Glottalised /p/ [ʔ͡p] and non-localised released /p/ [p] variants -- evaluation of speaker age. Tyneside group.

The evaluations of speaker perceived age (Fig. 3.42) of the localised variant do not corroborate the results of the production studies. The fact that [ʔ͡p] was found to be young-sounding might mean that, again, voice characteristics influenced listeners' judgements of speaker age. It has been noticed already that variants perceived to be male-sounding were also found to be young-sounding (see for example Figs. 3.34 & 3.36). This interaction, however, is not observed for the non-localised released /p/ variant, which was found to be male- and older-sounding than [ʔ͡p]. Since a wider spread of evaluations close to the mid-point on the scale characterises this variant, it could be argued that it was thought to be used most often by middle-aged speakers. This, on the other hand, reflects the findings of the production studies.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	30.403	6.288	4.835	<.01	**
Variant [ʔ͡p]	-12.315	5.851	-2.105	0.10641	
Speaker (B)	17.702	5.470	3.236	<.05	*

Table 3.51 Best-fit linear regression mixed-effects model of perception of speaker age for variants of /p/, where [ʔ͡p] is a localised variant compared against the non-localised released /p/. N=186; Listener=31. Tyneside group.

In evaluations of speaker age, the best-fit model included *variant* and *speaker* as fixed effects (Table 3.51). No significant difference between the evaluations of the two variants in terms of perceived speaker age was found.

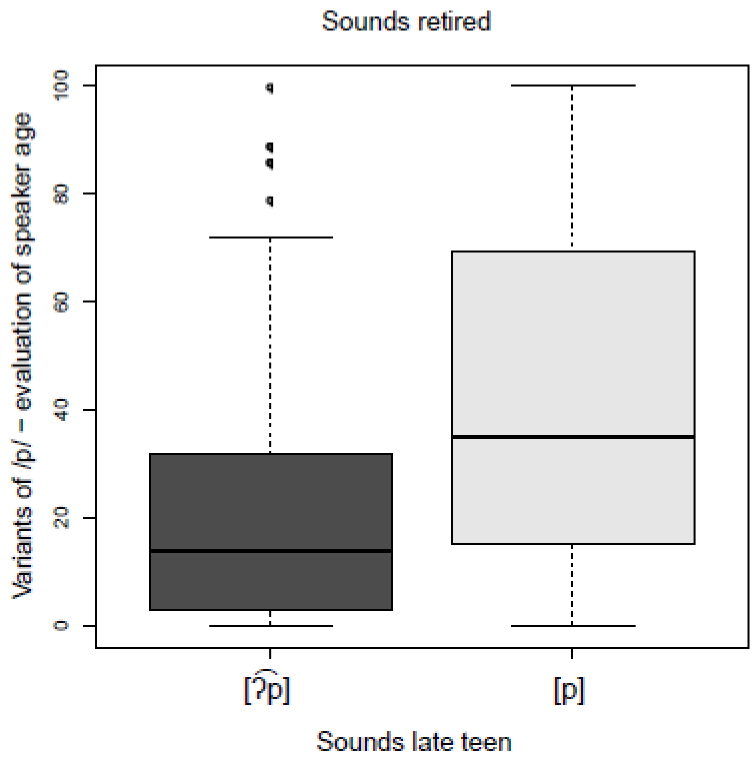


Fig. 3.43 Glottalised /p/ [ʔ͡p] and non-localised released /p/ [p] variants -- evaluation of speaker age. North-East group.

A closer look at Fig. 3.43 shows that in terms of speaker-age evaluations, the results for both groups are almost identical (cf. Fig. 3.42). [ʔ͡p] was rated as considerably younger sounding than [p] by the North-East group than was the case in the Tyneside group. If any age-gender interaction in the North-East group data can be mentioned, this can be done only in the case of [ʔ͡p] but not [p]. In this sense the two groups of respondents show matching results. Statistical results for perceived age of the speaker are reported in Table 3.52, and reveal no significant effects.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	31. 747	6.407	4.955	<.05	*
Variant [ʔ͡p]	-13.076	6.635	-1.971	0.143	
Speaker (B)	13.773	6.633	2.076	0.129	

Table 3.52 Best-fit linear regression mixed-effects model of perception of speaker age for variants of /p/, where [ʔ͡p] is a localised variant compared against the non-localised released /p/. N=198; Listener=33. North-East group.

As might be expected, no significant differences were reported for the Tyneside and North-East groups of listeners as far as the perceived age of the speaker for [ʔ͡p] or [p] was concerned.

The results for evaluation of speaker social class draw our attention to a difference in the perception of the two variants (Fig. 3.44). While the localised variant was found to be rather working class-sounding, ratings for the non-localised variant can be found much higher on the scale, which indicates that it was perceived as much less working class-sounding, perhaps middle-class sounding. The results show that social class is quite a salient parameter to listeners. However, the statistical results presented in Table 3.53 account for random effects in the data and show only a slight tendency towards perceiving [ʔ͡p] as more working class-sounding. In addition, the results show that when the words were produced by speaker B, listeners were more likely to perceive them as having been spoken by a speaker sounding less working-class ($p<0.05$). In the best-fit model to the data presented in Table 3.53, variant and speaker were entered as fixed effects.

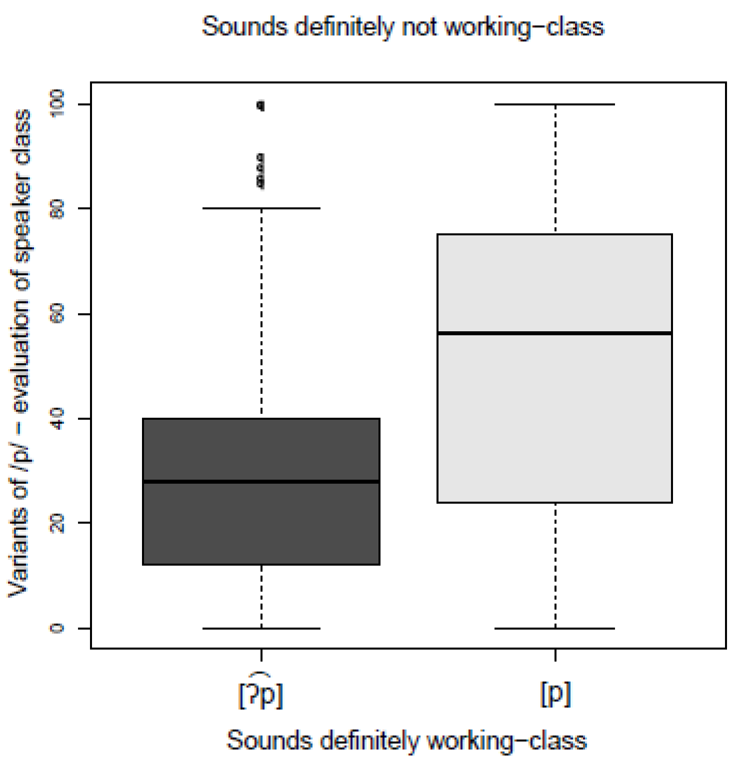


Fig. 3.44 Glottalised /p/ [ʔ͡p] and non-localised released /p/ [p] variants -- evaluation of speaker class. Tyneside group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	38.935	5.895	6.605	<.001	***
Variant [ʔ͡p]	-14.081	5.121	-2.749	<.1	
Speaker (B)	17.016	4.837	3.518	<.05	*

Table 3.53 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of /p/, where [ʔ͡p] is a localised variant compared against the non-localised released /p/. N=186; Listener=31. Tyneside group.

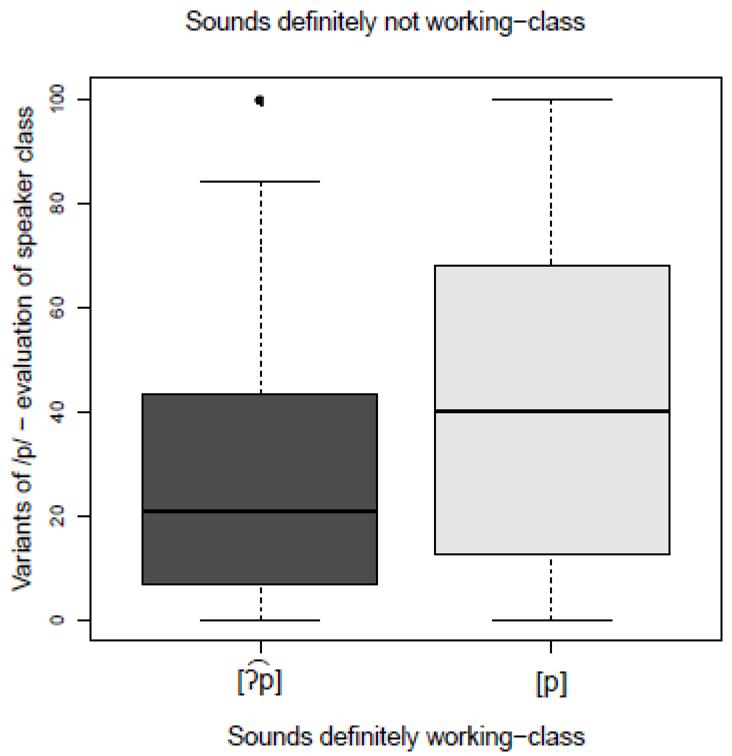


Fig. 3.45 Glottalised /p/ [ʔ͡p] and non-localised released /p/ [p] variants -- evaluation of speaker class. North-East group.

Finally, evaluations of speaker social class show that respondents in the two groups felt similarly about the localised and non-localised variants (Figs. 3.44 & 3.45). However, both [ʔ͡p] and [p] were rated as more working-class sounding by the North-East group of respondents. Table 3.54 reports no significant results in terms of perception of speaker social class in the North-East group of respondents.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	34.828	6.017	5.788	<.001	***
Variant [ʔ͡p]	-8.455	5.337	-1.584	0.205	
Speaker (B)	8.939	5.229	1.709	0.185	

Table 3.54 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of /p/, where [ʔ͡p] is a localised variant compared against the non-localised released /p/. N=198; Listener=33. North-East group.

In terms of perceived speaker social class, no statistically significant differences between the Tyneside and North-East listener groups were found for [ʔ͡p] and [p].

3.4.2.4 Perceptions of glottalised and released /k/

Table 3.55 presents patterns of use of /k/ variants in intervocalic contexts in Tyneside English.

Variants of word-medial /k/	Speakers
[ʔ͡k] – glottalised /k/, a localised variant	most often used by male speakers
[k] – released /k/, a non-localised variant	used across all speaker groups

Table 3.55 Realisations of intervocalic /k/ in Tyneside English (Beal et al., 2012; Docherty et al., 1997: 306; Milroy et al., 1994a; Milroy et al., 1994b).

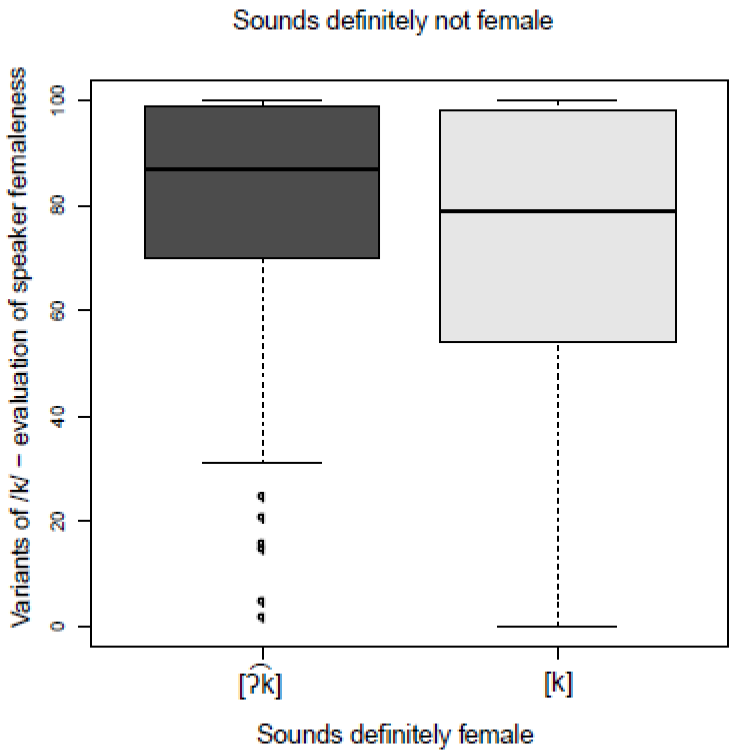


Fig. 3.46 Glottalised /k/ [ʔ͡k] and released /k/ [k] variants -- evaluation of speaker gender. Tyneside group.

Tyneside listeners rated the [ʔ͡k] and [k] differently in terms of the perceived gender of the speakers. The difference can be noticed especially when examining perceptions of femaleness (Fig. 3.46). While [ʔ͡k] was thought to sound definitely male, [k] was found to sound overall male; however, listeners perceived [k] as a less definitely male feature. It should be pointed out that evaluations of speaker gender of the glottalised and released /p/ and /k/ variants are in fact identical (cf. Fig. 3.40). Nevertheless, the statistical results presented in Table 3.56 show no significant effect of the variants in terms of the perceived femaleness of the speaker. The best-fit model was the simplest one, with *variant* as a fixed effect. The same model was used to test the perceived maleness of the speaker. The results were almost identical, with only a slight tendency towards evaluating [ʔ͡k] as less male-sounding.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	69.586	11.912	5.842	<.01	**
Variant [ʔ͡k]	9.919	16.566	0.599	0.581	

Table 3.56 Best-fit linear regression mixed-effects model of perception of speaker femaleness for variants of /k/, where [ʔ͡k] is a localised variant compared against the non-localised released /k/. N=183; Listener=31. Tyneside group.

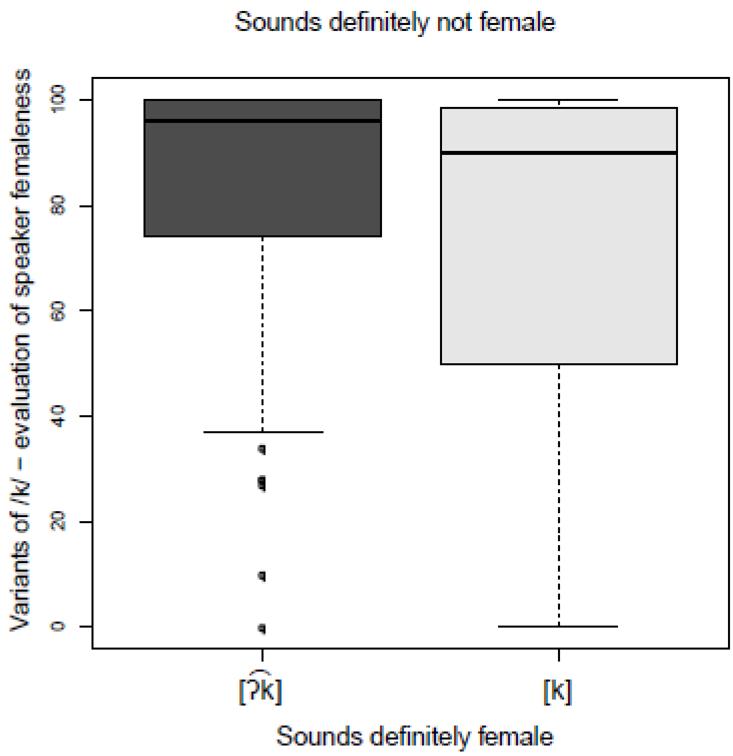


Fig. 3.47 Glottalised /k/ [ʔ͡k] and released /k/ [k] variants -- evaluation of speaker gender. North-East group.

As we saw in the results reported for glottalised and released /p/, also in the case of glottalised and released /k/ both groups of listeners evaluated [ʔ͡k] and [k] alike (Figs. 3.46 & 3.47). The glottalised variant was found to be less female sounding than the released variant (Fig. 3.46). As the same time, there was also a slight difference across the groups in their perceptions of the two variants, which were perceived as slightly less female-sounding by the North-East group than the Tyneside group. No statistical results were reported in terms of perceived speaker femaleness (Table 3.57) or maleness.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	69.586	11.912	5.842	<.01	**
Variant [ʔ͡k]	9.919	16.566	0.599	0.581	

Table 3.57 Best-fit linear regression mixed-effects model of perception of speaker femaleness for variants of /k/, where [ʔ͡k] is a localised variant compared against the non-localised released /k/. N=186; Listener=33. North-East group.

With respect to the perceived femaleness or maleness of [ʔ͡k] and [k] there were no statistically significant differences between the Tyneside and North-East groups of listeners.

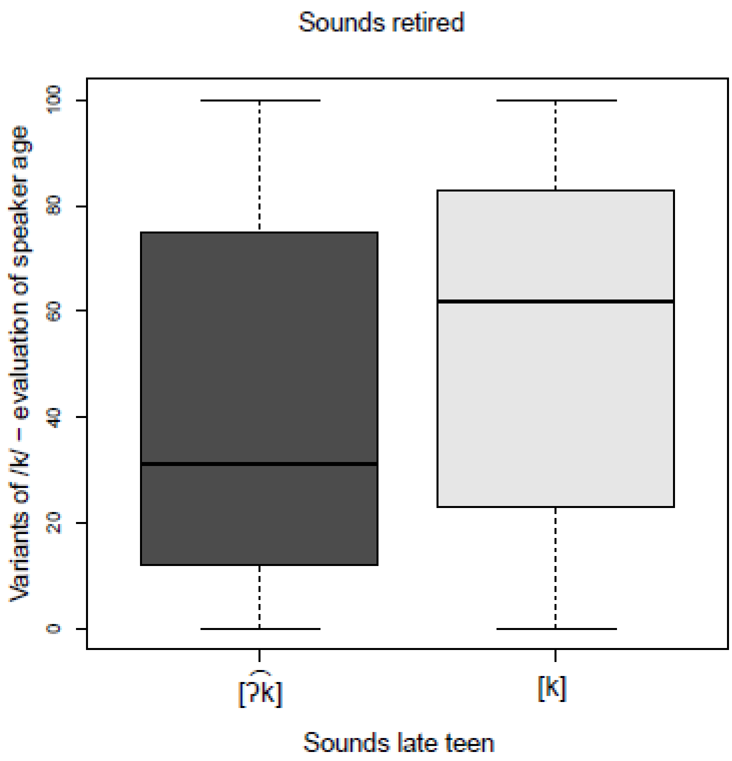


Fig. 3.48 Glottalised /k/ [ʔ͡k] and released /k/ [k] variants -- evaluation of speaker age. Tyneside group.

In terms of speaker age, [ʔ͡k] and [k] were found to be perceived as different-sounding by the Tyneside group of listeners (Fig. 3.48). [ʔ͡k] was evaluated as younger-sounding than [k]. Both variants are characterised by wide spreads of evaluation, [ʔ͡k] especially in the third quartile, and [k] in the second quartile. The results corroborate the findings of the production studies only to some extent. However, the statistical results presented in Table 3.58 show no significant effects. The best-fit model was the simplest model, with variant as the only fixed effect.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	54.806	9.984	5.489	<.01	**
Variant [ʔ͡k]	-14.140	13.759	-1.028	0.362	

Table 3.58 Best-fit linear regression mixed-effects model of perception of speaker age for variants of /k/, where [ʔ͡k] is a localised variant compared against the non-localised released /k/. N=183; Listener=31. Tyneside group.

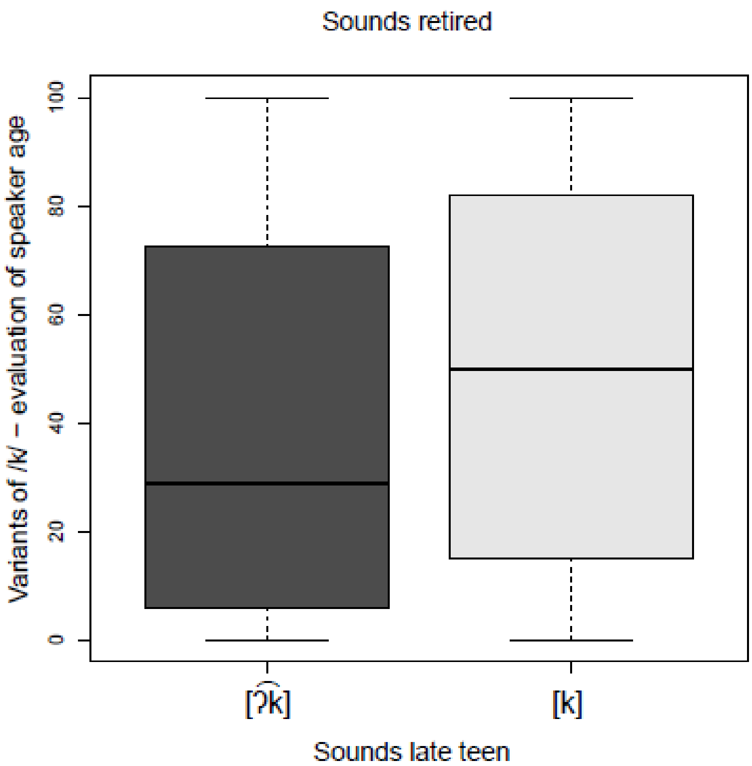


Fig. 3.49 Glottalised /k/ [ʔ͡k] and released /k/ [k] variants -- evaluation of speaker age. North-East group.

Evaluations of speaker age presented in Fig. 3.49 reveal that the two groups of respondents rated [ʔ͡k] virtually identically as younger-sounding, whereas the released variant was found to be older-sounding (cf. Fig. 3.48). Furthermore, a downward shift in evaluations by the North-East group can be noticed, especially in the case of the non-localised variant. In both groups the same statistical model was applied as the best-fit model to the data. Furthermore, perceptions of speaker age based on hearing the two variants varied statistically in neither of the listener groups.

When looking at Fig. 3.50, it seems that once again, listeners were quite sensitive to social class information encoded in the phonetic variants. While [ʔ͡k] was found to be working class-sounding, [k], was by comparison evaluated as much less working class-sounding. However, in this case differences between evaluations of speaker social-class were again not statistically significant (Table 3.59). While *variant*, *speaker* and *listener gender* were entered as fixed effects in the best-fit model, none of the dependent variables had a significant effect. Nevertheless, is should be noted that there was a slight tendency towards evaluating variants as more working-class sounding when the respondents were male.

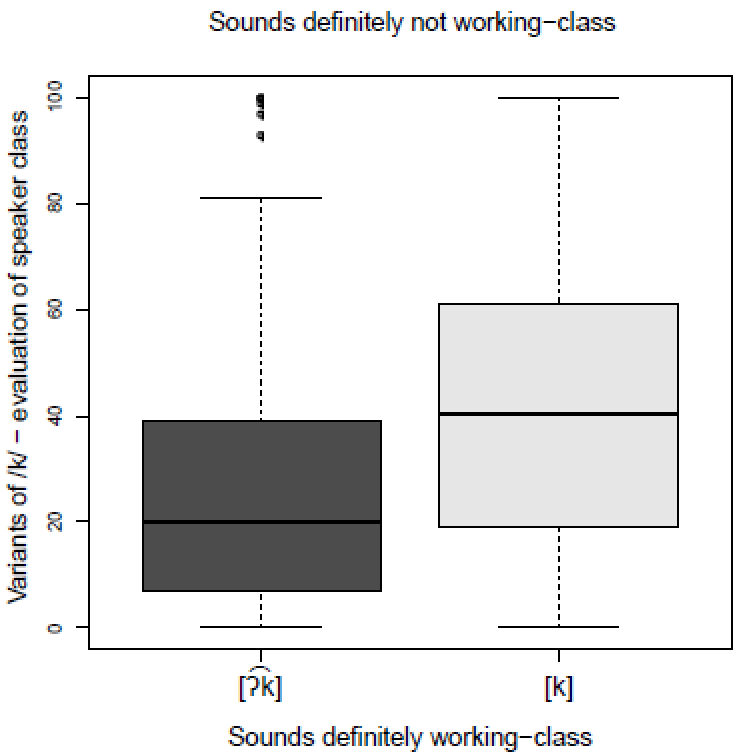


Fig. 3.50 Glottalised /k/ [ʔ͡k] and released /k/ [k] variants -- evaluation of speaker class. Tyneside group.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	15.284	19.348	0.790	0.509	
Variant [ʔ͡k]	0.657	13.563	0.048	0.965	
Speaker (B)	21.951	16.565	1.325	0.316	
Listener gender (Male)	-11.028	6.256	-1.763	<1	

Table 3.59 Best-fit linear regression mixed-effects model of perception of speaker social class for variants of /k/, where [ʔ͡k] is a localised variant compared against the non-localised released /k/. N=183; Listener=31. Tyneside group.

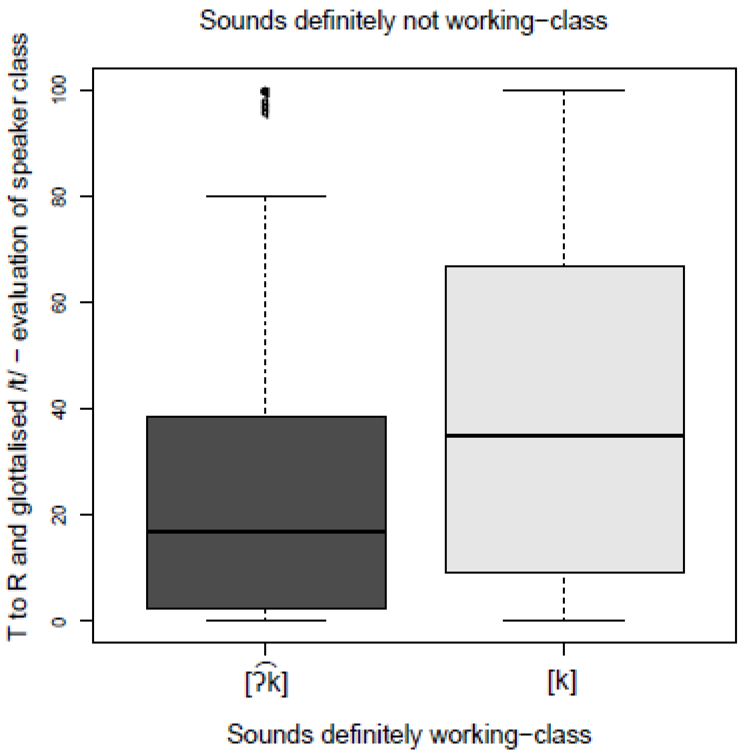


Fig. 3.51 Glottalised /k/ [ʔ͡k] and released /k/ [k] variants -- evaluation of speaker class. North-East group.

Figure 3.51 presents evaluations of perceived speaker social class. Again, as in previous cases, it seems that neither group of respondents differed in their evaluations of [ʔ͡k] and [k]. While the localised variant was rated as more working-class-sounding, the non-localised variant was found to be less working-class-sounding. However, again some group-based variability in terms of evaluations by the two groups has been found. Both variants were rated as slightly more working-class-sounding in the North-East group than by the Tyneside group. However, this difference was non-significant. Further, a linear regression test comparing the results for both groups did not yield significant results.

3.4.2.5 Perceptions of pre-aspirated and released /t/

Table 3.60 presents patterns of use of word-final post-vocalic /t/ in Tyneside English.

Variants of word-final post-vocalic /t/	Speakers
[^{h}t] – pre-aspirated /t/, a localised variant	characteristic of younger female speakers
[t] – released /t/, a non-localised variant	used across all speaker groups

Table 3.60 Realisations of word-final post-vocalic /t/ in Tyneside English (Beal et al., 2012; Docherty & Foulkes, 1999; Foulkes et al., 2005).

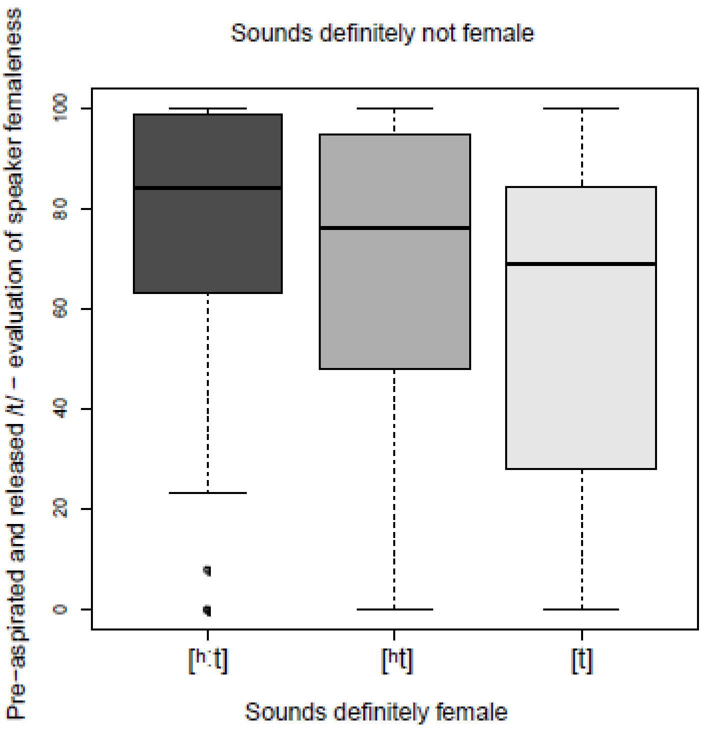


Fig. 3.52 Variants of pre-aspirated /t/, [$^{h:}$t] and [^{h}t] and non-localised released /t/ [t] -- evaluation of speaker gender. Tyneside group.

Two variants of pre-aspirated /t/, one of longer and more breathy pre-aspiration and the other of shorter and less breathy pre-aspiration, were contrasted with released /t/, which is a realisation found in Tyneside and other parts of the country. The reason for investigating two pre-aspirated variants of /t/ was to test how responsive the listeners are to pre-aspiration, which has been reported to be a characteristic of female speakers (Beal et al., 2012; Docherty & Foulkes, 1999; Foulkes et al., 2005). A closer examination of pre-aspirated and released realisations of /t/ reveals that (surprisingly) Tyneside respondents evaluated the most pre-aspirated variant as the least female-sounding, whereas they rated the non-localised variant as the most female-sounding of the three (Fig. 3.52). However, listeners did perceive the localised and non-localised ([t]) variants as different, as the non-localised variant was evaluated as the most female-sounding. Evaluations of the perceived maleness of the stimuli under investigation revealed that the three variants were found to be overall-male sounding. However, it should be pointed out that the [^{h}t] was more likely to be evaluated as less definitely male-sounding than the [^{hː}t].

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	60.08	7.64	7.864	<.001	***
Variant [^{hː}t]	13.42	11.55	1.162	0.297	
Variant [^{h}t]	7.74	11.51	0.672	0.531	

Table 3.61 Best-fit linear regression mixed-effects model of perception of speaker femaleness for pre-aspirated and released /t/, where [^{hː}t] and [^{h}t] are localised variants compared against the non-localised [t]. N=248; Listener=31. Tyneside group.

Regarding evaluations of speaker femaleness, the best-fit model to the data was the simplest model, with *variant* as a fixed effect (Table 3.61). Consistent with what might be expected by visual inspection of the graph, we cannot observe significant differences in the table. The same model was implemented to test speaker maleness, and identical results were reported.

Interestingly, listeners did not seem to be sensitive to any putative speaker gender information encoded in the pre-aspirated variants, which were reported in the production studies to be typically female. One possible explanation for this pattern is that the linguistic change has progressed and that word-final pre-aspirated /t/ is used by men and women alike.

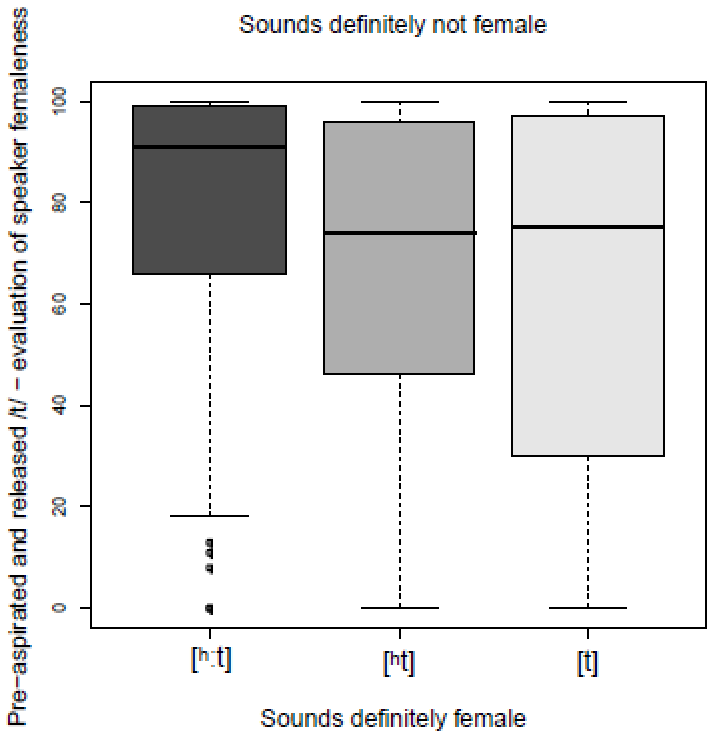


Fig. 3.53 Variants of pre-aspirated /t/, [^{hː}t] and [^{h}t] and non-localised released /t/ [t] – evaluation of speaker gender. North-East group.

Evaluations of speaker femaleness show similar results for both groups of respondents (Figs. 3.52 & 3.53). In the North-East group also [^{hː}t] was found to be less female-sounding in comparison to [^{h}t]. However, the released variant was rated as the most female-sounding of the three variants. Nevertheless, some differences between listener groups can be observed. For example, the general tendency among North-East listeners to evaluate variants as more definitely not female-sounding can also be observed in this case. Furthermore, it seems that the North-East listeners varied more in their perceptions of the non-localised and shorter pre-aspiration variants with respect to social class. In the best-fit model, *variant*, *speaker* and *listener gender* were entered as fixed effects (Table 3.62). However, no significant results in terms of perceived speaker femaleness or maleness were reported.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	73.648	10.668	6.903	<.01	**
Variant [hːt]	15.609	20.538	0.760	0.490	
Variant [ht]	-3.547	16.266	-0.218	0.838	
Speaker (B)	-16.554	25.155	-0.658	0.547	
Listener gender (Male)	-9.457	5.437	-1.740	<.1	

Table 3.62 Best-fit linear regression mixed-effects model of perception of speaker femaleness for pre-aspirated and released /t/, where [hːt] and [ht] are localised variants compared against the non-localised [t]. N=263; Listener=33. North-East group.

As far as the perceived femaleness and maleness of the speaker were concerned, there were no statistically significant results for the pre-aspirated and released /t/ variants in the data for the two groups of listeners.

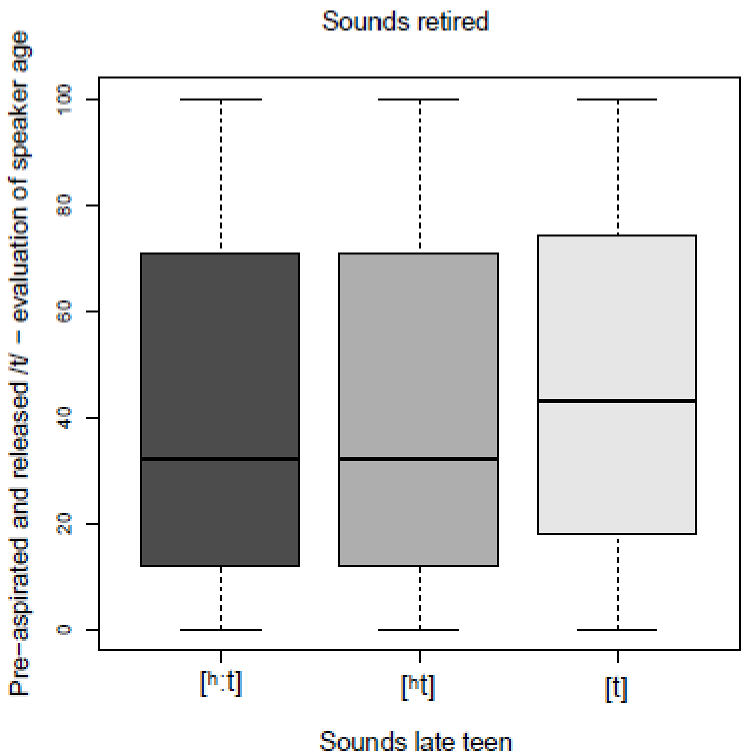


Fig. 3.54 Variants of pre-aspirated /t/, [hːt] and [ht] and non-localised released /t/ [t] -- evaluation of speaker age. Tyneside group.

As far as age ratings are concerned, the fully released /t/ was evaluated as somewhat older sounding in comparison to the pre-aspirated variants (Fig. 3.54). [ʰːt] and [ʰt] are characterised by a spread of evaluations in the third quartile. Overall, they were found to sound similar to the non-local variant with respect to perceived speaker age.

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	52.988	5.046	10.501	<.001	***
Variant [ʰːt]	-12.294	10.031	-1.226	0.300	
Variant [ʰt]	-12.633	10.327	-1.223	0.275	

Table 3.63 Best-fit linear regression mixed-effects model of perception of speaker age for pre-aspirated and released /t/, where [ʰːt] and [ʰt] are localised variants compared against the non-localised [t]. N=248; Listener=31. Tyneside group.

Regarding speaker age evaluations, the best-fit model to the data included variant as the only fixed effect (Table 3.63). As can be observed, no significant effects were found.

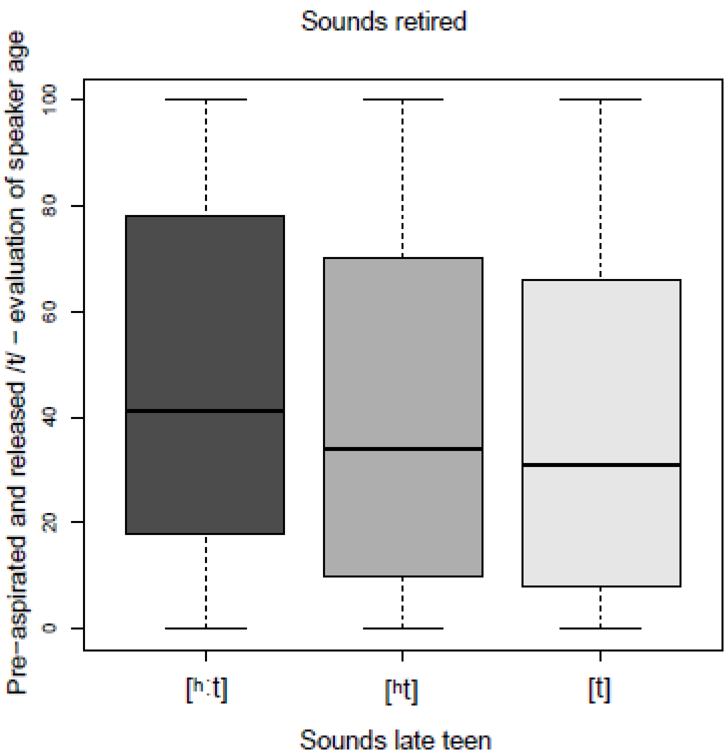


Fig. 3.55 Variants of pre-aspirated /t/, [ʰːt] and [ʰt] and non-localised released /t/ [t] -- evaluation of speaker age. North-East group.

Evaluations of speaker age present an overall picture that is similar to the examples discussed previously. Both groups of listeners rated the variants quite similarly (Figs. 3.54 & 3.55). Yet in comparison with the Tyneside group, an upward shift in the ratings of [^{hː}t], can be observed in the North-East group's data. A downward shift, on the other hand, can be seen for the ratings of the non-localised variant. No statistically significant differences in the evaluation of pre-aspirated and released /t/ by the North-East group were reported (Table 3.64).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intrcept)	34.747	4.787	7.259	<.01	**
Variant [^{hː}t]	-2.475	7.990	-0.310	0.782	
Variant [^{hː}t]	6.009	6.962	0.863	0.445	
Speaker (B)	31.183	9.648	3.232	<.1	

Table 3.64 Best-fit linear regression mixed-effects model of perception of speaker age for pre-aspirated and released /t/, where [^{hː}t] and [^{h}t] are localised variants compared against the non-localised [t]. N=263; Listener=33. North-East group.

In terms of the perceived age of the speaker there were no statistically significant differences among the pre-aspirated and released variants of /t/ in the data for the Tyneside and North-East groups of listeners.

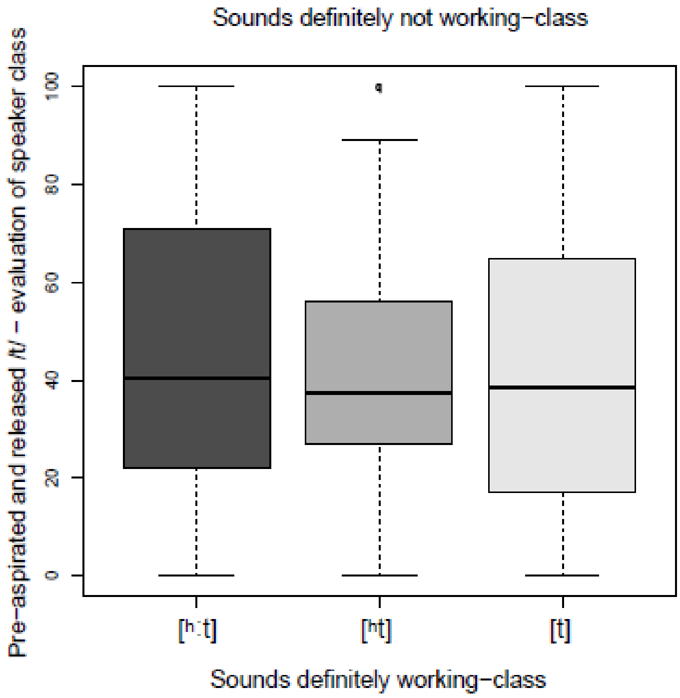


Fig. 3.56 Variants of pre-aspirated /t/, [hːt] and [ht] and non-localised released /t/ [t] -- evaluation of class. Tyneside group.

It will be noticed that variants of pre-aspirated and fully released /t/ were evaluated as rather middle class-sounding (Fig. 3.56). None of the variants was perceived as definitely working class-sounding. These findings are confirmed by the statistical results reported in Table 3.65, in which no significant differences between evaluations of the three variants in terms of social class were found. It can be observed, however, that there was a tendency towards evaluating the speaker as more working class-sounding when the respondents were male. This tendency has been reported before, and it seems that it might result from the fact that the sample was unbalanced in terms of listener gender. The best-fit model to the data included *variant* and *listener gender* as fixed effects (Table 3.65).

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	57.928	5.618	10.310	<.001	***
Variant [ʰːt]	-7.002	14.129	-0.496	0.632	
Variant [ʰt]	-13.083	14.129	-0.926	0.378	
Listener gender (Male)	-19.528	7.298	-2.676	<.05	*

Table 3.65 Best-fit linear regression mixed-effects model of perception of speaker social class for pre-aspirated and released /t/, where [ʰːt] and [ʰt] are localised variants compared against the non-localised [t]. N=248; Listener=31. Tyneside group.

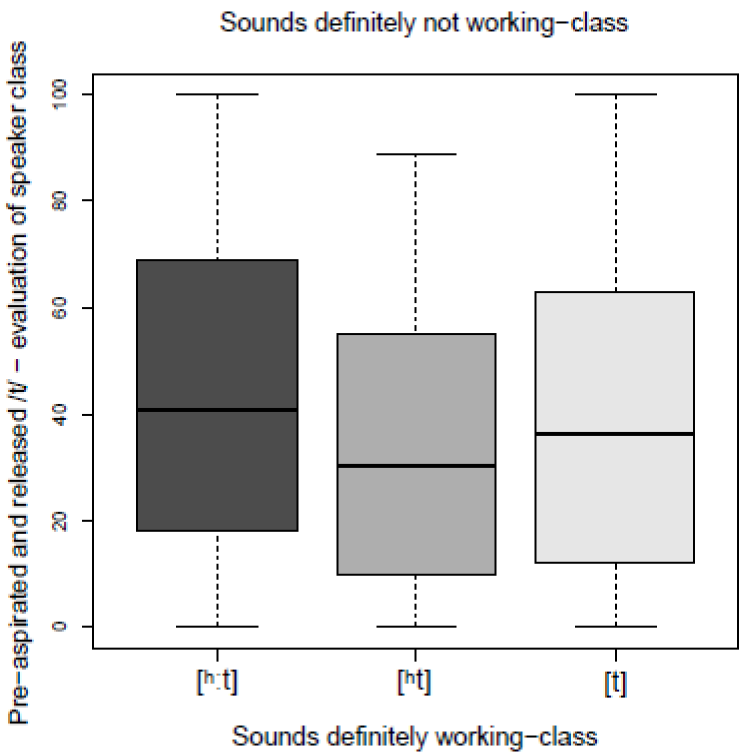


Fig. 3.57 Variants of pre-aspirated /t/, [ʰːt] and [ʰt] and non-localised released /t/ [t] -- evaluation of class. North-East group.

Finally, the results for speaker social class ratings among the North-East group of respondents are presented in Figure 3.57. Yet again, the two groups of respondents overall did not vary in their evaluations. Both the localised and the non-localised variants were found to be less working class-sounding. However, as reported in the previous cases, there was again a tendency here to evaluate the variants as slightly more working class-sounding, in comparison to the ratings given by the Tyneside listeners. The statistical results were non-significant. In addition, there were no significant differences between the Tyneside and North-East groups of respondents in terms of the perceived social class of the speaker.

3.5 Summary and discussion

As far as the results for vowel and consonantal variants are concerned, except for the NURSE fronted variant [øː] and T-to-R variable no other differences in terms of perceived speaker gender were found. Just as most of the vowel variants were reported in production studies to be most often used by male speakers, they were also evaluated by the listeners as having been produced by a male speaker. Therefore, it could be the case that for the listeners these variants carried social-indexical information about speaker gender and indexed male speech. Nevertheless, it should be pointed out that Tyneside listeners evaluated the fronted NURSE variant [øː] as less male sounding than the retracted variant [ɔː] (Table 3.21). This finding indicates that for Tyneside listeners this particular variant was in fact associated less with a male speaker.

The results for consonantal variants show [ɹ] as a clear example of a variant indexing female speech. In fact, the variant was evaluated as having been produced by a female speaker by both groups of respondents. This suggests that the association of the variant with female speech is common in all of the North-East of England and not only in Tyneside. It seems that, except for this variant, other consonantal variants indexed male gender to the listener. Even though the production studies reported some variants to be most often used by female speakers, for example pre-aspirated /t/, and others to be most often used by male speakers, for example glottalised /t/, the former were not associated with female speech by the listeners when compared with non-localised or other localised variants. Thus, it could be concluded that indexicality worked to a limited extent as far as perceptions of speaker gender were concerned.

Another explanation of the results found for vowels and consonants is the fact that localised variants are often associated with a male speaker (Labov, 1990). Thus, unless a variant has some strong indexical association with female speech, such as [ɹ] or even the fronted NURSE variant, it may be associated with a male speaker otherwise. Therefore, it could be that listeners were guided to a certain extent by this general association of localised variants with male speech when evaluating speaker gender.

Except for the fronted NURSE variant [øː] and the T-to-R variable, no statistically significant differences in terms of speaker age evaluations were reported. In fact speaker age was evaluated at chance level in the middle of the scale, and this indicates that listeners were uncertain as to how to evaluate speaker age. This would suggest that indexicality did not work when speaker age was evaluated. Therefore, it could be concluded that, overall, the variants under investigation were not associated by the listeners with a speaker of a particular age. However, it should be pointed out that [øː] was evaluated by the comparison group of respondents as younger sounding than the retracted [ɔː] and the non-localised [ɜː] variants. This might suggest that [øː] may be associated with younger speakers outside of Tyneside. Also, [ɹ] was evaluated as older sounding in comparison to [ʔ͡t]. Therefore, it is possible that [ɹ] was associated with an older female speaker, thus corroborating findings of the production studies and showing that the variant carries social-indexical information. Nevertheless, perhaps the impact of age-gender interaction mentioned earlier should not be entirely dismissed in the case of T-to-R.

Vowel analysis showed that, in terms of social class, localised variants were perceived by both groups of listeners as working-class sounding in comparison to the non-localised variants. It may be realistic to interpret these results in the light of the standard language ideology (SLI). One of the outcomes of SLI is that localised variants are often stigmatised. This would explain why localised variants investigated in this experiment were evaluated as having been produced by a working-class speaker. It seems that, irrespective of their exposure to the localised and non-localised variants, listeners were guided by the SLI when asked to evaluate the social class of the speaker. It may be that, upon hearing a localised variant, they evaluated it as less standard sounding than a non-localised variant. Nevertheless, these findings could be also accounted for to some extent by indexicality. To listeners from Tyneside and wider North-East localised variants may have been associated for the most part with working-class speakers and as a result they were evaluated by the listeners as such. Thus to the listeners localised variants carried social information about the speaker's social class. To be specific, the variants indexed 'working-classness' of the speaker. Another possible interpretation of the results could be indirectly through perceptual dialectology. The listeners could have been guided by perceptual dialectology and associated the variants with the North-East region. As a result of this regional association they further evaluated the localised variants as having been produced by a working-class speaker. Whilst some discrepancy between production and perception data can be observed in this case, it should be pointed out that a 100 per cent match between production and perception should not be expected. Production and perception studies are complementary but do not produce entirely matching results. Furthermore, it should be also pointed out that the production studies referenced in this thesis are not recent, perhaps adding to the divergence between results of the present experiment and results of the production studies resulting from changes in the patterns of usage which may have occurred.

The results for consonantal variants showed that only [ɹ] was strongly perceived by both groups of respondents as having been produced by a working-class speaker in comparison to [ʔ͡t]. This might suggest that this variant carries indexical information about the social class of the speaker in all of the North-East, including Tyneside. On the other hand, the fact that [ʔ] and [ʔ͡t] were evaluated by the North-East group as having been produced by a working-class speaker in comparison to [t] and no such perceptual differences were found for Tyneside listeners might mean that the two localised variants are encoded with different indexical information about speaker social class within and outside of Tyneside. Whilst [ʔ] and [ʔ͡t] seem to be salient markers of speaker working-class for North-East listeners, they are not for Tyneside listeners. This is a surprising result which needs further investigation.

As far as other consonantal variants under investigation are concerned, no statistically significant differences in terms of evaluations of social class were found. Thus, these variants do not seem to be salient markers of speaker social class for the respondents. Furthermore, the differences between evaluations of vowel and consonantal variants in terms of perceived social class of the speaker suggest that vowels tend to carry more indexical information about the speaker than consonants.

Milroy (2004: 170) argues that SLI may trigger development of ideologies supporting the use of non-standard variants. These non-standard varieties are in use despite the standard language culture we live in. This state of affairs is possible when or because speakers identify with the localised varieties. Being that indexicality is closely linked with speaker identity, it could be argued that listeners are able to recognise speaker indexical information if they belong to and identify with the group and the place. This could be seen especially in more subtle differences between the two groups of respondents in speaker class evaluations.

Since perceptual differences between the two groups of respondents are few and far between rather than having a patterned regularity, they cannot be accounted for within the framework of exemplar theory which explains speech perception through exposure. It was expected originally that due to high exposure to the dialect the Tyneside listeners would evaluate speaker-indexical information with more consistency. By contrast, listeners from other parts of the region were expected to provide consistently different evaluations of speaker-indexical information from the same phonetic variants. Yet because there were few differences between the two groups of respondents, this theoretical framework had to be abandoned. However, these differences could be explained by enregisterment. Variants which were evaluated differently in terms of speaker gender, age or social class by the Tyneside and North-East respondents are enregistered in Tyneside and the wider North-East with different meaning in terms of social information they carry about the speaker. In other words, the process of enregisterment as a process by which linguistic forms acquire non-linguistic meaning went slightly differently in Tyneside and wider North-East. What is more, also production studies investigating sociolinguistic patterns outside of Tyneside reported differences in these patterns when compared to patterns in Tyneside (Burbano-Elizondo, 2008; Llamas, 2001; 2007). If sociolinguistic patterns in other parts of the region were different, it is only natural to expect the enregisterment process to result in different outcomes, even if these are minor.

3.6 Experiment 1 Conclusion

The main aim of this experiment was to help us understand whether listeners are able to recognise social-indexical information encoded in phonetic variants when produced in a gender-ambiguous voice. The findings show that overall listeners evaluated the majority of variants as male sounding, although the spreads of evaluations visible in the graphs indicate that some of the variants were perceived as female sounding. Evaluations of speaker age show that, for the most part, listeners were not certain about the speaker's age and evaluated it in the middle of the scale. The results revealed also an age-gender interaction resulting from the experimental design (cf. Section 3.2.1), whereby speakers who were found to sound male were also evaluated as young-sounding. Finally, evaluations of perceived speaker social class showed differences between localised and non-localised variants. The results indicated that the localised variants were overall evaluated as having been produced by a working-class speaker. It seems that after removing the cue of gender-specific fundamental frequency the perceived social class of the speaker became the most salient social-indexical feature. From the evaluations of speaker social class, it is clear that listeners seemed to be quite sensitive to this type of indexical information.

Listeners with high previous exposure to the dialect under investigation were sensitive to speaker-indexical information at the segmental level. However, certain localised variants seemed to encode social-indexical information to a more consistent degree than others. While some of the results presented constitute very clear-cut cases, for example T-to-R, others reflect findings of the production studies only to a limited extent, for example the FACE vowel or pre-aspirated /t/. Furthermore, few differences between the Tyneside group and the North-East group of respondents were reported. With some exceptions, listeners with lower exposure to Tyneside English evaluated most of the social-indexical information carried by the variants in a similar manner to listeners who will have had high exposure to the dialect.

In comparison to the Tyneside group, the North-East group of respondents was more categorical in their evaluations of variants as male-sounding, as well as working-class sounding.

From the results of this experiment new questions have emerged. Previous research had shown that indexical information can be shifted by visual cues the listener is exposed to (Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999). Given that in Experiment 1 information about social class was retrieved with such consistency, could perceptions of speaker social class be manipulated by visual cues with information about the supposed speaker? Would different types of indexical information be affected by visual cues to the same degree? Experiment 2 described in Chapter 4 attempts to address these questions.

4 Experiment 2

4.1 Introduction

The aim of the experiment was to further investigate and clarify the results of Experiment 1. In Experiment 1 listeners were asked to evaluate speaker gender, age and social class on the basis of hearing a single phonetic variant produced by a gender-ambiguous sounding voice. As was shown in the previous chapter, listeners were particularly consistent when evaluating information about speaker social class, whereas evaluations of perceived speaker gender and age were not as clear. From these results a question emerged whether indexical information could be shifted by visual cues about the apparent speaker. If so, would visual cues shift perceptions of all types of indexical information, for example, social class and age? As will be recalled, in Experiment 1 listeners were not provided with any information about the speaker when evaluating the acoustic signal in terms of speaker social-indexical information, whereas in Experiment 2 listeners are primed with visual information about the supposed speaker. It was investigated how perceptions of the same speaker-indexical information would be affected when listeners were provided with some visuals with information about the supposed speaker. Therefore, in Experiment 2, visual cues to speaker gender, age and social class were added. This information was given to the listener gradually, in order to see whether each type of speaker-indexical information influenced perceptions of phonetic variants to any extent. It was anticipated that adding visual cues would help differentiate perceptions of speaker age and social class from those of speaker gender (cf. Experiment 1). Visual cues also attempted to shift listener's perceptions of speaker-indexical information. For example, seeing a young male face when hearing a phonetic variant might incline the listener to rate the variant differently to the answers given in Experiment 1. Thus, the hypothesis in Experiment 2 was that priming listeners with gendered faces should influence evaluations of speaker social class and speaker age. The hypothesis was supported by previous research which has confirmed that visual cues from the speaker's face provide additional information about the speaker and shift listener's perception of auditory information (Hay et al., 2006; Johnson et al., 1999). Mani and Schneider (2013) have shown that the speaker's face is a source of indexical information about the speaker. This information, in turn, is used by listeners during the speech perception process.

As expected, Experiment 1 also showed a recurring interdependence between perceived speaker age and gender. Some voices which were found to be male-sounding were also evaluated as young-sounding. Voices judged as more female sounding, by contrast, were often considered to be older-sounding (cf. Section 3.2.1). These interactions will be further investigated in this experiment.

Following the introduction, the experiment and types of visual images used in it are described. Next, the choice of statistical methods used to analyse results are described. Finally, the results are presented and discussed.

4.2 Method

4.2.1 Stimuli

The same stimuli as in Experiment 1 were used in the present experiment (see Section 3.2.1 for a full description). However, this experiment further investigates perceptions of the variants which gave clear results in Experiment 1, that is, variants of the NURSE vowel, T-to-R compared with glottalised /t/, and glottalled, glottalised and released /t/. The retracted NURSE variant [ɔː], most often used by older working-class males, was contrasted with the fronted variant [øː], which is most often used by young middle- and working-class females, and with an non-localised centralised [ɜː] (Beal et al., 2012; Watt, 1998; Watt & Milroy, 1999). T-to-R ([ɹ]), which is most often used by older working-class female speakers, is contrasted with glottalised /t/ ([ʔ͡t]) (cf. Section 3.4). Glottalised /t/, on the other hand, is most often found among older middle- and working-class males in Newcastle English (Beal et al., 2012: 39; Watt & Milroy, 1999). Glottalised /t/ ([ʔ͡t]) was contrasted with a glottalled /t/ ([ʔ]), which is most often found among younger middle-class female speakers (Beal et al., 2012; Milroy et al., 1994a; Milroy et al., 1994b; Watt & Milroy, 1999), and released /t/ ([t]), a non-localised variant which is used locally but also in a number of other varieties of British English. The GOAT vowel variants were removed because all of the localised variants under investigation were most often used by male speakers only, unlike in the case of the NURSE vowel variants.

4.2.2 Participants

Listeners were recruited in the same manner as for the previous experiment. Because the same audio stimuli were used in the present as well as in Experiment 1, it was ensured that participants who took part in the previous study did not participate in the present study. Thus participants in both tested groups, that is, from Tyneside and from the south of the country were new to the research.

Listeners were undergraduate and graduate students at the University of York and Newcastle University. Participants were paid £8 upon completion of the experiment in the first stage of data collection and £7 in the second stage.

As in Experiment 1, the Tyneside group of listeners fulfilled the condition of having high exposure to the dialect under investigation. The Tyneside group of respondents consisted of 27 participants, 15 of whom were from Newcastle. Other North-East localities included Gateshead and North and South Tyneside.

All participants reported having immediate family and/or friends from the North-East. With the exception of one participant who was 34 years old, all other participants were between 18 and 24 years of age. The majority reported having lived in Tyneside all their lives, that is, between 18 and 23 years.

In terms of social background, 11 participants reported being from middle-class households and 16 described themselves as being working-class.

As in the first experiment, the aim was to obtain a gender-balanced sample of participants, but yet again this proved to be difficult in practice. Thus, the sample consists of 7 males and 20 females. In general, it was mostly women who responded to calls for participation in experiments.

As a comparison group, it was decided to use listeners who had low exposure to any of the North-East dialects. The results in Experiment 1 showed that there were little differences between Tyneside listeners and listeners from the wider North-East in terms of perception of speaker-indexical information carried by the phonetic variants. This familiarity with the Tyneside variants could be the result of people's mobility, perhaps also TV and/or radio as well as other factors. It was also assumed that due to the choice of phonetic variants under investigation in the second study (those of the NURSE vowel, T-to-R, and glottalised /t/), there may be not very much difference in the results between a group of Tyneside listeners and listeners from elsewhere in the North-East. Therefore, it was decided to test listeners from the South of England in Experiment 2 to ensure that the comparison group was comprised of native speakers of British English with low exposure to the variety spoken in the North-East England.

The comparison group of participants consisted of 31 listeners who originated from the south of England. 8 listeners were from London or neighbouring boroughs. The majority of participants originated form Greater London, Sussex and Kent, but also Oxfordshire, Buckinghamshire, Berkshire and Surrey. Although participants came from a greater number of places in geographical terms than the Tyneside group of respondents, they were all from the south of the country, thus fulfilling the fundamental requirement of having lower exposure to Tyneside English. The majority of participants were up to two months into the first year of their undergraduate degrees and living away from home for the first time. 13 participants reported having a flatmate/s, a parent or further relatives as their close contacts from the North-East. However, this information might be somewhat unreliable because when asked if they had ever lived in the North-East some participants confirmed that they considered York to be a North-Eastern town.

All participants in this group were between 18 and 24 years of age, and so age-wise they were similar to the Tyneside group of respondents. This group of respondents also was not gender-balanced, as it consisted of 19 males and 12 females. Interestingly, in this group of listeners it was mostly men who responded to the call for participation. Only one person reported being working-class and the remaining 31 participants reported being middle-class.

4.2.3 Social class images

Different sets of images were used in this experiment to evaluate perceived speaker social class and to evaluate perceived speaker age.

In the first two blocks in the evaluation of the social class of the supposed speaker, participants were instructed to pick one of the two pictures. In block 1 the pictures showed examples of two types of housing, associated with working- and middle-class people. In block 2, information about the gender of the supposed speaker was added and a gendered face (male or female) was projected against a background showing the two types of housing. In block 3, where information about age of the supposed speaker was added to the information about social class and gender, younger or older gendered face was projected against a background showing the two types of housing and listeners were instructed to pick one of the four pictures.

The first set of images (block 1) presents social class images of housing that participants were expected to associate with working-class and middle-class people, respectively.

Fig. 4.1 Block 1. Images showing backgrounds associated with middle- and working-class people. The images that participants saw were displayed in randomised order.

The second set of images (block 2) presents a male or female face of the same age against a backdrop showing housing that participants were expected to associate with working-class and middle-class people, respectively.

Fig. 4.2 Block 2. A female face against backgrounds associated with working- and middle-class people. The images that participants saw were displayed in randomised order.

Fig. 4.3 Block 2. A male face against backgrounds associated with working- and middle-class people. The images that participants saw were displayed in randomised order.

The third set of images (block 3) shows younger and older male and female faces. As before, the faces are presented against housing associated with working- and middle-class people. While the young faces appear to be those of people in their early twenties, the older faces are middle-aged looking.

Fig. 4.4 Block 3. A younger and older female face against backgrounds associated with working- and middle-class people. The images that participants saw were displayed in randomised order.

Fig. 4.5 Block 3. A younger and older male face against backgrounds associated with working- and middle-class people. The images that participants saw were displayed in randomised order.

4.2.4 Age images

In the evaluation of the age of the supposed speaker, participants were instructed to pick one of the three pictures in the male and female conditions. The first face showed a man or woman in his/her early 20s, the second a man or a woman in their 40s, and the third a man or a woman in their 60s. It was ensured that in all of the sets the faces showed the same age range. In Experiments 2 and 3 male and female faces will be referred to as the “supposed speaker”.

Figures 4.6 - 4.9 present face composites used for evaluations of speaker age.

Fig. 4.6 A set of female faces for speaker age rating.

Fig. 4.7 A set of female faces for speaker age rating.

Fig. 4.8 A set of male faces for speaker age rating.

Fig. 4.9 A set of male faces for speaker age rating.

The faces shown in this and the previous sections (4.2.3) were created in EvoFIT (Frowd et al., 2006; Frowd, 2015). Even though EvoFIT was originally used in forensics, it can in fact, have a number of other applications. For example, it can be used in sociolinguistic, psycholinguistic or psychological research. EvoFIT can be of great assistance in a broad array of perceptual experiments where, for example, the speaker's ethnic background, age, masculinity/femininity, social class and other indexical information needs to be assessed from speech. One of the main advantages of the software is that it allows the user to obtain a highly controlled end product. Composite faces created with EvoFIT can be of different shapes and sizes but also different ethnicities. Speaker characteristics such as age, masculinity, attractiveness, honesty, health or texture can be manipulated to a required degree (Frowd et al., 2006). Furthermore, facial features such as lips or jaws can be made more or less prominent (Frowd et al., 2013).

At the same time, the same face can be "aged" to obtain an age continuum from a late teen to a 60 year-old. EvoFIT allows the user to adjust the look of eyebrows, "eyebags", nose and nostrils, as well as jawline to reflect speaker age (Frowd et al., 2006: 44). Aging a composite face can be fully controlled and achieved also by additional manipulation of face health.

Thus, EvoFIT was used to generate the faces used in both types of images in the present experiment (that is, the age and social class images). While the age images used male and female face composites morphed in EvoFIT (Frowd et al., 2006; Frowd, 2015), the class images integrated morphed faces with other visual cues. Black and white composites were photoshopped (Adobe Photoshop CS6, 2013) into black and white backgrounds cueing working- and middle-class responses. Similar visual scales were applied by Hay et al. (2006) and Squires (2013).

4.2.5 Pre-testing social class images

To validate the social-class images used in this experiment and to ensure that the complete pictures were in fact associated with middle- and working-class people, images were rated by an independent group of volunteers as follows.

Three short surveys were administered online using the SurveyGizmo software (SurveyGizmo, 2014). Images were rated in terms of social class by three groups of respondents. Each successive survey permitted the narrowing down of types of images suitable for the study and the selection of images yielding the best results.

The three surveys were completed by 20, 14 and 10 respondents from Tyneside and the North-East respectively. The images were not tested on any other group of respondents originating from a different region in the country. Tables 4.1 and 4.2 provide information about participants' age and gender.

Number of respondents	Age of respondents
8	18-24
21	25-35
5	34-45
1	45-54

Table 4.1 Number of respondents who participated in the three online surveys and their ages.

		Gender	
	Age	Male	Female
Survey 1	18-24	1	2
	25-34	5	9
	35-44	1	2
Survey 2	18-24	1	2
	25-34	1	7
	35-44	1	2
Survey 3	18-24	1	2
	25-34	1	5
	45-54		1

Table 4.2 A breakdown of participants' age range and gender in each successive survey.

Participants were asked to rate pictures on a scale from 0 to 100 and to provide any additional comments they may have had. Participants were also advised to pay attention to the backgrounds shown in the pictures rather than faces themselves.

A total of 11, 15 and 14 images were rated in the first, second and third survey. Only images perceived by 85% or more respondents as middle-class or working-class were selected for the scale. The cut off point for selecting the images was at 85% of expected ratings.

Overall, it seemed that certain types of backgrounds were more easily associated with working- or middle-class. Images showing shops, interiors, for example, living rooms, different makes of cars, housing etc. Although all of the above seemed to evoke class associations, images of houses evoked these associations in the most consistent way. Therefore, it was decided that pictures of housing associated with working- and middle-class people should be used in the experiments.

Furthermore, pictures of houses, unlike images of interiors, are good pictures to use in perceptual research since they contain a single focal point (Farnand, 2013). According to guidelines for pictorial stimuli used in perceptual experiments developed by Farnand (2013), pictures with one focal point, unlike complex pictures or pictures with no focal point, yield the best results in terms of attention consistency.

Farnand (2013: 176) also points out that one of the features attracting viewers' attention is a single human face, either natural or artificial, which becomes a focal point of an image. Even though adding composites of human faces to pictures showing different types of housing in fact turned pictorial stimuli into images with two focal points, with a composite face being one and an example of housing being the other, it was believed that the viewer would be able to decode the image's meaning easily, owing to the fact that the backgrounds were chosen for their simplicity, i.e. they were not excessively 'busy'. What is more, as has been already mentioned, viewers were instructed to pay attention to backgrounds when evaluating the pictorial stimuli.

4.2.6 Procedure

The same audio stimuli that were used in Experiment 1 (cf. Section 3.2.1) were used in Experiment 2. When audio was played, the pictures were simultaneously displayed on the screen. Upon hearing a stimulus, listeners were asked to select one of the images showing the supposed speaker. If class images were shown on screen, listeners were to select one of two or four images; if, on the other hand, an age scale was presented, listeners were to select one of three images. In block I of class evaluations listeners saw two images of housing; the first one associated with working-class people and the second one associated with middle-class people. Listeners had to choose which of the two pictures was more applicable. In block II of class evaluations listeners saw the same images of housing. However, this time a male or female face was projected over the housing images. Listeners either had to select a male face on the background of middle- or working-class housing or a female face on the background of middle- or working-class housing. Finally, in block III listeners saw the same housing but this time the gendered faces were younger and older. Listeners had to choose one of four pictures showing a younger and older man on the background of middle- or working-class housing or a younger and older woman on the background of middle- or working-class housing.

In age evaluations listeners were shown images of three men of different apparent ages or three women of different apparent ages. In each case listeners were to pick one image.

As can be seen each stimulus was rated in terms of perceived speaker age and social class. In both instances, participants were primed to see the speaker as male or female.

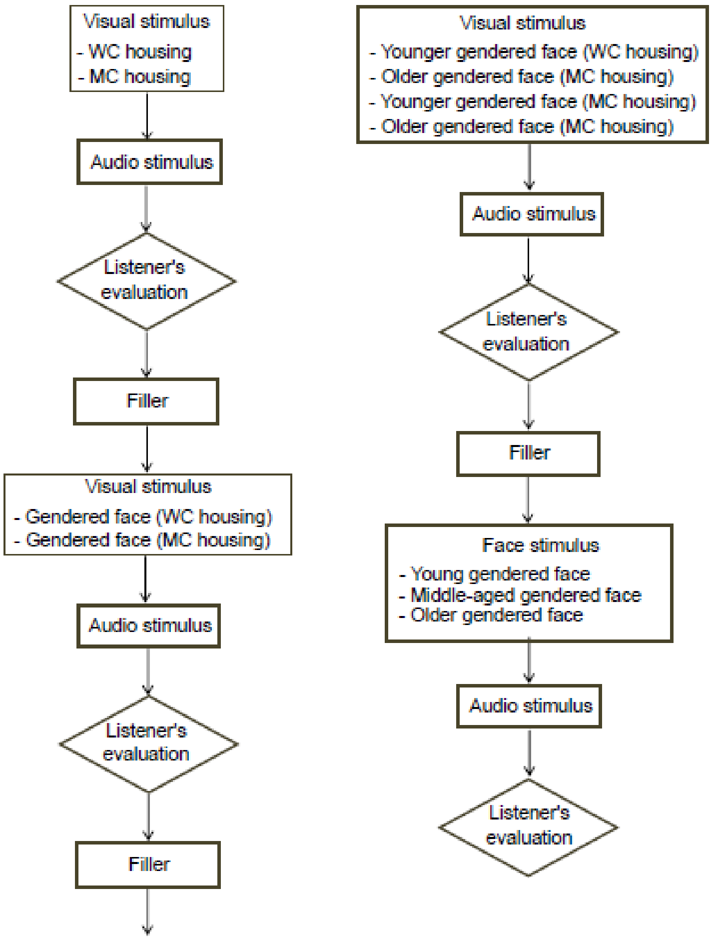


Fig. 4.10 Experimental procedure. The flowchart illustrates the process of social class evaluation of the speaker (first three blocks) and age gender evaluation of the speaker (the last block).

No other images, such as pictures of stimuli, as per Experiment 1, were shown. Even though pictorial representations of stimuli were shown in Experiment 1, it was decided to exclude them from Experiment 2, with only one exception. Since Experiment 2 is a priming experiment, there was a risk of additional images cancelling out priming effects exerted by images showing the supposed speaker. However, a pictorial representation was shown with the retracted NURSE variant [ɔː]. This was done to ensure that the listeners with high exposure to the features of the dialect under investigation understood the word as expected, rather than as a different word.

As before, the experiment was conducted in laboratory conditions and administered in SurveyGizmo (SurveyGizmo, 2014). At the beginning of the experiment, there was a training session, after which participants were given time to ask questions. A sequence of 391 single-word stimuli and fillers were presented over headphones at a comfortable hearing level. The entire session took about 75 minutes, including three breaks during the experiment. As in Experiment 1, participants were asked to complete Sudoku puzzles during the breaks.

Stimuli were administered in a similar manner to that used in Experiment 1. Images were displayed on the screen and audio files were played with a delay of a second.

At the end of the main experiment, participants were shown all the images used for the class evaluations and were asked to group them according to perceived social class (middle vs. working). Even though the images had been pre-rated, this step ensured that the present group of participants associated images with ‘middle-‘ and ‘working classness’ in the same way as the previous groups of participants.

4.3 Results

The section presents and discusses the results of perceived speaker social class and age for the variants of the NURSE vowel, T-to-R, as well as, glottalled, glottalised and released /t/. As was the case in Experiment 1 also here glottalised /t/ is investigated in two different environments (cf. Sections 2.8 & 3.4).

Logistic regression mixed-effects models were applied to investigate the perceived social class of the speaker. The dependent factor was the *perceived class of the speaker* (working-class or middle-class) and the fixed effects included the following:

- The phonetic variant (number of variants depended on the variable)
- The gender of the face showing the supposed speaker (male or female)
- Speaker (A or B)
- Listener (participant)
- The gender of the listener (male or female)
- The social class of the listener (working-class or middle-class)

In the first block of social class analysis (cf. Section 4.2.3), (in the full model variant), *speaker, gender of the listener* and *social class of the listener* were entered as fixed effects. In the second and third block of social class evaluations *gendered face of the supposed speaker* was added as a fixed effect. Furthermore, in block 3, where information about age of the supposed speaker was added to the information about social class and gender, a split analysis was performed. The perceived class of the speaker was analysed separately from perceived speaker age. Furthermore, an interaction between the specific *face genders (male/female) of the supposed speaker* and *phonetic variants* was tested.

For evaluations of perceived speaker age, ordinal regression mixed-effects models were performed. Even though it could be argued that age is conceptually a linear notion, for the purpose of this experiment age was treated as a discrete variable. As has been mentioned in Section 4.2.4, listeners were instructed to choose one of three pictures of the supposed speaker, where the images represented younger, middle and older-looking faces. Since there were only three discrete points on the scale, (younger, middle and older), it would be difficult to construct a continuous scale using these three points. Therefore, it was decided to treat the age of the supposed speaker as three discrete choices instead and apply ordinal regression. Nevertheless, to account for the fact that age is a continuous variable conceptually speaking, a linear regression mixed-effects model was also applied to test for any possible differences and similarities in the output of the ordinal and linear tests. The tests showed no significant results. It is believed that this resulted from using a three-point scale, which was in fact too short to be able to function as a linear scale.

In the ordinal regression for perceived speaker age, the dependent factor was the *perceived age of the speaker* (young, middle-aged or older) and the remaining fixed effects were the same as in the logistic regression applied for evaluations of perceived speaker social class in this experiment. An interaction between the *face gender* (male/female) of the supposed speaker and the *phonetic variant* was tested as well.

4.3.1 Evaluations of perceived speaker social class

4.3.1.1 Block 1

Fig. 4.11 A screenshot of the images used in block 1.

4.3.1.1.1 Variants of the NURSE vowel

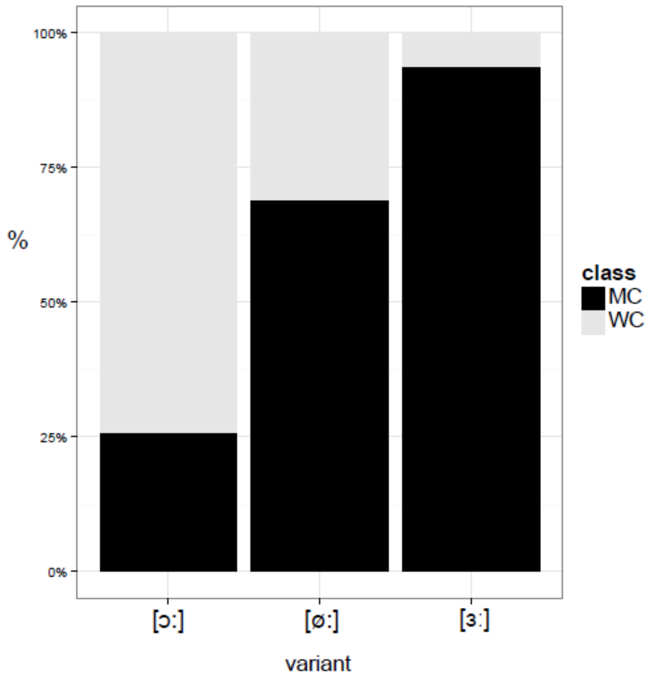


Fig. 4.12 Perception of speaker social class for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class information about the supposed speaker. Tyneside group.

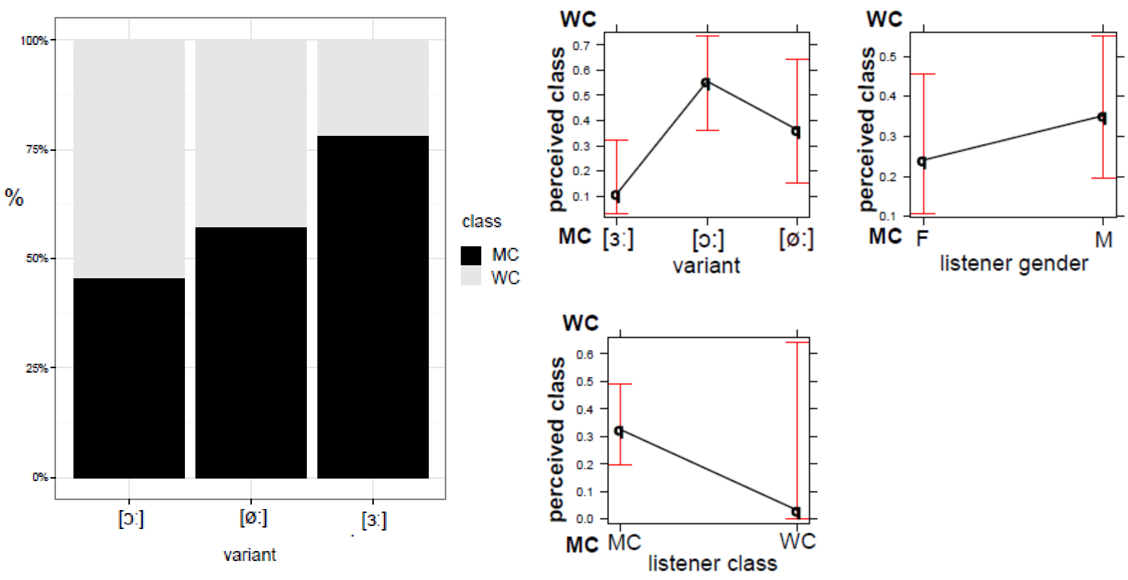
As has been already mentioned, in the first block of evaluation of speaker social class (Fig. 4.12), listeners were presented with images representing middle- and working-class housing. The retracted variant [ɔː], which is most often used by older working-class males (Beal et al., 2012, Watt & Milroy, 1999, Watt, 1998) was judged by listeners to be overall working-class sounding. The fronted variant [øː], which is described as most often used by young middle- and working-class females but also by older working-class females (Beal et al., 2012, Watt & Milroy, 1999, Watt, 1998), was rated as overall middle-class sounding. Finally, the non-localised [ɜː], was judged as definitely middle-class sounding. It seems that findings of this study agree with results reported by production studies as far as the evaluation of perceived speaker social class of the [ɔː] variant is concerned. Nevertheless, the situation is not as clear in the case of the [øː] variant. While the variant was rated as definitely working-class sounding, these findings do not reflect results reported in production studies (Watt & Milroy, 1999).

In the best-fit model (Table 4.3) entering *gender* and *social class of the listener* or *speaker* as fixed effects did not contribute to the model and so these factors were removed from it.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-3.455	1.086	-3.181	<.01	**
Variant [ɔː]	5.398	1.197	4.509	<.001	***
Variant [øː]	2.657	1.076	2.469	<.05	*
Listener gender (Male)	-1.753	0.767	-2.285	<.05	*

Table 4.3 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class information about the supposed speaker. Random slopes (1 + variant | listener) and random intercepts (1|audio). Number of observations (N)=243; Listener=27. Tyneside group.

A closer look at the output of logistic regression reveals that compared to variant [ɜː], there was a tendency for variants [ɔː] and [øː] to be evaluated as having been produced by a working-class speaker (Table 4.3). Thus, if the variant was [ɔː] or [øː], it was more likely to be evaluated as having been produced by a working-class speaker compared to variant [ɜː]. The effect was highly significant for both variants (p<0.001 and p<0.5 respectively). As a result, there was a significant correlation between *variants* and the *social class of the speaker*, which is the predicted variable. Furthermore, an effect of gender of the listener was observed. When the gender of the listener was male, there was a tendency to evaluate the variants as having been uttered by a middle-class speaker (p<0.05). Nevertheless, it should be kept in mind that the sample of respondents was not balanced for gender, which might contribute to the effect of gender of the listener.



a) **b)**

Fig. 4.13 a) & b) Perception of speaker social class for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class information about the supposed speaker. Southern group.

The two graphs, (a) and (b), in Fig. 4.13 show evaluations of speaker social class. Whilst Fig. 4.13 (a) shows a raw count of the responses, Fig. 4.13 (b) shows all effects occurring in the data thus reflecting findings of the statistical analysis. Fig. 4.13 (b) shows predicted probability for class evaluation along with error bars which give some idea of the uncertainty of the estimate. In the comparison group of respondents the localised variants of the NURSE vowel were evaluated quite similarly as in the main group. The retracted variant [ɔː] seemed to be cueing slightly more working-class ratings than the fronted variant [øː]. As expected, however, the listeners differentiated between the localised variants and the non-localised one. It can be noticed that [ɜː] was evaluated as overall having been produced by a middle-class speaker in comparison to the two localised variants.

In statistical analysis (Table 4.4), the *gender* and *class of the listener* were not significant; yet they contributed to the model. Thus, in the best-fit model, *variant, gender* and *social class of the listener* were entered as fixed effects.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-2.350	0.792	-2.965	<.01	**
Variant [ɔː]	2.323	0.816	2.845	<.01	**
Variant [øː]	1.536	0.932	1.648	<.1	
Listener gender (Male)	0.538	0.574	0.937	0.348	
Listener class (WC)	-2.687	2.043	-1.315	0.188	

Table 4.4 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class information about the supposed speaker. Random slopes (1 + variant | listener) and random intercepts (1 | audio). N=288; Listener=32. Southern group.

When compared with the results obtained in the Tyneside group of listeners it can be noticed that in both groups the retracted variant had a tendency to be evaluated as having been uttered by a working-class speaker. However, this tendency for the retracted [ɔː] to be evaluated as working-class-sounding in comparison to the centralised [ɜː] was weaker in the Southern group of listeners than in the Tyneside group of listeners ($p<0.01$ and $p<0.001$ respectively). As far as the fronted variant [øː] is concerned, it was more likely to be rated as having been spoken by a working-class speaker when compared to the centralised [ɜː]. However, this difference was significant only among the Tyneside group of respondents ($p<0.05$), while in the Southern group it was merely a tendency that did not achieve significance.

The perceptual differences between non-localised and localised variants in the comparison group would mean that participants with low exposure to features of the dialect under investigation were sensitive to [ɔː] as sounding different from the other localised and non-localised variant. Nevertheless, more subtle speaker social-indexical information encoded in the localised variants, which would allow listeners to differentiate between them, seems to have been out of reach for the majority of respondents in the comparison group.

The findings might imply that, as expected, listeners with high exposure to the dialectal features of Tyneside English were more sensitive to speaker-indexical information carried by the phonetic variants. When comparing the results in each of the groups, however, it could be argued that the differences between the groups were not large. Thus, it seems that the group of listeners with low exposure to the dialect evaluated the variants similarly. The explanation might be that speaker social class is a type of social-indexical information listeners are very sensitive to. Thus, as could be predicted, a localised non-standard variant would be evaluated as more working-class sounding especially when compared with a non-localised standard variant even if the listeners were not particularly familiar with the variety.

In fact, a statistical comparison of the results among the Tyneside group and the Southern group presented in Table 4.5 showed that the *dialect exposure* was not statistically significant, and that [ɔː] and [øː] were rated as more likely to have been produced by a working-class speaker in comparison to the centralised [ɜː] ($p<0.001$ and $p<0.01$, respectively).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-2.621	0.756	-3.465	<.001	***
Variant [ɔː]	3.546	0.751	4.719	<.001	***
Variant [øː]	1.902	0.733	2.595	<.01	**
Dialect exposure	-0.247	0.532	-0.465	0.642	

Table 4.5 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class information about the supposed speaker. N=531, Listener=59. Tyneside and Southern groups.

4.3.1.1.2 T-to-R

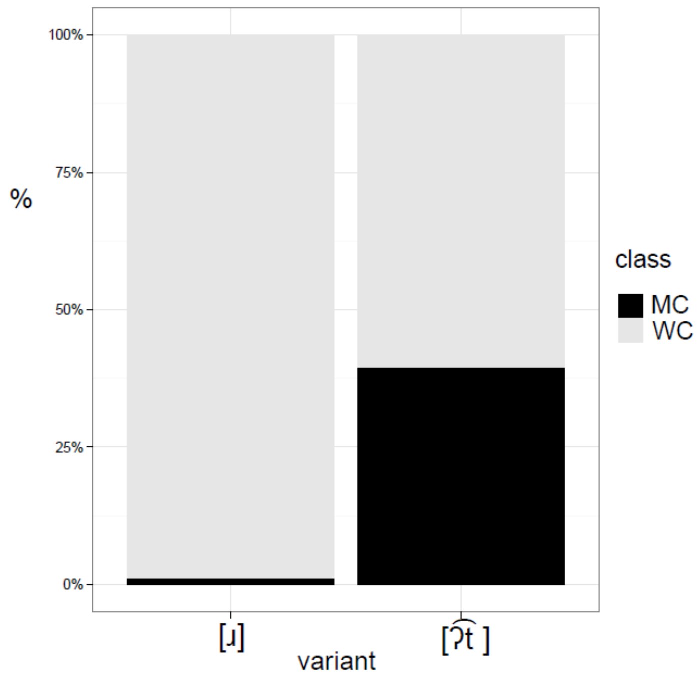


Fig. 4.14 Perception of speaker social class for [ɹ] and [ʔ͡t]. Listeners were exposed to social class information about the supposed speaker. Tyneside group.

It will be recalled that [ɹ] is a characteristic feature of the speech of older working-class females, although it can be also found to a lesser extent among younger working-class females (Watt & Milroy, 1999). [ʔ͡t], on the other hand, is most often found among older middle- and working-class male Tyneside speakers (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy et al., 1994a; 1994b; Watt & Milroy, 1999).

Figure 4.14 shows that [ɹ] was evaluated as working-class sounding by the Tyneside group of listeners. At the same time, [ʔ͡t] was found to be somewhat more working-class sounding than middle-class sounding. In fact, these findings reflect the results of production studies quoted above.

As was the case with the variants of the NURSE vowel, here also the largest model included interactions. The best-fit model to the data was simple and included *variant* and *listener class* as the fixed effects (Table 4.6). It should be mentioned that as a result of the design of the experiment, which needed to account for words sounding gender-ambiguous but natural, some of the words were produced by one of the speakers only (cf. Section

3.2.1). This was the case for [ɹ] compared with [ʔ͡t]. As a result of the lack of speaker variation, speaker voice was excluded from the analysis in this case. In addition, it can be noticed that (except for *variant*) none of the fixed effects contributed to the model.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	0.570	0.443	1.286	0.198	
Variant [ɹ]	8.904	3.639	2.447	<.05	*
Listener class (WC)	-0.128	0.608	-0.211	0.832	

Table 4.6 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class information about the supposed speaker. N=162; Listener=27. Tyneside group.

In terms of statistical results for the Tyneside group of respondents, [ɹ] was more likely to be evaluated as having been produced by a working-class speaker ($p<0.05$) than was [ʔ͡t].

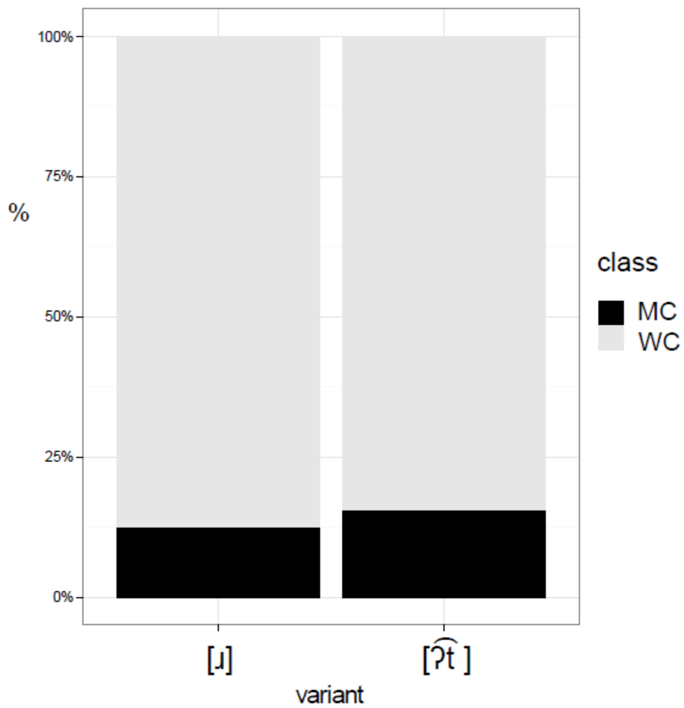


Fig. 4.15 Perception of speaker social class for [ɹ] and [ʔ͡t]. Listeners were exposed to social class information about the supposed speaker. Southern group.

In the Southern group, the two variants under investigation ([ɹ] and [ʔ͡t]) were found to be overall equally working-class sounding (Figure 4.15). This was not the case for T-to-R among the Tyneside group of respondents, who rated [ɹ] as having been produced almost exclusively by a working-class speaker. [ʔ͡t], on the other hand, was found to be more middle-class sounding in the Tyneside group. It could be concluded that listeners with low exposure to the dialect perceived both localised variants as more likely to be used by working-class speakers.

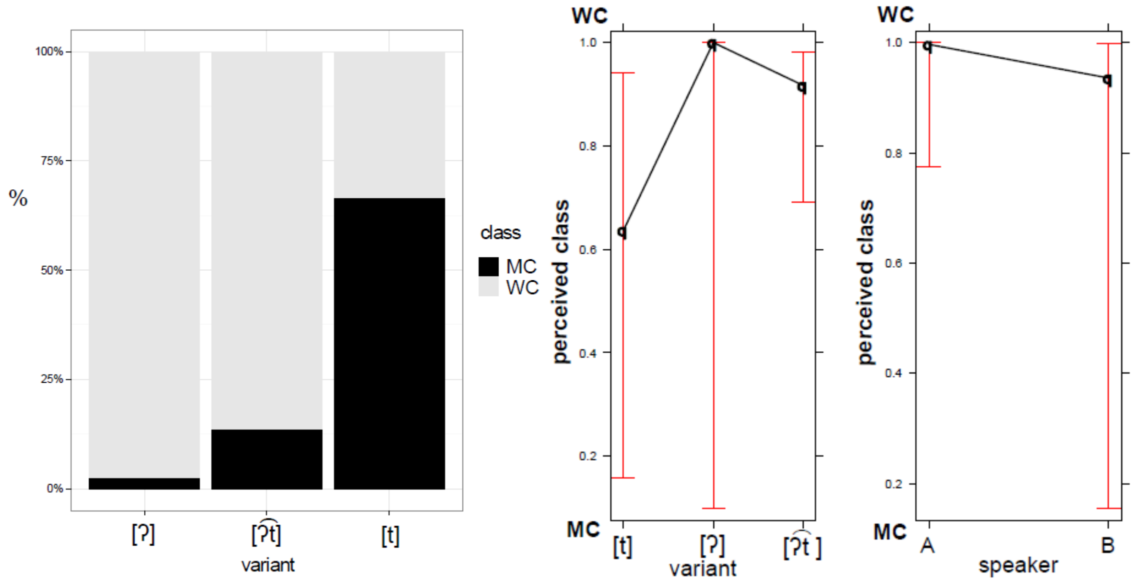
Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Intercept	5.801	5.858	0.990	0.322	
Variant [ɹ]	-2.867	5.523	-0.519	0.604	

Table 4.7 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class information about the supposed speaker. N=192; Listener=32. Southern group.

The best-fit model to the data was similar to the one for the Tyneside group of listeners (Table 4.7). The results of statistical analysis for [ɹ] and [ʔ͡t] in the Southern group confirm results observed in Figure 4.15. While for the Tyneside group of listeners [ɹ] was found to be significantly more likely to be evaluated as having been spoken by a working-class speaker than [ʔ͡t] ($p<0.5$) (cf. Table 4.6), the results for [ɹ] in the Southern group of listeners yielded no statistical significance, suggesting at the same time that both variants were perceived as having been uttered by a working-class speaker.

As far as *exposure* to dialect is concerned, a comparative analysis between the Tyneside and Southern groups did not find the variable to be significant. [ɹ] did not yield significant results either.

4.3.1.1.3 Glottalled, glottalised and released /t/



a) b)

Fig. 4.16 a) & b) Perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class information about the supposed speaker. Tyneside group.

Figure 4.16 (a) shows a raw count of the class evaluations whereas Fig. 4.16 (b) shows all effects occurring in the data. It was decided to present both figures because the raw count in Fig. 4.16 (a) does not reflect statistical results presented in Table 4.8. When looking at the graph bar statistically significant differences between the variants are expected. However, statistical results as well as Fig. 4.16 (b) show that it is not the case. Fig. 4.16 (b) shows predicted probabilities for social class with error bars thus giving some idea of the uncertainty of the estimate. The discrepancy between the graph bar and statistical results and all effects graph most probably results from the non-significant tendency of perceiving words produced by speaker B as having been produced by a middle-class speaker.

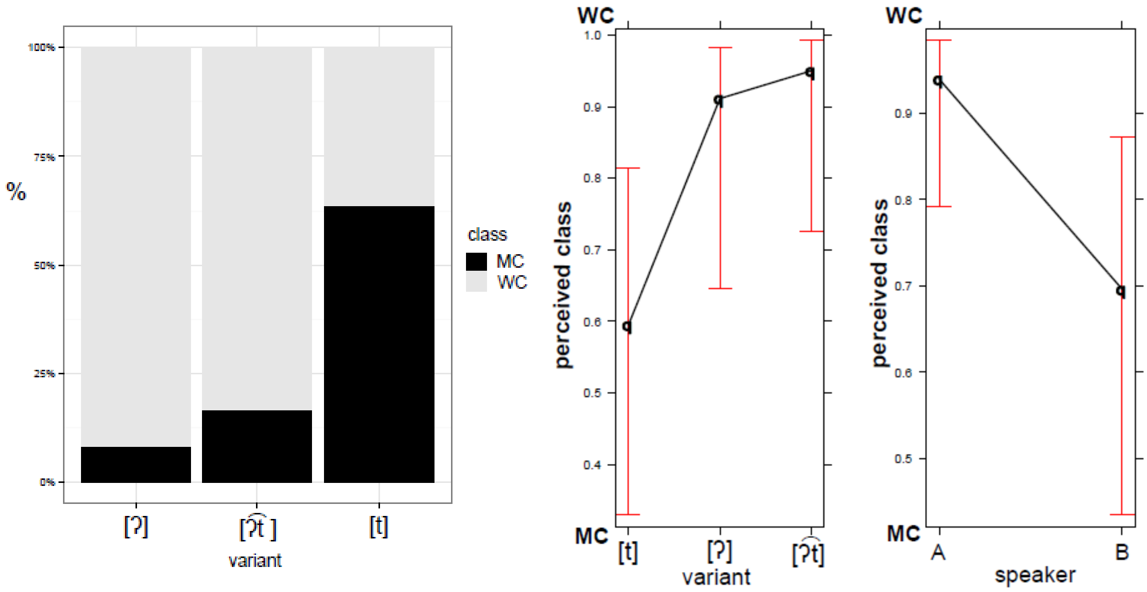
As will be recalled, [ʔ͡t] is most often used by older middle- and working-class males in Newcastle English. In the same environments, [ʔ] is used by younger middle-class females (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy et al., 1994a; 1994b). Foulkes et al. (2010: 350) point out a strong association of this variant with female speakers among Newcastle speakers.

As can be noticed in Table 4.8, in the Tyneside group [ʔ], [ʔ͡t] and [t] were rated as having been uttered by a working-class speaker, with [ʔ] perceived to sound even more working-class than [ʔ͡t] or [t] (Fig. 4.16). Thus, these results, and especially the results for the [ʔ], stand in opposition to the findings of the production studies (Beal et al., 2012; Docherty & Foulkes, 1999; Milroy et al., 1994a; 1994b).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	1.883	1.730	1.088	0.276	
Variant [ʔ]	9.458	6.457	1.465	0.142	
Variant [ʔ͡t]	1.840	1.334	1.379	0.167	
Speaker (B)	-2.975	1.586	-1.875	<.1	

Table 4.8 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class information about the supposed speaker. The model excludes interactions between predictors. N=243; Listener=27. Tyneside group.

In the best-fit model to the data, *variant* and *speaker* were entered as fixed effects (Table 4.8). As can be seen from the statistical analysis in Table 4.8, *phonetic variant* had no significant effect on evaluation of the social class of the supposed speaker. It seems that there were no differences in terms of speaker social class between the three variants. In addition, there was a tendency to evaluate variants as having been spoken by a middle-class speaker if the words were produced by speaker B (cf. Section 3.2.1).



a) **b)**

Fig. 4.17 a) & b) Perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class information about the supposed speaker. Southern group.

Figures 4.17 (a) and (b) present evaluations of [ʔ], [ʔ͡t] and [t] with respect to perceived speaker social class among the Southern group of participants. As was the case with Tyneside respondents, also here Fig. 4.17 (a) shows a raw count of the class evaluations whereas Fig. 4.17 (b) shows all effects occurring in the data. As before, Fig. 4.17 (b) shows predicted probabilities for social class with error bars to give some idea of the uncertainty of the estimate. It will be noticed that whilst Fig. 4.17 (b) matches statistical results, Fig. 4.17 (a) and statistical analysis show somewhat different results. This discrepancy could result from the fact that the graph bar presents a raw count of speaker class evaluations whereas statistical analysis takes all effects in the data into account. Furthermore, the effect of *speaker* could account for it as well, whereby words produced by speaker B were evaluated as having been produced by a middle-class speaker. Also the order of the variants in both graphs is different probably because of *speaker* effect, whereby in Fig. 4.17 (a) [ʔ] was evaluated as the most working-class sounding whereas in Fig. 4.17 (b) it was [ʔ͡t].

Interestingly, it seems that evaluations of speaker social class for the localised variants did differ from those of the Tyneside group, since [ʔ͡t] was found to be the most working-class sounding of the three variants.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	1.232	0.896	1.375	0.169	
Variant [ʔ]	1.947	1.188	1.639	0.101	
Variant [ʔ͡t]	2.565	0.902	2.842	<.01	**
Speaker (B)	-1.908	0.842	-2.264	<.05	*

Table 4.9 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class information about the supposed speaker. The model excludes interactions between predictors. N=288; Listener=32. Southern group.

The best-fit model to the data included *variant* and *speaker* as fixed effects and was significantly better from other models (p<0.05) (Table 4.9). A closer look at the statistical results for the Southern group reveals that only [ʔ͡t] was significantly different from [t] (p<0.01) (Table 4.9). In addition, words produced by speaker B were significantly more likely to be evaluated as having been produced by a middle-class speaker (p<0.05). As has been already mentioned, this might result from the design of the study (cf. Section 3.2.1).

It is interesting to see that the Southern group of respondents differentiated between localised and non-localised variants in terms of speaker social class, whereas for the Tyneside group of respondents, no effect of variant was observed (cf. Table 4.8). As could be expected, a comparison between the Tyneside and Southern groups of listeners revealed that the *dialect exposure* did not yield significant results (Table 4.10). At the same time [ʔ͡t] was perceived as having been spoken by a working-class speaker. There was also an effect of the *speaker* which was already reported for the Southern group.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	1.266	0.889	1.424	0.154	
Variant [ʔ]	4.213	2.935	1.435	0.151	
Variant [ʔ͡t]	1.897	0.680	2.787	<.01	**
Speaker (B)	-2.073	0.791	-2.619	<.01	**
Dialect exposure	0.034	0.405	0.085	0.932	

Table 4.10 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class information about the supposed speaker. N=531; Listener=59. Tyneside and Southern groups.

4.3.1.2 Block 2

In the next block of the experiment, participants were provided with more information about the supposed speaker. Thus, in addition to information about social class of the supposed speaker in the form of pictures showing middle- and working-class housing, listeners were primed with images of gendered faces and so in the male condition, listeners saw a male face, whereas in the female condition they saw a female face.

Fig. 4.18 A screenshot of the scale used in the second block. Female condition.

Fig. 4.19 A screenshot of the scale used in the second block. Male condition.

4.3.1.2.1 The NURSE vowel

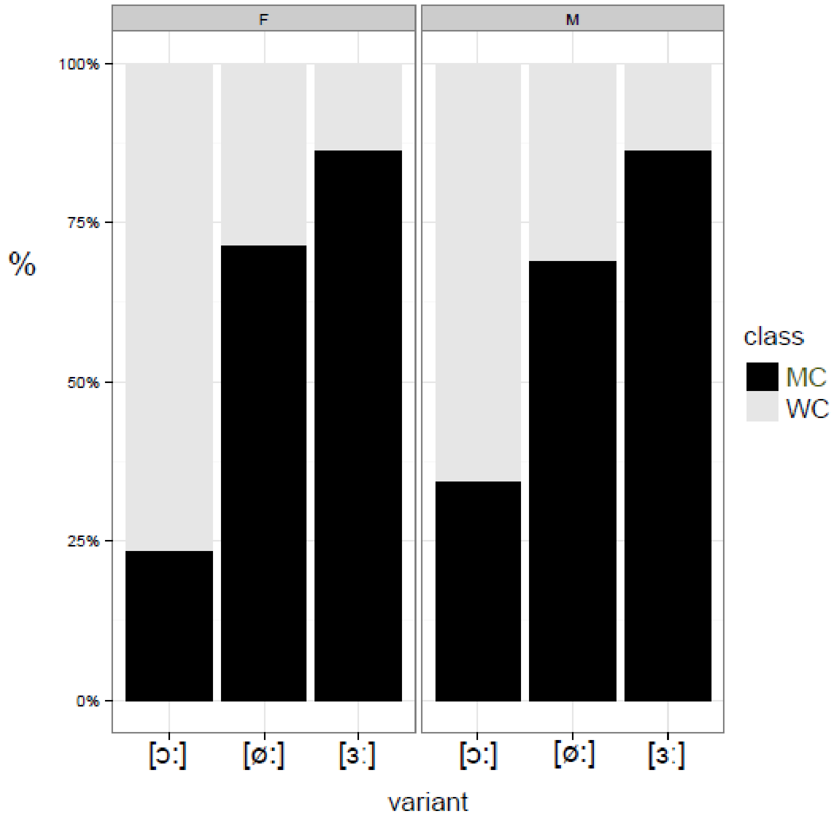


Fig. 4.20 Perception of speaker social class for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class and gender information about the supposed speaker. Tyneside group.

In block 2, in the Tyneside group of respondents the perceptions of the NURSE variants remained the same as in block 1 (Fig. 4.20). Thus, introducing information about gender of the supposed speaker did not result in a shift in evaluations of perceived speaker social class. Again, [ɔː] was evaluated as overall having been spoken by a working-class speaker when listeners were primed with a female face. About one fourth of the [ɔː] stimuli were evaluated as middle-class sounding. Furthermore, a slight shift in favour of judging the variant as having been produced by a working-class speaker in comparison to block 1 can be noticed when listeners were primed with a female face. It is worth pointing out that the opposite can be seen when listeners were primed with a male face; a slight shift towards evaluating the variant as middle-class sounding can be observed. The fronted [øː] variant was found to be overall middle-class sounding when listeners were exposed to both female and male faces. About one third of the stimuli were thought to have come from a working-class speaker when listeners were exposed to either male or female faces. However, it does not seem that the gender of the faces had any meaningful impact on listeners' classification of the variant as working- or middle-class. The centralised variant [ɜː] was evaluated as overall middle-class sounding, irrespective of whether listeners were primed with a male or female face.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-2.468	0.496	-4.976	<.001	***
Variant [ɔː]	3.427	0.521	6.575	<.001	***
Variant [øː]	1.337	0.562	2.376	<.05	*

Table 4.11 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class and gender information about the supposed speaker. N=486; Listener=27. Tyneside group.

As far as statistical analysis is concerned, the simplest model was the best-fit model to the data (Table 4.11). Because none of the fixed effects except for *variant* were either significant or contributed to the model, they were excluded from analysis. The same held true for an interaction between *variant* and *gendered face* of the supposed speaker. Since no effect of gender was found among the Tyneside group of respondents, it is fair to say that exposing listeners to a female or male face did not influence evaluations of speaker perceived social class to any meaningful extent. Phonetic variants, however, predicted the social class of the speaker with a significant result. Both variants [ɔː] and [øː] were found to be significantly more working-class sounding than the centralised [ɜː] ($p<0.001$ and $p<0.05$ respectively), with variant [ɔː] in the lead.

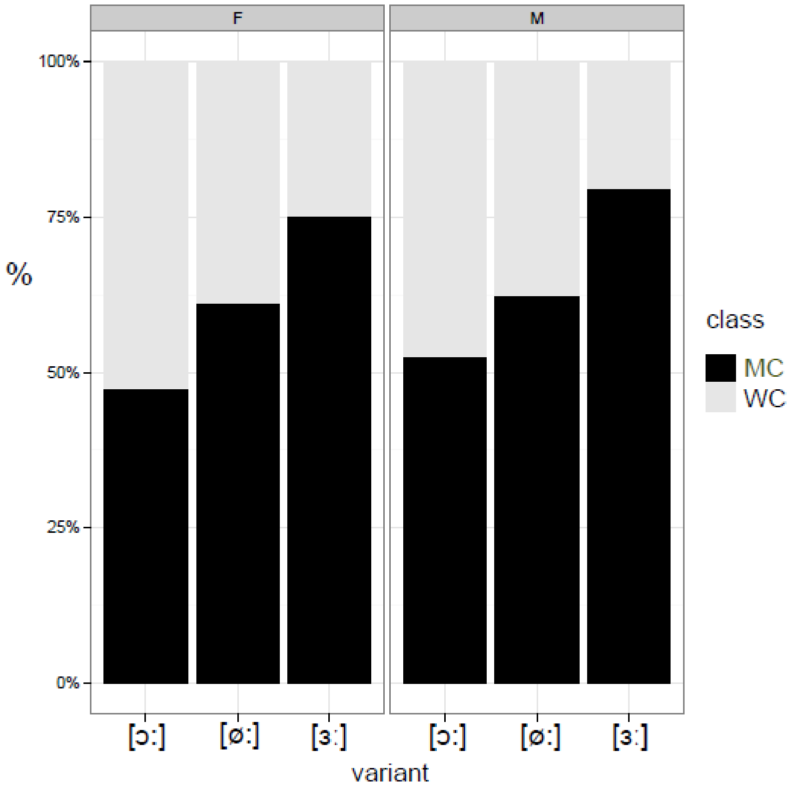


Fig. 4.21 Perception of speaker social class for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class and gender information about the supposed speaker. Southern group.

When looking at Figure 4.21, it seems that for the Southern group of respondents, the effect of cueing with a gendered face was similar to the effect obtained among the main group of respondents. In other words, also here the gender of the supposed speaker did not change evaluations of social class of the variants in any meaningful way. While the centralised [ɜː] was found to have been uttered by a middle-class speaker, the localised [ɔː] and [øː] were evaluated as working-class-sounding (Table 4.12).

In the best-fit model to the data, variant, gender of the face of the supposed speaker and speaker were entered as fixed effects (Table 4.12). As can be noticed, no effects of the gendered face or speaker were observed. However, both of these fixed effects contributed to the model, and so their results were also reported.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.100	0.366	-3.001	<.01	**
Variant [ɔː]	1.765	0.458	3.849	<.001	***
Variant [øː]	1.241	0.628	1.977	<.05	*
Face gender (Male)	-0.205	0.199	-1.031	0.302	
Speaker (B)	-0.735	0.392	-1.875	<.1	

Table 4.12 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class and gender information about the supposed speaker. N=576; Listener=32. Southern group.

Even though the graphs with results for variants of the NURSE vowel in both groups show between-group differences with respect to evaluation of the retracted variant [ɔː] in comparison with the fronted [øː] and centralised [ɜː] (Figs. 4.20 & 4.21), when comparing statistical results obtained in the Tyneside and Southern groups it can be noticed that [ɔː] and [øː] were statistically significant in both groups. Moreover, the significance was at the same level. For both groups, retracted [ɔː] was more likely to be evaluated as having been uttered by a working-class speaker than the centralised [ɜː] ($p<0.001$). As far as the fronted variant [øː] is concerned, it too was evaluated identically by both groups of listeners, and in comparison to the centralised [ɜː] it was more likely to be thought to have come from a working-class speaker ($p<0.05$). Furthermore, the effect for the *face gender* of the supposed speaker was not significant for either of the groups. Thus, it seems that after taking into account variation in the data, the results for the groups of listeners with high and low exposure to the dialect were similar. Furthermore, when we compare the results obtained in the first block of analysis with the results in the second block, it can immediately be noticed that after adding information about the gender of the face of the speaker, the results for the Tyneside group of respondents remained constant. However, they changed in the comparison group. It seems that after introducing information about the gender of the supposed speaker, evaluations in the Southern group became similar to evaluations in the group with high exposure to the dialect.

As could be expected from the results above, in a statistical comparison of social class evaluations among the Tyneside and Southern groups, *dialect exposure* did not yield statistically significant results. The variants were evaluated exactly the same as in the analyses for the Tyneside and Southern groups.

4.3.1.2.2 T-to-R

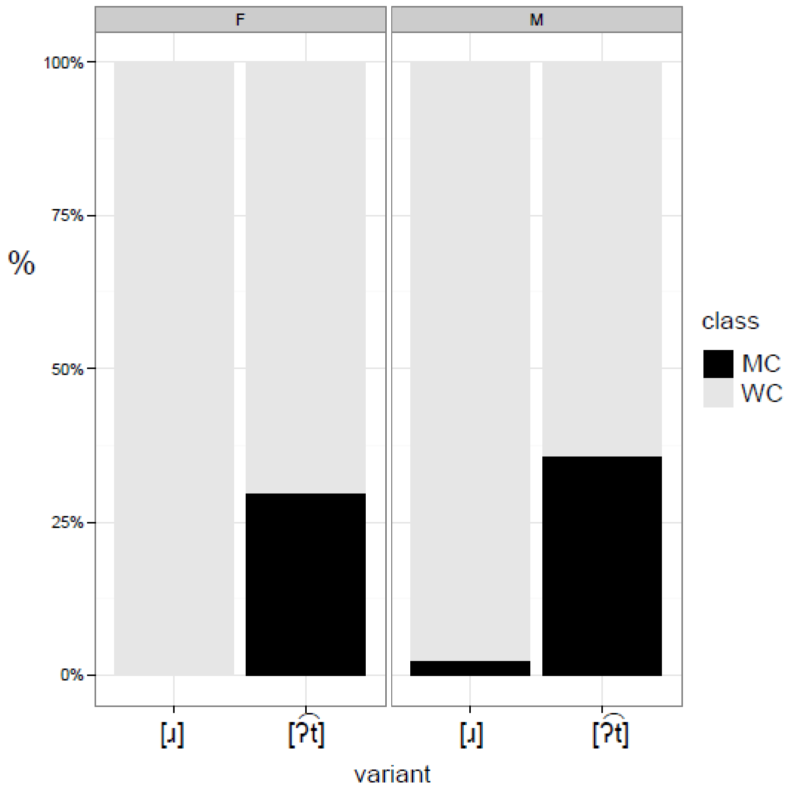


Fig. 4.22 Perception of speaker social class for [ɹ] and [ʔ͡t]. Listeners were exposed to social class and gender information about the supposed speaker. Tyneside group.

In the second block of evaluation of speaker perceived social class, Tyneside listeners evaluated [ɹ] as having been produced by a working-class speaker upon being primed with a picture of a female face (Fig. 4.22). In fact, the results were almost identical when listeners were primed with a male face. As far as evaluations of [ʔ͡t] are concerned, listeners found the variant to be overall more working-class sounding. At the same time, roughly a third of all [t͡ʔ] tokens were rated as middle-class sounding. It will be noticed that in block 2 of evaluations of speaker social class, [t͡ʔ] was perceived as only slightly more working-class sounding than in block 1 (cf. Fig. 4.14).

Variant, gendered face of the supposed speaker and an interaction between these two predictors were entered as fixed effects in the best-fit model (Table 4.13). Because in the case of variants of T-to-R there was only one speaker (cf. Section 4.3.1.1.2), the speaker factor was excluded from the analysis. As can be seen, none of the fixed effects had a significant effect. Nevertheless, the best-fit model to the data was the one including interactions between predictors.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	1.413	0.542	2.606	<.01	**
Variant [ɹ]	25.537	99.053	0.258	0.796	
Face gender (Male)	-0.450	0.4283	-1.051	0.293	
Variant [ɹ]:face gender (Male)	-15.566	99.018	-0.157	0.875	

Table 4.13 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [t͡ʔ]. Listeners were exposed to social class and gender information about the supposed speaker. N=324; Listener=27. Tyneside group.

It is interesting to observe the difference in the results of logistic regression analysis for the Tyneside group in blocks 1 and 2 of evaluation of the social class of the speaker. Unlike in block 1, when information about speaker gender was added, in block 2 there was no effect of variant (Table 4.13). At the same time, the results show no effect for the face of the supposed speaker or the interaction between gender of the face and phonetic variants.

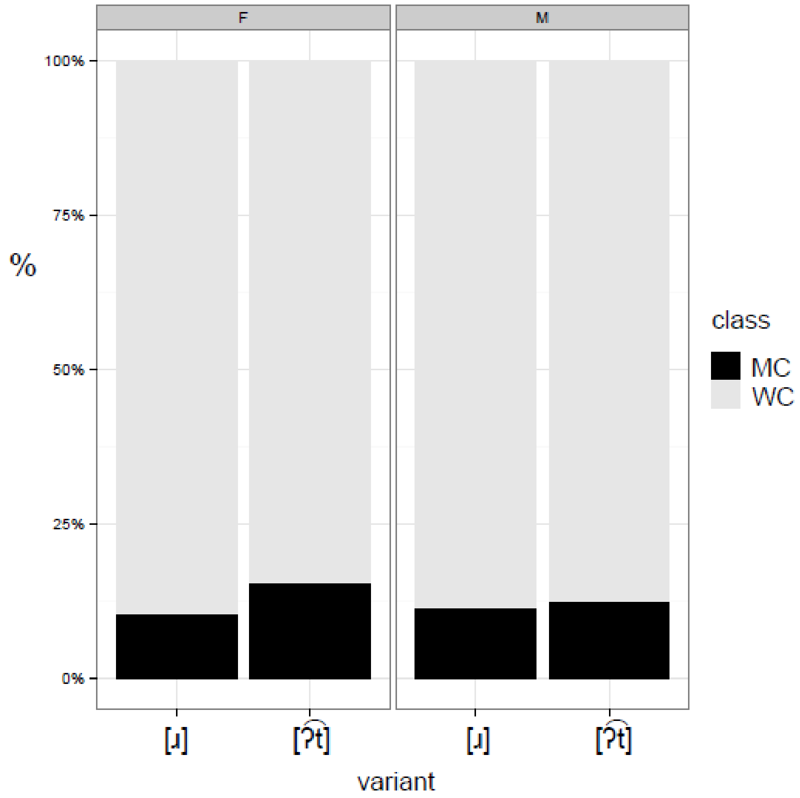


Fig. 4.23 Perception of speaker social class for [ɹ] and [ʔ͡t]. Listeners were exposed to social class and gender information about the supposed speaker. Southern group.

As was the case in the previous examples, also for T-to-R adding information about gender of the supposed speaker did not change perceptions of the variants in terms of speaker social class (Figure 4.23). Both [ɹ] and [ʔ͡t] were evaluated as thought to have come from a working-class speaker. Once more, this would suggest that there were perceptual differences between the variants among this group of respondents. However, a closer look at the statistical results will provide more insight. In the best-fit model to the data ($p<0.01$), *variant* and *gender of the listener* were entered as fixed effects (Table 4.14).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	4.399	0.960	4.580	<.001	***
Variant [ɹ]	0.083	0.567	0.148	0.882	
Listener gender (Male)	-2.327	0.907	-2.564	<.05	*

Table 4.14 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class and gender information about the supposed speaker. The model excludes interactions between predictors. N=384; Listener=32. Southern group.

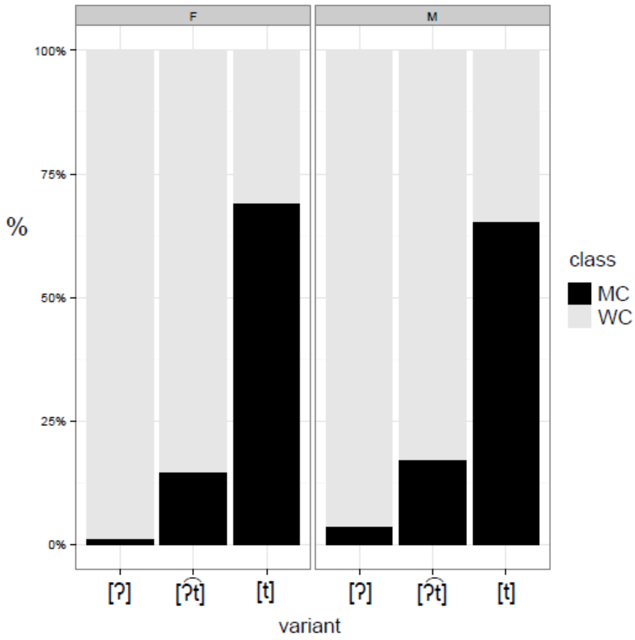
Statistical analysis reveals that the Southern group of listeners found no differences between [ɹ] and [ʔ͡t] as far as perception of social class of the speaker was concerned (Table 4.14). These results were similar to the results in the first block for the same group. At the same time, the results for the Southern group were no different from results for the Tyneside group in block 2, where no significant effect for variant was reported. Furthermore, in none of the groups of listeners did the gender of the face of the supposed speaker influence evaluations of speaker social class.

A statistical comparison of the Tyneside and Southern groups in Table 4.15 reveals an effect of the *dialect exposure* ($p<0.05$), whereby when the listener belonged to the group with higher exposure to Tyneside English, there was a tendency to perceive variants as middle-class sounding. [ɹ] was perceived as having been spoken by a working-class sounding speaker ($p<0.01$).

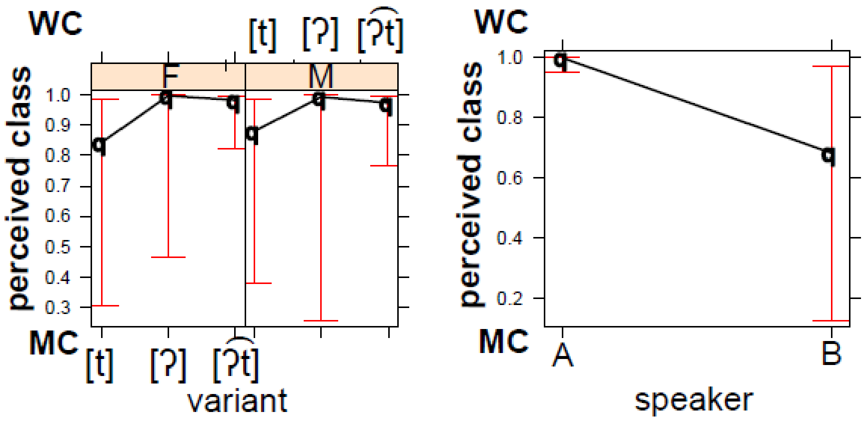
Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	2.593	0.481	5.391	<.001	***
Variant [ɹ]	5.674	1.910	2.971	<.01	**
Dialect exposure	-1.364	0.659	-2.070	<.05	*

Table 4.15 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class and gender information about the supposed speaker. The model excludes interactions between predictors. N=708; Listener=59. Tyneside and Southern groups.

4.3.1.2.3 Glottalled, glottalised and released /t/



a)



b)

Fig. 4.24 a) & b) Perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class and gender information about the supposed speaker. Tyneside group.

Figure 4.24 (a) shows a raw count of the class evaluations whereas Fig. 4.24 (b) shows all effects occurring in the data as well as predicted probability for class evaluation and error bars which give some idea of the uncertainty of the estimate. It was decided to present both figures because the raw count in Fig. 4.24 (a) does not reflect statistical results presented in Table 4.16 adequately. It seems that differences between the variants should result in statistically significant results but because of the *speaker* effect, which can be seen in Fig. 4.24 (b), they do not.

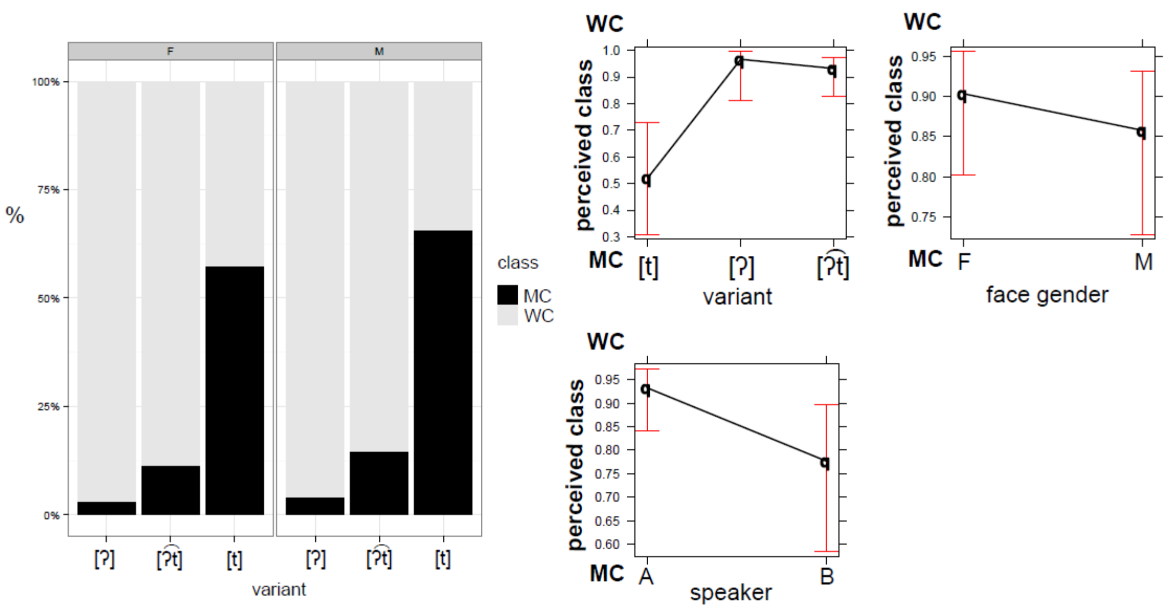
As far as the Tyneside group of listeners is concerned, it does not seem that adding speaker gender prompts resulted in any differences in the perception of the variants in terms of social class of the speaker (Table 4.16). Consequently, the results of blocks 1 and 2 seem to match each other as the effect of *variant* was statistically non-significant. The effect of *face gender* was not significant either and, as such, did not appear to contribute to evaluations of the phonetic variants (Table 4.16). At the same time, however, *speaker* had a significant effect. While in the previous block of the analysis, (block 1), *speaker* was not significant, in the present analysis the effect of speaker B is clear. Words produced by speaker B were more likely to be evaluated as having been produced by a middle-class speaker than words produced by speaker A ($p<0.01$). This might result from the design of the study although this effect was not designed in (cf. Section 3.2.1).

Even though there was no statistically significant effect of the interaction between *face gender* and *phonetic variant*, nor effects for any of the fixed factors, a comparison of the models indicated that the interaction between predictors, as well as all fixed effects significantly contributed to the model ($p<0.01$) (Table 4.16). The interaction between *face gender* and *variant* is also presented in Figures 4.24 (a) and (b) and the *speaker* effect is presented in Figure 4.24 (b).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	4.157	1.991	2.087	<.05	*
Variant [ʔ]	5.079	3.847	1.320	0.186	
Variant [ʔ͡t]	2.453	1.509	1.625	0.104	
Face gender (Male)	0.319	0.464	0.689	0.490	
Speaker (B)	-5.668	1.815	-3.123	<.01	**
Listener gender (Male)	-1.615	0.992	-1.628	0.103	
Listener class (WC)	0.944	0.781	1.209	0.226	
Variant [ʔ]:face gender (Male)	-1.809	1.414	-1.279	0.200	
Variant [ʔ͡t]:face gender (Male)	-0.785	0.832	-0.944	0.345	

Table 4.16 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class and gender information about the supposed speaker. N=486; Listener=27. Tyneside group.

After inclusion of visual cues to the supposed speaker's gender, the results of the Southern group's evaluation of social class of the three variants remained somewhat similar to those for block 1, in which only cues to the supposed speaker's social class were supplied. It is interesting to see that the Southern group of respondents seemed to find differences between the two localised variants and the non-localised variant (Table 4.17 & Figs. 4.25 (a) & 4.25 (b)). While both localised variants were more likely to be evaluated as having come from a working-class speaker by comparison with the non-localised variant, there was a stronger class effect for [ʔ͡t] than [ʔ] ($p<0.001$ and $p<0.01$ respectively). It will be recalled that no significant effect for *variant* was reported by the Tyneside group of listeners. As before, speaker voice information had a significant effect, where words produced by speaker B were more likely to be evaluated as having been produced by a middle-class speaker (<0.05). It should be also pointed out that no effect of the *gendered face* of the supposed speaker was found among the Southern group of respondents.



a) **b)**

Fig. 4.25 a) & b) Perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class information and information about gender of the supposed speaker. Southern group.

As was the case with the Tyneside respondents also here Fig. 4.25 (a) shows a raw count of the class evaluations whereas Fig. 4.25 (b) shows all effects occurring in the data as well as predicted probabilities for social class with error bars. It was decided to present both figures because together they correspond with statistical results presented in Table 4.17 more adequately. Nevertheless, it will be noticed that the results for [ʔ] and [ʔ͡t] in the graph visualisations and statistical analysis differ. Statistical analysis takes into account all effects occurring in the data together whereas both figures present the effects in isolation, hence the discrepancy. Therefore, the description of the results presented above relies on the statistical analysis.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Intercept	0.921	0.657	1.402	0.160	
Variant [ʔ]	3.265	1.117	2.923	<.01	**
Variant [ʔ͡t]	2.528	0.623	4.059	<.001	***
Face gender (Male)	-0.440	0.273	-1.611	0.107	
Speaker (B)	-1.385	0.540	-2.565	<.05	*

Table 4.17 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class and gender information about the supposed speaker. Model excludes interactions between predictors – they did not contribute to the model. N=576; Listener=32. Southern group.

A combined analysis of evaluation of speaker social class for the Tyneside and Southern groups showed that there was no effect of the *dialect exposure* (Table 4.18). Furthermore, [ʔ] and [ʔ͡t] were evaluated as having been produced by a working-class speaker ($p<0.05$ and $p<0.01$ respectively) and the effect of the *speaker* reported in separate analyses was also found.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Intercept	2.467	0.886	2.784	<.01	**
Variant [ʔ]	3.523	1.397	2.521	<.05	*
Variant [ʔ͡t]	2.111	0.728	2.898	<.01	**
Face gender (Male)	-0.195	0.279	-0.697	0.485	
Speaker (B)	-2.525	0.720	-3.507	<.001	***
Listener gender (Male)	-1.050	0.445	-2.360		*
Listener class (WC)	0.400	0.523	0.765	0.444	
Variant [ʔ]:Face gender (Male)	-0.449	0.726	-0.619	0.536	
Variant [ʔ͡t]: Face gender (Male)	-0.140	0.462	-0.304	0.760	
Dialect exposure	-0.824	0.486	-1.695	<.1	

Table 4.18 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class and gender information about the supposed speaker. N=1062; Listener=59. Tyneside and Southern groups.

4.3.1.3 Block 3

Fig. 4.26 A screenshot of the scale used in the third block. Female condition.

Fig. 4.27 A screenshot of the scale used in the third block. Male condition.

4.3.1.3.1 The NURSE vowel

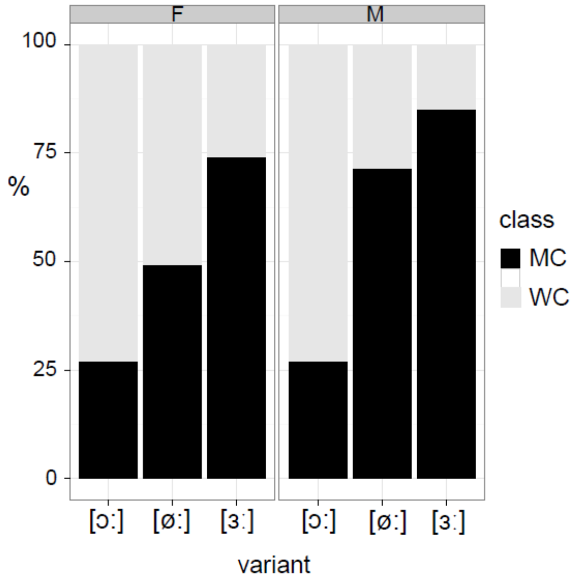
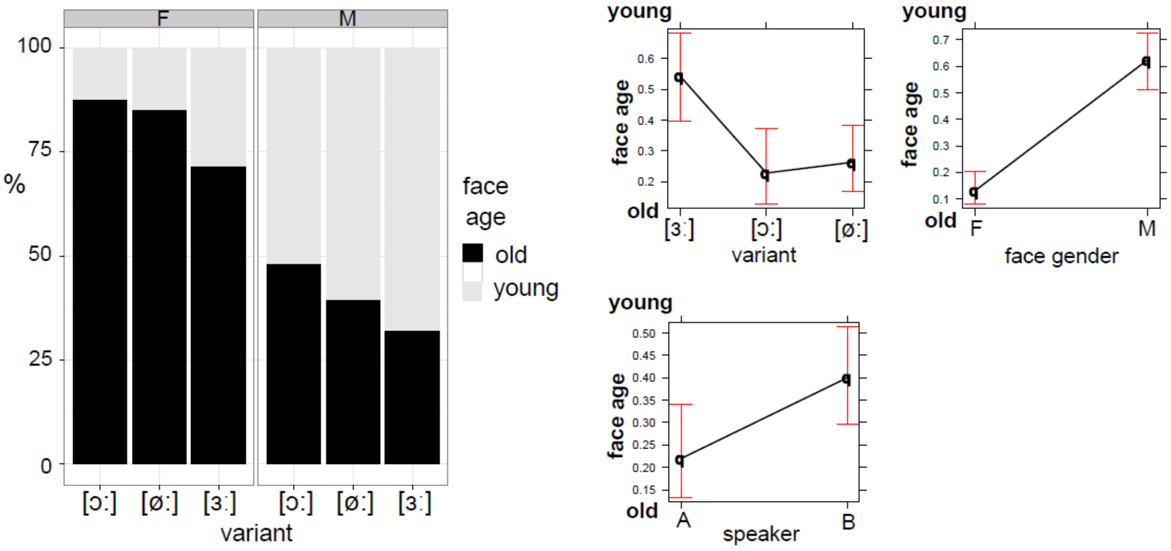


Fig. 4.28 Perception of speaker social class for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class, gender and age information about the supposed speaker. Tyneside group.



a) **b)**

Fig. 4.29 a) & b) Perception of speaker age for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class, gender and age information about the supposed speaker. Tyneside group.

At the last stage of evaluation of perceived speaker social class listeners were additionally provided with information about the supposed speaker's age, which was the final piece of information. As was the case with data visualisation for the variants of the NURSE vowel and variants of /t/ in the previous blocks also here an additional graph in Fig. 4.29 is presented for the same reasons as before. In the best-fit model to the data investigating perceptions of speaker social class, *variant* and *gendered face* of the supposed speaker were entered as fixed effects. Although the interaction between the *gendered face* of the speaker and *variant* was not significant, it contributed to the model (Table 4.19). The statistical results presented in Table 4.19 and the visualisation of the model in Fig. 4.28 show that evaluations of the perceived social class of the speaker were identical with the results in block 2 (cf. Table 4.11). In both blocks of analysis [ɔː] was evaluated as more likely to have been spoken by a working-class speaker ($p<0.001$). The same holds for [øː], except that the effect was less strongly significant ($p<0.05$). Additionally, in block 3 when information about speaker age was added, there was an effect of the gendered face of the supposed speaker. When the face was male, an increase in evaluations of variants as having been spoken by a middle–class speaker was observed ($p<0.05$).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.531	0.517	-2.958		**
Variant [ɔː]	2.567	0.607	4.22	<.001	***
Variant [øː]	1.558	0.617	2.524	<.05	*
Face gender (Male)	-0.948	0.478	-1.983	<.05	*
Variant [ɔː]:face gender (Male)	0.948	0.599	1.583	0.113	
Variant [øː]: face gender (Male)	-0.108	0.596	-0.183	0.855	

Table 4.19 The best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=486; Listener=27. Tyneside group.

Figures 4.28 and 4.29 (a) and (b) show that when listeners were primed with a female face, the retracted variant [ɔː] was rated as strongly working-class and older sounding than the centralised variant [ɜː], which was found to be middle-class sounding and somewhat younger in comparison. However, priming listeners with a male face resulted in a similar evaluation of social class and the [ɔː] variant was found equally working-class sounding, but there was a decrease in perceived speaker age. As far as evaluations of the [øː] variant are concerned, listeners rated it roughly equally middle-class and working-class sounding as well as older when a female face was shown. When the face of the supposed speaker was male, on the other hand, the variant was judged to be more likely to have come from a younger middle-class speaker. Finally, the central variant [ɜː] was rated as strongly middle-class sounding and having been produced by an older speaker when listeners were primed with a female face. It was also judged to have come from a middle-class speaker when the face of the supposed speaker was male. This time, however, [ɜː] was more likely to be evaluated as having been spoken by a younger speaker. Furthermore, it should be stressed that overall, priming listeners with younger and older male and female faces resulted in a split of evaluations between the two age groups but also in some changes in perception of speaker social class. The pattern observed here is that priming listeners with a female face resulted in variants being perceived as having been produced by an older speaker. On the other hand, priming with a male face had the opposite effect and the variants were perceived as having been produced by a younger speaker. In addition, it is interesting to observe that both in the female and male conditions the centralised variant [ɜː] was apparently thought to have been spoken by a younger speaker in comparison to the other variants. The above observations are supported by the results of the statistical analysis for age (Table 4.20). In the best-fit model to the data, *variant* and *gendered face of the speaker* as well as *speaker* and *listener gender* were entered as fixed effects.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.875	0.341	-5.501	<.001	***
Variant [ɔː]	-1.446	0.460	-3.143	<.01	**
Variant [øː]	-1.206	0.398	-3.031	<.01	**
Face gender (Male)	2.394	0.264	9.040	<.001	***
Speaker (B)	0.867	0.313	2.765	<.01	**
Listener gender (Male)	1.085	0.457	2.37	<.05	*

Table 4.20 Best-fit logistic regression mixed-effects model of perception of speaker age for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=486; Listener=27. Tyneside group.

As can be seen in Table 4.20, both localised variants were evaluated as more likely to have been uttered by an older speaker ($p<0.01$). An effect for the gendered face of the speaker can be also observed. When the face was male, there was a tendency to evaluate the words as having been spoken by a younger speaker ($p<0.001$). In addition, the effect of speaker was also significant. Words produced by speaker B were more likely to be rated as young-sounding ($p<0.01$).

A common finding in the age analysis in block 3 is the effect of the gendered face of the speaker. As has been observed, when the face of the speaker was male, there was a tendency to evaluate variants as having been uttered by a younger speaker.

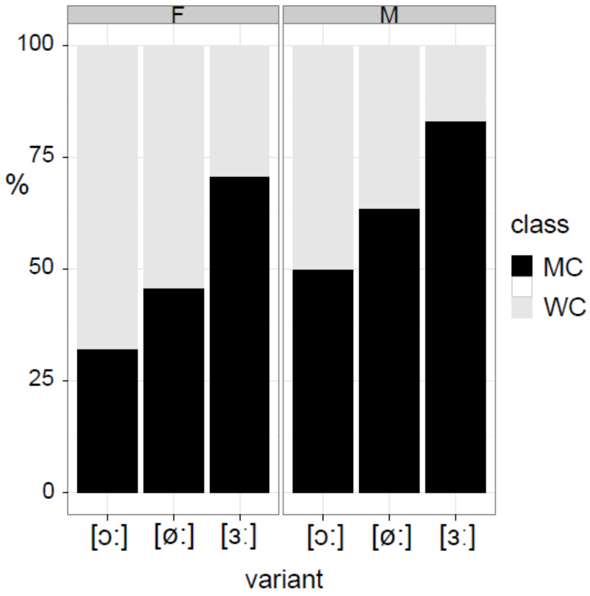


Fig. 4.30 Perception of speaker social class for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class, gender and age information about the supposed speaker. Southern group.

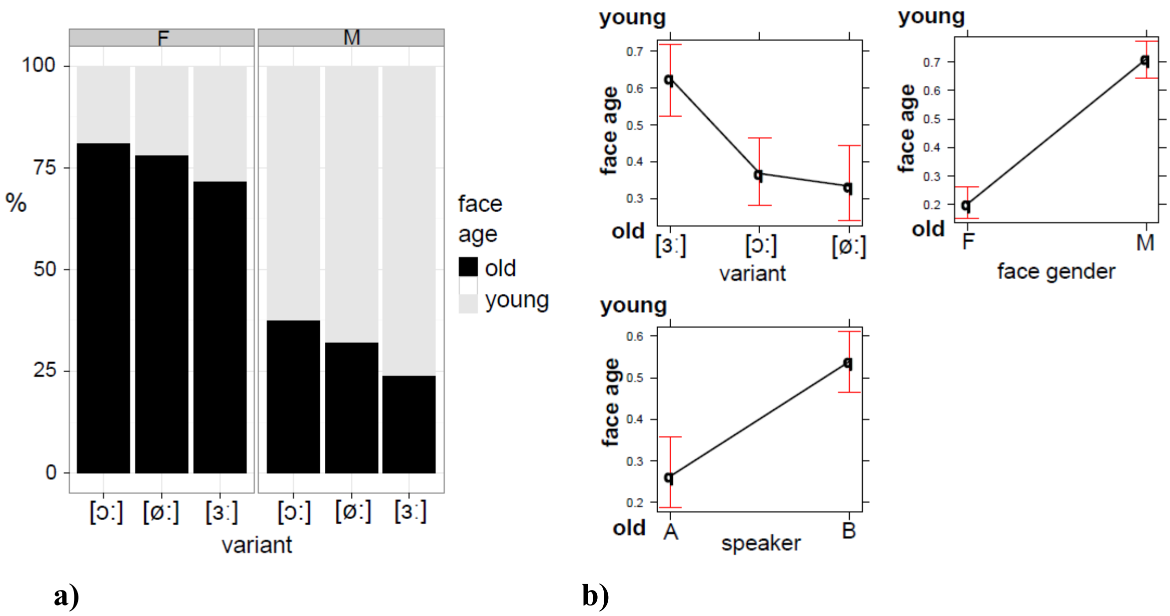


Fig. 4.31 a) & b) Perception of speaker age for the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Listeners were exposed to social class, gender and age information about the supposed speaker. Southern group.

Figures 4.30 and 4.31 (a) and (b) present evaluations of perceived speaker social class and age returned by the Southern group of participants. Also here an additional graph in Fig. 4.31 is presented for the same reasons as before. [ɔː] was evaluated as having been produced by a working-class speaker when the face of the supposed speaker was female. In comparison, when the face of the supposed speaker was male, [ɔː] was evaluated almost equally working- and middle-class sounding. The same variant in the male condition in the Tyneside group was perceived overall to be working-class sounding. [øː], on the other hand, was found to be only slightly working-class sounding in the female condition, whereas in the male condition, the proportions were reversed and it was evaluated as slightly middle-class sounding. Finally, [ɜː] was evaluated as overall middle-class sounding when the supposed speaker's face was female, as well as when it was male. Evaluations of perceived speaker age across the two groups of respondents, that is Southern and Tyneside, show a very similar pattern, where priming with a male face resulted in evaluating the variants as having come from a younger speaker and as having come from an older speaker when the face of the supposed speaker was female. Even though the differences in evaluations of each of the variants in terms of speaker age were very small or even minimal in the female and male condition, it can be noticed that in both conditions [ɜː] was perceived as having been uttered by a speaker sounding younger than the speaker producing the two localised variants. This was also true of the Tyneside group.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.400	0.578	-2.419	<.05	*
Variant [ɔː]	2.361	0.707	3.337	<.001	***
Variant [øː]	1.634	0.706	2.312	<.05	*
Face gender (Male)	-0.956	0.245	-3.892	<.001	***

Table 4.21 Best-fit logistic regression mixed-effects model of perception of speaker social class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=576; Listener=32. Southern group.

In the analysis of perceptions of speaker social class the best-fit model to the data included variant and gendered face of the speaker as fixed effects (Table 4.21). The statistical results in the two groups of respondents were yet again almost identical. [ɔː] and [øː] were evaluated as more likely to have been produced by a working-class speaker ($p<0.001$ and p<0.05, respectively). These results matched those of the Tyneside group. The only difference between the groups of listeners was the effect of gendered face of the speaker, which was stronger for the Southern group of respondents (p<0.001). When comparing the results for blocks 2 and 3 in the Southern group, it will be noticed that the only difference was the effect of the gendered face of the speaker, which was significant in block 3 (p<0.001), whereas it was not present in the previous block (cf. Table 4.12).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.415	0.246	-5.741	<.001	***
Variant [ɔː]	-1.060	0.301	-3.517	<.001	***
Variant [øː]	-1.211	0.358	-3.381	<.001	***
Face gender (Male)	2.280	0.224	10.150	<.001	***
Speaker (B)	1.189	0.272	4.359	<.001	***

Table 4.22 Best-fit logistic regression mixed-effects model of perception of speaker age for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=576; Listener=32. Southern group.

As far as analysis of perceived speaker age in the Southern group is concerned, the best-fit model to the data included *variant, gendered face* of the speaker and *speaker* as fixed effects (Table 4.22). The results for evaluation of speaker age show that among the Southern group of participants [ɔː] and [øː] were more likely to be evaluated as having come from an older speaker. The effect was also larger than in the Tyneside group (p<0.001 and p<0.01, respectively). As far as the gendered face of the speaker is concerned, the effect was the same across the respondent groups (p<0.001). Thus, when the face of the supposed speaker was male the variant was more likely to be perceived as having been uttered by a younger speaker. Finally, the effect of speaker was larger for the Southern group (p<0.001) than for the Tyneside group (p<0.01).

Tables 4.23 and 4.24 show analyses of speaker social class and age for the Tyneside and Southern groups combined. While there was no effect of *dialect exposure* on evaluations of social class, the tendencies in evaluations of the variants were similar to those in the individual analyses for the Tyneside and Southern groups (Table 4.23). [ɔː] and [øː] were found to have been produced by a working-class speaker ($p<0.001$). There was also the effect of the *face* of the supposed speaker, whereby when the face was male, variants were evaluated as more middle-class sounding ($p<0.001$).

There was no effect of the *dialect exposure* on evaluations of speaker age either (Table 4.24). As a result, [ɔː] and [øː] were evaluated similarly as in the separate analyses for the two groups of respondents. [ɔː] and [øː] were evaluated as having been produced by an older speaker. Furthermore, when the words were accompanied by a male face of the supposed speaker, they were perceived as produced by a younger speaker ($p<0.001$). The effect of the *speaker* was also reported ($p<0.001$).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.456	0.356	-4.086	<.001	***
Variant [ɔː]	2.554	0.402	6.345	<.001	***
Variant [øː]	1.408	0.400	3.513	<.001	***
Face gender (Male)	-0.805	0.150	-5.373	<.001	***
Dialect exposure	0.150	0.291	0.514	0.607	

Table 4.23 Best-fit logistic regression mixed-effects model of perception of speaker class for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=1062; Listener=59. Tyneside and Southern groups.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.323	0.224	-5.893	<.001	***
Variant [ɔː]	-1.190	0.257	-4.630	<.001	***
Variant [øː]	-1.215	0.267	-4.546	<.001	***
Face gender (Male)	2.332	0.173	13.454	<.001	***
Speaker (B)	1.049	0.205	5.104	<.001	***
Dialect exposure	-0.378	0.226	-1.670	0.095	

Table 4.24 Best-fit logistic regression mixed-effects model of perception of speaker age for variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=1062; Listener=59. Tyneside and Southern groups.

4.3.1.3.2 T-to-R

Also for T-to-R, adding information about the age of the speaker induced a shift in evaluations of speaker social-indexical information (Fig. 4.32). [ɹ] was found to be working-class and older-sounding when listeners were primed with a female face. In comparison, [ʔ͡t] in the female condition was also found to be overall working-class sounding, but there was an almost 50:50 split between young and old in terms of perceptions of speaker age. This would mean that [ʔ͡t] was perceived as equally likely to have been spoken by an older or a younger speaker. When the face of the supposed speaker was male, on the other hand, [ɹ] was also perceived as having been spoken by a working-class speaker, who was judged more likely to be younger. [ʔ͡t] was evaluated as overall working-class sounding and to have been spoken by a young speaker.

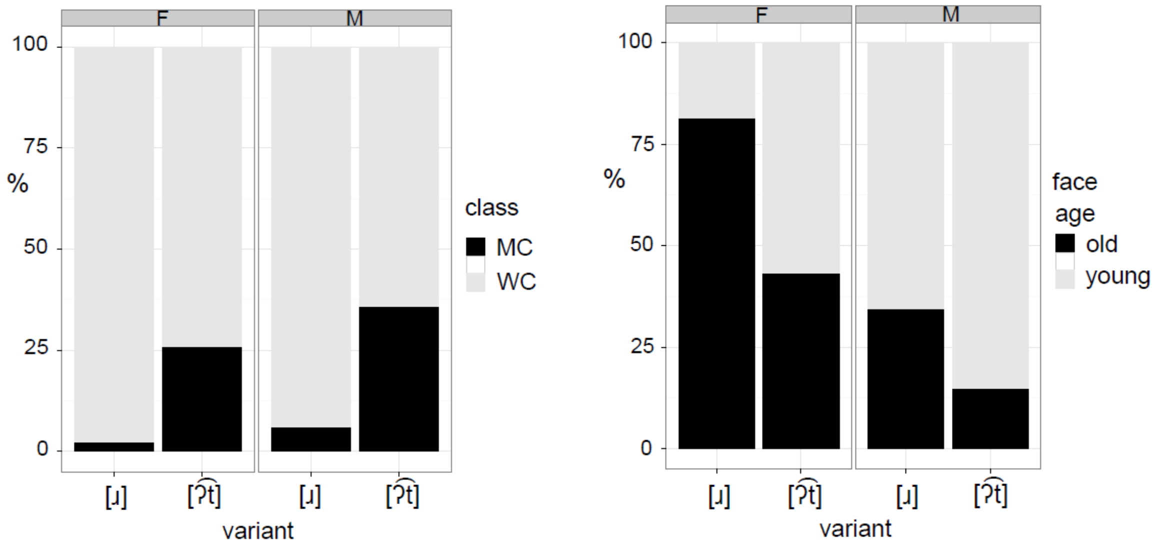


Fig. 4.32 Perception of speaker social class and age for [ɹ] and [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Tyneside group.

The best-fit model to the data included *variant* and *gendered face* of the supposed speaker as fixed effects (Table 4.25). The statistical results show that evaluations of the perceived social class of the speaker were slightly different from the results in block 2 (cf. Table 4.13). It seems that adding information about speaker age resulted in a shift in evaluations of perceived speaker social class, as a result of which [ɹ] was found to be significantly more likely to be perceived as having been uttered by a working-class speaker ($p<0.05$), whereas in block 2, there was no effect of the [ɹ] variant. However, the effect of *speaker gendered face* remained non-significant even though there was a slight tendency towards evaluating variants as having been spoken by a middle-class speaker when the face of the supposed speaker was male.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	1.382	0.425	3.249	<.01	**
Variant [ɹ]	3.340	1.494	2.236	<.05	*
Face gender (Male)	-0.662	0.354	-1.868	0.061	

Table 4.25 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. The model excludes interactions between predictors. N=324; Listener=27. Tyneside group.

In the best-fit model investigating perceptions of speaker age, as per the model investigating perceptions of speaker social class, *variant, gendered face* of the speaker and an interaction between the two were entered as fixed effects (Table 4.26).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	0.292	0.314	0.931	0.352	
Variant [ɹ]	-2.129	0.524	-4.063	<.001	***
Face gender (Male)	1.558	0.400	3.893	<.001	***
Variant [ɹ]:face gender (Male)	1.091	0.622	1.752	<1	

Table 4.26 Best-fit logistic regression mixed-effects model of perception of speaker age for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=324; Listener=27. Tyneside group.

The results in Table 4.26 show that [ɹ] was more likely to be evaluated as having been produced by an older speaker (p<0.001). The effect of the gendered face of the supposed speaker was also significant. When the face was male, there was a decrease in perceived age of the speaker.

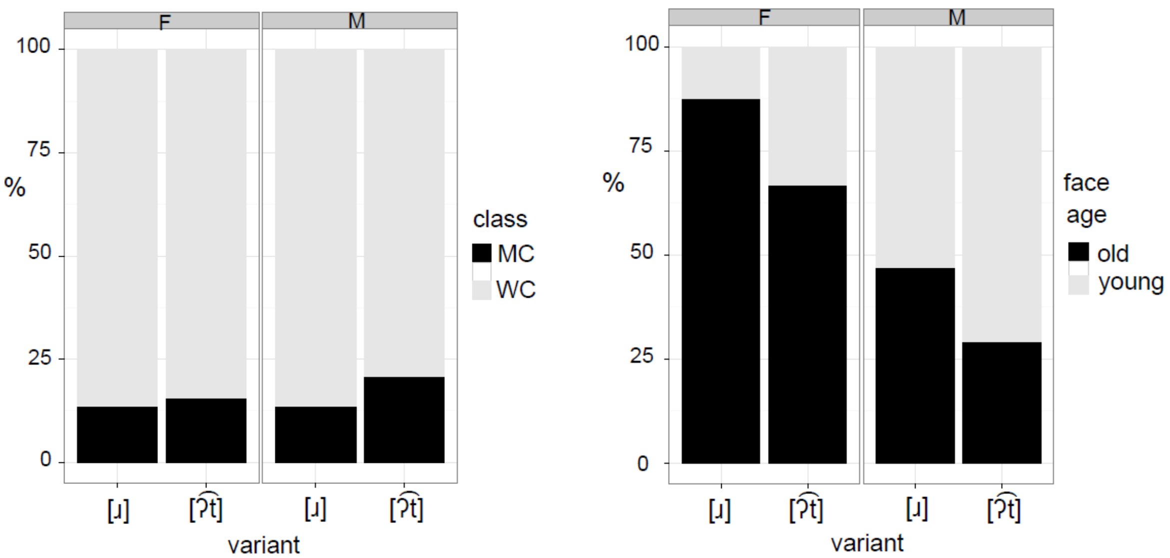


Fig. 4.33 Perception of speaker social class and age for [ɹ] and [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Southern group.

In the data for the Southern group of participants presented in Fig. 4.33 there were virtually no differences between perceptions of [ɹ] and [ʔ͡t] when the face of the supposed speaker was female and when it was male. Both variants ([ɹ] and [ʔ͡t]) were more likely to be evaluated as having been spoken by a working-class speaker. As far as perceptions of speaker age are concerned, when the face of the supposed speaker was female, [ɹ] was found to be more likely to have been spoken by an older speaker, whereas for [ʔ͡t] this tendency was lower, in the sense that for over a quarter of the listeners the speaker appeared to be young. However, when listeners were exposed to a male face, [ɹ] was almost equally likely to be heard as having been spoken by a younger or older speaker, whereas [ʔ͡t] was perceived as overall having been produced by a young speaker.

The best-fit model to the data was simple and included variant and gendered face of the speaker as fixed effects (Table 4.27). When compared to the results for the Tyneside group, it can be noticed that variant had no effect in the Southern group unlike what was reported by the Tyneside group. At the same time, the gender of the face of the supposed speaker was not significant in any of the groups.

Fixed effects	Estimate	Std. Error	Z	Pr(>|z|)	Sig.
(Intercept)	2.161	0.448	4.822	<.001	***
Variant [ɹ]	0.376	0.522	0.719	0.472	
Face gender (Male)	-0.243	0.313	-0.778	0.436	

Table 4.27 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. The model excludes interactions between predictors. N= 384; Listener=32. Southern group.

As far as the results of perceived speaker age are concerned, the best-fit model to the data was in fact identical to the model investigating perceptions of speaker social class presented above. Table 4.28 shows the results for evaluations of perceived speaker age among the Southern group of participants. Overall, [ɹ] was more likely to be evaluated as having come from an older speaker ($p<0.001$). In addition, the gendered face of the supposed speaker also had a significant effect on speaker age ratings. When the face was male, variants were more likely to be heard as having been produced by a younger speaker ($p<0.001$).

Fixed effects	Estimate	Std. Error	Z	Pr(>|z|)	Sig.
(Intercept)	-0.899	0.271	-3.315	<.001	***
Variant [ɹ]	-1.050	0.269	-3.897	<.001	***
Face gender (Male)	2.020	0.266	7.596	<.001	***

Table 4.28 Best-fit logistic regression mixed-effects model of perception of speaker age for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. The model excludes interactions between predictors. N= 384; Listener=32. Southern group.

Overall, there seems to be a difference in perceptions of the social class of the speaker using [ɹ] and [ʔ͡t] between the Tyneside and Southern groups of respondents. As has been shown, among the group with higher exposure to the dialect, [ɹ] was more likely to be evaluated as working-class, whereas no such result was reported by the Southern group. However, the results for evaluations of perceived speaker age did not differ between the two groups. The same effect as for the NURSE variants in terms of the influence of face gender on the perceived speaker age was observed (cf. Figs. 4.28 & 4.29 & Table 4.20). Nevertheless, there were differences between perceptions of the variants across the two groups upon exposing listeners to a female face. For example, [ʔ͡t] was rated as overall older-sounding in the Southern group, whereas it was found to be almost equally likely to have been spoken by either an older or a younger speaker in the Tyneside group.

It seems that listeners with high-exposure as well as low-exposure to the dialect were quite sensitive to information about speaker social class: so much so, in fact, that in some cases there were no differences between the two groups.

A comparison of the results for the social class for the Tyneside and the Southern groups shows that there was no effect of the *dialect exposure* (Table 4.29). [ɹ] was evaluated similarly as in the Tyneside group (cf. Table 4.25) and was perceived as having been spoken by a working-class speaker.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	1.922	0.396	4.843	<.001	***
Variant [ɹ]	1.787	0.643	2.776	<.01	**
Face gender (Male)	-0.434	0.235	-1.848	<.1	
Dialect exposure	-0.327	0.490	-0.667		

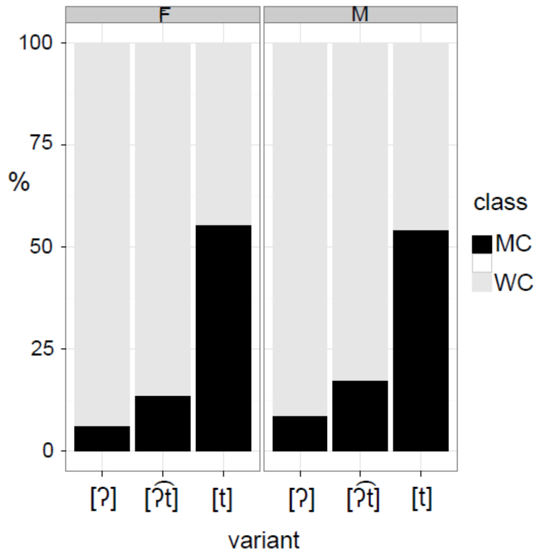
Table 4.29 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. N= 1062; Listener=59. Tyneside and Southern groups.

A comparison of the age results for the Tyneside and Southern groups (Table 4.30) shows a statistical effect for the *dialect exposure* ($p<0.001$), whereby listeners with higher exposure to Tyneside English were more likely to evaluate the variants as younger sounding. [ɹ] was evaluated as having been produced by an older speaker ($p<0.001$).

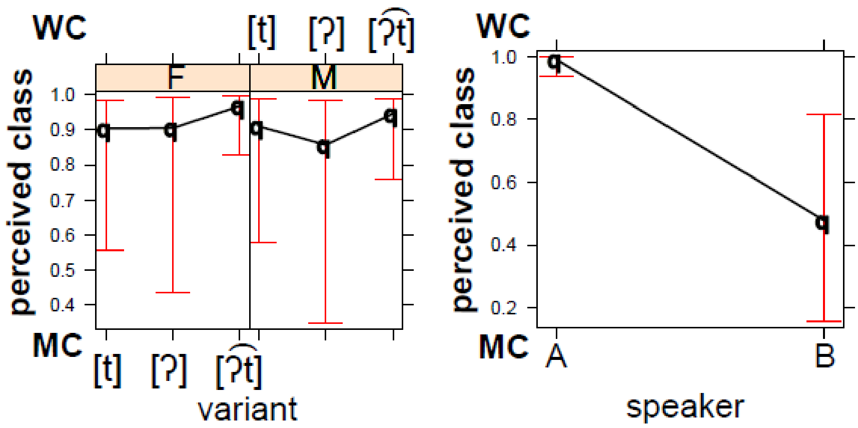
Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-0.635	0.226	-2.804	<.01	**
Variant [ɹ]	-1.664	0.337	-4.927	<.001	***
Variant [ɹ]:Face gender (Male)	0.728	0.396	1.838	<.1	
Dialect exposure	0.794	0.224	3.543	<.001	***

Table 4.30 Best-fit logistic regression mixed-effects model of perception of speaker age for [ɹ] compared with [ʔ͡t]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=708; Listener=58; Tyneside and Southern groups.

4.3.1.3.3 Glottalled, glottalised, and released /t/



a)



b)

Fig. 4.34 a) & b) Perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Tyneside group.

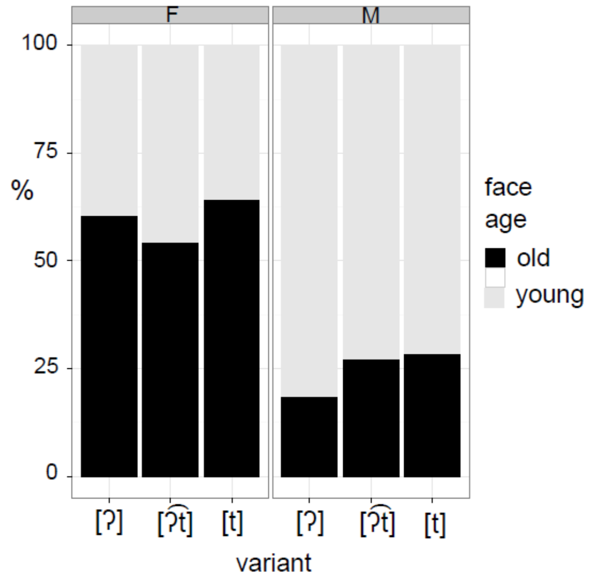


Fig. 4.35 Perception of speaker age for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Tyneside group.

The results presented in Table 4.31 and Figs. 4.34 (a) and (b) show evaluations of class and age for [ʔ], [ʔ͡t] and [t]. As was the case in blocks 1 and 2 also here additional graph is shown. Whilst the bar graph (Fig. 4.34 (a)) shows a raw count of the speaker class evaluations, the other graph (Fig. 4.34 (b)) shows all effects occurring in the data along with predicted probabilities for social class with error bars. The order of the variants in both graphs is different probably because of *speaker* effect, whereby in Fig. 4.34 (a) [ʔ] was evaluated as slightly more working-class sounding than [ʔ͡t] whereas in Fig. 4.34 (b) evaluation was the opposite. Similar evaluations were found in the Southern group in block 1 (cf. Fig. 4.17 (a) & (b)).

Furthermore, when looking at Fig. 4.34 (a) it seems that statistically significant results could be expected. However, Table 4.31 shows that it is not the case as no differences between the variants were found. The best-fit model in evaluations of perceived speaker social class included *variant, gendered face* of the speaker, and *speaker* as fixed effects. In addition, the model tested for an interaction between *gendered face* of the speaker (male/female) and *variant* ([ʔ], [ʔ͡t] and [t]) (Table 4.31). The results show two significant factors, *speaker* and *listener gender*. When words were produced by speaker B, there was a tendency to evaluate them as having been spoken by a middle-class speaker ($p<0.01$). Additionally, *listener gender* was statistically significant ($p<0.01$). When a listener was male he was more likely to evaluate the variants as having been produced by a middle-class speaker. However, as was found in previous cases when the gender of the listener yielded significant results, it should be pointed out that the sample of listeners was unbalanced for listener gender. Therefore, this result might not be meaningful. Most importantly, however, there was no effect of the variant. In addition, gendered face of the supposed speaker did not have any effect on evaluations of perceived speaker social class. Figs. 4.34 (a) and (b) show the effect of the interaction between *variant* and *face gender* and Fig. 4.34 (b) shows the effect of *speaker*.

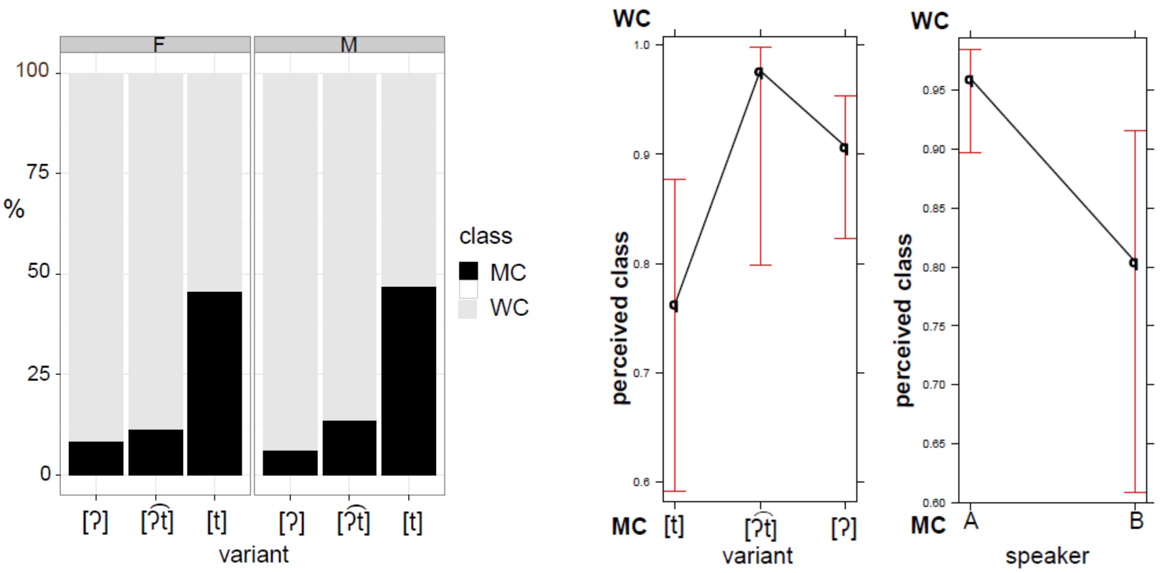
Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	4.761	1.586	3.002	<.01	**
Variant [ʔ]	0.021	1.921	0.011	0.991	
Variant [ʔ͡t]	1.192	1.210	0.985	0.324	
Face gender (Male)	0.080	0.402	0.201	0.840	
Speaker (B)	-4.638	1.426	-3.252	<.01	**
Listener gender (Male)	-1.769	0.611	-2.892	<.01	**
Variant [ʔ]:face gender (Male)	-0.551	0.800	-0.688	0.491	
Variant [ʔ͡t]:face gender (Male)	-0.640	0.739	-0.867	0.386	

Table 4.31 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=486; Listener=27. Tyneside group.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-0.525	0.487	-1.080	0.280	
Variant [ʔ]	0.395	0.631	0.627	0.530	
Variant [ʔ͡t]	0.282	0.624	0.453	0.650	
Face gender (Male)	1.837	0.233	7.873	<.001	***
Listener gender (Male)	-0.697	0.382	-1.822	<.1	

Table 4.32 Best-fit logistic regression mixed-effects model of perception of speaker age for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Model excludes interactions between predictors as they did not contribute to the model. N=486; Listener=27. Tyneside group.

Table 4.32 presents the best-fit model for evaluations of perceived speaker age. *Variant*, *gendered face* of the speaker and *listener gender* were entered as fixed effects. While none of the variants had a statistical effect, gendered face of the speaker did (p<0.001). When the face was male, a decrease in perceived speaker age was observed.



a) b)

Fig. 4.36 a) & b) Perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Southern group.

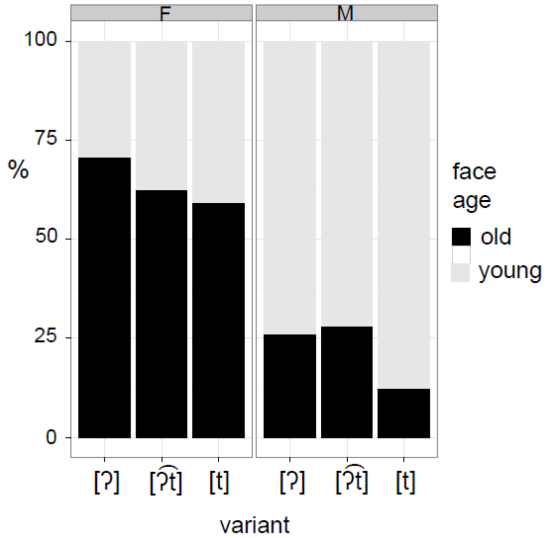


Fig. 4.37 Perception of speaker age for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Southern group.

The above figures, Figs. 4.36 (a), 4.36 (b) and 4.37, show evaluations of perceived speaker social class and age among the Southern group of respondents. For the same reasons as already explained in the previous cases also here an additional graph is presented. Because Fig. 4.36 (a) shows a raw count of class evaluations and does not take into account additional factors such as, for example *speaker* effect, it does not fully reflect findings of the statistical analysis (Table 4.33). The discrepancy most certainly results from the fact that *speaker* effect is included in the statistical analysis. Fig. 4.36 (b), on the other hand, shows *speaker* effect present in the data.

As was the case with the Tyneside group also here the order of the variants in both graphs is different most probably because of *speaker* effect. Whilst Fig. 4.36 (a) shows that [ʔ] was evaluated as the most working-class sounding of the three variants, Fig. 4.36 (b) shows [ʔ͡t] as the most working-class sounding.

As far as the perceived age of the speaker is concerned, a pattern similar to that seen in the previous examples can be observed. When listeners saw a male face, they tended to perceive the variants as overall younger-sounding. In both conditions (male and female) the non-localised variant was evaluated as the most likely to be evaluated as young sounding, while the two localised variants, on the other hand, were perceived as young-sounding but somewhat less so in comparison to the non-localised variant.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	1.951	0.600	3.252	<.01	**
Variant [ʔ]	2.559	1.340	1.910	<1	
Variant [ʔ͡t]	1.109	0.462	2.400	<.05	*
Speaker (B)	-1.758	0.529	-3.321	<.001	***

Table 4.33 Best-fit logistic regression mixed-effects model of perception of speaker social class for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. Model excludes interactions between predictors. N=576; Listener=32. Southern group.

The best-fit model investigating perceptions of speaker class included *variant* and *speaker* as fixed effects (Table 4.33). The table above presents the results of statistical analysis of perceptions of speaker social class of [ʔ], [ʔ͡t] and [t] among the Southern group of respondents. It will be noticed that [ʔ͡t] was more likely to be rated as having been produced by a working-class speaker than was [t] ($p<0.05$). In addition, no effect of face gender was found, which means that the variants were evaluated identically in the male and female conditions. However, an effect of the *speaker* was reported. When words were produced by speaker B, listeners were more likely to rate them as having been uttered by a middle-class speaker ($p<0.001$). Speaker B was more likely to be evaluated as middle-class sounding by the Southern group of respondents ($p<0.001$) than by the Tyneside group of respondents ($p<0.01$).

In a separate analysis performed on evaluations of perceived speaker age, in the best-fit model, *variant, gendered face* of the speaker, *speaker*, and *gender* of the listener were entered as fixed effects (Table 4.34). In addition, the model tested for an interaction between *gendered face* of the speaker and *variants*.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-1.400	0.717	-1.953	<.1	
Variant [ʔ]	0.700	0.800	0.876	0.381	
Variant [ʔ͡t]	0.723	0.622	1.162	0.245	
Face gender (Male)	2.507	0.418	5.997	<.001	***
Speaker (B)	1.317	0.625	2.108	<.05	*
Listener gender (Male)	-0.535	0.233	-2.291	<.05	*
Variant [ʔ]:face gender (Male)	-0.261	0.558	-0.468	0.640	
Variant [ʔ͡t]:face gender (Male)	-0.863	0.535	-1.611	0.107	

Table 4.34 Best-fit logistic regression mixed-effects model of perception of speaker age for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=576; Listener=32. Southern group.

In evaluations of perceived speaker age among the Southern group, none of the variants, nor the interaction between variant and gendered face of the speaker, yielded significant effects (Table 4.34). Instead, the gendered face of the speaker had a significant effect on perceptions of speaker age ($p<0.001$). In fact, the same effect as for variants of the NURSE vowel and T-to-R in both groups of respondents was observed for variants of /t/ in the Tyneside and Southern groups (cf. Tables 4.20, 4.22, 4.26, 4.28 & 4.32). In addition, there was an effect of the *speaker*, whereby variants produced by speaker B were more likely to be evaluated as young-sounding ($p<0.05$). Finally, an effect of listener gender was also reported. When the listener was male, he was more likely to evaluate variants as having been spoken by an older speaker ($p<0.05$). However, it should be pointed out that the Southern group of respondents was also unbalanced for gender, with the majority of the group being male respondents.

A statistical comparison of the results for social class among the Tyneside and Southern groups showed no effect of the *dialect exposure*. The *variant* did not yield statistically significant effect either.

As far as perceptions of speaker age are concerned, a combined analysis of speaker age for the Tyneside and Southern groups showed an effect of the *dialect exposure*, whereby among listeners with high exposure to Tyneside English there was a tendency to evaluate the variants as younger-sounding (p<0.05) (Table 4.35). When the face gender was male, listeners with higher exposure tended to evaluate the variants as older-sounding than the other groups of respondents (p<0.05). However, no effect of the *variant* was reported.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
(Intercept)	-0.365	0.826	-0.442	0.658	
Variant [ʔ]	0.659	0.839	0.786	0.431	
Variant [ʔ͡t]	0.828	0.692	1.197	0.231	
Face gender (Male)	2.260	0.295	7.639	<.001	***
Speaker (B)	0.349	0.700	0.499	0.618	
Listener gender (Male)	-0.537	0.210	-2.557	<.05	*
Variant [ʔ]:face gender (Male)	0.048	0.410	0.117	0.906	
Variant [ʔ͡t]:face gender (Male)	-0.767	0.384	-1.996	<.05	*
Dialect exposure	-0.701	0.281	-2.496	<.05	*

Table 4.35 Best-fit logistic regression mixed-effects model of perception of speaker age for [ʔ], [ʔ͡t] and [t]. Listeners were exposed to social class, gender and age information about the supposed speaker. N=1062; Listener=59. Tyneside and Southern groups.

4.3.2 Evaluations of perceived speaker age

Perceived speaker age was evaluated using scales constructed in EvoFIT (Frowd et al., 2006; Frowd, 2015) for this purpose. As was the case when evaluating speaker social class, listeners were here also primed with a male or a female face. The aim was to obtain a male and female face looking young, middle-aged and old. Thus, the youngest face was of a person in his or her early 20s or late teens. The middle picture showed a face in its early 40s, and the oldest face in its late 50s or early 60s. The faces were supposed to be approximately twenty years apart. Furthermore, male and female faces in each of the age groups were morphed using the same settings. Two sets of male and two sets of female faces were used in the experiment. Listeners were not provided with any information about the age of faces in the pictures.

Both groups of participants, that is the Tyneside and Southern participants, were separately tested on perception of age of the composite faces after the experiment proper. There were two reasons to test the faces independently by having their ages rated by the respondents. Firstly, it was necessary to know how respondents in the 18-24 age group perceived the ages of the faces close in age to their own age, as well as faces perceived to be older than participants themselves. Knowing these things was necessary for interpretation of the results of the experiment. Secondly, even though the same settings were applied when morphing the faces, some of the faces, and especially male ones, gave an impression of being younger-looking than the female faces. The results of the independent age rating of the faces used in the experiment are presented in Table 4.36.

	Ranges of evaluations of face age - Tyneside group	Median evaluations of face age - Tyneside group	Ranges of evaluations of face age - Southern group	Median evaluations of face age - Southern group
Young female face 1	17-26	21	18-30	22
Young female face 2	15-30	20	16-30	21
Young male face 1	17-25	20	17-26	20
Young male face 2	16-25	20	18-25	20
Middle-aged female face 1	27-49	33	30-41	35

Middle-aged female face 2	25-50	35	26-45	38
Middle-aged male face 1	28-48	35	28-40	35
Middle-aged male face 2	27-45	30	24-45	35
Older female face 1	37-65	50	45-65	50
Older female face 2	40-65	55	45-70	50
Older male face 1	40-65	50	39-65	50
Older male face 2	37-65	48	35-70	50

Table 4.36 Range and median evaluations of age of the male and female faces morphed in EvoFIT. Ratings provided by the Tyneside and Southern groups of respondents.

As was the case with the middle-aged looking faces, the older male and female faces were also evaluated by both groups of participants as younger looking than was expected by the experimenter. A possible explanation could be that the young age of respondents made them sensitive to young faces and they were able to evaluate them as was expected. However, the other faces were older than the respondents themselves, and so they may have found it difficult to rate them as expected. Nevertheless, knowing how the faces were perceived by participants was necessary to interpret the experimental results. Furthermore, the results show that the earlier concern over the female faces looking older than the male faces could be ignored, as the differences did not seem to be substantial.

The following section presents and discusses the results of evaluations of perceived speaker age for all the variants researched in the present experiment. Results obtained from the group of listeners with high exposure to the phonetic features under investigation - that is, the Tyneside listeners - are compared with the results obtained from listeners with low-exposure to the phonetic features, i.e. the Southern listeners. First, the results for the variants of the NURSE vowel are presented, followed by the results for T-to-R and finally variants of /t/.

4.3.2.1 The NURSE vowel

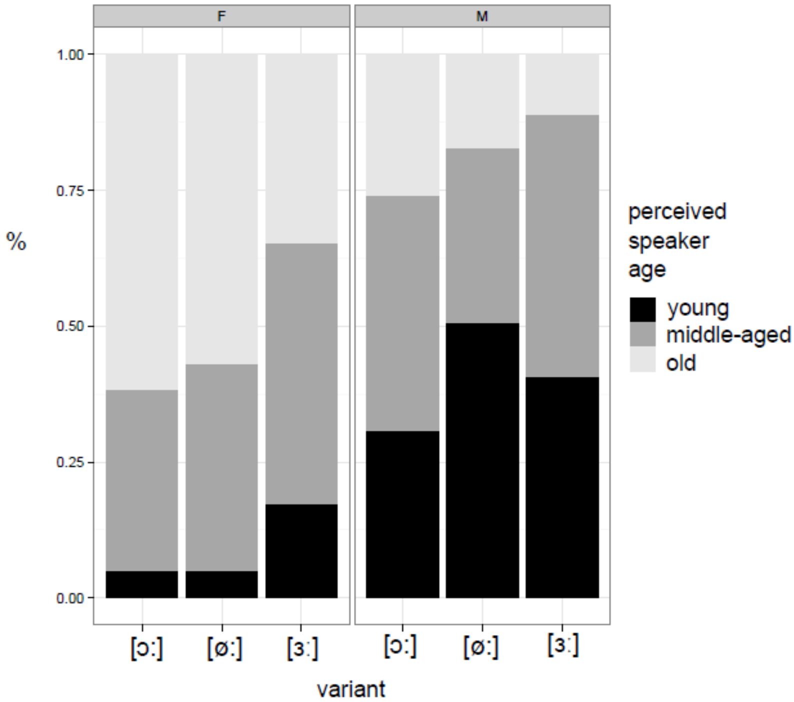


Fig. 4.38 Perception of speaker age of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Tyneside group.

As has been already mentioned, the retracted variant [ɔː] is most often used by older male Tyneside English speakers, while the fronted variant [øː] is most frequently used by younger but also older female speakers. When listeners were primed with a female face, [ɔː] was rated as overall older sounding, but also middle-aged sounding (Fig. 4.38). The same holds for [øː], for which the difference between the two age groups was even smaller. When listeners were primed with a male face, on the other hand, [ɔː] was rated as the most middle-aged sounding but also young- and somewhat old-sounding. [øː], on the other hand, was rated as the most young-sounding, less as middle-aged sounding and somewhat old-sounding. The centralised [ɜː] variant was rated as overall middle-aged sounding in both the male and female conditions. However, the variant was found to be old-sounding and only somewhat young-sounding when listeners were primed with a female face. By contrast, when primed with a male face, listeners found the variant to be young- and marginally old-sounding. Again, it can be noticed that, depending on the variant under investigation, a change in the gender of the supposed speaker had a different degree of impact on the ratings for perceived speaker age. While the change was the most significant for [øː], it also affected the two remaining variants.

The best-fit model to the data is presented in Table 4.37. This model was simple; it had three main effects and tested for an interaction between *variants* of the NURSE vowel and the specific *face gender* (male/female) of the supposed speaker.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
variant [ɔː]	1.189	0.387	3.069	<.01	**
variant [øː]	1.036	0.363	2.849	<.01	**
Face gender (Male)	-1.322	0.309	-4.269	<.001	***
variant [ɔː]:face gender (Male)	-0.484	0.437	-1.106	0.268	
variant [øː]:face gender (Male)	-1.212	0.447	-2.710	<.01	**

Table 4.37 Best-fit ordinal regression mixed-effects model with interactions between predictors. Perceptions of speaker age of the variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. N=486; Listener=27. Tyneside group.

From Table 4.37 we can learn that in general, [ɔː] and [øː] were reported as having been produced by an older speaker more often than was the case for [ɜː]. This difference was statistically significant for both localised variants ($p<0.01$). In addition, [ɔː] was more likely to be rated as slightly older-sounding than [øː]. Furthermore, having a male face accompany the words resulted in a decrease in the reported speaker age. The effect was statistically significant ($p<0.001$). As far as the interaction between the specific face genders (male/female) and NURSE variants is concerned, it can be noticed that when a male face was shown, listeners tended to evaluate [øː] and [ɔː] as younger-sounding than the centralised [ɜː]. However, the effect was statistically significant only for [øː] ($p<0.01$). Thus, it can be further inferred that in the male condition the fronted variant [øː] was found to be the youngest-sounding, the retracted variant [ɔː] was evaluated as older-sounding than the fronted one and, finally, the centralised [ɜː] as the oldest of the three. There was no effect of *speaker*.

In other words, given the gender-ambiguous voice, when a male face was seen, the variant was perceived as younger-sounding because F0 was towards the upper end of the typical range for male speakers. Thus, gender of the face on its own predicts the age. Also, the variants [ɔː] and [øː] on their own predict the perceived speaker age as older.

Let us now compare the results for the Tyneside group of participants with those for the Southern group (Fig. 4.39). It seems that in the Southern group of respondents, middle-aged was the most likely response when the listeners were exposed to the variants of the NURSE vowel when words were accompanied by a female face of the supposed speaker. On the other hand, when the face of the supposed speaker was male, the variants tended to be rated as younger-sounding. It can be noticed that the interaction between the gender of the supposed speaker and age of the speaker for the NURSE variants was reversed in the present group of respondents. While Tyneside listeners rated variants of the NURSE vowel as the most old-sounding when the supposed speaker was female, Southern listeners rated the variants as the most young-sounding when the supposed speaker was male.

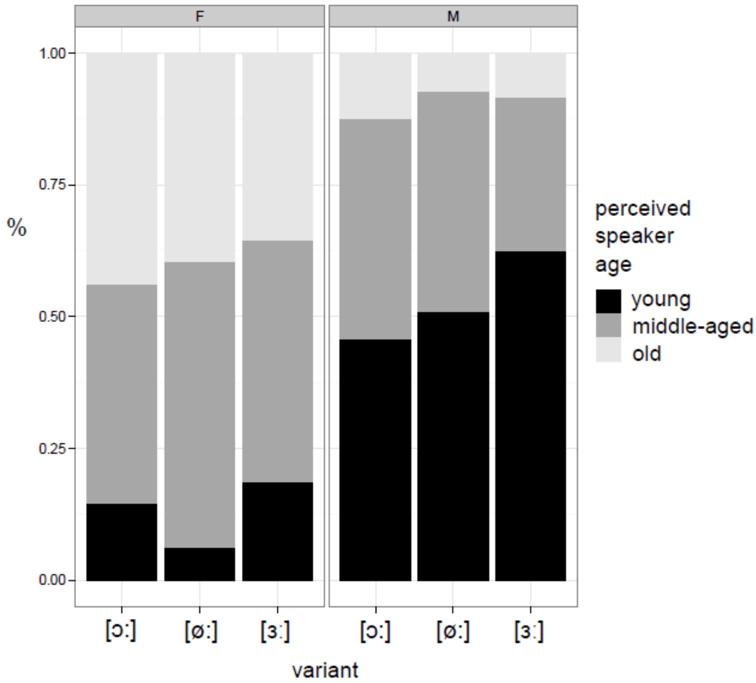


Fig. 4.39 Perception of speaker age of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel. Southern group.

In the Southern group, the best-fit model to the data tested for the effect of individual *variants*, as well as the interaction between individual NURSE *variants* and *gender of the face* of the supposed speaker (Table 4.38). The only significant factor was the effect of the gendered face of the supposed speaker. As expected, when the words were accompanied by a male face, a decrease in the perceived age of the speaker was reported. The effect of gendered face of the supposed speaker was the only similarity between the Tyneside and Southern groups of respondents. For both groups it was highly significant ($p<0.001$). It will be recalled that this effect has been already reported across all variables in block 3 of social class evaluation (cf. Section 4.3.1.3).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Variant [ɔː]	0.601	0.360	1.669	0.095	
Variant [øː]	0.460	0.326	1.409	0.158	
Face gender (Male)	-2.202	0.194	-11.343	<.001	***

Table 4.38 Best-fit ordinal regression mixed-effects model with interactions between predictors. Perceptions of speaker age of the variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. N=576; Listener=32. Southern group.

A comparison of the Tyneside and Southern groups of respondents revealed that [ɔː] was more likely to be perceived as having been uttered by an older speaker (p<0.01) (Table 4.39). Among listeners with higher exposure to Tyneside English there was a tendency to evaluate the variants as older sounding (p<0.01).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Variant [ɔː]	0.776	0.283	2.735	<.01	**
Variant [øː]	0.438	0.260	1.681	<.1	
Face gender (Male)	-2.035	0.139	-14.607	<.001	***
Dialect exposure	0.564	0.193	2.914	<.01	**

Table 4.39 Best-fit ordinal regression mixed-effects model with interactions between predictors. Perceptions of speaker age of the variants of the NURSE vowel, where [ɔː] and [øː] are compared against [ɜː]. N=1062; Listener=59. Tyneside and Southern groups.

4.3.2.2 T-to-R

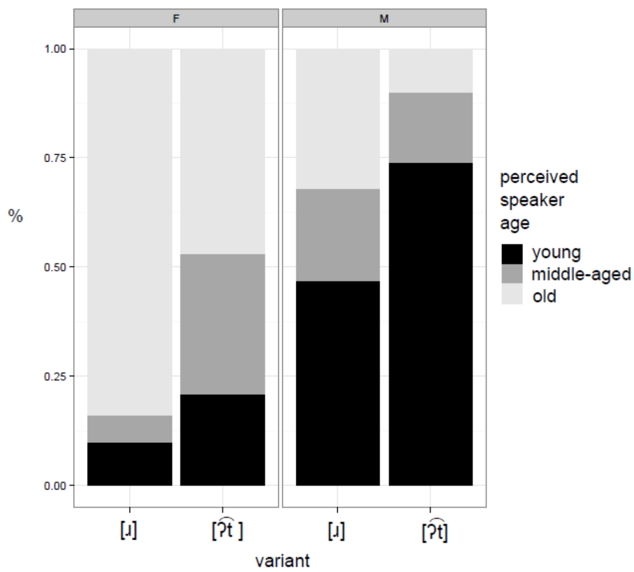


Fig. 4.40 Perception of speaker age for [ɹ] and [ʔ͡t]. Tyneside group.

This section discusses the results for T-to-R for the Tyneside group of respondents. It will be recalled that in Tyneside English [ɹ] is used mostly by older females but also by younger ones. [ʔ͡t], on the other hand, is most frequent among older males. As was the case in Experiment 1, in the present Experiment the results for [ɹ] and [ʔ͡t] were the most clear-cut. [ɹ] was evaluated as definitely older-sounding when listeners were primed with a female face (Fig. 4.40). However, the results were not as categorical when listeners were primed with a male face because the variant was found to be largely young-sounding but also somewhat old- and middle-aged sounding. [ʔ͡t], by contrast, was rated as definitively young-sounding when listeners were primed with a male face. Yet it was rated as old, middle-aged and young-sounding when listeners were primed with a female face.

The best-fit model to the data included *variant* and *gender of the face* of the supposed speaker entered as fixed effects but also as an interaction (Table 4.40). Even though the interaction between predictors was not significant, it contributed to the model. Thus its results were reported in the table below.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Variant [ɹ]	2.738	0.687	3.985	<.001	***
Face gender (Male)	-2.678	0.397	-6.739	<.001	***
Variant [ɹ]:face gender (Male)	-1.304	0.685	-1.903	<.1	

Table 4.40 Best-fit ordinal regression mixed effects model with interactions between predictors. Perceptions of speaker age of [ɹ] compared with [ʔ͡t]. N=324; Listener=27. Tyneside group.

The statistical model shows that both the effect of *gender of the face* of the supposed speaker and the effect of *variant* were significant ($p<0.001$) (Table 4.40). Overall, there was a tendency to report [ɹ] as having been produced by an older speaker (in comparison to [ʔ͡t]) ($p<0.001$). In addition, when the words were accompanied by gendered faces, seeing a male face led to a decrease in estimated age. The results for the interaction between *gender of the face* (male/female) and *phonetic variants* show that even though the interaction was non-significant, when a male face was presented listeners showed a tendency to evaluate [ɹ] as younger-sounding than when the face of the supposed speaker was female. This tendency has been already reported for variants of the NURSE vowel in this section (cf. Tables 4.37 & 4.38).

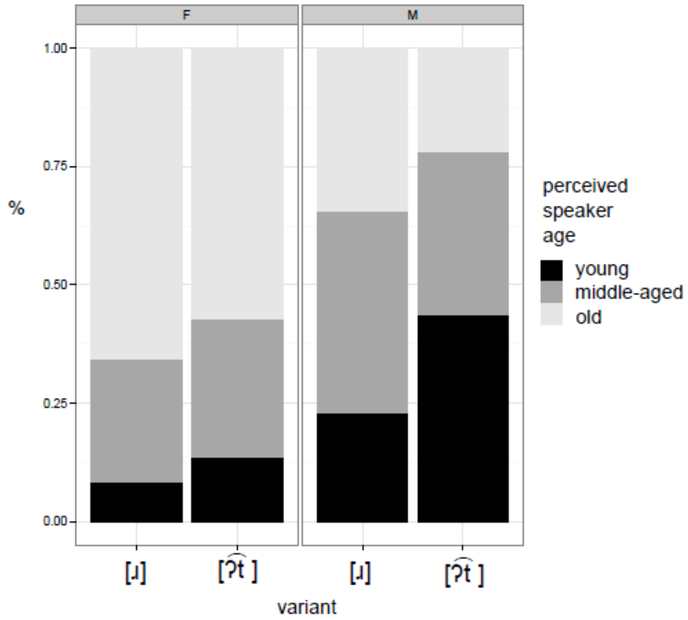


Fig. 4.41 Perception of speaker age for [ɹ] and [ʔ͡t]. Southern group.

The results for the Southern group of respondents (Fig. 4.41) show that [ɹ] and [ʔ͡t] were rated as old-sounding when the supposed speaker was female. In comparison, the two variants were rated as younger-sounding, that is young or middle-aged, when the condition was male. However, these two variants were perceived differently in the male condition. While [ɹ] was evaluated as mostly middle-aged sounding, [ʔ͡t] was evaluated as young-sounding. These findings do not reflect the patterns reported for the Tyneside respondents, who had a more diverse perception of the two variants. It seems that in their evaluations of speaker age, Southern listeners in particular were guided by voice features rather than the speaker social-indexical information carried by the variants.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Variant [ɹ]	0.755	0.262	2.879	<.01	**
Face gender (Male)	-1.719	0.228	-7.516	<.001	***

Table 4.41 Best-fit ordinal regression mixed-effects model with interactions between predictors. Perceptions of speaker age of [ɹ] compared with [ʔ͡t]. N=384; Listener=32. Southern group.

Table 4.41 presents results of the best-fit model to the data for the Southern group. This model tested only for the effect of the *variant* and *gendered face* of the speaker. As can be noticed, [ɹ] was more likely to be rated as having been spoken by an older speaker. The same tendency was found among the Tyneside respondents. However, the significance of the effect among the Southern group of respondents was lower than that found for the Tyneside group of respondents ($p<0.01$ and $p<0.001$, respectively). Unlike in the Tyneside group, in the Southern group the interaction between variant and gender of the face of the supposed speaker was neither significant nor a contribution to the model. Therefore, it was not reported in the best-fit model. Another similarity between the groups was the significant effect of the gendered face of the speaker on evaluations of perceived speaker age ($p<0.001$) (cf. Tables 4.37, 4.38 & 4.41).

As expected, a comparison of the Tyneside and Southern groups of respondents did not reveal an effect for the *dialect exposure* (Table 4.42). The same tendency for [ɹ] as in the individual analyses was found, whereby [ɹ] was found to have been produced by an older speaker ($p<0.001$).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Variant [ɹ]	1.184	0.188	6.275	<.001	***
Face gender (Male)	-2.199	0.185	-11.868	<.001	***
Dialect exposure	-0.694	0.364	-1.906	<.1	

Table 4.42 Best-fit ordinal regression mixed-effects model with interactions between predictors. Perceptions of speaker age of [ɹ] compared with [ʔ͡t]. N=702; Listener=59; Tyneside and Southern groups.

4.3.2.3 Glottalled, glottalised and released /t/

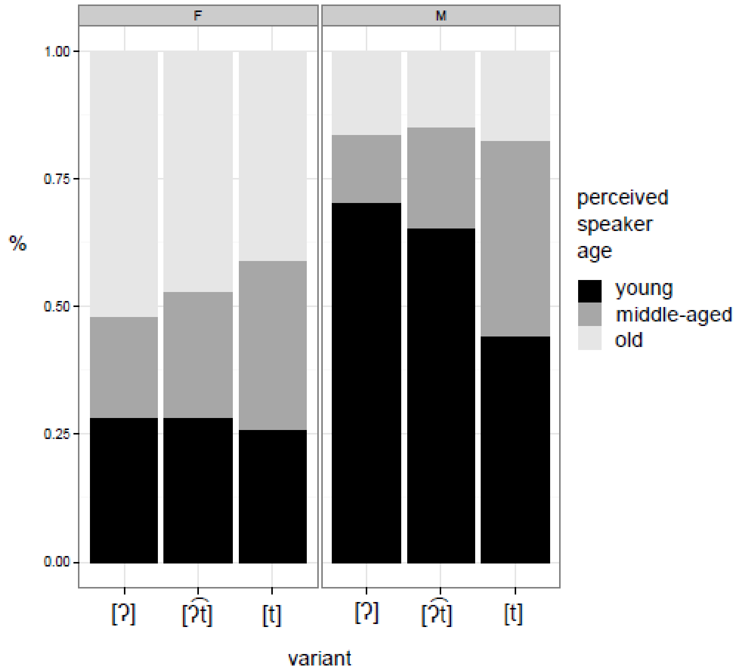


Fig. 4.38 Perception of speaker age for [ʔ], [ʔ͡t] and [t]. Tyneside group.

For the Tyneside group, the results for the variants of /t/ seem to be somewhat inconclusive. While [ʔ] was rated as overall old-sounding when listeners were primed with a female face, it was also found to be somewhat young- and middle-aged sounding. The results were the opposite when listeners were primed with a male face: the variant was rated as overall young-sounding. The same pattern can be noticed for the [ʔ͡t]. Finally, listeners did not respond in a consistent way in their ratings of released /t/ with respect to the perceived age of the supposed talker when they were primed with a female or male face. It seems, especially in the case of the localised variants, that priming with a female face resulted in more old ratings, while priming with a male face gave the opposite result.

The best-fit model tested for interactions between *variant* and *face gender* of the supposed speaker (Table 4.43). As can be seen, speaker as well as *gender* and *social class of the listener* were excluded from the best-fit model.

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Variant [ʔ]	0.395	0.559	0.707	0.479	
Variant [ʔ͡t]	0.113	0.486	0.234	0.815	
Face gender (Male)	-1.103	0.312	-3.527	<.001	***
Variant [ʔ]:face gender (Male)	-1.700	0.521	-3.262	<.01	**
Variant [ʔ͡t]:face gender (Male)	-0.892	0.465	-1.918	0.055	

Table 4.43 Best-fit ordinal regression mixed-effects model with interactions between predictors. Perceptions of speaker age of [ʔ], [ʔ͡t] and [t]. N=486; Listener=27. Tyneside group.

The statistical analysis shows that *phonetic variants* did not have a significant effect. [ʔ] was evaluated as slightly older-sounding as compared to [ʔ͡t] and [t]. None of the results achieved statistical significance, however. Thus, the phonetic variants under investigation did not seem to exert the expected effect. However, *face gender* had a significant effect on evaluations of perceived speaker age. When words were accompanied by a male face listeners were more likely to evaluate the variant as having been produced by a younger speaker, which was the case in the previous examples as well (cf. Tables 4.37, 4.38, 4.40 & 4.41). This effect was statistically significant ($p<0.001$). Finally, testing for interactions between predictors showed that when a male face was shown, listeners were more likely to evaluate [ʔ] and [ʔ͡t] as having been produced by a younger speaker. While the effect for [ʔ] was significant ($p<0.01$), this was not the case for [ʔ͡t].

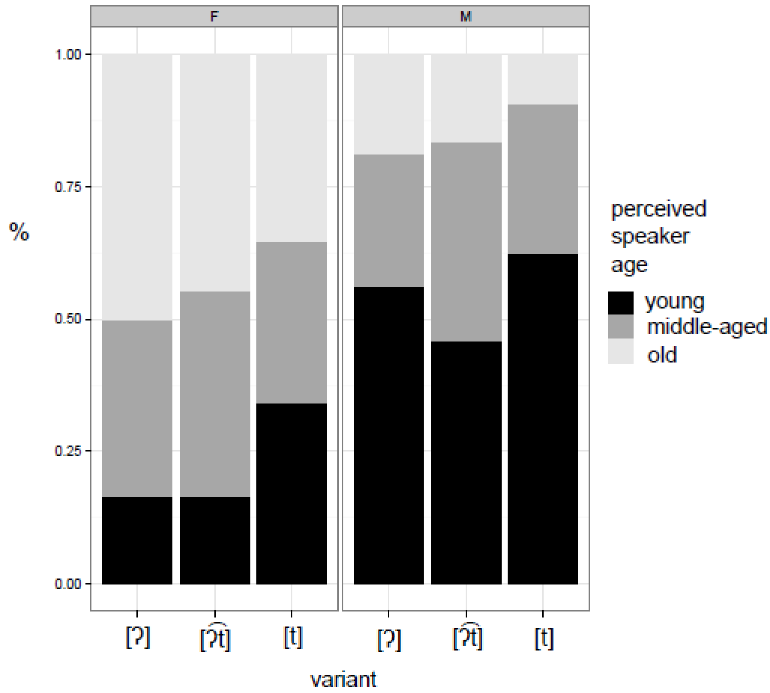


Fig. 4.39 Perception of speaker age for [ʔ], [ʔ͡t] and [t]. Southern group.

In the Southern group variants were more likely to be rated as old- or middle-aged sounding when the supposed speaker was female, but the same variants were rated as young-sounding when the supposed speaker was male (Fig. 4.43). This pattern was consistent across the two groups of respondents as well. It can be also noticed that compared with the Tyneside group of respondents, the spread of middle-aged evaluations in the Southern group of respondents is larger than in the Tyneside group.

The best-fit model to the data included the *face gender* of the supposed speaker and *phonetic variant* as predictors (Table 4.44).

Fixed effects	Estimate	Std. Error	z	Pr(>|z|)	Sig.
Variant [ʔ]	0.679	0.445	1.526	0.127	
Variant [ʔ͡t]	0.751	0.362	2.076	<.05	*
Face gender (Male)	-1.747	0.183	-9.524	<.001	***

Table 4.44 Best-fit ordinal regression mixed-effects model with interactions between predictors. Perceptions of speaker age of [ʔ], [ʔ͡t] and [t]. N=450; Listener=32. Southern group.

Unlike in the Tyneside group of respondents, among the Southern respondents there was a tendency to evaluate [ʔ͡t] as produced by an older speaker ($p<0.05$). However, no significant effect was found for [ʔ]. As expected, the effect of the *gendered face* of the supposed speaker was the same as in the previous cases and was highly significant ($p<0.001$) (cf. Tables 4.37, 4.38, 4.40, 4.41 & 4.43). However, there was no interaction between *variant* and *gender of the face*, which was another difference between the two groups of listeners.

A comparative ordinal regression between the Tyneside and Southern groups did not report any differences as far as the perceived age of the speaker was concerned. The e*xposure* to the dialect did not yield significant results.

4.4 Summary and discussion

When comparing evaluations of speaker social class based on the vowels, that is, the NURSE variants, across the three blocks in both groups of respondents, it can be noticed that the evaluations are very similar if not identical. Despite this fact, explanations for the results found in the two groups are different. Whilst high exposure to the dialect influenced social class evaluations in the Tyneside group of respondents, evaluations provided by the Southern group seem to be driven by standard language ideology (SLI). It seems that upon hearing a localised variant, listeners with lower exposure to the dialect associated such a variant with a working-class speaker, not necessarily a local, Tyneside, one. One of the outcomes of standardisation is reducing the importance of language varieties other than the standard variety (Milroy, 2007: 138). As a result, non-standard varieties, and in the UK context urban non-standard varieties, become stigmatised (Milroy, 2001: 548). Listeners who were brought up in the standard language culture associate non-standard localised variants with lower-class speakers. Even though both Tyneside and Southern listeners were brought up within this ideology, it seems that two different mechanisms could be used to explain the results obtained in the two groups of respondents. As has been mentioned also the Tyneside listeners evaluated the localised variants as having been uttered by a working-class speaker. However, in their case evaluations were driven by indexicality. Social-indexical information carried by the variant was recognised by the listeners as information indexing a distinct social group. In other words, the Tyneside listeners recognised localised phonetic variants under discussion as variants indexing working-class Tyneside speakers, while the southern listeners' perceptions might have been restricted to assessments of class, not localness.

The differences between the groups of respondents and the explanatory mechanisms are more clearly visible in social class evaluations based on consonants. As far as evaluations of consonants are concerned, a pattern was observed which was similar to the one described for vowels. The Southern listeners did not find any perceptual differences between [ɹ] and [͡ʔt] across the three blocks, which would suggest that they associated both variants with working-class speakers. The Tyneside listeners, on the other hand, evaluated [ɹ] as having been produced by a more working-class speaker in comparison to [͡ʔt]. At the same time, the Tyneside listeners did not find any perceptual differences between [ʔ], [͡ʔt] and [t] across the three blocks of social class evaluation. By contrast, Southern listeners evaluated [͡ʔt] as having been produced by a working-class speaker in comparison to [t] but also [ʔ] (block 2). As was the case with vowels also here findings reported by the comparison group could be accounted for by the SLI. The results showed that localised variants which listeners were unfamiliar with ([ɹ] and [͡ʔt]) were associated with a speaker of lower-class and evaluated as having been produced by a working-class speaker. As far as evaluations provided by the Tyneside listeners are concerned, these might have been driven by indexicality. It has been mentioned in the Results section that due to their greater familiarity with the dialect, the Tyneside listeners were sensitive to more subtle indexical information about speaker social class. Whilst variants under investigation were recognised as indexing speakers belonging to the group of Tyneside speakers, the variants were also recognised as indexing speaker's social class along with any differences in speaker social backgrounds as signalled by particular variants. What this means is that speaker-indexical information was salient to the listeners familiar with the dialect, whereas listeners less familiar with it, except for perceiving the localised variants as having been spoken by a working-class speaker, were not able to access speaker-indexical information at the same level as the Tyneside listeners.

As was the case in Experiment 1, also in Experiment 2 evaluations of social class provided by the comparison group could be indirectly explained by perceptual dialectology whereby localised and non-standard variants are associated with a geographical region. It is possible that the Southern group of respondents associated the localised variants with the North-East of England and through this association further evaluated practically all localised variants as having been produced by a working-class speaker without differentiating between the variants, as was the case with [ɹ] and [ʔ͡t].

The differences between the main and comparison groups of respondents could be further accounted for by the theory of enregisterment (Agha, 2005). Linguistic forms which differ from standard forms may become linked with context, that is enregistered with meaning, and as a result become interpretable with social information (Johnstone, 2014). "Enregistered dialect is a set of forms that are associated by people in a particular community or area, with a place or group of people" (Johnstone, 2014: 2). Therefore, what is being indexed by an enregistered form is specific to place and community. The same form may mean something in one place but mean nothing or something else in another place. Results found in the two groups of respondents are an illustration of such a case. Listeners familiar with the dialect were aware of the differences between non-localised and localised variants as well as differences between localised variants. By contrast, Southern listeners were not aware of the differences between localised variants. Assuming that the Southern group of listeners had some knowledge of Tyneside English, it could be argued that to them localised variants were not enregistered at the same level as they were to Tyneside listeners. It is also possible that listeners in the comparison group did not have knowledge of Tyneside English and to them localised variants were not enregistered with meaning at all.

As was the case in Experiment 1, even more so here it could be argued that listeners recognise speaker indexical information if they identify with the speakers of the dialect as well as the place where the dialect is spoken. Again, differences between the two groups of respondents were subtle and could be observed in evaluations of speaker social class.

In the present experiment the two experimental groups varied more in terms of their exposure to Tyneside English than experimental groups in Experiment 1. However, as has been mentioned, perceptual differences between the two groups of participants were inconsistent. Whilst there were some differences between the groups in terms of social class evaluations for the consonants, there were no statistically significant differences between the groups for the NURSE variants. Moreover, age evaluations of consonants showed almost no differences between the groups. These inconsistencies and lack of a regular pattern show that exposure to the dialect was not a key factor influencing participants' responses. Therefore, exemplar theory cannot be used to justify the results.

Evaluations of speaker age in block 3 in Section 4.3.1.3 as well as in Section 4.3.2 show that face gender of the supposed speaker had a statistically significant influence on perception of speaker age in both groups of respondents. Seeing a male face made listeners perceive the variant as having been produced by a younger speaker, whereas seeing a female face made the listeners perceive the variant as having been produced by an older speaker. It will be recalled that in Experiment 1, in which listeners were exposed only to audio, evaluations of speaker age were often in the middle of the scale. This suggested that the listeners were not certain as to how to rate the age of the speaker. In Experiment 2, where listeners were exposed to audio but also to visual information about the supposed speaker, the results were quite different and the variants were evaluated as having been uttered by a younger or older speaker depending on the gender of the face shown in the picture. This may mean that gendered face of the supposed speaker overrode the ambiguity in voice as regards sex. Thus the results found in the two experiments provide further evidence that unimodal and multimodal speech perception may produce different results.

The results also provide evidence that it is not only audio information which the listener tunes into during the process of speech perception. Whilst the unimodal approach to speech perception is the traditional way of investigating the phenomenon, it does not reflect the natural process of speech perception, during which the listener relies on audio as well as visual cues. Therefore, the results of Experiment 2 improve our understanding of how a specific type of speaker-indexical information could be affected or shifted by a visual cue showing certain information about the speaker.

4.5 Experiment 2 Conclusion

Overall, this experiment produced a number of interesting and important findings. The main finding is that the gendered face of the supposed speaker did not impact evaluations of speaker social class. However, the gendered face of the supposed speaker had influence on age ratings.

As far as evaluations of perceived speaker age are concerned, it can be concluded that the present findings confirm the results of Experiment 1, in which voices evaluated as younger-sounding were frequently found to be also more male-sounding, while voices rated as more female-sounding were often recognised as older-sounding. In the present experiment, the effect of *face gender* of the supposed speaker was found, whereby priming listeners with a male face skewed perceptions of speaker age towards being younger-sounding. Priming with a female face resulted in the opposite outcome, with the speaker being rated as older-sounding. The effect was consistent across both groups of participants. Yet, as has been pointed out, listeners familiar with the dialect seemed to access speaker-indexical information to a certain degree, despite the effect of *face gender*.

Evaluations of perceived speaker social class show that certain localised variants seemed to encode social-indexical information to a more consistent degree than others. When provided with information about speaker social background (block 1), listeners' evaluations matched the results of production studies, for the most part. However, adding information about speaker gender to information about speaker social class (block 2) gave mixed results. In some cases, as with the variants of the NURSE vowel or glottalled, glottalised and released /t/, no statistical differences emerged. One of the main findings was that the effect of the gendered face of the speaker was not significant in terms of evaluations of perceived speaker social class. Finally, adding information about speaker age to the puzzle (block 3) also failed to exert an influence on perceptions of speaker social class. Thus, it seems that as far as evaluations of speaker social class were concerned, visual information about the speaker did not override information from the acoustic signal.

Further questions arose from the results of this experiment. For example, if some of the speaker-indexical information is not affected by exposing the listener to visual cues about the apparent speaker, would the effect be different if, for example, the apparent speaker was non-White? The following experiment attempts to address this question and investigates perceptions of speaker age and gender when exposing the listener to visual cues about the ethnicity of the speaker.

5 Experiment 3

5.1 Introduction

The two previous experiments in this thesis investigated perceptions of speaker gender, age and social class on the basis of phonetic variants. The second experiment introduced visual cues to age, gender and social class of the apparent speaker. The results showed that some speaker-indexical information was not shifted as a result of exposure to the visual cues. However, the question arose whether a similar effect would be observed if the apparent speaker ethnicity was non-White. Therefore, in the third, and final experiment a new socio-indexical factor is introduced which has not been investigated so far, that is the apparent ethnicity of the speaker. The aim of this experiment is to investigate what effect the apparent ethnicity has on judgements of speaker-indexical information. In other words, how perception of speaker gender is affected when apparent ethnicity changes.

By testing ethnicity this experiment also attempts to give us more information about how speaker-indexical information is interlinked when memory representations are formed; if a phonetic variant is associated, for example, with older women, is it in fact associated with older women or older White women? By priming listeners with a face of the supposed speaker this experiment investigated how using an "ethnic" face instead of an "Anglo" face would weaken the strength of memory representations. It also investigated how listeners' prior expectations about hearing a variant affected their gender-ethnicity choices.

It was hypothesised that listeners with high exposure to localised phonetic variants would project their memory representations of variants associated with local White male and female speakers onto local male and female speakers of apparent ethnic background. In other words, it was investigated how inseparable speaker social-indexical information such as age, gender and ethnicity are in terms of memory representations listeners store in their long-term memory and whether there is a transfer of associations onto a supposed speaker of different ethnicity. It was tested whether a variant associated, for example, with a White female speaker was also associated with a female speaker when the apparent ethnicity was other than White, or if a variant associated with an older White male speaker was also associated with an older male speaker when the apparent ethnicity was other than White.

As far as the two previous experiments are concerned, theoretical justification for conducting them was based on previous research into perception and identification of speaker-indexical information (Hay et al., 2006a; Hay & Drager, 2010; Foulkes et al., 2010; Niedzielski, 1999). This experiment links to the study of perceived speaker ethnicity by Purnell et al. (1999) but investigates it from a different perspective.

5.2 Method

5.2.1 Ethnicities

Three ethnicities were investigated in this study: White, Asian[3] and Black. The results for the two ethnic minorities were compared and contrasted with results obtained for the White Anglo ethnicity. The choice of the two ethnic minorities was based on the 2011 Census data provided by the Office for National Statistics (ONS, 2011). It should be pointed out that Tyneside does not belong to areas with high numbers of ethnic minority inhabitants in comparison to other parts of England. Overall, the North-East of England has a smaller ethnic minority population than elsewhere in the country. For example, in 2011 the Asian and Black ethnic minority population in Newcastle upon Tyne, Gateshead, South and North Tyneside was 3.6 % and 2.3 % in all of the North-East. By comparison, in the West Midlands it was 12.2 %, and 25.3 % in Greater London (ONS, 2011). In the same year 13% of new-born babies in Newcastle but only 5% in all of the North-East were born to non-citizens (ONS, 2011). Nevertheless, there is a steady tendency for the ethnic population to grow in comparison to the White population.

As was the case in Experiment 2, in the present experiment only black and white images of the face of the supposed speaker were shown. Even though, more ethnic minorities inhabit Tyneside than the two chosen for the study, it would be difficult to show

[3] Asian in the UK context refers to people originally from South Asia, that is, Pakistan, India, Bangladesh and Sri Lanka.

differences between, for example, Indian, Pakistani, Bangladeshi (Urdu or Sikh) and other South Asian individuals. The same problem was faced in the case of Black African, Black Caribbean and other Black individuals. Whilst the choice of two ethnicities is a simplification, the distinctions between them are clearly defined. Thus, in the present experiment the British Black ethnicity includes Black, African, Caribbean and Black British. As far as Asian ethnicity is concerned, it includes Indian, Pakistani and Bangladeshi. Similarly, White British population includes people of English, Irish, Scottish or European origin, etc.

The table below (Table 5.1) presents a comparison of the three ethnicities in percentages.

Area	White	Asian	Black
England	85.31%	5.55%	3.48%
Tyneside	92.62%	6%	1.94%

Table 5.1 A comparison of White, Asian and Black ethnicities in England and Tyneside. Census data (ONS, 2011).

5.2.2 Images

As was the case in Experiment 2, Experiment 3 was also a priming experiment which consisted of two separate tasks. Listeners were asked to evaluate the perceived gender and age of the speaker. This time, however, listeners were primed with images showing faces of men and women of three ethnicities: White, Asian and Black. The reason for showing images was to make the listeners see the images as the speaker. Thus, the images will be also referred to as the "supposed speaker". Two independent sets of images were applied in this study to evaluate perceived speaker gender and age. For the sake of consistency with Experiment 2, picture representations of the supposed speaker were created using EvoFIT (Frowd et al., 2006; Frowd, 2015). As has been mentioned, the pictures were in black and white. No clothing, jewellery, piercing, traditional or religious body ornamentation was shown. Three sets of male and female faces of White, Asian and Black ethnicities were morphed to categorise the gender of the speaker. In addition, three sets of male and female faces of the three ethnicities were created to evaluate the age of the speaker. Young, middle-aged and older faces were morphed in the same manner as in the previous experiment. Male and female faces of ethnic origin used for gender categorisation were morphed in such a way that they looked middle-aged. The reason for presenting listeners with a face of the supposed speaker looking neither very young nor very old was to limit the effect of interaction between perceived age and gender of the speaker voice resulting from the manipulation applied to audio stimuli (cf. Section 3.2.1).

Fig. 5.1 A male and female face of White ethnicity.

Fig. 5.2 A male and female face of Asian ethnicity.

Fig. 5.3 A male and female face of Black ethnicity.

Fig. 5.4 A set of White female faces for speaker age rating.

Fig. 5.5 A set of Asian female faces for speaker age rating.

Fig. 5.6 A set of Black female faces for speaker age rating.

Fig. 5.7 A set of White male faces for speaker age rating.

Fig. 5.8 A set of Asian male faces for speaker age rating.

Fig. 5.9 A set of Black male faces for speaker age rating.

5.2.3 Procedure

The same audio stimuli as in Experiment 1 and 2 were used in Experiment 3. In the present experiment perceptions of the NURSE vowel and T-to-R were investigated. Stimuli were administered in an identical manner to that used in the two previous experiments. Pictures of ethnic faces appeared on screen and the audio stimulus was played with a delay of a second. Upon hearing a stimulus, listeners were instructed to select one of the images showing the face they thought was that of the speaker. In this experiment speaker gender and age were evaluated. If a male and female speaker were shown on the screen, listeners were asked to select one of the two images. If, on the other hand, three male or three female faces of different apparent ages were presented, listeners were asked to select one of the three images. Faces shown with each of the stimuli were of the same ethnicity. Figure 5.10 illustrates the experimental procedure. As was the case in Experiment 2, also in the present experiment in the case of the retracted NURSE variant [ɔː] a picture showing the word being played was presented. Images were included to ensure the listener heard the variant as intended by the researcher. With the exception of variant [ɔː], no other images of audio stimuli were shown (cf. Section 3.2.3).

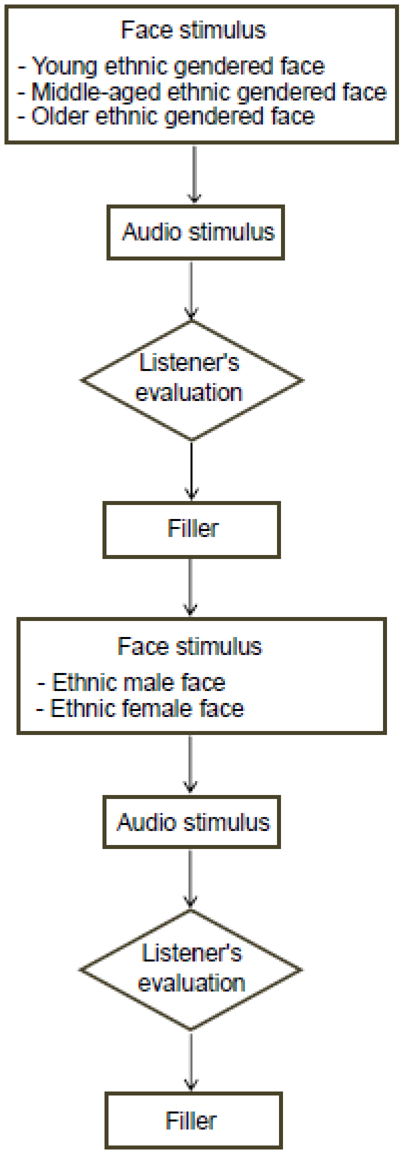


Fig. 5.10 Experimental procedure. The top of the flowchart illustrates the process of age evaluation of the speaker; the bottom of the chart shows the process of gender evaluation of the speaker.

As in the previous experiments the present experiment was conducted in laboratory conditions and administered in SurveyGizmo (SurveyGizmo, 2015). The experimental procedure remained the same as well. At the beginning of the experiment, all ethnic faces were shown to the participant. It was decided to introduce this step because images of the faces were black and white, which in turn could lead, for example, to misidentifying some of the Asian faces as, for example, Native American or Black. Next, the training session followed. A total of 270 single-word stimuli and fillers were presented over headphones at a comfortable hearing level, one at a time. Listeners heard each stimulus only once. The entire session took about 40 minutes, including three breaks during the experiment.

As was the case in Experiment 2, at the end of the present experiment participants were shown all the ethnic faces used for age rating and gender categorisation and were asked to estimate the age of each face. Although faces were morphed using the same algorithm and were evaluated during the process it was important to establish how listeners perceived the faces and whether corresponding ages across the three ethnicities were perceived similarly. Results of the evaluations are presented in Table 5.10.

5.2.4 Participants

Listeners were recruited in a different manner from that used in the two previous experiments. They came from a variety of backgrounds. While some participants were staff and students at Newcastle University, the majority of participants were approached on the streets of Newcastle. As in Experiment 1 and 2, here Tyneside listeners were in the group with high exposure to the dialect. However, in the present experiment there was no comparison group and the results and discussion focus on the Tyneside listeners. As before, it was ensured that participants had not participated in any of the previous studies presented in the thesis and were new to the research. However, in terms of participant age there was more variation than in the previous experiments. 31 volunteers, aged 18-55, participated in the study. 14 listeners were female and 17 listeners were male. As far as the social background of participants was concerned, 7 participants were middle-class and the rest declared themselves to be working-class.

As has been mentioned, participants belonged to a wide variety of professions. None of the participants reported having a hearing impairment or suffering from a cold.

Given that the present experiment investigated the ethnicity of the speaker, it might be an interesting idea to compare responses provided by listeners from various ethnic backgrounds. Nevertheless, it was not possible to do so in the present study. With the exception of one participant who was British Asian, all other participants were White British.

Participants were paid £7 upon completing the experiment.

5.3 Results

The section presents the results of the perceptual experiment. Variants of the NURSE vowel as well as the T-to-R variable (cf. Sections 2.6 & 3.4) are discussed in terms of perceived speaker gender and age when the displayed facial image was British White, Asian or Black.

As was the case in Experiment 2 (cf. Sections 3.3 & 4.3), in the present experiment mixed-effects logistic and ordinal regression tests were applied. The former test was used to investigate *the perceived gender of the speaker* and the latter to examine *the perceived age of the speaker*.

Fixed effects in logistic regression were the following:

- The phonetic variant (number of variants depended on the variable)
- Speaker (A or B)
- Listener (participant)
- The gender of the listener (male or female)
- The social class of the listener (working-class or middle-class)

In the ordinal regression, additionally the *gender of the face* shown in the image was entered as a fixed effect and an interaction between *variant* and *face gender* was tested.

In addition, a series of logistic and ordinal regression mixed-effects tests attempting to find statistical differences between the three ethnicities shown in images were performed. In these tests an extra fixed effect was entered, which was *ethnicity*.

In the first part of the Results section perceptions of the gender are presented and followed by perceptions of the age in the second part of the section. In both cases the acoustic stimuli were accompanied by images showing the supposed ethnic speaker.

5.3.1 Evaluations of perceived speaker gender

The following section presents results for evaluations of speaker gender when listeners were presented with images showing faces of three ethnicities. Results for the variants of the NURSE vowel are followed by results for T-to-R.

5.3.1.1 The NURSE vowel

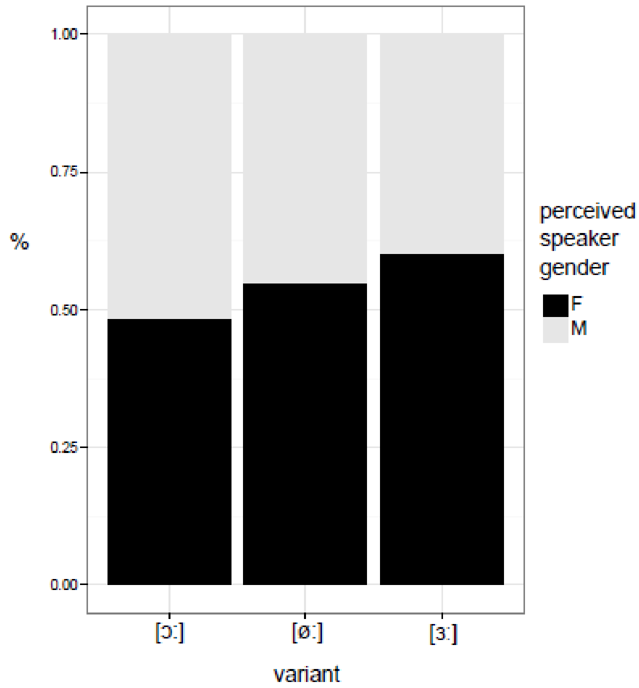


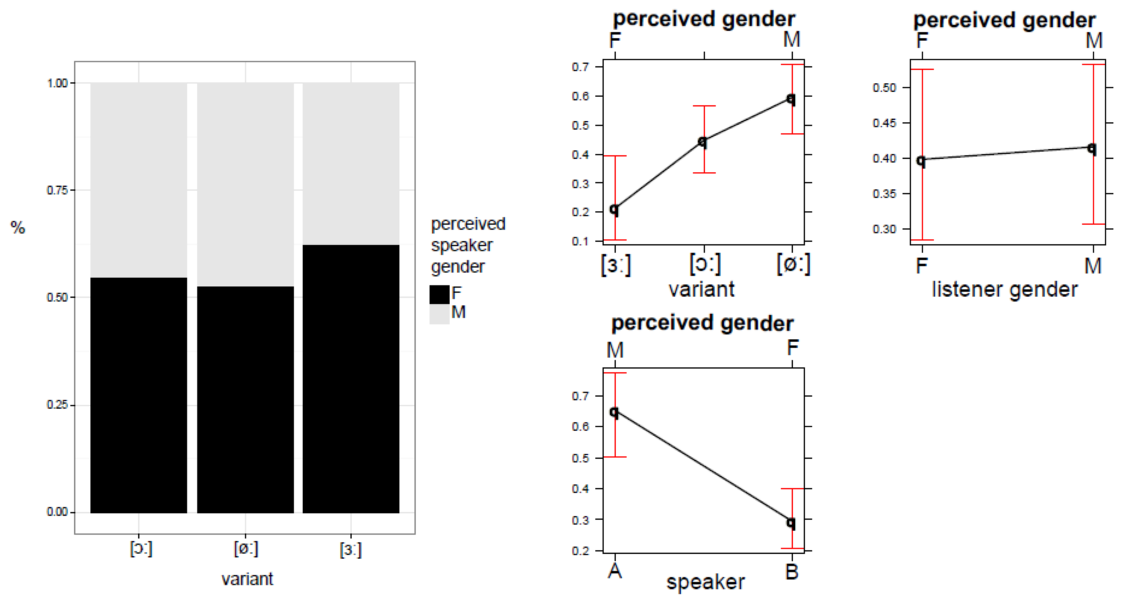
Fig. 5.11 Perceptions of speaker gender of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel, where the supposed speaker was White British.

Fig. 5.11 presents evaluations of speaker gender for the variants of the NURSE vowel when the supposed speaker was British White. As can be observed, the gender evaluations were at chance level, indicating that none of the variants was perceived by listeners to be used particularly often by male or female speakers. While this result could be expected in the case of the non-localised variant [ɜː], it seems that the localised variants did not cue gender responses either as male or female.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	-0.366	0.558	-0.656	0.511	
Variant [ɔː]	1.131	0.642	1.760	<.1	
Variant [øː]	1.123	0.740	1.518	0.129	
Speaker B	-1.036	0.576	-1.799	<.1	

Table 5.2 Best-fit logistic regression mixed-effects model of the perceived gender of the supposed British White speaker for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. Random effects were random slopes (1 + variant | listener) and random intercepts (1 | audio). Number of observations (N)=279; Listener=31.

When evaluating speaker gender when the ethnic face of the supposed speaker was British White, the best-fit logistic regression mixed effects model to the data included *variant* and *speaker* as fixed effects. As can be noticed from Table 5.2 perceptions of [ɔː] and [øː] did not differ from [ɜː] in terms or perceived gender. There was a tendency for [ɔː] and [øː] to be perceived as having been spoken by a male speaker. Nevertheless, the evaluations were at chance level and neither gender, that is neither female nor male gender were in the lead. While the results were statistically non-significant, such results might suggest that the manipulation applied to the tokens worked to a certain extent. Even though listeners were exposed to stimuli produced by male speakers whose fundamental frequency was shifted to make the words sound gender-ambiguous, gender-correlated phonetic variants triggered gender categorisations to a less meaningful degree than expected. As has been mentioned, there was only a tendency for the retracted variant [ɔː] to be perceived as having been spoken by a male speaker. As far as the fronted variant [øː] was concerned, it was following in the footsteps of [ɔː] and was overall perceived as having been produced by a male speaker. There was also a tendency for words produced by speaker B to be evaluated as having been uttered by a female speaker. However, this tendency was non-significant.



a) **b)**

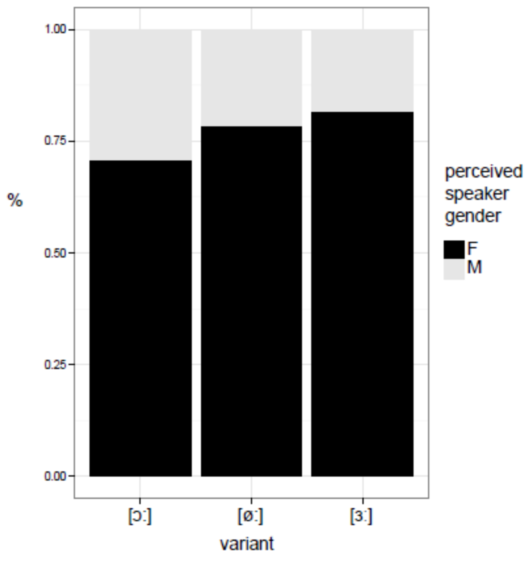
Fig. 5.12 a) & b) Perceptions of speaker gender of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel, where the supposed speaker was British Asian.

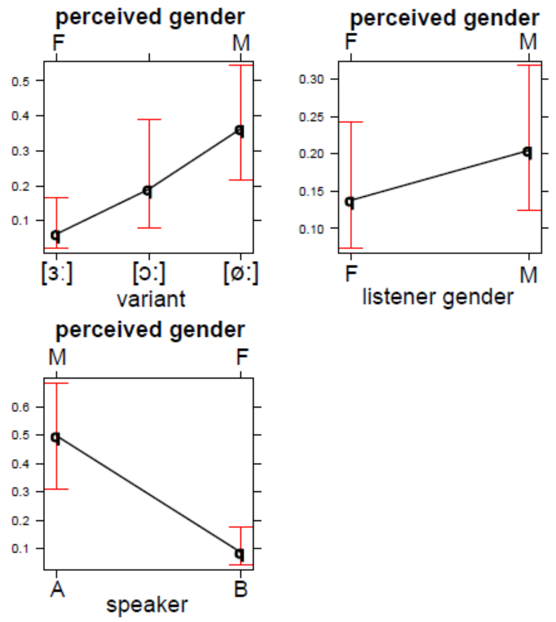
Figures 5.12 (a) and (b) and Table 5.3 show evaluations of speaker gender when the ethnic face was Asian. Whilst the bar graph (Fig. 5.12 (a)) shows a raw count of the speaker gender evaluations, the other graph (Fig. 5.12 (b)) shows all effects occurring in the data as well as predicted probabilities for gender with error bars which give some idea of the uncertainty of the estimate. When looking at Fig. 5.12 (a) it seems that no statistically significant results could be expected. However, Table 5.3 shows that it is not the case as differences between the variants were found. This discrepancy most probably results from the *speaker* effect. It can be observed that variant [ɔː] was perceived as having been produced by a male speaker when the supposed speaker was British Asian ($p<0.05$). The effect was even stronger for variant [øː] ($p<0.01$). These results stand in opposition to results found when the supposed speaker was White, where none of the localised variants had a significant effect (cf. Table 5.2). As already mentioned, there was also an effect of the *speaker* variable as words produced by speaker B were more likely to be evaluated as having been uttered by a female speaker ($p<0.001$). The *gender* of the participant was not significant, yet it contributed to the model.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	-0.339	0.451	-0.750	0.453	
Variant [ɔː]	1.089	0.492	2.214	<.05	*
Variant [øː]	1.687	0.518	3.257	<.01	**
Listener gender (Male)	0.073	0.327	0.224	0.822	
Speaker B	-1.501	0.396	-3.787	<.001	***

Table 5.3 Best-fit logistic regression mixed effects model of perceived gender of the supposed British Asian speaker for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=279; Listener=31.

Figures 5.13 (a) and 5.13 (b) and Table 5.4 show the results of the evaluation of gender when the supposed speaker was British Black. Fig. 5.13 (b) is shown for the same reasons as previously. Also here some discrepancy between the results shown in Fig. 5.13 (a) and statistical analysis can be observed. As before, this discrepancy is a result of the *speaker* effect which was not taken into account in Fig. 5.13 (a). The best-fit model of the data included *variant, listener gender* and *speaker* as fixed effects. As can be noticed the localised variants [ɔː] and [øː] were more likely to be perceived to have been produced by a male speaker. However, the effect was significant only for the fronted variant [øː] ($p<0.01$). It seems that the effect of the *phonetic variant* when the supposed speaker was British Black was the opposite of the findings reported in the case when the supposed speaker was British White (cf. Table 5.2) but similar to the findings reported when the supposed speaker was British Asian (cf. Table 5.3). As can be recalled, the overall tendency to view the localised variants as having been spoken by a male speaker was confirmed. Nevertheless, in the present case the fronted variant [øː] was perceived as definitely having been spoken by a male speaker ($p<0.01$), whereas in the case of the supposed British White speaker there was no significant result. Even though no effect of the *listener gender* was found, it contributed to the model. There was also an effect of the *speaker*. Words produced by speaker B were more likely to be perceived as having been uttered by a female speaker ($p<0.001$).





a) b)

Fig. 5.13 a) & b) Perceptions of speaker gender of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel, where the supposed speaker was British Black.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	-1.408	0.466	-3.019	<.01	**
Variant [ɔː]	1.261	0.658	1.914	<.1	
Variant [øː]	2.148	0.782	2.746	<.01	**
Listener gender (Male)	0.475	0.377	1.259	0.208	
Speaker B	-2.343	0.649	-3.608	<.001	***

Table 5.4 Best-fit logistic regression mixed effects model of perceived gender of the supposed British Black speaker for variants of the NURSE vowel, where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=279; Listener=31.

As the last step of analysis of the NURSE vowel variants, the three ethnicities were compared in Fig. 5.14, which shows that when the apparent ethnicity was White or Asian the results were at chance level in terms of the perceived speaker gender. However, when the apparent ethnicity was Black, there was a tendency to evaluate the variants as more likely to be have been produced by a female speaker.

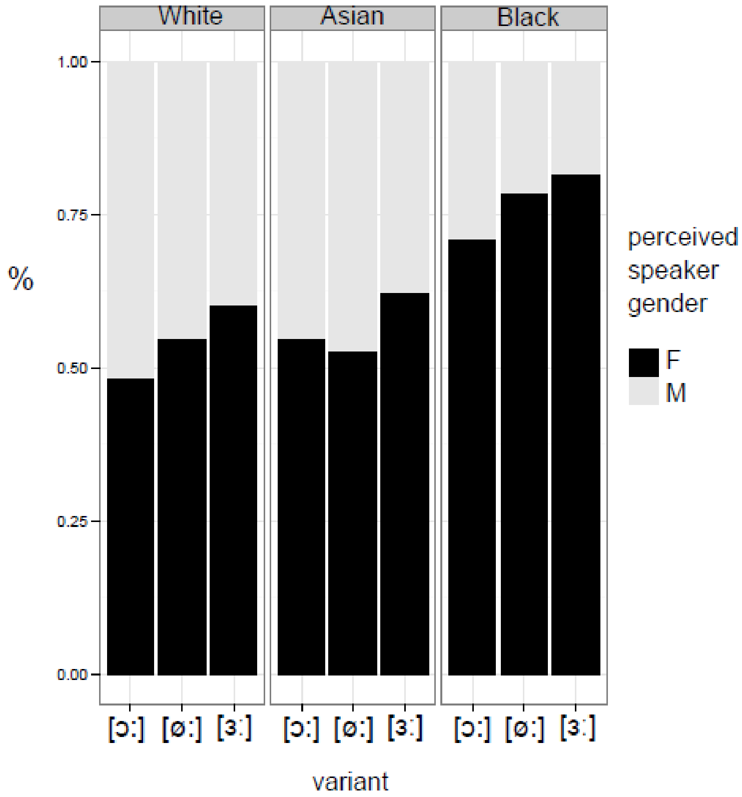


Fig. 5.14 Perceptions of speaker gender of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel for all three ethnicities.

The results for all three ethnicities were also compared in a single logistic regression test to state whether any of the ethnicities of the supposed speaker had a significant effect on categorisations of speaker gender. Even though there was a tendency for the NURSE variants to be evaluated as having been produced by a female speaker when the ethnicity of the supposed speaker was Black or Asian, the results presented in Table 5.5 indicate that none of the three ethnicities (White, Asian, or Black) had a significant effect on gender categorisations. At the same time, there was a non-significant tendency to perceive both localised variants [ɔː] and [øː], as having been uttered by a male speaker.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	22.405	131.727	0.170	0.865	
Variant [ɔː]	0.455	0.880	0.517	0.605	
Variant [øː]	0.386	0.839	0.461	0.645	
Ethnicity Asian	-0.096	152.930	-0.001	0.999	
Ethnicity Black	-24.214	131.728	-0.184	0.854	

Table 5.5 Best-fit logistic regression mixed effects model of perceived gender of the supposed speaker of the three ethnicity variants of the NURSE vowel. Where [ɔː] and [øː] are localised variants compared against the non-localised [ɜː]. N=837; Listener=31.

5.3.1.2 T-to-R

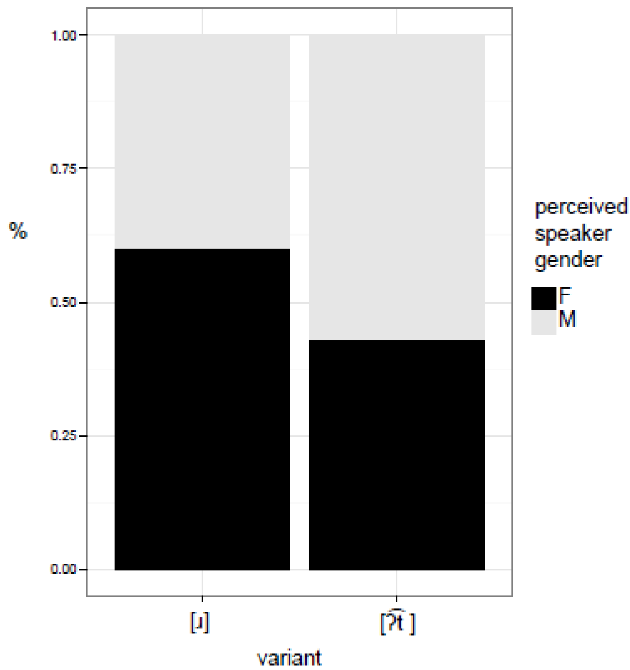


Fig. 5.15 Perception of speaker gender for [ɹ] and [ʔ͡t], where the supposed speaker was British White.

Fig. 5.15 presents evaluations of perceived speaker gender in the case of T-to-R when the supposed speaker was British White. There was a slight tendency to evaluate [ɹ] as having been spoken by a female speaker and to perceive [ʔ͡t] as having been produced by a male speaker. Nevertheless, as can be seen in Table 5.6, these trends in evaluations of speaker gender were not significant in statistical terms. In this sense the results were similar to the findings reported for the NURSE variants when the supposed speaker was British White. At the same time, a non-significant tendency among male listeners to perceive [ɹ] as having been produced by a male speaker was found.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	-0.102	0.541	-0.189	0.85	
Variant [ɹ]	-0.931	0.628	-1.484	0.137	
Listener gender (Male)	0.848	0.513	1.653	<.1	

Table 5.6 Best-fit logistic regression mixed effects model of perceived gender of the supposed British White speaker for [ɹ] and [ʔ͡t]. N=186; Listener=31.

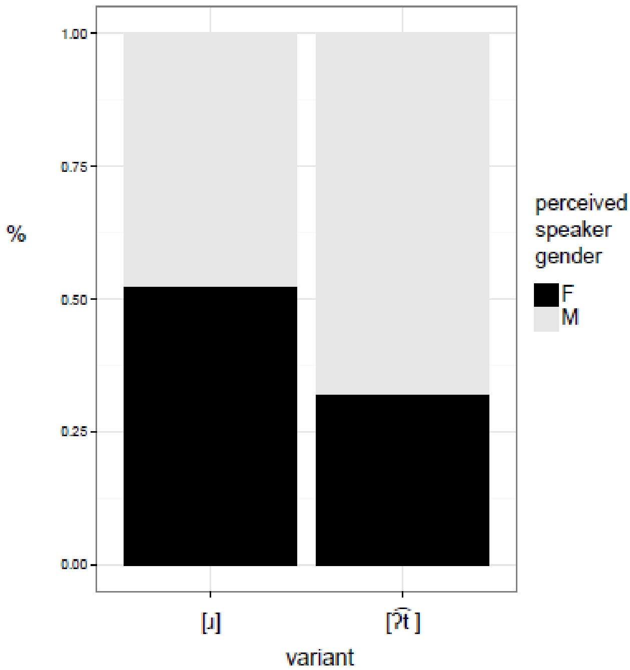


Fig. 5.16 Perception of speaker gender for [ɹ] and [ʔ͡t], where the supposed speaker was British Asian.

Figure 5.16 shows that perceptions of [ɹ] and [ʔ͡t] when the supposed speaker was Asian exhibit a similar pattern to the one observed when the supposed speaker was White. While [ʔ͡t] was evaluated as having been produced by a male speaker, [ɹ] was characterised by an even split of male and female evaluations. As can be seen in Table 5.7, evaluations of the perceived gender when the displayed facial image was that of a British Asian, present similar results to findings reported when the facial image was British White (cf. Table 5.6). A non-significant tendency towards perceiving [ɹ] in comparison to [ʔ͡t] as having been uttered by a female speaker was observed.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	1.080	0.531	2.032		*
Variant [ɹ]	-1.234	0.637	-1.936	<.1	

Table 5.7 Best-fit logistic regression mixed effects model of perceived gender of the supposed British Asian speaker for [ɹ] and [ʔ͡t]. N=186; Listener=31.

Figure 5.17 shows evaluations of perceived speaker gender when the supposed speaker was British Black. In this condition there was a tendency to perceive [ɹ] as having been produced by a female speaker whilst [ʔ͡t] was not perceived as having been uttered by a female or male speaker in particular. The statistical results support these observations (Table 5.8). Overall, the statistical results were similar to the ones reported in the White and Asian conditions.

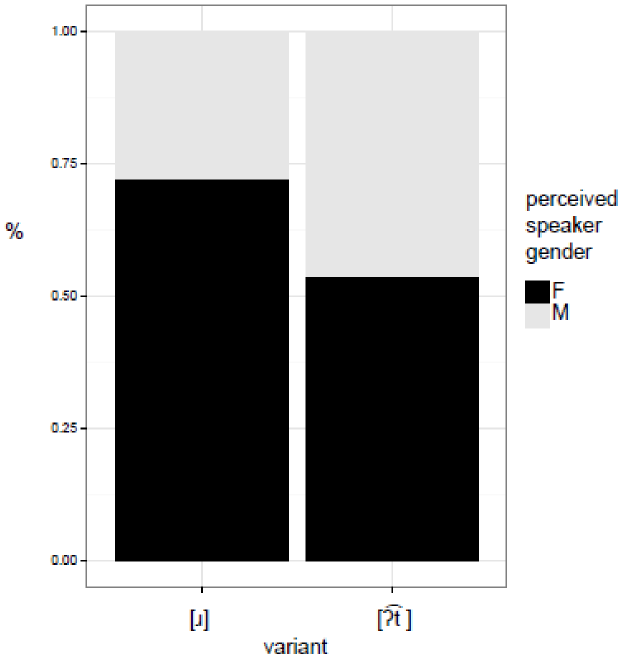


Fig. 5.17 Perception of speaker gender for [ɹ] and [ʔ͡t], where the supposed speaker was British Black.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	-0.164	0.333	-0.492	0.623	
Variant [ɹ]	-1.128	0.583	-1.935	<.1	

Table 5.8 Best-fit logistic regression mixed effects model of perceived gender of the supposed British Black speaker for [ɹ] and [ʔ͡t]. N=186; Listener=31.

A comparison of the evaluations of the perceived gender for [ɹ] and [ʔ͡t] when the ethnicity of the supposed speaker was White, Asian and Black is presented in Fig. 5.18 and Table 5.9. As can be observed, [ɹ] was perceived to have been spoken by a female speaker ($p<0.05$). With respect to the effect of the ethnicity of the supposed speaker, none was observed. Nevertheless, there was a tendency to evaluate variant [ɹ] as having been spoken by a female speaker when the apparent ethnicity was Black and as having been produced by a male speaker when the apparent ethnicity was Asian (Fig. 5.18 & Table 5.9). There was also a non-significant tendency to perceive the variants as having been uttered by a male speaker when the listener was male.

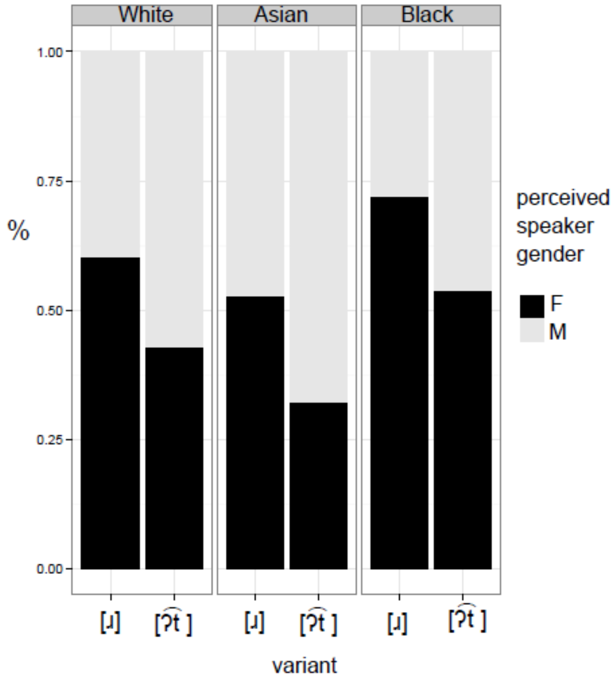


Fig. 5.18 Perception of speaker gender for [ɹ] and [ʔ͡t] for all three ethnicities.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
(Intercept)	21.228	3060.532	0.007	0.994	
Variant [ɹ]	-1.076	0.541	-1.988	<.05	*
Ethnicity Asian	0.018	4348.795	0.000	1.000	
Ethnicity Black	-22.321	3060.532	-0.007	0.994	
Listener gender (Male)	0.671	0.361	1.859	<.1	
Listener class (WC)	0.712	0.444	1.600	0.109	

Table 5.9 Best-fit logistic regression mixed effects model of perceived gender of the supposed speaker of the three ethnicities for [ɹ] and [ʔ͡t]. N=558; Listener=31.

5.3.2 Evaluations of perceived speaker age

At the end of the experiment listeners were asked to evaluate the perceived age of the ethnic faces used in the experiment. The results are provided in the table below.

Ethnicity	Ranges of evaluations of face age	Median evaluations of face age
Young White female face	17-30	22
Middle-aged White female face	29-45	36
Older White female face	39-65	52
Young Indian female face	17-35	24
Middle-aged Indian female face	26-48	31
Older Indian female face	31-65	49
Young Black female face	16-28	20
Middle-aged Black female face	22-45	35
Older Black female face	29-65	45.5
Young White male face	17-27	20
Middle-aged White male face	27-50	38
Older White male face	43-65	54

Young Indian male face	18-26	22
Middle-aged Indian male face	29-45	35
Older Indian male face	39-59	50
Young Black male face	15-30	21
Middle-aged Black male face	23-48	38
Older Black male face	38-65	51
White female face (gender evaluation)	29-55	38
White male face (gender evaluation)	29-55	40
Indian female face (gender evaluation)	28-55	42
Indian male face (gender evaluation)	30-50	40
Black female face (gender evaluation)	21-49	40
Black male face (gender evaluation)	22-60	40

Table 5.10 Ranges and median evaluations of age of the male and female ethnic faces morphed in EvoFIT. Ratings provided by the participants.

The following section focuses on evaluations of perceived age of the speaker. Evaluations of the NURSE variants for the three ethnicities are presented and followed by evaluations of T-to-R for the three ethnicities. The results of the perceived age of the speaker across the three ethnicities show that, as expected, listeners were sensitive to age-gender interaction resulting from manipulation of the stimuli (cf. Section 3.2.1). This interaction was already found in Experiments 1 and 2. As was the case in Experiment 2, in the present experiment having a male face accompany the words played to listeners resulted in a decrease in the estimated age of the speaker.

5.3.2.1 The NURSE vowel

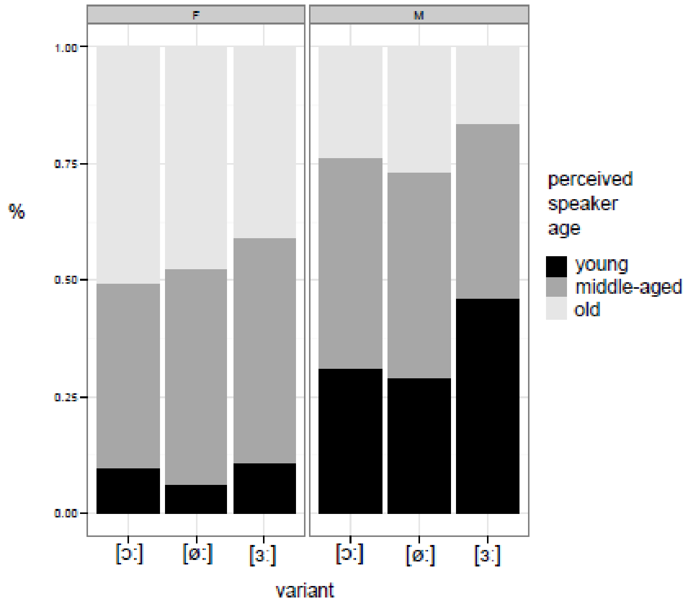


Fig. 5.19 Perceptions of speaker age of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel, where the supposed speaker was British White.

It seems that variants of the NURSE vowel were perceived as having been produced by a middle-aged speaker when words were accompanied by a male face of the supposed British White speaker (Fig. 5.19). In addition, the interaction between the gender and age of the speaker was observed, especially in the case of the non-localised [ɜː], which was found to be mostly young-sounding. The two localised variants [ɔː], [øː] were perceived as overall produced by a middle-aged speaker. As far as evaluations of the perceived age of the speaker when the displayed facial image was that of a British White female are concerned, overall the localised variants were evaluated as having been produced by an older speaker. These results also indicate that in the female condition the age-gender interaction influenced listeners' perceptions of the phonetic variants under investigation (cf. Section 3.2.1). While the two localised variants were overall perceived as having been spoken by an older female speaker, the non-localised variant was evaluated as overall middle-aged sounding.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.5071	0.4469	1.135	0.256	
Variant [øː]	0.5384	0.4300	1.252	0.211	
Face gender (Male)	-1.4878	0.1778	-8.367	<.001	***

Table 5.11 Best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against the non-localised [ɜː] and the supposed speaker was British White. N=558; Listener=31.

In the best-fit model to the data, *variant* and *gender* of the face of the supposed speaker were entered as fixed effects. It can be noticed that there was a slight tendency to perceive the two localised NURSE variants, [ɔː] and [øː], as having been uttered by an older speaker by comparison with the non-localised variant [ɜː] (Table 5.11). However, the tendency was not significant. Furthermore, an effect of gender of the face of the supposed speaker was found.

In order to minimise the effect of age-gender interaction resulting from manipulation applied to phonetic stimuli, statistical analysis of perceived age of the speaker was also performed separately for each of the genders (cf. Section 3.2.1). This enabled investigating if there was an effect of the variant within each gender.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.392	0.552	0.710	0.478	
Variant [øː]	0.310	0.529	0.586	0.558	

Table 5.12 Best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was a British White female. N=279; Listener=31.

The best-fit model to the data in the female, as well as the male condition when the speaker was British White was the simplest model, and had *variant* as a fixed effect. The results

(Tables 5.12 & 5.13) show that there was a tendency to perceive the two localised variants as having been produced by an older speaker. This tendency was overall lower when the supposed speaker was a British White female than when the supposed speaker was a British White male. Nevertheless, the results in the male and female condition were not statistically significant, which was also the case in the main analysis (cf. Table 5.11).

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.699	0.481	1.453	0.146	
Variant [øː]	0.848	0.471	1.801	<.1	

Table 5.13 Best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was a British White male. N=279; Listener=31.

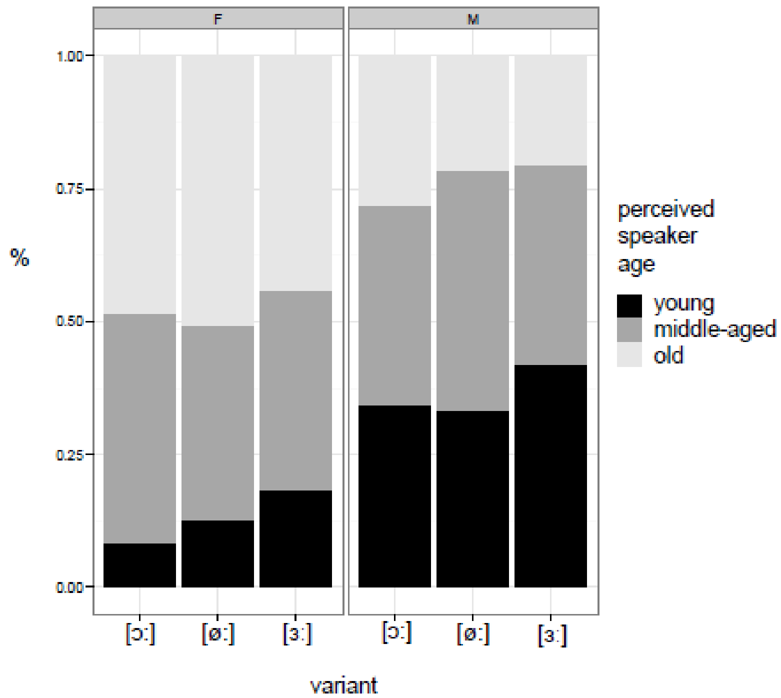


Fig. 5.20 Perceptions of speaker age of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel, where the supposed speaker was British Asian.

Figure 5.20 presents evaluations of speaker age for the variants of the NURSE vowel when the gendered face of the supposed speaker was British Asian. When the face of the speaker was female, listeners tended to perceive the variants as having been uttered by an older speaker. Otherwise the variants were evaluated as middle-aged sounding. When the face of the speaker was male, the localised variants were perceived as middle-aged sounding and younger sounding. The non-localised variant [ɜː], on the other hand, was evaluated as younger and middle-aged sounding. As can be observed, the age-gender interaction was once more recognised by listeners in this case (cf. Section 3.2.1).

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.411	0.342	1.200	0.230	
Variant [øː]	0.310	0.335	0.925	0.355	
Face gender (Male)	-1.376	0.179	-7.65	<.001	***

Table 5.14 Best-fit ordinal regression mixed effects model showing perceptions of age of the supposed speaker for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was British Asian. N=558; Listener=31.

The best-fit model to the data for evaluations of perceived speaker age for the NURSE variants when the facial image was Asian, included *variant* and *gender* of the face of the supposed speaker as fixed effects (Table 5.14). The results reveal a non-significant tendency of the localised NURSE variants to be perceived as having been produced by an older speaker. The same pattern could be observed for variants of the NURSE vowel when the supposed speaker was of British White ethnicity. While the effect for the *variant* was non-significant, there was a tendency to perceive the retracted variant [ɔː] as only slightly older in comparison to the fronted variant. In addition, as we saw in the case of the supposed British White speaker, in the case of the supposed British Asian speaker there was a significant effect of the *face gender* of the supposed speaker. Upon hearing the words accompanied by a British White male face, listeners were more likely to evaluate the variants as younger-sounding (<0.001).

The results of perceived age of the speaker in a separate analysis for each gender are presented below in Tables 5.15 and 5.16.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.333	0.314	1.060	0.289	
Variant [øː]	0.364	0.347	1.047	0.295	

Table 5.15 The best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was a British Asian female. N=279; Listener=1.

Whilst in the female condition the best-fit model to the data was the simplest model with *variant* as a fixed effect (Table 5.15), in the male condition the model included *variant*, *listener gender* and *speaker* as fixed effects (Table 5.16). As far as the phonetic variants are concerned, in the female condition there was a tendency for all three variants to be perceived as having been produced by an older speaker. In the male condition, on the other hand, the retracted variant [ɔː] was perceived as older-sounding, while the fronted variant [øː] as younger-sounding. However, these tendencies were not statistically significant. Additionally, in the male condition the effect of the *listener gender* ($p<0.05$) and *speaker* ($p<0.05$) were observed. Male respondents were more likely to evaluate the variants as having been uttered by an older speaker. Words produced by speaker B were also more likely to be perceived as having been spoken by an older speaker.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.231	0.459	0.503	0.614	
Variant [øː]	-0.216	0.490	-0.443	0.658	
Listener gender (Male)	0.948	0.460	2.061	<.05	*
Speaker (B)	0.977	0.394	2.479	<.05	*

Table 5.16 Best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was a British Asian male. N=279; Listener=31.

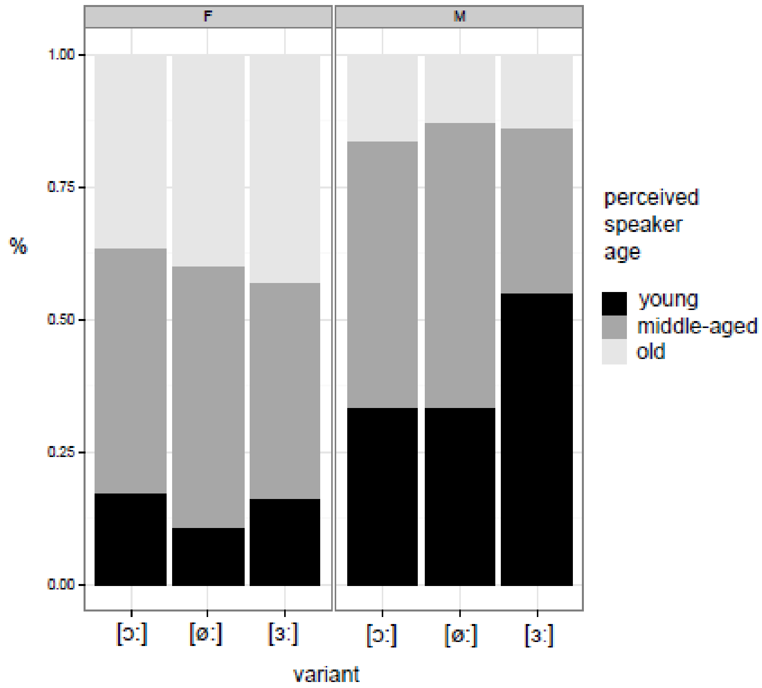


Fig. 5.21 Perceptions of speaker age of the [ɔː], [øː] and [ɜː] variants of the NURSE vowel, where the supposed speaker was Black British.

Figure 5.21 presents results of perceived speaker age for the variants of the NURSE vowel when the supposed speaker was British Black. It seems that localised variants were found to be overall middle-aged sounding in the male as well as female conditions. The non-localised variant [ɜː], on the other hand, was perceived as having been produced by a younger speaker in the male condition and as having been produced by an older and middle-aged speaker in the female condition.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	-0.278	0.351	-0.792	0.428	
Variant [øː]	-0.000	0.299	-0.001	0.998	
Face gender (Male)	-2.280	0.323	-7.041	<.001	***
Variant [ɔː]:face gender (Male)	1.180	0.425	2.774	<.01	**
Variant [øː]:face gender (Male)	0.806	0.422	1.906	0.056	

Table 5.17 Best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was Black British. N=558; Listener=31.

In evaluations of speaker age when the supposed speaker was of Black ethnicity an interaction between *variant* and *gender* of the face of the supposed speaker were entered as fixed effects (Table 5.17). As can be noticed, when the supposed speaker was of Black ethnicity there was a tendency among the listeners to perceive the localised variants [ɔː] and [øː] as younger-sounding in comparison to the non-localised variant [ɜː]. However, this tendency was not statistically significant. It is worth pointing out that the pattern found in this case was the reverse of other cases, that is of British White and Asian ethnicities, where a tendency to perceive [ɔː] and [øː] as older sounding in comparison to [ɜː] was found.

Furthermore, there was also an effect of the *face gender* of the supposed Black speaker. Upon hearing a word and seeing a supposed British Black male speaker listeners were more likely to rate the variant as having been uttered by a younger speaker. The same pattern was found for the NURSE variants when the supposed speaker was of British White or Asian ethnicity, as well as for the T-to-R variable across all three ethnicities.

The present model also tested for interaction between *phonetic variant* and *gender* of the face of the supposed speaker. The results showed that when a Black male face was shown, listeners were more likely to evaluate both localised variants ([ɔː] and [øː]) as older sounding. However, the results were significant only for the retracted variant [ɔː] ($p<0.01$).

Tables 5.18 and 5.19 present separate analyses of perceived speaker age for female and male supposed Black speaker.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	-0.310	0.390	-0.793	0.428	
Variant [øː]	-0.003	0.315	-0.011	0.991	

Table 5.18 Best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was a British Black female. N=279; Listener=31.

In the female condition the best-fit model was the simplest model with *variant* as a fixed effect (Table 5.18), whereas in the male condition *listener gender* was also entered as a fixed effect (Table 5.19). In the female condition there was a tendency to perceive both [ɔː] and [øː] as having been uttered by a younger speaker. In the male condition, on the other hand, this tendency was reversed such that [ɔː] and [øː] had a significant effect ($p<0.05$). In addition, in the male condition the *listener gender* had a significant effect as well ($p<0.05$). When the listener was male he was more likely to perceive the localised variants as having been produced by an older speaker.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.918	0.381	2.405	<.05	*
Variant [øː]	0.816	0.364	2.242	<.05	*
Listener gender (Male)	0.913	0.418	2.183	<.05	*

Table 5.19 Best-fit ordinal regression mixed effects model showing perceptions of age for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against non-localised variant [ɜː] and the supposed speaker was a British Black male. N=279; Listener=31.

As the last step of analysis of perceived speaker age for the NURSE variants, a comparison of evaluations for all three ethnicities was carried out. The best-fit ordinal regression model to the data presented in Table 5.20 included *variant* and *ethnicity* of the supposed speaker as fixed effects. The results show that there was no effect of *variant*. However, when the supposed speaker was of Asian or Black ethnicity, there was a tendency to perceive the variants as having been produced by a younger speaker. This trend was, however, significant only in the case of Black ethnicity ($p<0.001$).

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɔː]	0.366	0.346	1.058	0.289	
Variant [øː]	0.413	0.310	1.331	0.183	
Ethnicity Asian	-0.080	0.115	-0.693	0.488	
Ethnicity Black	-0.407	0.115	-3.534	<.001	***

Table 5.20 Best-fit ordinal regression mixed effects model showing perceptions of age for the three ethnicities for the variants of the NURSE vowel, where localised variants [ɔː] and [øː] were compared against the centralised variant [ɜː]. N=1674; Listener=31.

5.3.2.2 T-to-R

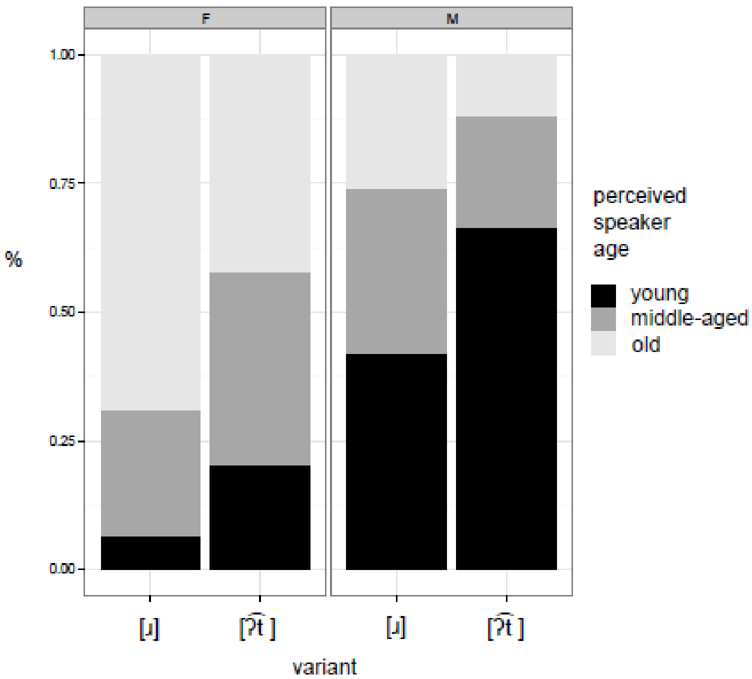


Fig. 5.22 Perception of speaker age for [ɹ] and [ʔ͡t], where the supposed speaker was British White.

Figure 5.22 presents evaluations of perceived speaker age for [ɹ] and [ʔ͡t] when words were accompanied by a gendered face of the supposed British White speaker. As can be noticed, [ʔ͡t] was evaluated as definitely younger-sounding in the male condition. [ɹ] was also found to be somewhat younger-sounding. In the female condition, [ɹ] was found to be definitely older-sounding while [ʔ͡t] was evaluated as somewhat older-sounding. It can be observed that these results in the two conditions (male and female) were mirror images of one another. This could imply that listeners were sensitive to the age-gender interaction, but not much else.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	1.287	0.344	3.742	<.001	***
Face gender (Male)	-2.341	0.253	-9.226	<.001	***

Table 5.21 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t], where the supposed speaker was White British. N=372; Listener=31.

In evaluations of perceived speaker age for [ɹ] and [ʔ͡t] when the supposed speaker was British White, the best-fit model included *variant* and *gender* of the face of the supposed speaker as fixed effects (Table 5.21). The results of the ordinal regression mixed effects show that [ɹ] was more likely to be perceived by listeners as having been uttered by an older speaker ($p<0.001$) in comparison with [ʔ͡t]. Furthermore, as expected, the gender of the face of the supposed speaker also had an effect on perceptions of speaker age ($p<0.001$).

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	1.319	0.556	2.373	<.05	*
Listener gender (Male)	-1.099	0.437	-2.516	<.05	*

Table 5.22 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t], where the supposed speaker was British White female. N=186; Listener=31.

As can be seen in Table 5.22, which shows the results of age evaluation when the supposed speaker was British White female, [ɹ] was more likely to be perceived as having been produced by an older speaker ($p<0.05$). In the male condition, [ɹ] was also perceived as having been spoken by an older speaker ($p<0.01$) (Table 5.23). Furthermore, the effect of *listener gender* was found in the female condition ($p<0.01$). When the listener was male he was more likely to evaluate the variants as having been produced by a younger speaker.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	1.200	0.412	2.912	<.01	**

Table 5.23 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t], where the supposed speaker was British White male. N=186; Listener=31.

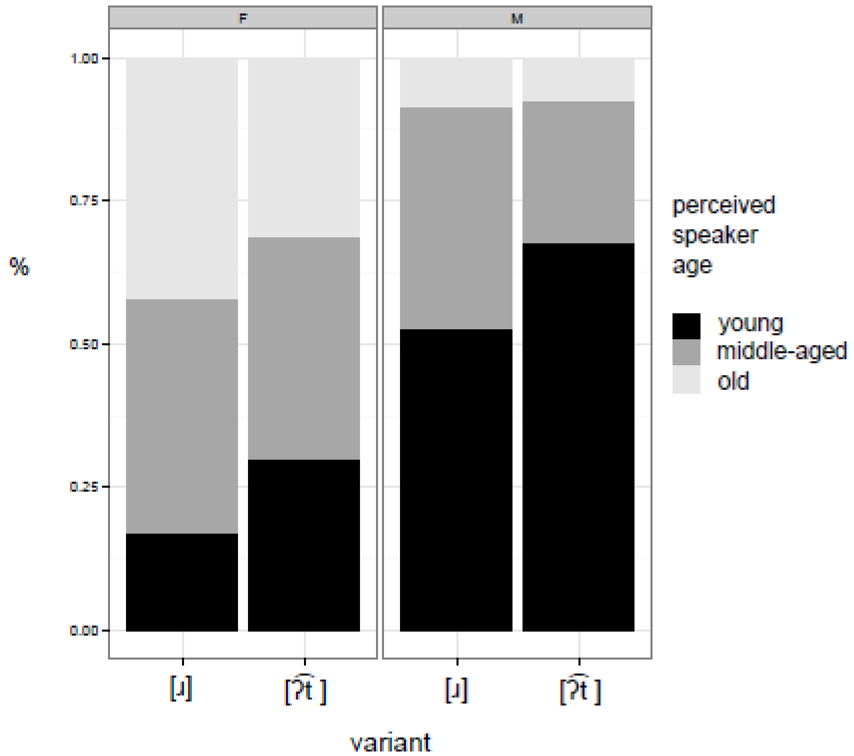


Fig. 5.23 Perception of speaker age for [ɹ] and [ʔ͡t], where the supposed speaker was British Asian.

Fig. 5.23 presents evaluations of speaker age for [ɹ] and [ʔ͡t] when the supposed speaker was British Asian. It can be observed that the results were quite similar to ones found for the British White supposed speaker (Fig. 5.20). While in the male condition [ʔ͡t] was evaluated as having been spoken by a younger speaker and [ɹ] was found to be younger and middle-aged sounding, in the female condition [ɹ] was found to be older and middle-aged sounding, whereas [ʔ͡t] was younger- and middle-aged sounding. In the female condition, however, these evaluations were divided almost equally between the two age groups, and none of the age groups was in the lead.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	0.713	0.222	3.210	<.01	**
Face gender (Male)	-2.125	0.243	-8.736	<.001	***

Table 5.24 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t] when the supposed speaker was British Asian. N=372; Listener=31.

Table 5.24 presents results for perceived speaker age of [ɹ] and [ʔ͡t] when the supposed speaker was British Asian. Overall, [ɹ] was perceived as having been spoken by an older speaker ($p<0.01$) in comparison to [ʔ͡t]. Furthermore, as in the case of the British White supposed speaker, in the case of the supposed British Asian speaker there was a significant effect of the *gender* of the face. When the face of the supposed speaker was male, there was a tendency to evaluate variants as having been uttered by a younger speaker ($p<0.001$).

When the perceived age of the speaker was investigated separately for the Asian female as well as the male supposed speaker, the best-fit model included *variant* as a fixed effect (Tables 5.25 & 5.26). The results show that in both conditions [ɹ] was more likely to be perceived as having been spoken by an older speaker ($p<0.05$). However, the effect of the variant was less significant than in the main analysis (cf. Table 5.24).

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	0.665	0.303	2.192	0.05	*

Table 5.25 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t] when the supposed speaker was British Asian female. N=186; Listener=31.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	0.869	0.388	2.237	<.05	*

Table 5.26 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t] when the supposed speaker was British Asian male. N=186; Listener=31.

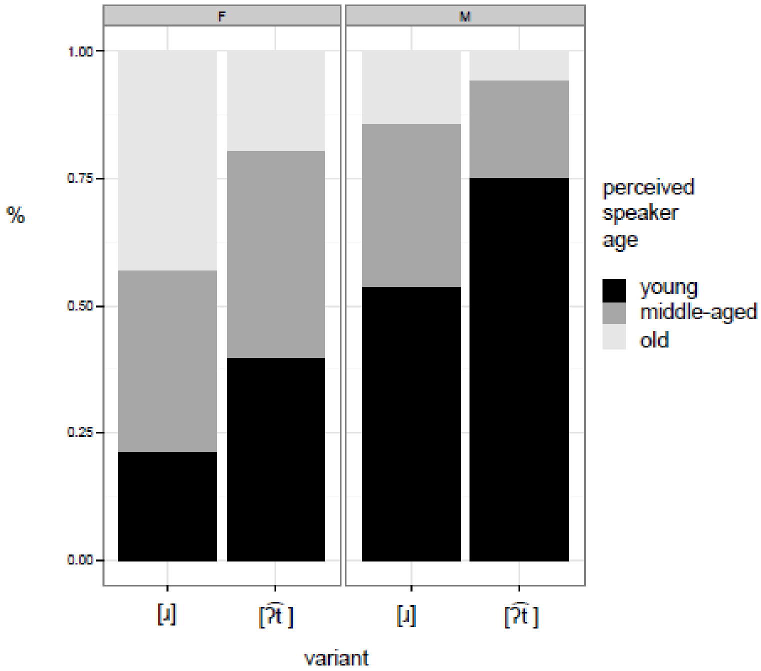


Fig. 5.24 Perception of speaker age for [ɹ] and [ʔ͡t], where the supposed speaker was British Black.

Fig. 5.24 presents the results of the age evaluations of [ɹ] and [ʔ͡t] when the supposed speaker was British Black. As was the case when the supposed speaker was White or Asian, in the male condition [ʔ͡t] was perceived to be younger sounding. [ɹ] was also found to be younger-sounding, but to a lesser extent than [ʔ͡t]. In the female condition, on the other hand, [ʔ͡t] was evaluated as middle-aged and younger-sounding and [ɹ] as older- and middle-aged sounding.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	1.084	0.220	4.912	<.001	***
Face gender (Male)	-1.657	0.226	-7.324	<.001	***

Table 5.27 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t] when the supposed speaker was Black British. N=372; Listener=31.

As was the case in evaluations of age of [ɹ] and [ʔ͡t] when the supposed speaker was British, White and Asian, in the present case when the supposed speaker was British Black, an effect of the phonetic variant was observed. [ɹ] was perceived as having been uttered by an older speaker (p<0.001) as compared to [ʔ͡t]. Additionally, an effect of the *gender* of the face of the supposed British Black speaker was found (p<0.001).

As can be seen in Tables 5.28 and 5.29 in the further analysis, when the supposed speaker was a British Black male or a British Black female, [ɹ] was more likely to be perceived as having been spoken by an older speaker (p<0.01). The effect of the *variant* was less significant than in the main analysis, however (cf. Table 5.27).

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	1.227	0.407	3.013	<.01	**

Table 5.28 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t] when the supposed speaker was British Black male. N=186; Listener=31.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	1.142	0.390	2.928	<.01	**

Table 5.29 Best-fit ordinal regression mixed effects model of perceived age for [ɹ] and [ʔ͡t] when the supposed speaker was British Black female. N = 186; Listener=31.

As the final step of analysis a comparison of evaluations of the perceived speaker age across the three ethnicities was carried out. The best-fit model to the data included *variant* and *ethnicity* of the supposed speaker as fixed effects (Table 5.30). As can be noticed, [ɹ] was perceived as having been spoken by an older speaker ($p<0.001$). In addition, an effect of Asian and Black ethnicity was found. When the supposed speaker was British Asian or Black the variants were perceived as having been produced by a younger speaker ($p<0.001$). *Listener gender* contributed to the model even though it did not have a statistically significant effect.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)	Sig.
Variant [ɹ]	0.833	0.174	4.772	<.001	***
Ethnicity Asian	-0.592	0.143	-4.138	<.001	***
Ethnicity Black	-0.797	0.145	-5.495	<.001	***
Listener gender (Male)	-0.441	0.287	-1.534	0.125	

Table 5.30 Best-fit ordinal regression mixed-effects model showing perceptions of age for [ɹ] and [ʔ͡t] for the three ethnicities. N=1116; Listener=31.

5.4 Summary and discussion

This experiment showed that ethnicity of the gendered face of the supposed speaker did not affect evaluations of speaker gender. In the case of the NURSE variants the Tyneside listeners did not find statistically significant differences between the variants in terms of perceived speaker gender. Thus it could be concluded that indexicality did not guide listeners in their evaluations of the speaker gender upon hearing a NURSE variant. It will be recalled that also in Experiment 1 the Tyneside respondents found no differences between the NURSE variants when asked to evaluate perceived femaleness of the speaker (Table 3.20). Therefore, it seems that indexicality in the case of the NURSE variants was not indicating gender.

However, when the analysis was performed separately for each ethnicity, the results for the variants of the NURSE vowel showed that whilst there was no effect of phonetic variant when the supposed speaker was White, [ɔː] and [øː] were perceived as having been uttered by a male speaker when the supposed speaker was Asian or Black. Specifically, the fronted variant [øː] was perceived as having been spoken by a male speaker in comparison to the centralised and retracted variants when the supposed speaker was Asian or Black ($p<0.01$), whereas the retracted variant [ɔː] was perceived as having been spoken by a male speaker when the supposed speaker was Asian ($p<0.05$). Whilst evaluations of [øː] were the opposite to the findings reported in Experiment 1 where in evaluations of perceived maleness the retracted variant was rated as having been produced by a male speaker ($p<0.05$) in comparison to [øː] and [ɜː], they were also opposite to the findings of the production studies.

These results could be explained by the fact that people have a tendency to associate localised variants with male speakers (Labov, 1990). Nevertheless, the fact that this was the case only when speaker ethnicity was Asian or Black is an interesting finding. Perhaps hearing the localised variants accompanied by a male face of ethnic minority status activated the association of localised variants with a male speaker to a higher degree than in the case of White male faces. It could be the case that this association also neglected indexical information altogether.

As far as [ɹ] and [ʔ͡t] are concerned, Tyneside listeners evaluated [ɹ] as having been produced by a female speaker in comparison to [ʔ͡t]. This would imply that listeners familiar with the dialect were guided by indexicality in their evaluations. These results agree with results of the two earlier experiments where [ɹ] was found to index female speech. At the same time, the results provide us with another important finding. This experiment showed that indexical information carried by the variants was not interlinked with the speaker's ethnicity. Instead, it seems that indexical information about speaker gender was associated with a female rather than a White female.

Separate analyses of gender categorisation of [ɹ] and [ʔ͡t] for each of the ethnicities agreed with the main analysis and revealed a non-significant tendency to perceive [ɹ] as having been spoken by a female speaker when the supposed speaker was White or non-White.

Because the results of gender evaluations for the NURSE variants and T-to-R are fairly inconsistent they cannot be accounted for by exemplar theory which assumes that the amount of exposure to a linguistic feature enables feature recognition during speech perception. Even though listeners were in a group of people with high exposure to the dialect, it seems that their evaluations did not reflect indexical information carried by the NURSE variants as no statistically significant differences between these variants were identified. No statistically significant differences between the NURSE variants were found also in the case of age evaluations, which once again, goes against exemplar theory.

As far as perceived age of the speaker is concerned when the apparent ethnicities were compared, the results varied between variables. For the variants of the NURSE vowel there was no effect of ethnicity except for the Black ethnicity where listeners tended to perceive variants as produced by a younger speaker ($p<0.001$). For [ɹ] and [ʔ͡t], on the other hand, there was an effect of Asian and Black ethnicity where listeners were more likely to perceive the variants as produced by a younger speaker ($p<0.001$). An effect of variant was also found, whereby [ɹ] was perceived as having been spoken by an older speaker ($p<0.001$). Furthermore, when each ethnicity was investigated separately, a non-significant tendency was found to perceive the localised NURSE variants as having been spoken by an older speaker when the apparent ethnicity was White or Asian. When the ethnicity of the supposed speaker was Black, there was a (non-significant) tendency to perceive the localised NURSE variants as having been produced by a younger speaker.

Statistical results for the perceived age of the supposed speaker were almost identical for [ɹ] and [ʔ͡t] when the ethnicities were investigated separately. The overall pattern was that in all three groups [ɹ] was perceived to have been uttered by an older speaker than was the case for [ʔ͡t]. The only difference between the three groups was the strength of this effect, which was identical for the supposed British White and Black speaker ($p<0.001$) and slightly weaker for the British Asian supposed speaker ($p<0.01$).

It will be noticed that indexical information about speaker age carried by the NURSE variants was not identified by the listeners since no statistically significant differences between the variants were found. This finding is also consistent with findings reported in Experiments 1 and 2. In addition, evaluations of speaker age are linked with multimodal speech perception which is explained below.

As with Experiment 2, multimodal speech perception was also investigated in Experiment 3. The results showed that visual information about speaker ethnicity shifted neither perceptions of speaker gender nor age. At the same time, however, the present experiment gave similar results for age evaluations to results of Experiment 2, where gendered faces shown in the images shifted perceptions of speaker age. As expected, there was a decrease in perceived speaker age when the face in the picture was male and an increase in the perceived speaker age when a female face accompanied the words. This effect was identical for the NURSE variants as well as the T-to-R variable across all ethnicities. As can be noticed, results of this experiment provide us with two important findings. The first finding tells us that visual information about the supposed speaker's ethnicity is unlikely to shift perceptions of speaker-indexical information, in this case information about speaker age or gender. The second finding tells us that visual information about gender and ethnicity of the supposed speaker are not interlinked, which is an interesting non-interaction. In fact, it was gender of the face alone that shifted perceptions of speaker-indexical information.

5.5 Experiment 3 Conclusion

The aim of this experiment was to investigate the effect of apparent speaker ethnicity on perceptions of speaker gender and age. The main finding of this experiment is that the ethnicity of the supposed speaker did not shift perceptions of speaker age or gender.

The results of combined analyses show that even though there was a tendency to evaluate the variants as having been produced by a female speaker when the apparent ethnicity was Black (variants of the NURSE vowel and T-to-R), these results yielded no statistical significance. Thus, the ethnicity of the speaker did not influence perceptions of speaker gender. The other finding resulting from gender evaluations of T-to-R is that indexical information about speaker gender was not interlinked with information about speaker ethnicity but associated with gender only.

As far as perceived age of the speaker is concerned, ethnicity of the supposed speaker did not influence evaluations of speaker age. However, the gendered face of the supposed speaker alone shifted perceptions of speaker age. This finding was consistent with the results for speaker age in Experiment 2.

6 General Discussion

6.1 Introduction

This chapter brings together the findings of the three experiments and discusses them in the light of standard language ideology (see Chapter 2).

This research set out to investigate perceptions of speaker-indexical information from gender-ambiguous speech which is an innovative approach to such investigation. This was done by carrying out three consecutive perceptual experiments which investigated perceptions of speaker age, gender and social class. A summary of the experiments and discussion of the results are presented in the following sections.

6.2 Experiments

In the first experiment in the series, three types of indexical information were investigated: speaker gender - that is, how "male" or how "female" the speaker sounded - speaker age, and social class. Unlike the two following experiments, Experiment 1 was not a priming experiment. Listeners heard audio samples, on the basis of which they were asked to evaluate speaker-social information using a Visual Analogue Scale. Experiment 1 established that overall listeners were able to extract speaker-indexical information from phonetic segments alone when the speaker sounded gender-ambiguous. Even though the consistency with which the information was recovered seemed to depend on the social-indexical feature being evaluated, as well as on the variant, evidence is provided that speaker social-indexical information is identifiable from gender-ambiguous speech.

Experiments 2 and 3 are priming experiments. The aim of Experiment 2 was to clarify the results of Experiment 1 by investigating perceptions of speaker social-indexical information that is age and social class, when the listener is presented with images designed to cue the supposed speaker's age, gender, and social class. Experiment 2 was performed in the interests of investigating how perceptions of the same speaker-indexical information as in Experiment 1 might be shifted when providing the listener with visual information about the supposed speaker. For example, upon seeing a young female face when hearing a phonetic variant, the listener might rate the variant differently from the answer s/he gave in response to the same stimulus in Experiment 1. Listeners evaluated speaker age, gender and social class from examples of the phonetic variants under investigation. There were two visual conditions, a female one and a male one, and the same phonetic variants were heard in each. Listeners were asked to evaluate the age and social class of the speaker in both conditions. While in evaluations of speaker age there was only one block in which listeners were given information about the gender of the supposed speaker, in evaluations of social class there were three separate blocks. In block 1, the visual stimuli presented working- and middle-class housing; in block two, a gendered face of the supposed speaker was shown against the housing background. In block three, visual information about speaker age was also added in the images.

In Experiment 3, a new social factor, ethnicity, was introduced. The aim of this experiment was to investigate perceptions of speaker-indexical information such as age and gender when listeners were primed with visual cues to the supposed speaker's ethnicity. White, Asian and Black ethnicities were used in the experiment. As in the two previous experiments, here listeners evaluated perceived speaker social-indexical information from examples of the gender-ambiguous-sounding phonetic variants under investigation. During speaker gender categorisation, in each of the ethnic conditions (White, Asian or Black), listeners were presented with a male and female face of the supposed ethnic speaker when the audio file was played. Furthermore, listeners were asked to evaluate the age of the ethnic speaker, which was done in a manner similar to that used in Experiment 2. The same phonetic variants were heard in the male and in the female face conditions across the three ethnicities.

Because exposing the listener to visual cues about the apparent White speaker gave mixed results in terms of shifting perceptions of speaker indexical information, Experiment 3 investigated whether different results would be obtained if the apparent speaker was ethnic. Experiment 3 also attempted to shed light on how inseparable and interlinked different types of speaker-indexical information are when memory representations are formed. Do listeners who are White and live in a community where the majority of language users are also White associate a phonetic variant primarily, for example, with older women or older White women? Additionally, this experiment tested how listeners' expectations about the voice of the speaker affected listeners' gender choices.

The following sections discuss the results for speaker social class, gender and age evaluations in the three experiments described above.

6.3 Perceptions of speaker social class

In Experiment 1, listeners were the most consistent in terms of evaluations of perceived speaker social class. They also responded in line with expectations. Listeners in the Tyneside group, with higher exposure to the dialect under investigation, and the North-East group, with lower exposure to the dialect, were very close in their ratings of the perceived social class of the speaker. Furthermore, patterns identified by listeners reflected the ones reported in the production studies. This would imply that, even though a number of participants reported feeling uncomfortable having to evaluate the social class of the speaker, the Tyneside and North-East listeners were quite responsive to this type of indexical information. Nonetheless, when the two groups were investigated separately, the statistical results revealed that occasionally the North-East listeners evaluated some variants of the GOAT vowel, and glottalled, glottalised and released /t/, as more likely to have been spoken by a working-class speaker than did the Tyneside listeners.

The similarities in the results obtained for the Tyneside and North-East groups of respondents can potentially be explained by the fact that similar phonetic variants are used in parts of the North-East other than the Tyneside area. However, these variants are marked somewhat differently in terms of social information encoded in them, which results from the frequency of use and differences in sociolinguistic patterns (Beal et al., 2012). Furthermore, the results indicated that even though the North-East listeners were not speakers of Tyneside English, and their exposure to the dialect was presumably lower than that of the Tyneside respondents, the North-East listeners were still able to access speaker social class information. The explanation could be standard language ideology. It was mentioned in Chapter 2 that according to standard language ideology language users are aware of what standard forms sound like, even if they do not use them. Through the system of education and use and endorsement of the standard by national institutions, people of different social standings are able to differentiate between standard and non-standard forms (Milroy, 2004). Standardisation also diminishes the importance of other language varieties which are not the standard variety itself (Milroy, 2007: 138). This in turn leads to stigmatisation of non-standard varieties (Milroy, 2001: 548) and association of these varieties with lower-class speakers. The question of whether the users speak the standard variety is a separate one. As a result, localised varieties are quite powerful at indexing social status (Garrett et al., 2011: 58). It seems that there are reasons to believe that the two groups of participants in Experiment 1 based their answers on their knowledge of Received Pronunciation (RP) which is a standard accent of England. As has been mentioned, class ratings among both groups of listeners were quite similar, which would imply that listeners with high exposure to the dialect, as well as listeners with lower exposure, were able with ease to access information about the variants used by the speakers as being localised, and thus evaluated these strongly localised accent features negatively. However, had the comparison group consisted of participants from a region of the country further away from Tyneside in geographical terms, it is plausible that the differences between the results obtained in the two groups would have been larger.

Even though the two groups of respondents in Experiment 1 rated social class of the speakers similarly, some statistical differences between the two groups were found, whereby the group with lower exposure to the dialect evaluated some of the variants as more likely to have been uttered by a working-class speaker than the group with high exposure to the dialect. Whilst standard language ideology refers to correctness rather than status of the speaker and applies to grammar, lexis and dialect rather than pronunciation, Kerswill (2007) shows that the distinction between accent and dialect is not clear in speakers' minds. Therefore, a possible explanation of the above mentioned tendency observed in the North-East listeners may be that they were guided by the SLI to a higher degree than the Tyneside listeners, thus associating localised accent features with a working-class speaker. This would suggest that in addition to basing their judgments on the standard/non-standard dichotomy, exposure to the dialect also influenced their evaluations. In the sense that while listeners with lower exposure were stricter in their evaluations of speaker social class because they relied mostly on the information about localness of the variants, listeners with higher exposure could access more detailed indexical information encoded in the variants used by the speakers.

Nevertheless, we should not rule out the possibility that listeners with lower exposure to the dialect were also able to access some of the indexical information carried by the variants under investigation. For example, whilst to the North-East listeners the variant indexed a working-class speaker, to the Tyneside listener the variant indexed a Tyneside speaker from the working class. Following this line of argumentation, it could be argued that listeners with lower exposure to the dialect than the comparison group should be even stricter in their evaluations of speaker social class, since their judgements would be based on standard language ideology to a greater extent and less on the subtleties of the indexical information carried by the variant. The differences between the Tyneside and North-East groups of respondents could be also accounted for by enregisterment (Agha, 2005). Linguistic forms which are different from standard forms may become enregistered with (non-linguistic) meaning and become interpretable with social information (Johnstone, 2014). However, what the forms mean depends on a specific community or area. Therefore, the same linguistic form may be interpreted differently in two different communities or areas. A case in point are the Tyneside and North-East groups of respondents. Variants which were evaluated differently in terms of perceived social class but also gender and age by the Tyneside and North-East participants seem to have been enregistered with different social information in Tyneside and other parts of the North-East region. In other words, the same linguistic forms are linked with non-linguistic information about the speaker which is different in Tyneside and elsewhere in the North-East. Furthermore, this variation in enregisterment of linguistic forms in Tyneside and the wider North-East should in fact be expected if, for the same variants, production studies reported other sociolinguistic correlations outside of Tyneside (Burbano-Elizondo, 2008; Llamas, 2001; 2007).

In Experiment 2, the situation was somewhat different in terms of the comparison group. This group was comprised of listeners from the south of England. As such, these listeners were expected to have lower exposure to Tyneside English than the North-East comparative group in Experiment 1.

The situation in Experiment 2, which was a priming experiment, was similar to that in Experiment 1. Evaluations of speaker social class were overall similar in the Tyneside and Southern groups of respondents. The main difference in Experiment 2 between the Tyneside listeners (with higher exposure to the dialect) and the Southern listeners (with lower exposure to the dialect) was the fact that in block 1 they did not differentiate between localised variants, which to listeners with lower exposure may have sounded localised and, as a result, automatically cued working-class evaluations. On the matter of social class evaluations in block 1, it could also be observed that some of the localised and non-localised variants were not always distinguished. This was the case for the NURSE variants or variants of /t/ in block 1, when listeners were provided with information about the social class of the supposed speaker. This would imply that having low exposure to the dialect prevented listeners from distinguishing more detailed social-indexical information and thereby differentiating between these variants. Instead, this group of listeners was able to apply the basic distinction between localised and non-localised variants. Also these results could be explained with reference to the standard language ideology (SLI) since, as has been already mentioned, the distinction between dialect and accent is not clear in speakers' minds (Kerswill, 2007). As an outcome of SLI, localised variants are often attributed to lower-class speakers. Perhaps even more so than the North-East respondents also the Southern respondents were guided by SLI in their evaluations. Thus, on the basis of the results in Experiment 2, it could be argued that both groups of respondents relied on their knowledge of differences between localised and non-localised forms when evaluating the supposed speaker's social class. With reference to evaluations of social class, this knowledge may have its roots in conservative ideologies such as standard language ideology promoted by the state (Agha, 2004; Milroy, 2001). While localised variants may be evaluated positively in terms of solidarity, they are evaluated negatively in terms of prestige. In other words, the variety is evaluated on how 'nice' or 'ugly' and how 'good' or 'bad' it sounds with reference to the standard variety (Garrett et al., 2011: 61). Thus, even though language speakers would describe a person using localised forms as friendly, they probably would see him/her as a lower-class speaker. The results of Experiments 1 and 2 show that these attitudes and (resulting from them) evaluations of localised forms are true even though, as Garrett et al. (2011) claim, social perceptions of non-standard localised dialects and accents are changing. These changes mentioned by Garrett et al. may have their roots in dialect and accent levelling, which are occurring in several parts of the UK (Kerswill, 2003; Llamas, 2000; Watt, 2002). The outcome of dialect or accent levelling is a reduction of localised forms. Even though speakers may still want to indicate their regional belonging, they may not want to indicate localness as much (Watt, 2002). Nevertheless, even if levelling occurs and people tend to use a regional standard, the influence of SLI is present and affects evaluations of localised variants in terms of prestige. Listeners seem to respond to this type of indexical information even if they are users of the localised phonetic forms.

Nevertheless, the already mentioned differences between the Tyneside and Southern respondents in evaluations of, particularly, localised variants may suggest that because the former respondents were more familiar with the dialect they were also sensitive to more subtle indexical-information about speaker social class such as the differences between localised variants. Thus it could be argued that the Tyneside listeners in their evaluations of speaker class were guided by indexicality. These listeners recognised social-indexical information carried by the variant as information indexing a Tyneside speaker (area belonging) of a working-class background (group belonging).

However, as was the case in Experiment 1, also here higher sensitivity to more subtle differences between the variants presented by the listeners with higher exposure to the dialect and the lack of thereof shown by the listeners with lower exposure to the dialect could be further explained by enregisterment. Therefore, it could be the case that to the Southern listeners localised Tyneside variants were not enregistered with meaning which would imply that this group of participants were not familiar with Tyneside English. It could be also that to the Southern listeners localised variants were not enregistered at the same level as they were to the Tyneside listeners. This on the other hand would imply that the Southern listeners were familiar with Tyneside English to a limited extent.

However, statistical comparisons of the Tyneside and Southern groups indicated that the exposure to the dialect was not a significant factor for the most part. This would suggest that differences between the two groups were not large enough to show in statistical analyses.

Experiment 2 also investigated perception of speaker social class when listeners were primed with an image of a male or a female face. This was part of the investigation of multimodal speech perception. The overall lack or rare occurrence of statistical effects of the gendered face on the evaluations of social class of the speaker in blocks two and three imply that the Tyneside listeners (with higher exposure to the dialect) and Southern (with lower exposure to the dialect) were sensitive to social class information about the speaker in a way that seemed difficult to manipulate experimentally. While the gendered face of the speaker in blocks two and three of Experiment 2 did not have an effect on evaluations with respect to speaker social class of T-to-R and variants of /t/, it did have some effect on evaluations of the NURSE variants in block three, when images cued information about social class, gender and age of the supposed speaker. It was hypothesised that showing a female face could have resulted in fewer variants being evaluated as working-class. Overall, however, it seems that the social class information carried by the variants was more salient than visual information cueing speaker gender, and as such, gender information did not shift class evaluations. At the same time, it could be noticed from the statistical results that introducing visual information about speaker gender in block two overall did not result in a change of evaluations of perceived speaker social class from block 1. These findings did not provide evidence in support of visual information about the supposed speaker shifting perceptions of the acoustic information (Niedzielski, 1999; Hay & Drager 2010; Drager, 2011). So far, then, we do not have evidence that multimodal speech perception differs from unimodal speech perception.

As far as exemplar theory is concerned, the results across the three experiments showed that respondents' evaluations of speaker social class, age or gender could not be explained by the theory. Exemplar theory is a model of speech perception based on exposure and even though exposure to the dialect mattered – this could be observed in the subtle differences between the groups with higher (the Tyneside groups) and lower (the North-East and Southern groups) exposure to the dialect in Experiments 1 and 2 – these differences lacked in consistency and were not statistically significant. In fact, on the whole there were a lot of similarities between the main and comparison groups in how they evaluated speaker social class, age and gender. In addition, some speaker indexical information was hardly accessed by the Tyneside listeners. Such examples include speaker gender in Experiments 1 and 3 as well as speaker age in Experiments 1, 2 and 3. These results might in fact go against exemplar theory. Despite their high exposure to the dialect and despite having the exemplars frequently updated, the listeners failed to access some of the social information.

Even though there is no denying the fact that social information is accessed and evaluated during speech perception, perhaps there may be a different explanation for how it is stored and retrieved. Labov, for example, argues that social information may be stored in a more abstract form rather than being stored in exemplars and that it may be separated from other semantic information (Labov, 2006: 512). Perhaps this could serve as an explanation for the issue of why different types of social information were accessed with varying consistency.

While this section focused primarily on perceptions of speaker social class, a more in-depth discussion of evaluations of speaker gender and age is presented below.

6.4 Perceptions of speaker gender

Sex and gender are two separate notions in sociolinguistics. While sex takes into account differences between men and women resulting from biological differences, for instance, the length of larynx or vocal cords in men and women varies in terms of size, which in turn affects fundamental frequency. Gender, on the other hand, is a social construct grounded in biological sex which defines the social roles of men and women. Examples of how gender may be constructed socially are different phonetic variants used by men and women. While some variants may be used by females to indicate their gender and social class affiliation, different variants may be used by males for the same purpose.

As far as evaluations of the perceived maleness and femaleness of the speaker in Experiment 1 are concerned, listeners in general did not evaluate as female variants which were identified in production studies as most frequently used by female speakers. Overall, all phonetic variants under investigation were judged to have been spoken by a male speaker. This could be explained by the fact that even though the voices sounded gender-ambiguous, it is often the case that localised variants are attributed to male speakers.

Regarding the results across the Tyneside and North-East groups of respondents, speaker gender was evaluated by both groups similarly. Despite the fact that the two groups of listeners agreed in their evaluations of the femaleness or maleness of the speaker, a statistical comparison of the two groups revealed that in some cases the Tyneside respondents, in comparison to the North-East respondents, were less likely to evaluate the variants as having been spoken by a male speaker. This was indicated by, for example, the perceived maleness of the NURSE variants but also in the perceived femaleness and maleness of T-to-R. This would mean that listeners were sensitive to gender differences between the variants only quite sporadically. In addition, differences in the evaluations of variants can be observed in the graphs whereby some of the variants used mostly by female Tyneside speakers are evaluated as more categorically not female-sounding by the North-East group of respondents than by the Tyneside group. This might imply again that Tyneside listeners were able to access more detailed indexical information about the speaker. It could be argued that whilst the North-East listeners were guided in their evaluations by the fact that male speakers are believed to use localised variants more often than female speakers, to the Tyneside listeners the variants under investigation indexed a male speaker. Indexicality also seems to have worked in the case of [ɹ], which was found to be statistically more likely to be thought to have been uttered by a female speaker in comparison to [ʔ͡t]. Furthermore, this variant seemed to index a female speaker to both groups of respondents. Thus, it seems that [ɹ] was associated with female speech in all of the North-East. As a matter of fact this variant is used most often by female speakers in all of North-East England (Beal et al., 2012: 40). However, with the exception of [ɹ] other consonantal variants were overall evaluated by both groups of listeners as having been uttered by a male speaker. Thus it could be concluded that speaker gender was indexed by the variants to a limited extent. Therefore, the results reported above may be explained for the most part by the general belief that male speakers tend to use more non-standard and strongly localised accent features in their speech than female speakers (Labov, 1990; Watson & Clark, 2015: 44). Female speakers, on the other hand, are believed to use more standard forms and to avoid stigmatised ones. The explanation of this generalisation is that women are usually more aware of the dependence of social status on the language one speaks. Because women tend to be the primary care givers to their children, by using "proper" forms it would be expected that they would tend to use standard features to ensure that their offspring are not socially disadvantaged. In fact, women's use of stigmatised forms is socially disapproved of or even condemned (Watt, 1998: 87). On the other hand, men's use of localised forms seems to be socially approved. Nevertheless, even though speakers of localised varieties are aware of these social boundaries exerted on language, as well as the stigmatisation imposed on their dialects by SLI, they may still identify with the place and may want to index this belonging (Milroy, 2004). This can be seen, for example, in the choice of phonetic variants among Tyneside speakers. Nevertheless, the results of Experiment 1 indicate that listeners in the Tyneside and North-East groups were also guided by the social constraints on language use by males and females when evaluating the perceived maleness and femaleness of the speaker. Overall, it seems that the influence of social constraints on language use has a stronger effect on the perception of the localised variants than indexical information about speaker gender carried by the variants.

The results of Experiment 3 showed that upon comparing the three ethnicities, no effect of ethnicity on speaker gender categorisation was found. Nevertheless, the [ɹ] variant was more likely to be perceived as having been spoken by a female speaker, which is in line with the results of the production studies. This allows us to conclude that the respondents acted on the social-indexical information carried by the variant. It will be recalled that the same finding was reported in Experiment 1. Furthermore, when ethnicities were investigated separately, the results indicated that, notably, there was no effect of *phonetic variant* when the face of the supposed speaker was White. However, the localised variants were evaluated as having been spoken by a male when the face of the supposed speaker was Asian or Black. For example, the fronted NURSE variant, [øː], was likely to be evaluated as having been uttered by a male when the face of the supposed speaker was Asian or Black. In addition, the retracted NURSE variant, [ɔː], was perceived as having been spoken by a male when the face of the supposed speaker was Asian. There was also a non-significant tendency for the localised NURSE variants to be evaluated as having been spoken by a male when the face of the supposed speaker was White or Black. These results show that for some reason indexicality did not guide the listeners in the evaluations of the NURSE variants. It seems that, overall, the listeners may have relied on the general belief that localised variants are most often used by male speakers instead. In contrast, [ɹ] was perceived to have been uttered by a female when the supposed speaker was Black. In addition, a non-significant tendency was observed toward evaluating the variant as having been spoken by a female speaker when the face of the supposed speaker was White or Asian. As can be noticed, localised NURSE variants were evaluated as having been produced by a male speaker. In addition, while categorisations of variants [ɔː] and [ɹ] overall reflect findings of the production studies, the categorisation of variant [øː] as having been spoken by a male stands in opposition to the findings of the production studies, which reported the variant to be used most often by female speakers. Thus, it seems that as in Experiment 1, in Experiment 3 listeners were overall guided by social constraints on language use when evaluating speaker gender. Except for [ɹ], which is a variant often used by females and widespread in North-East England, it seems that the participants were led by the general view that localised variants are attributed to male speakers. It could be argued that the ethnicity of the supposed speaker influenced speaker gender categorisations only to a limited extent. However, a pattern did not seem to exist.

Thus, the results above would imply that overall the ethnicity of the supposed speaker did not influence gender categorisations of investigated phonetic variants in any systematic manner.

6.5 Perceptions of speaker age

It was mentioned in the previous section (Section 6.4) that in addition to indexing speaker gender, different phonetic variants tend to be used by males and females to show belonging to a particular social class. However, indexical information encoded in the variants used by language users can be even more detailed and may carry, for example, information about speaker age.

As was pointed out in Section 3.2.1, as a result of manipulation of the audio stimuli there was an interaction between the perceived age and gender of the speaker, in the sense that, in the judgement of the phoneticians evaluating the stimuli, shifting fundamental frequency made the stimuli sound as though they had been produced by either an older female or a younger male.

Evaluating speaker age in Experiment 1 seemed to pose a similar problem to the listeners as did evaluating speaker gender (cf. Section 6.4). Overall, variants were rated in the middle range of the scale, which would imply that listeners found the voices to be in general mature- and middle-aged-sounding or they did not know how to evaluate them. The only exceptions were the majority of the glottalled and glottalised variants, which were found to have been produced by a younger speaker. This might mean that listeners did not associate the variants with speakers of any specific ages. Thus it seems that indexicality did not guide listeners' responses. Furthermore, statistical tests found no differences between the Tyneside and North-East groups of respondents in terms of their evaluations of speaker age.

As expected, the age-gender interaction that resulted from manipulation of the audio stimuli could be noticed in the age ratings of, for example, a variant of the GOAT vowel [ʊə] and T-to-R. Variants perceived to be more female-sounding were also evaluated as older-sounding, whereas variants judged to have been produced by a male were also rated as younger-sounding. It could be the case that the voice quality or possibly formant frequency (FF) information overshadowed indexical information about speaker age carried by the variant. The fact that no statistically significant differences between the two groups of respondents were found could mean that listeners in both groups were equally responsive to the age-gender interaction occurring in the stimuli (cf. Section 3.2.1).

Nevertheless, when the two groups were investigated separately, occasionally the North-East group seemed to be more prone to relying on the voice quality information (see, for example, the localised variants of FACE in Chapter 3). This might mean that because of having lower exposure to the dialect, the North-East listeners were at an even bigger disadvantage than the Tyneside listeners when accessing indexical information carried by the variants.

As far as perceived speaker age in Experiment 2 is concerned, the results show that while in general the face gender of the supposed speaker had almost no influence on evaluations of perceived speaker social class, it did in fact impact on the ratings of perceived speaker age. Thus, gender of the face of the supposed speaker predicted age. The same pattern was found for the Tyneside and Southern groups of participants. It should be noted that the perceived age of the speaker may have been related to the apparent age of a synthetic face that the listener was exposed to. As was evident from the results presented in Chapter 4, some (or perhaps a large amount of) variation in age ratings was most probably cued by the presence of a gendered face accompanying a gender-ambiguous voice. It is likely that upon hearing a gender-ambiguous voice and seeing a female face, listeners thought the speaker was older. Seeing a man’s face, on the other hand, seems to have made listeners think the speaker was younger. Again, this presumably resulted from the voice quality of the audio stimuli. What seems to have happened is that the information from gendered faces overshadowed information carried by the phonetic variants themselves. Thus, listeners perceived the same voice qualities as belonging to an older female and a younger male speaker. This was reflected in the evaluations of the perceived speaker age of the variants in the male and female conditions. For example, when listeners heard the variants accompanied by a male face, they seemed to perceive the variants to sound different with respect to speaker age. The general tendency was to evaluate variants as older-sounding when the face of the supposed speaker was female. The opposite tendency was observed when the face of the speaker was male, whereby the stimuli tended to be evaluated as younger-sounding. Not only were these results consistent with the results in Experiment 1 but they would also imply that priming with the gender of the supposed speaker had an effect. Thus, the findings of this experiment provide additional evidence in favour of visual information shifting perceptions of indexical information.

The results in Experiment 2 also show noticeable differences between the Tyneside group of respondents (with higher exposure to the dialect under investigation) and the Southern group of respondents (with lower exposure). For example, while none of the NURSE variants themselves yielded statistically significant results among the Southern group of listeners, among the Tyneside group of respondents the localised variants [ɔː] and [øː] were evaluated differently from the non-localised variant [ɜː]. Overall, it seems that when evaluating speaker age the Southern listeners (as well as the North-East listeners in Experiment 1) relied on age-gender interaction more than the Tyneside listeners did. Thus, even though the age-gender interaction issue should not be neglected, it seems that speaker-indexical information could be accessed in spite of the interaction.

Furthermore, investigations of speaker age perceptions in Experiments 1 and 2 provided an important finding and forwarded our knowledge about multimodal speech perception from gender-ambiguous voice. Experiment 1 used unimodal speech perception, where the listener was exposed only to audio information, while Experiment 2 used multimodal speech perception where the listener was exposed to audio information along with visual information about the supposed speaker. The two experiments found different results for speaker age evaluations. Whilst in the experiment using unimodal speech perception listeners evaluated speaker age at the midpoint of the scale, showing that they were uncertain about how to rate the speaker age, in the experiment implementing multimodal speech perception, we can observe different results. As has been already mentioned, the variant was evaluated as having been uttered by a younger or older speaker depending on the gender of the face shown in the picture. When the face in the picture was male, the variant was evaluated as having been produced by a younger speaker and when the face in the picture was female, the variant was evaluated as having been uttered by a female speaker. These results provide further evidence that unimodal and multimodal speech perception may produce different outcomes. Whilst unimodal speech perception was the traditional and simplified model of speech research (Bernstein, 2012: 39; Rosenblum, 2005: 51), it does not reflect the natural process in which also the sense of vison is involved. In a natural act of speech perception the listener can hear and see the speaker and so s/he relies on audio cues as well as visual ones. Therefore, this thesis provides us with an important and innovative finding which furthers our understanding of the process of speech perception. It helps us to better understand how seeing the talker impacts what we hear. A number of studies show that looking at the speaker improves speech perception across different listening conditions. Speech perception is enhanced in listening adverse conditions (Buchwald et al., 2009) as well as good listening conditions (Sakamoto et al., 2014). Experiments 1 and 2 described in the thesis showed that specific type of visual information about the speaker may shift perception of some of speaker-indexical information.

However, it should be also pointed out that no effect of visual information about the supposed speaker was found when speaker social class was evaluated in Experiment 2. Seeing a gendered face of the supposed speaker did not change listeners' evaluations of speaker social class. Therefore, it seems that only some types of information may be affected by visual cues.

Multimodal speech perception was also investigated in Experiment 3. Whilst visual information about the ethnicity of the supposed speaker shifted neither perceptions of speaker age nor gender, visual information about speaker gender alone influenced perceptions of speaker age. This finding was consistent with findings reported in Experiment 2.

In Experiment 3, speaker age was evaluated in male and female conditions across three ethnicities: White, Asian and Black. As far as evaluations of perceived speaker age for the NURSE variants are concerned, when the three ethnicities were compared the results indicated no *ethnicity* effect on age evaluations except for the Black ethnicity. In the case of [ɹ] and [ʔ͡t], the statistical results revealed that there was an effect for Asian, as well as Black, ethnicity of the supposed speaker. Overall, seeing an Asian or Black ethnic face in comparison to seeing a White face made listeners more likely to evaluate the variants as having been uttered by a younger speaker. This might have resulted from the fact that Tyneside listeners were not as accustomed to hearing the voices of Black or Asian speakers as those of White ones.

As was the case in Experiment 2, in Experiment 3 gender of the face of the supposed speaker predicted the age. Across all ethnicities matching a gender-ambiguous-sounding voice with an ethnic male face resulted in variants being perceived as having been uttered by a younger speaker, whereas matching a gender-ambiguous-sounding voice with an ethnic female face resulted in the variants being evaluated as more likely to have been uttered by an older speaker. These findings would imply that the interaction between gender and age of the speaker was not dependent on the ethnicity of the supposed speaker. Instead, it was a more universal effect which listeners were sensitive to, irrespective of the ethnic background of the speaker. These results were consistent with the findings of the previous experiment.

However, it should be pointed out that while separate results for each of the ethnicities revealed that none of the NURSE variants yielded a significant effect, the results for T-to-R indicated that overall [ɹ] was perceived as significantly more likely to have been uttered by an older speaker. This would mean that even though the face gender of the supposed speaker had an effect on the perceived age of the speaker, indexical information about speaker age carried by the phonetic variant was also decoded by listeners.

As has been mentioned, there was some effect for ethnicity of the supposed speaker when the three ethnicities were compared. However, when each of the ethnicities was investigated separately, the results found for the variants under investigation, as well as the effect for face gender of the supposed speaker, were very similar across the three ethnicities, which would imply that speaker ethnicity in fact did not influence the perception of speaker social-indexical information to any meaningful extent.

The results of the three experiments discussed in this chapter provide consistent evidence that certain types of speaker social-indexical information, for example social class information, can be accessed by listeners at the segmental level. It seems that removing the cue of gender-specific fundamental frequency made speaker social class the most salient indexical feature. At the same time, age judgements seemed to depend on perceived speaker gender, and perhaps the other way round. Moreover, it is obvious from the results that not all types of information can be accessed with the same consistency. As could be seen, gender was one of them. As far as speaker age information is concerned, manipulations applied to audio stimuli posed a certain limitation on findings of experiments presented in this thesis, in the sense that to a certain degree the voice quality of the stimuli overshadowed speaker-indexical information carried by the phonetic variants. Nevertheless, it should be stressed at this point that it would be difficult to obtain gender-ambiguous, yet natural-sounding stimuli which would not carry any information about voice age. At the same time investigations of speaker age presented us with an important finding regarding multimodality in speech perception.

Perhaps different results would be obtained in terms of the perception of speaker age from listeners older than the current groups of respondents, especially in Experiments 1 and 2. Even though the VAS scales used in Experiment 1 had clearly defined endpoints, it could be safely assumed that the general perception of age and what or who is thought to be young or old varies between age-groups (Harnsberger et al., 2008: 66).

Conclusions to each of the experiments, as well as the summary of the findings of the research discussed in this thesis are presented in the following chapter.

7 Conclusions

7.1 Main findings

Apart from providing additional evidence in support of the interaction of linguistic and non-linguistic information in speech perception this innovative and original research resulted in three other findings. Firstly, whilst previous studies on perception of speaker-indexical information have established that listeners are able to identify speaker-indexical information from single phonetic units (Hay & Drager, 2010; Hay et al., 2006a; Johnson et al., 1999; Niedzielski, 1999; Purnell et al., 1999) this research supported findings of the Christchurch group (Hay & Drager, 2010; Hay et al., 2006a) and took our understanding of the process a step further. The innovative approach used in this research involved the application of gender-ambiguous voice in the identification of speaker-indexical information. From the findings reported in this thesis we know not only that listeners are able to retrieve speaker-indexical information from phonetic segments but also that the process is successful when these phonetic segments sound gender-ambiguous.

This finding is of great importance to sub-disciplines of linguistics such as sociolinguistics, phonetics and sociophonetics as well as any other field investigating speech perception. It also has significant implications for speech recognition systems.

To establish whether the amount of exposure to the dialect under investigation was a decisive factor in retrieving speaker-indexical information from the variants, listeners with high exposure to the dialect were compared with listeners with lower exposure to the dialect. The results indicated that there were some interesting effects as the Tyneside listeners were able to access the subtleties of speaker-indexical information carried by the variants. This was the case in evaluations of speaker social class in Experiments 1 and 2 where Tyneside listeners were less categorical in their evaluations of variants as working-class sounding. Furthermore, unlike the Tyneside and North-East listeners, Southern listeners did not differentiate between localised variants which might indicate that the exposure to the dialect had impact on accessing indexical information. Also in gender evaluation in Experiment 1 Tyneside listeners were less likely to evaluate the variants as having been spoken by a male speaker in comparison to the North-East listeners. In evaluations of speaker age in Experiments 1 and 2 Tyneside listeners managed to access speaker-indexical information despite the age-gender interaction, whereas the North-East and Southern listeners relied more on the interaction in their evaluations of speaker age. It seems that even though exposure to the dialect was not a key factor in retrieving speaker-indexical information from the variants, it impacted evaluations of speaker-indexical information in the case of Tyneside listeners.

Nevertheless, indexical information was overall retrieved and evaluated with similar consistency in three groups of respondents. This might suggest that for British listeners exposure may not be a decisive factor in the process of recognition of speaker-indexical information, and especially in recognition of speaker social class (Trudgill & Giles, 1976).

However, the findings reported in the thesis showed that different types of indexical information were retrieved by the listeners with varying consistency (Experiment 1). While information about speaker social class was retrieved with considerable consistency, information about speaker gender was retrieved with much less consistency (with the exception of T-to-R), as was information about speaker age, which in addition was overshadowed by the age-gender interaction resulting from the experimental design. These results might suggest that perhaps not all types of social-indexical information are equally important to the listener. While social class information seems to be more important in the process of speech perception, information about speaker age and gender may not be as important. Furthermore, it was also observed that certain variants encoded speaker-indexical information more consistently than others. This was shown by the fact that indexical information encoded in these variants was not as well recognised by the listeners as information encoded in other variants. Perhaps it could be said that such variants did not index social information with as much strength as others. It should be mentioned that the results were similar in the group with high and somewhat lower exposure to the dialect under investigation.

Furthermore, this research offered a multivariate investigation of perception of speaker-indexical information. A number of variables were researched which was usually not the case in the previous studies.

The second main finding concerned multimodality in speech perception and established that human perception of speech can change depending on whether the listener only hears the speaker or whether s/he hears and sees the speaker. Experiment 2 addressed the shortcomings of traditional speech processing studies which focus on the audio channel and showed that seeing a gendered face of the supposed speaker made the listeners change their evaluations of speaker age. The results presented in this thesis show that exposure to a gendered face resulted in a shift in age evaluations of the speaker whereby exposure to a male face resulted in a decrease of the perceived speaker age, whereas exposure to a female face had the opposite effect and resulted in an increase of the perceived age. As mentioned in Section 3.2.1 shifting fundamental frequency made the stimuli sound as though they had been produced by either an older female or a younger male. In the light of the above results it seems that gendered face overrode the ambiguity in voice with regard to sex. The results were also consistent with findings reported in Experiment 1. Nevertheless, it is interesting to observe that within the male and female conditions variants were evaluated differently for age which would suggest that listeners accessed information about speaker age despite the age-gender interaction.

Unlike in the case of age, the manipulation did not shift perceptions of speaker social class. Previous studies reported that perceptions of speech may shift under the influence of visual information (Hay & Drager, 2010; Hay et al., 2006a; Hay et al., 2006b; Niedzielski, 1999). Where, for example, exposing the listener to a soft toy associated with a specific country or exposing the listener to a national or regional label (Hay et al., 2006a; Niedzielski, 1999) results in a shift in perceptions of the audio stimulus (Hay & Drager, 2010). Similarly, exposing the listener to visual cues about the age and social class of the supposed speaker also results in a shift in perceptions of the audio stimulus (Hay et al., 2006b). However, the results reported in this thesis suggest that perceptions of speaker social class may not change when listeners are provided with information about the gender of the supposed speaker (Experiment 2). In other words, the gendered face of the supposed speaker did not have an effect on perceptions of speaker social class. Furthermore, visual prompts to social class or age of the apparent speaker did not shift perceptions of speaker social class either. Except for occasional changes in the strength of the effect of the *variant*, the tendencies remained the same in the group of respondents with high and lower exposure to the dialect. As can be noticed, the results for perception of speaker social class reported in this thesis did not support findings of earlier studies where visual cues resulted in a shift of perceptions of audio stimulus. A possible explanation of the different results obtained for evaluations of speaker social class and age with regard to whether speaker indexical-information was shifted by visual cues about the supposed speaker may be that the result of manipulation depends on the type of indexical information. This might again suggest that social class information is salient to the listener and is not easily manipulated.

As can be noticed, investigating multimodal speech perception resulted in several findings which are of crucial importance to the sub-disciplines of sociolinguistics or psycholinguistics but also psychology or any other discipline investigating speech perception.

The third main finding showed that exposing the listener to visual cues with information about the ethnicity of the supposed speaker did not influence perceptions of speaker gender or age (Experiment 3). This finding also concerned multimodal speech perception. Speaker gender was evaluated at chance level and listeners were consistent in their evaluations upon being exposed to the White and non-White (Asian or Black) face of the supposed speaker. These results did not support the hypothesis which assumed that listeners with high exposure to localised phonetic variants would transfer their memory representations of variants associated with local White male and female speakers onto local non-White male and female speakers. Because there was no effect of apparent ethnicity, it could be concluded that information about speaker gender seems to not be interlinked with information about speaker ethnicity. Nevertheless, a (non-significant) tendency was observed to evaluate the variants as having been produced by a female speaker when the face of the apparent speaker was Black. This tendency might suggest that perhaps the listeners were less used to Black people speaking the variety, which could result from the fact the Black minority is not large in Tyneside.

Apart from speaker gender, speaker age was also investigated when the listener was exposed to faces of supposed speakers of different ethnicities. The results reported in this thesis showed that speaker age was not influenced by the ethnicity of the supposed speaker but by his/her gender alone. Separate analyses for each gender showed that unlike in the case of speaker gender, speaker age was influenced by the non-White face of the supposed speaker. Upon seeing a Black face, and occasionally an Asian face, listeners tended to evaluate the variants as having been spoken by a younger speaker. This finding, on the other hand, supports the hypothesis since it seems that information about speaker age was interlinked with information about speaker ethnicity giving different results for Asian and Black speakers in comparison to White ones. The explanation might be similar to that for gender, namely the listeners may not have been used to Asian or Black speakers as much as White speakers using the dialect. In addition, as was the case in previous experiments, an effect of gendered face was found. From the results reported for perceived gender and age when listeners were exposed to speakers of different ethnicities, it could be presumed that, similarly to Experiment 2, the effect of manipulation depended upon the type of indexical information investigated.

Furthermore, since in Experiments 2 and 3 listeners saw the face image before they heard the audio, it could be concluded on the basis of the results that listeners' prior expectations about the speaker did not affect their social class or gender-ethnicity choices.

The following section lists some limitations of the present research.

7.2 Research limitations

This section highlights some limitations of the methodology used in the thesis.

The stimuli used in the experiments were manipulated by shifting the fundamental frequency of the original recordings to obtain gender-ambiguous speech. However (cf. Section 3.2.1), along with changing F0, formant frequencies were also adjusted to changes in F0. This was an automatic process applied by the software. Even though the aim of receiving natural- and gender-ambiguous samples was achieved, it could be argued that manipulations could have been more controlled, in the light of interpreting the results of the experiments.

The age-gender interaction resulting from the manipulation of the stimuli was also a limitation to this research (cf. 3.2.1). As was mentioned in the previous chapters, voice quality seemed to overshadow indexical information about speaker age that was carried by the variants. Even though age information was to some extent still accessible to the listeners, which could be seen when analyses of the age results for the female and male conditions were performed separately (cf. Experiment 2), it did not seem to be as readily accessible as social class information. The question is, however, whether it is possible to construct a voice which does not carry any specific age information. In this regard, it seems that constructing a gender-ambiguous voice might be easier to achieve. Furthermore, these limitations provide additional evidence in support of the view that different types of indexical information are interlinked in the speaker's voice, which also affects the processing of these information types by listeners.

Another limitation was the smaller number of variables investigated in Experiments 2 and 3 in comparison to Experiment 1. However, the character of the experiments did not allow the usage of larger numbers of stimuli in the interest of limiting the duration of the experiments. Nevertheless, investigating perceptions of other variables would be of interest.

The following section points out some future research areas identified on the basis of the present findings.

7.3 Future research

The relation between speaker-indexical information and linguistic information should be investigated further to advance our knowledge about speaker perception.

This research investigated perceptions of gender-ambiguous speech which was achieved by manipulating F0 and automatic adjustment of FF to changes in F0 (cf. Section 3.2.1). The main goal was to achieve audio samples sounding natural as well as gender-ambiguous. However, the question which arises is whether speaker-indexical information could be retrieved from stimuli in which only F0 was manipulated. Although the effect of somewhat unnatural sounding voice would be achieved, such study would provide the answer to the question of whether indexical information could be still retrieved from phonetic segments. It should be expected that some types of information, for example, speaker social class should be accessed by the listener.

Further investigation of the effect of apparent ethnicity is also required. Experiment 3 showed no effect of apparent ethnicity on speaker gender evaluation. Yet it showed the effect of apparent Asian and Black ethnicity on speaker age evaluation. Therefore, the question which arises from these results is the following: would apparent ethnicity affect gender evaluations and would the results actually show an interlink between gender and ethnicity if the experiment was carried out in a community where the ratio of non-White population to White population was higher, for example in London, or somewhere where the ratio was even lower than in Tyneside? Taking into consideration the fact that there was a non-significant tendency to evaluate the speaker as female when the apparent ethnicity was Black and the fact that Black ethnicity constituted only 1.94% of Tyneside population in comparison with White and Asian, which constituted 92.62% and 6%, respectively, it could be a possible assumption that when ethnic minorities constitute a considerable proportion of the population, the effect of ethnicity does not exist. Perhaps it is the case that the effect could be observed when the ethnic minority constitutes only a small fraction of the population. Further research would shed more light on the question and provide us with information whether the lack of the ethnicity effect on gender evaluations depends on the ratio of ethnicities within the population or whether it is a general finding true in any circumstances.

7.4 Closing remarks

By using gender-ambiguous voices, this thesis has shed new light on our understanding of the perception of speaker indexical information. The results of this research provide evidence in support of earlier findings that listeners are able to recognise speaker-indexical information encoded in phonetic variants. However, by using gender-ambiguous voices this research has improved our understanding of the process. Thus, the innovative results indicate that gender-ambiguity in the speaker's voice does not prevent the listener from recovering speaker-indexical information from phonetic variants.

This thesis also pushed forward our understanding of the role of visual information in speech perception. The innovative results showed that some speaker information may be shifted by visual cues about the supposed speaker. We now have more evidence that there may be perceptual differences between unimodal and multimodal speech perception.

The findings of the experiments presented in this thesis provide more evidence in support of the fact that listeners attend to acoustic information in speech at different levels. They help us understand how linguistic information interacts with non-linguistic information. As far as non-linguistic information in speech is concerned, it involves biological, as well as socially constructed information. While there may be an overlap between some biological and social information, for example, gender, which is in fact grounded in biology, indexical information is socially constructed, providing cues to the speaker's gender, social class as well as age. Overall, it seems that listeners rely on speaker-indexical information quite extensively, so much so that they are able access it even when part of the biological information about the speaker is not present in the acoustic signal.

8 Appendices

8.1 Appendix A: List of stimuli

Variable	**Variant 1**	**Variant 2**	**Variant 3**
The FACE vowel:	*pay* [ɪə]	*pay* [eː]	
The FACE vowel:	*name* [ɪə]	*name* [eː]	
The FACE vowel:	*way* [ɪə]	*way* [eː]	
The GOAT vowel	*grow* [ɵː]	*grow* [oː]	
The GOAT vowel	*stone* [iːə]	*stone* [oː]	
The GOAT vowel	*dough* [ɵː]	*dough* [oː]	
The GOAT vowel	*go* [ʊə]	*go* [oː]	
The GOAT vowel	*home* [iːə]	*home* [oː]	
The GOAT vowel	*toe* [ʊə]	*toe* [oː]	
The NURSE vowel	*nurse* [ɔː]	*nurse* [ɜː]	
The NURSE vowel	*turn* [ɔː]	*turn* [ɜː]	
The NURSE vowel	*fur* [ɔː]	*fur* [ɜː]	
The NURSE vowel	*blur* [øː]	*blur* [ɜː]	
The NURSE vowel	*stir* [øː]	*stir* [ɜː]	
The NURSE vowel	*fur* [øː]	*fur* [ɜː]	
T-to-R	*get off* [ɹ]	*get off* [ʔ͡t]	

T-to-R	*put on* [ɹ]	*put on* [ʔ͡t]	
T-to-R	*shut up* [ɹ]	*shut up* [ʔ͡t]	
Variants of /t/	*bottle* [ʔ]	*bottle* [ʔ͡t]	*bottle* [t]
Variants of /t/	*city* [ʔ]	*city* [ʔ͡t]	*city* [t]
Variants of /t/	*wanted* [ʔ]	*wanted* [ʔ͡t]	*wanted* [t]
Variants of /p/	*copy* [ʔ͡p]	*copy* [p]	
Variants of /p/	*happy* [ʔ͡p]	*happy* [p]	
Variants of /p/	*hippie* [ʔ͡p]	*hippie* [p]	
Variants of /k/	*lucky* [ʔ͡k]	*lucky* [k]	
Variants of /k/	*local* [ʔ͡k]	*local* [k]	
Variants of /k/	*look up* [ʔ͡k]	*look up* [k]	
Pre-aspirated /t/	*bat* [$^{h:}$t]	*bat* [t]	
Pre-aspirated /t/	*strut* [$^{h:}$t]	*strut* [t]	
Pre-aspirated /t/	*mat* [^{h}t]	*mat* [t]	
Pre-aspirated /t/	*kit* [^{h}t]	*kit* [t]	

8.2 Appendix B: White and non-White ethnicities in the UK

Area	White	Asian	Black
England	85.31%	5.55%	3.48%
Tyneside	92.62%	6%	1.94%
Blackburn with Darwen UA	69.05%	26.52%	0.63%
Greater Manchester (Met County)	83.75%	8.11%	2.76%
Bolton	81.81%	12.39%	1.68%
Manchester	66.5%	12.08%	8.64%
Sheffield	83.62%	5.64%	3.63%
Bradford	67.35%	24.89%	1.77%
Kirklees	79.09%	14.99%	1.87%
Leeds	85%	5.73%	3.45%
Leicester UA	50.39%	31.85%	6.24%
Nottingham UA	71.44%	9.07%	7.26%
West Midlands (Met County)	70.07%	15.88%	5.99%
Birmingham	57.89%	22.53%	8.98%
Sandwell	69.9%	16.86%	5.95%
Wolverhampton	67.93%	14.83%	6.94%
Luton UA	54.57%	26.37%	9.8%
Inner and outer London	59.69%	12.1%	13.32%
Milton Keynes UA	79.99%	5.6%	6.88%
Slough	45.53%	33.76%	8.64%
Bristol, City of UA	83.89%	3.62%	6.01%

8.3 Appendix C: Information sheet for Experiments 1, 2 & 3

Welcome to the survey!

Thank you for taking the time to participate in the study. This study is being conducted by Ania Kubisz who is a Ph.D. student at the University of York. The study is supervised by Dr Carmen Llamas and Dr Dominic Watt. Our contact details are listed below.

Feel free to contact Ania Kubisz at ania.kubisz@york.ac.uk at any time with questions or comments relating to the study.

Please read the following information carefully before commencing the study.

What is the research about?

This research is investigating the properties of the accents of the North East of England.

Who can participate?

This study has been designed to be completed by native speakers of British English. Your participation will help us understand the aspects of the accents of the North East of England.

What does the study involve?

You will be asked to listen to a set of recordings and answer questions about the voices you hear.

The study takes about 40 minutes including two breaks.

What will happen to the data I provide?

The answers that you provide will be stored alongside the data from other participants and will not be traceable to you. All data are stored securely and only group results will be published.

Are there any risks to taking part?

No risks in taking part have been identified.

Do I have to take part?

No, participation is entirely voluntary. You are free to quit at any time before the end of the study, and your data will be destroyed and will not be used in the research. To quit, just exit your browser or navigate to a different website.

What about confidentiality?

Your identity and personal information will be kept strictly confidential. Your Internet Protocol address (IP address) will <u>not</u> be recorded.
An IP address is a series of numbers unique to your computer. They facilitate the access of your computer to the Internet.

<u>Are there any benefits to participating?</u>

You will be participating in an exciting research project that will help us understand some of the changes occurring in the English language.
Furthermore, I can provide you with the overall findings of the study. If you are interested in receiving this information then please contact me at the email address below.

Do not hesitate to direct any questions to:

<u>Principal researcher</u>

Ania Kubisz,

Dept. of Language & Linguistic Science,

University of York,
Heslington,
York.
YO10 5DD

Telephone: 07775183999

Email: ania.kubisz@york.ac.uk

<u>Research supervisors</u>

Dr Carmen Llamas,
Dept. of Language & Linguistic Science,
Linguistic
University of York,
Heslington,
York.
YO10 5DD

Dr Dominic Watt,
Dept. of Language &
Science,
University of York,
Heslington,
York.
YO10 5DD

Telephone: 01904 322618 Telephone: 01904 322671

Email: carmen.llamas@york.ac.uk Email: dominic.watt@york.ac.uk

Many thanks,

Ania Kubisz

By clicking the arrow and beginning the study, you confirm that you:

Have read and understood the above information

Understand that the information you provide will be held in confidence by the researcher, and that your name or identifying information about you will not be mentioned in any publication

Understand that you can withdraw at any time before the end of the study if you no longer wish to take part in the survey, and that in such a case all your data will be destroyed

Agree to participate in the study

After you've clicked on the arrow, you'll be asked to answer a few questions about yourself.

8.4 Appendix D: Consent form for Experiments 1, 2 & 3

*This study has been reviewed and approved by the Departmental Ethics Committee of the Department of Language and Linguistic Science at the University of York. If you have any questions regarding this, you can contact the head of the Ethics Committee, Dominic Watt (**email:** dominic.watt@york.ac.uk; **Tel:** (01904) 322671).*

If you have further questions regarding this study, please feel free to contact:

Ania Kubisz

Department of Language and Linguistic Science

University of York, Heslington, York, YO10 5DD

tel: 07775183999

email: ak970@york.ac.uk

Supervisors details:

Dr Carmen Llamas and Dr Dominic Watt

Department of Language and Linguistic Science
University of York
Heslington, York, UK
YO10 5DD
email: carmen.llamas@york.ac.uk,
dominic.watt@york.ac.uk

Title of project: Accent features in varieties of English in the North East of England.

Lead researcher: Ania Kubisz

Consent form

This form is for you to state whether or not you agree to take part in the study. Please read and answer every question. If there is anything you do not understand, or if you want more information, please ask the researcher.

Have you read and understood the information leaflet about the study? Yes ❒ No ❒

Have you had an opportunity to ask questions about the study and have these been answered satisfactorily? Yes ❒ No ❒

Do you understand that the information you provide will be held in confidence by the research team, and your name or identifying information about you will not be mentioned in any publication? Yes ❒ No ❒

Do you understand that you may withdraw from the study at any time without giving any reason, and that in such a case all your data will be destroyed? Yes ❒ No ❒

Do you agree to allowing recordings to be used in the future for secondary research on language? Yes ❒ No ❒

Do you agree to take part in the study? Yes ❒ No ❒

Do you agree to the researcher keeping your contact details after the end of the current project, in order that she may contact you in the future about possible participation in other studies? Yes ❒ No ❒
(You may take part in the study without agreeing to this).

Please give your contact details for sending summary of results.
(Only fill this section if you would like a summary of the results emailed to you).

Your name (in BLOCK letters): ______________________________

Your email address:

____

Your signature: ______________________________

Researcher's name: ______________________________

Date: ______________________________

8.5 Appendix E: Experiment 2 Participants' details

All participants reported having immediate family and/or friends from the North-East. With the exception of one participant who was 34 years old, all other participants were between 18 and 24 years of age. The majority reported having lived in Tyneside all their lives (that is, between 18 and 23 years, with the exception of the last 1 to 8 months when they started their undergraduate degree at the University of York). Two participants were further in their studies and reported having lived in York for 18 months and 2 years respectively. One participant reported having lived in Newcastle for the last 16 years, and one for 34 years.

Participants came from a number of fields of study, such as linguistics and combined degrees with linguistics, psychology, history, nursing, economics and biology.

In terms of social background, 11 participants reported being from middle-class households and 16 described themselves as being working-class. Eight participants reported speaking one or more foreign languages.

As in the first study, the aim was to obtain a gender-balanced sample of participants, but yet again this proved to be difficult in practice. Thus, the sample consists of 7 males and 20 females. In general, it is mostly women who respond to calls for participation in experiments, a fact which also seems to be true in other disciplines, for example in psychology.

As a comparison group, it was decided to use listeners who had low exposure to any of the North-East dialects. Thus, instead of testing listeners from the wider North-East, the comparison group was comprised of native speakers of British English who had low exposure to the variety spoken in the North-East England. The reason behind this is that due to the choice of phonetic variants under investigation in the second study (those of the NURSE vowel, T-to-R, and glottalised /t/), there may be not very much difference in the results between a group of Tyneside listeners and listeners from elsewhere in the North-East. It was assumed that using participants who are speakers of varieties of English from outside of the North-East would show more differences.

The comparison group of participants consisted of 31 listeners. Listeners in this group originated from the south of England and were not familiar with the dialect under investigation. 8 listeners were from London or neighbouring boroughs. The majority of participants originated form Greater London, Sussex and Kent, but also Oxfordshire, Buckinghamshire, Berkshire and Surrey. Participants were from localities as far east as Margate in Kent and Colchester in Suffolk, as far west as Devon, as far north as East Northamptonshire and all the way south (for example, from Hastings and Brighton in East Sussex and Chichester in West Sussex). Although participants originated from a greater number of places in geographical terms than the Tyneside group of respondents, they were all from the south of the country, thus fulfilling the fundamental requirement of not being familiar with Tyneside English. Participants in the comparison group had never lived in the North-East of England. The majority of participants were up to two months into the first year of their undergraduate degrees and living away from home for the first time. Only one participant reported having lived in York for 2 years. 13 participants reported having a flatmate/s, a parent or further relatives as their close contacts from the North-East. However, this information might be somewhat unreliable because when asked if they had ever lived in the North-East some participants confirmed that they considered York to be a North-Eastern town.

All participants in this group were between 18 and 24 years of age, and so age-wise they were similar to the Tyneside group of respondents.

This group of respondents also was not gender-balanced, as it consisted of 19 males and 12 females. Interestingly, in this group of listeners it was mostly men who responded to the call for participation. The reason behind this could be that the majority of participants were students of economics, electronics and combined degrees with economics and electronics, (e.g. music technology) which are fields usually chosen by men. The third field of study in terms of numbers of participants was psychology. The majority of participants in this field of study were female. Only two participants reported studying linguistics, and they were also female.

In this group of participants only one person reported being working-class. The remaining 31 participants reported being middle-class. In this group, 7 participants reported having knowledge of one or more foreign languages.

In addition to the call for participation stating clearly that only participants from the South of the country were being sought for, it was confirmed with each of the volunteers individually that they met the criteria for participating in the study. One participant was deleted from the analysis for not fulfilling the main criterion, as (s)he came from Wrexham in Wales.

9 References

Abbate, C. S., Ruggieri, S. & Boca, S. (2013). Automatic Influences of Priming on Prosocial Behavior. *Europe's Journal of Psychology*, 9(3). [Online]. Available at: http://ejop.psychopen.eu/article/view/603/html [Accessed 10 November 2015].

Ahearn, E. P. (1997). The use of visual analog scales in mood disorders: A critical review. *Journal of Psychiatric Research*, 31(5); 569-579.

Aitken, R. C. B. (1969). Measurement of Feelings Using Visual Analogue Scales. *Journal of the Royal Society of Medicine,* 62(10), 989-993.

Adobe Audition 3.0 User Guide. 2007. *Adobe Systems Incorporated*

Adobe Photoshop CS6 Extended. (2013). *Adobe Systems Incorporated.*

Agha, A. (2003). The social life of cultural value. *Language and Communication*, 23, 231–273.

Agha, A. (2004). Registers of Language. In A. Duranti (Ed). *A Companion to Linguistic Anthropology*. Oxford: Blackwell, pp.23-45.

Assmann, P.F., Nearey T. M. & Dembling, S. (2006). Effects of frequency shifts on perceived naturalness and gender information in speech. *Proceedings Ninth Int. Conf. Spoken Language Processing, pp. 889-892, Pittsburgh, PA, September 17-21, 2006.*

Assmann, P. F. & Nearey, T. M. (2007). Relationship between fundamental and formant frequencies in voice preference. *Acoustical Society of America,* 122(2), 35-43.

Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Baayen, R. H. (2013). Multivariate statistics. In R. J. Podesva & D. Sharma, (Eds). Research Methods in Linguistics. Cambridge: Cambridge University Press, pp. 337-372.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.

Balconi, M. & Ferrari, C. (2012). Subliminal and Supraliminal Processing of Facial Expression of Emotions: Brain Oscillation in the Left/Right Frontal Area. *Brain Sciences, 2*, 85-100.

Bargh, J. A. (2006). Bypassing the Will: Toward Demystifying the Nonconscious Control of Social Behavior. In R.R. Hassin, J.S. Uleman & J.A. Bargh (Eds.). *The New Unconscious.* Unconscious processing. Oxford University Press, pp. 37-60.

Bargh, J. A., Chaiken, S., Raymond, P. & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology,* 32, 185–210.

Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action. *Journal of Personality and Social Psychology*, 71(2), 230-244.

Bargh, J. A. & Chartrand, T. L. (2014). The Mind is in the Middle: A Practical Guide to Priming and Automaticity Research. In H. T. Reis & C.M. Judd (Eds). *Handbook of Research Methods in Social and Personality Psychology.* New York: Cambridge University Press (2nd edition), pp. 373-403.

Barr, D. J., Levy, R., Scheepers, C. & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.

Barreda, S. & Nearey, T. M. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *The Journal of the Acoustical Society of America* , 131(1), 466–477.

Bates, D. & Maechler, M. (2015). lme4: Linear mixed effects models using S4 classes. R package version 3.2.2. Retrieved from http//CRAN.R-project.org.

Beal, J., Burbano Elizondo, L. & Llamas, C. (2012). *Urban North-Eastern English: Tyneside to Teesside.* Edinburgh: Edinburgh University Press.

Bernstein, L. E. (2012). Visual speech perception. In G. Bailly, P. Perrier, E. Vatikiotis-Beteson (Eds.). *Audiovisual Speech Processing*. Cambridge: Cambridge University Press, pp. 21-39.

Biemans, M. 2000. *Gender Variation in Voice Quality.* Unpublished: University of Utrecht. PhD.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in gender stereotyping. *Journal of Personality and Social Psychology,* 70, 1142–1163.

Bucholtz, M. & Hall, K. (2010). Locating Identity in Language. In C. Llamas & D. J. L. Watt, (Eds). *Language and identities.* Edinburgh: Edinburgh University Press, pp. 18-28.

Buchstaller, I., Corrigan. K. P., Holmberg, A., Honeybone, P. & Maguire, W. (2013). T-to-R and the Northern Subject Rule: questionnaire-based spatial, social and structural linguistics. *English Language and Linguistics,* 17(1), 85–128.

Buchwald, A. B., Winters, S. J., Pisoni, D. B. (2009). Visual speech primes open-set recognition of spoken words. *Language and Cognitive Processes,* 24(4), 580-610.

Burbano-Elizondo, L. (2008). *Language variation and identity in Sunderland*. Unpublished: University of Sheffield. PhD.

Cheshire, J. (1999). Spoken standard English. In T. Bex & R. Watts, (Eds). *Standard English: The Widening Debate.* London: Routledge, pp. 129-148.

Corrigan, K. P. (2012). GOAT vowel variants in the diachronic electronic corpus of Tyneside English (DECTE). In T. Nevalainen & E. Closs Traugott (Eds). *The Oxford Handbook of the History of English*. Oxford University Press, pp. 90-93.

Clopper, C., Conrey B., Pisoni D.B. (2005). Effects of Talker Gender on Dialect Categorization. *Journal of Language and Social Psychology,* 24, 182-206.

Coleman, R. O. (1971). Male and Female Voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, 14, 565-577.

Coleman, R. O. (1976). A Comparison of the Contributions of Two Voice Quality Characteristics to the Perception of Maleness and Femaleness in the Voice. *Journal of Hearing and Speech Research*, 19, 168-180.

Couper, M. P., Tourangeau, R., Conrad, F. G. & Singer, E. (2006). Evaluating the Effectiveness of Visual Analog Scales: A Web Experiment. *Social Science Computer Review* 24 (2); 227-245.

Davis, C. & Kim, J. (2006). Audio-visual speech perception off the top of the head. *Cognition,* 100, B21–B31.

Dijksterhuis, A. (2013). Automaticity. In D.E. Carlston, (Ed). *The Oxford Handbook of Social Cognition. New York: Oxford University Press,* pp. 239-256.

Dijksterhuis, A., Aarts, H., Bargh, J. A. & van Knippenberg, A. (2000). On the Relation between Associative Strength and Automatic Behavior. *Journal of Experimental Social Psychology*, 36, 531-544.

Dijksterhuis, A., Aarts, H., Smith P. K. (2006). The Power of the Subliminal: On Subliminal Persuasion and Other Potential Applications. In R.R. Hassin, J.S. Uleman & J.A. Bargh (Eds). *The New Unconscious.* Oxford University Press, pp.77-106.

Dijksterhuis, A. & van Knippenberg, A. (1998). The relation between perception and behavior or how to win game of Trivial Pursuit. *Journal of Personality and Social Psychology*, 74(4), 865-877.

Docherty, G. J., Foulkes, P., Milroy, J., Milroy, L. & Walshaw, D. (1997). Descriptive adequacy in phonology: A variationist perspective. *Journal of Linguistics 33*, 275–310.

Docherty, G. J. & Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. In P. Foulkes & G. Docherty, (Eds). *Urban Voices: Accent Studies in the British Isles*. London: Arnold, 46–71.

Drager, K. (2005). From bad to bed: The relationship between perceived age and vowel perception in New Zealand English. Te Reo, 48, 55–68.

Drager, K. (2011). Speaker Age and Vowel Perception. *Language and Speech* 54(1) 99–121.

Eckert, P. (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Oxford: Blackwell.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics,* 12(4), 453–476.

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, 41, 87–100.

Eckert, P. & Podesva, R. (2011). Sociophonetics and sexuality: Toward the symbiosis of sociolinguistics and laboratory phonology. *American Speech*, 86(1). 6-13.

Farnand, S. (2013). *Designing pictorial stimuli for perceptual image difference experiments.* Unpublished: Rochester Institute of Technology. PhD.

Ferguson, M. J. (2008). On Becoming Ready to Pursue a Goal You Don't Know You Have: Effects of Nonconscious Goals on Evaluative Readiness. *Journal of Personality and Social Psychology*, 95(6), 1268–1294.

Foulkes, P., Docherty, G. J., Khattab, G. & Yaeger-Dror, M. (2010). Sound judgements: perception of indexical features in children's speech. In D. Preston & N. Niedzielski, (Eds). *A Reader in Sociophonetics*. Berlin: de Gruyter, 327-356.

Foulkes, P., Docherty, G. J., Watt, D. (2005). Phonological Variation in Child-Directed Speech. *Language*, 81(1), 177-206.

Frowd, C.D. (2015). Facial Composites and Techniques to Improve Image recognizability. In T. Valentine & J. Davis (Eds). *Forensic Facial Identification*. Wiley-Blackwell: pp. 43-70.

Frowd, C., Bruce, V., McIntyre, A. H., Ross, D., Fields, S., Plenderleith, Y. & Hancock, P. J. B. (2006). Implementing Holistic Dimensions for a Facial Composite System. *Journal of Multimedia*, 1(3), 42-51.

Frowd, C., Skelton, F., Hepton, G., Holden, L., Minahil, S., Pitchford, M., McIntyre, A., Brown, C. & Hancock, P. J. B. (2013). Whole-face procedures for recovering facial images from memory. *Science and Justice*, 53, 89-97.

Gal, S. (1989). Language and political economy. *Annual Review of Anthropology,* 18, 345-367.

Gal, S. (2006). Contradictions of standard language in Europe: Implications for the study of practices and publics. *Social Anthropology*, 14(2), 163–181.

Garrett, P.,Coupland, N. & Wiliams, A. (2003). *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance.* Cardiff: University of Wales Press.

Garrett, P., Selleck, C. & Coupland, N. (2011). English in England and Wales: Multiple ideologies. In T. Kristiansen & N. Coupland, (Eds). *Standard Languages and Language Standards in a Changing Europe*. Olso: Novus Press, pp. 57-65.

Gelfer, M. P. & Mikos, V. A. (2005). The Relative Contributions of Speaking Fundamental Frequency and Formant Frequencies to Gender Identification Based on Isolated Vowels. *Journal of Voice*, 19(4), 544-554.

Glaser, W. R. (1992). Picture naming. *Cognition,* 42, 61-105.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.

Haddican, B., Foulkes, P., Hughes, V. & Richards, H. 2013. Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change* 25, 371–403.

Harnsberger, J. D., Shrivastav, R., Brow, W. S., Rothman, H., Jr. & Hollien, H. (2008). Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age. *Journal of Voice*, 22(1), 58-69.

Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4), 865–892.

Hay, J., Nolan, A. & Drager, K. (2006a). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, 23, 351-379.

Hay, J., Warren, P. & Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of* Phonetics, 34, 458–484.

Hayes, M. H. S. & Patterson, D. G. (1921). Experimental Development of the Graphic Rating Method. *Psychological Bulletin* 18, 98–99.

Hillenbrand, J. & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, 71(5), 1150–1166.

Hubbard, D. J. & Assmann, P. F. (2013). Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency. *The Journal of the Acoustical Society of America,* 133(4), 2367–2376.

Irvine, J. T. & Gal, S. (2009). Language Ideology and Linguistic Differentiation. In A. Duranti (Ed). Linguistci Anthropology: A Reader. Wiley-Blackwell (2nd edition), pp. 402-434.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485–499.

Johnson, K., Strand, E. A. & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics,* 27, 359-384.

Johnstone, B. (2010). Locating Language in Identity. In C. Llamas & D. J. L. Watt, (Eds). *Language and identities.* Edinburgh: Edinburgh University Press, pp. 29-36.

Johnstone, B. (2014). *Beyond "enregistered dialect": Form and meaning in time and space.* [Online presentation]. Available at: https://prezi.com/xlal2k2vpgxn/beyond-enregistered-dialect-form-and-meaning-in-time-and/ [Accessed 30 September 2016].

Johnstone, B., Andrus, J. & Danielson, A. E. (2006). Mobility, Indexicality, and the Enregisterment of "Pittsburghese". *Journal of English Linguistics*, 34(2), 77-104.

Johnstone, B. & Kiesling, S. F. (2008). Indexicality and experience: Exploring the meanings of /aw/-monophthongization in Pittsburgh. *Journal of Sociolinguistics,* 12(1), 5–33.

Jones, M. & Llamas, C. (2008). Fricated realisations of /t/ in Dublin and Middlesbrough English: an acoustic analysis of plosive frication and surface fricative contrasts. *English Language and Linguistics*, 12(3), 419-443.

Joseph, E. J. (2010). Identity. In Language and identities. C. Llamas and D. Watt, (Eds). *Language and identities.* Edinburgh: Edinburgh University Press, pp. 9-17.

Kerswill, P. (2007). Standard and non-standard English. In D. Britain (Ed). *Language in the British Isles.* Cambridge: Cambridge University Press, pp. 34-51.

Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. In D. Britain and J. Cheshire (Eds). *Social Dialectology. In honour of Peter Trudgill.* Amsterdam: Benjamins, pp. 223-243.

Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America,* 87(2). 820-857.

Kraljic, T., Samuel, A. G. (2006). Generalization in perceptual learning for speech. Psychonomic Bulletin & Review, 13(2), 262-268.

Kraljic, T., Samuel, A. G., Brennan, S. E. (2008). First Impressions and Last Resorts. How Listeners Adjust to Speaker Variability. Association for Psychological Science, 19(4), 332-338.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change,* 2, 205-254.

Labov. W. (2001). Principles of Linguistics Change. Social Factors. Vol. 2. Wiley-Blackwell.

Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics, 34,* 500–515.

Lacerda, F. (1998). An Exemplar-Based Account of Emergent Phonetic Categories. *Institute of Linguistics, Dept. of Phonetics, Stockholm University, S-106 91 Stockholm, Sweden.*

Ladefoged, P. (1996). *Elements of acoustic phonetics*. Chicago; London: University of Chicago Press.

Lancaster, B. (2007). *The North East, England's most distinctive region?* In B. Lancaster, D. Newton and N. Vall (eds.). An Agenda for Regional History. Newcastle upon Tyne: Northumbria University Press, pp.23-41.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America*, 59, 675-678.

Laver, J. 2002. *Principles of phonetics*. Cambridge.

Likert, R. (1932). The method of constructing an attitude scale. In M. Fishbein (Ed). *Attitude theory and measurement*. London: Wiley, pp. 90–95.

Llamas, C. (2000). Middlesbrough English: Convergent and divergent trends in a 'part of Britain with no identity'. *Leeds Working Papers in Linguistics,* 8, 123-148.

Llamas, C. (2001). *Language Variation and Innovation in Teesside English*. Unpublished: The University of Leeds. PhD.

Llamas, C. (2007). "A place between places": Language and identities in a border tow. *Language in Society,* 36, 579–604.

Llamas, C. & Watt, D. (2014). Scottish, English, British?: Innovations in attitude measurement. *Language and Linguistics Compass,* 8(11), 610-617.

Luce, P. & Lyons, E.A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, 26(4), 708-715.

Mackenzie Beck, J. (1997). Organic variation on the vocal apparatus. In W. J. Hardcastle & J. Laver (Eds.). *The handbook of phonetic sciences.* Oxford: Blackwell, pp. 256-197.

Magnuson, J. S. & Nusbaum, H. C. (2007). Acoustic Differences, Listener Expectations, and the Perceptual Accommodation of Talker Variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 391–409.

Maguire, W. N. (2007). *What is a merger and can it be reversed? The origin, status and reversal of the "NURSE-NORTH merger" in Tyneside English.* Unpublished: Newcastle University. PhD.

Mani, N. & Schneider, S. (2013). Speaker Identity Supports Phonetic Category Learning. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 623–629.

Marsh-Richard, D. M., Hatzis, E. S., Mathias, C. W., Venditti, N. & Dougherty, D. M. (2009). Adaptive Visual Analog Scales (AVAS): A Modifiable Software Program for the Creation, Administration, and Scoring of Visual Analog Scales. *Behav Res Methods*, 41(1), 99–106.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Meng, H., Oviatt, S., Potamianos, G. & Rigoll, G. (2009). Introduction to the Special Issue on Multimodal Processing in Speech-Based Interactions. IEEE Transactions on Audio, Speech, and Language Processing, 17 (3), pp. 409-410.

Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics,* 5(4), 530-555.

Milroy, J. (2007). The ideology of the standard language. In C. Llamas, L. Mullany & P. Stockwell, (Eds). *The Routledge Companion to Sociolinguistics.* London: Routledge, pp. 133-139.

Milroy, L. (1999). Standard English and language ideology in Britain and the United States. In *Standard English: The Widening Debate.* T. Bex & R. Watts, (Eds). London: Routledge, pp. 173-206.

Milroy, L. (2000). Britain and the United States: Two Nations Divided by the Same Language (and Different Language Ideologies). *Journal of Linguistic Anthropology,* 10(l), 56-89.

Milroy, L. (2004). Language ideologies and linguistic change. In Carmen Fought (Ed.) *Sociolinguistic Variation: Critical Reflections*. Oxford: Oxford University Press, pp. 161–177.

Milroy, J. & Milroy, L. (1985). Authority in language. Investigating language prescription and standardisation. London: Routledge & Kegan Paul.

Milroy, J. & Milroy, L. (1997). Exploring the social constraints on language change. In S. Eliasson & E. H. Jahr (Eds). *Trends in Sociolinguistics: Studies and Monographs 100. Language and its Ecology: Essays in Memory of Einar Haugen*. Berlin, New York: Mouton de Gruyter, pp. 75-102.

Milroy, J., Milroy L. & Hartley, S. (1994a). Local and supra-local change in British English: the case of glottalisation. *English World-Wide,* 15, 1-33.

Milroy, J., Milroy, L., Hartley, S. & Walshaw, D. (1994b). Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. *Language Variation and Change,* 6, 327-357.

Montgomery, C. (2015). Perceptual Ideology across the Scottish/English Border. In D. Watt & C. Llamas (Eds). *Language, Borders and Identity*. Edinburgh: Edinburgh University Press, pp. 118-136.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47, 379-390.

Munson, B. (2007). The Acoustic Correlates of Perceived Masculinity, Perceived Femininity, and Perceived Sexual Orientation. *Language and Speech,* 50(1), 125–142.

Munson, B. (2011). The influence of actual and imputed talker gender on fricative perception, revisited (L). *Journal of the Acoustical Society of America.,* 130(5), 2631– 2634.

Munson, B. & Babel, M. (2007). Loose Lips and Silver Tongues, or, Projecting Sexual Orientation Through Speech. *Language and Linguistics Compass,* 1(5), 416–449.

Murry, T. & Singh, S. (1980). Multidimensional analysis of male and female voices. *Journal of the Acoustical Society of America*, 68(5), 1294-1300.

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, 18(1), 62-85.

O'Callaghan, C. (2012). Perception and Multimodality. In: E. Margolis, R. Samuels & S. P. Stich (Eds.). *Oxford Handbook of Philosophy and Cognitive Science.* New York: Oxford University Press, pp. 92-117.

Office for National Statistics. (2011). *Census 2011.* London: Census gives insights into characteristics of the North East's population. [Online]. Available at: http://www.ons.gov.uk/ons/dcp29904_291540.pdf [Accessed 11 October 2015].

Pearce, M. (2009). A Perceptual Dialect Map of North East England. *Journal of English Linguistics,* 20(10) , 1-31.

Pearce, M. (2012). Folk accounts of dialect differences in Tyne and Wear. *Dialectologia et Geolinguistica*, 20, 5-25

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds). *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins, pp. 137–158.

Podesva, R. (2011). The California vowel shift and gay Identity. *American Speech*, 86(1), 32-51.

Preston, D. R. (1993). Folk dialectology. In D. R. Preston (Ed). *American Dialect Research.* Amsterdam: John Benjamins, pp. 333-378.

Public Health England. (2011). Black and ethnic minorities. [Online]. Available at http://www.nepho.org.uk/topics/Black%20and%20ethnic%20minorities [Accessed 9 October 2015].

Pufahl, A. & Samuel, A.G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology*, 70, 1–30.

Purnell, T., Idsardi, W. & Baugh, J. (1999). Perceptual and phonetic experiments in American English dialect identification. *Journal of Language and Social Psychology,* 18, 10–30.

Redinger, D. & Llamas, C. (2015). Multilingual Luxembourg: Language and Identity at the Romance/Germanic Language Border. In D. Watt & C. Llamas (Eds). *Language, Borders and Identity*. Edinburgh: Edinburgh University Press, pp. 169-185.

Remez, R. E. (2012). The Puzzles of multimodal speech perception. In G. Bailly, P. Perrier, E. Vatikiotis-Beteson (Eds.). *Audiovisual Speech Processing*. Cambridge: Cambridge University Press, pp. 4-20.

Remez, R.E., Fellowes, J. M. & Rubin, P.E. (1997). Talker Identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance,* 23, 651-666.

Rosenblum, L. D. (2005). Primacy of Multimodal Speech Perception. In: D. B. Pisoni, & R. E. Remez (Eds.). *The Handbook of Speech Perception*. Oxford: Blackwell, pp. 51-78.

Sakamoto, S., Hasegawa, G., Abe, T., Ohtani, T., Suzuki, Y. & Kawase, T. (2014). Contribution of detailed parts around talker's mouth for audio-visual speech perception. *The Journal of the Acoustical Society of America* (135), 2414. [Online]. Available at: http://dx.doi.org.ezproxy.york.ac.uk/10.1121/1.4878010 [Accessed 11 October 2016].

Sankoff, G. (2005). Cross-sectional and longitudinal studies in sociolinguistics. In U. Ammon, N. Dittmar, K. J. Mattheier & P. Trudgill, (Ed). *Sociolinguistics: An International Handbook of the Science of Language and Society*. Berlin: Mouton de Gruyter, pp. 1003–1013.

Silverstein, M. (1979). Language structure and linguistic ideology. In R. Clyne, W. Hanks, C. Hofbauer, (Eds). *The Elements: A Parasession on Linguistic Units and Levels*. Chicago: Chicago Linguistic Society, pp. 193-247.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23, 193-229.

Schryver de M., Vandist, K. & Rosseel, Y. (2009). How many exemplars are used? Explorations with the Rex Leopold I model. *Psychonomic Bulletin & Review*, 16(2), 337-343.

Shields, B. J., Cohen, D. M., Harbeck-Weber, C., Powers, J. D., & Smith, G. A. (2003) Pediatric Pain Measurement Using a Visual Analogue Scale: A Comparison of Two Teaching Methods. *Clinical Pediatrics,* 42, 227-234.

Stenner, M.-P., Bauer, M., Sidarus, N., Heinze, H.-J., Patrick Haggard, P. & Dolan, R. J. (2014). Subliminal action priming modulates the perceived intensity of sensory action consequences. Cognition, 130, 227-235.

Stern, R. A. & Bachman, D. L. (1991). Depressive symptoms following stroke. *American Journal of Psychiatry*, 148, 351-356.

Strand, E. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18, 86 -100.

Sumner, M., Kyung Kim, S., King. E. & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1-13.

SurveyGizmo. (2014). [Online] Available at: http://www.surveygizmo.com/

Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. In P. Warren & C. I. Watson, (Eds). *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, University of Auckland, New Zealand. December 6-8, Copyright, Australian Speech Science & Technology Association Inc., pp. 421-426.

Szakay, A. (2008). *Ethnic Dialect Identification in New Zealand. The Role of Prosodic Cues*. VDM Verlag Dr. Mueller.

Teinonen, T., Aslin, R. N., Alku, P. & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition,* 108, 850-855.

Torrance, G. W., Feeny, D. & Furlong, W. (2001). Visual Analog Scales: Do They Have a Role in the Measurement of Preferences for Health States? *Med Decis Making,* 21, 329–334.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. London: Cambridge University Press.

Trudgill, P. (1986). *Dialects in Contact*. Wiley-Blackwell.

Trudgill, P. (1999). Standard English: what is isn't. In *Standard English: The Widening Debate.* T. Bex & R. Watts, (Eds). London: Routledge, pp. 117-128.

Trudgill, P. & Giles, H. (1976). Sociolinguistics and linguistic value judgements: correctness, adequacy and aesthetics. Trier, L.A.U.T.

Vall, N. (2011). *Cultural region. North east England 1945-2000.* Manchester and New York: Manchester University Press.

Viereck, W. (1968). A diachronic-structural analysis of a northern English urban dialect. *Leeds Studies in English*, 65-79. [Online]. Available at: http://www.leeds.ac.uk/lse/lse.html [Accessed 27 May 2014].

Wagner, S. E. (2012). Age Grading in Sociolinguistic Theory. *Language and Linguistics Compass*, 6(6), 371–382

Walker, A. (2007). The effect of phonetic detail on perceived speaker age and social class. *Proceedings 16th ICPhS*, Saarbrücken, 1453-1456.

Watson, K. & Clark, L. (2015). Exploring listeners' real-time reactions to regional accents. *Language Awareness*, 24(1), 38–59.

Watt, D. (1998). *Variation and Change in the Vowel System of Tyneside English*. Unpublished: Newcastle University. PhD.

Watt, D. (2000). Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change* 12, 69–101.

Watt, D. (2002). 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics*, 6(1), 44-63.

Watt, D. & Allen, W. (2003). Tyneside English. *Journal of the International Phonetic Association,* 33, 267-271.

Watt, D. & Milroy, L. (1999). Patterns of variation and change in three Newcastle vowels: is this dialect levelling? In P. Foulkes & G. Docherty, (Eds). *Urban Voices: Accent Studies in the British Isles*. London: Arnold, pp. 25–46.

Wewers, M. E. & Lowe, N. K. (1990). A Critical Review of Visual Analogue Scales in the Measurement of Clinical Phenomena. *Researchin Nursing & Health,*13, 227-236

Wheeler, S. C. & Berger, J. (2007). When the Same Prime Leads to Different Effects. *Journal of Consumer Research*, 34, 357-368.

Woolard, K. A. (2008). Why *dat* now?: Linguistic-anthropological contributions to the explanation of sociolinguistic icons and change. *Journal of Sociolinguistics,* 12(4), 432–452.

Woolard, K & Schieffelin, B. (1994). Language Ideology. *Annual Review of Anthropology,* 23, 55-82.

Zealley, A. K. & Aitken, R. C. B. (1969). Measurement of Mood. *Journal of the Royal Society of Medicine*, 62(10), 993-996.