The Merging of the GOAT and THOUGHT Vowels in Tyneside English: Evidence from Production and Perception

A thesis submitted to the
School of English Literature, Language and Linguistics
for the degree of Doctor of Philosophy

Jasmine Warburton

September 2020

Abstract

This thesis presents a sociolinguistic investigation into the potential merging of the GOAT and THOUGHT vowels in contemporary Tyneside English. Following Watt's (1998) observation that these two vowels are homophonous for some Tyneside speakers, the main study examines the extent to which GOAT and THOUGHT are found to overlap in both speech production and speech perception. Given the lack of previous research on this sound change, evidence from sociolinguistic interviews and experimental perception tasks are evaluated here in order to determine whether there is a current GOAT-THOUGHT merger-in-progress in Tyneside.

The speech production data is made up of conversational interviews with 28 speakers, obtained from the *Diachronic Electronic Corpus of Tyneside English*. An acoustic analysis was performed on this interview data in order to assess the merging of GOAT and THOUGHT in phonetic space. Both static (one-measurement) and dynamic (vowel trajectory) data were collected, and several different measures of merger were employed in the analysis. Results showed that many speakers exhibit considerable overlap of GOAT and THOUGHT in the vowel space due to a lowering of the GOAT vowel, while findings from Pillai scores and Generalised Additive Mixed Models indicated that young females have the greatest overlap between the two vowels.

For the investigation into how GOAT and THOUGHT are perceived in Tyneside, 43 listeners took part in a vowel perception experiment. Discrimination and identification tasks were carried out so that listeners could be assessed on their ability to distinguish and categorise these vowels. Using measures of accuracy, sensitivity, and reaction time, results showed that listeners were significantly less accurate and significantly slower in responding to GOAT-THOUGHT trials than to trials containing control stimuli. Such findings suggest some degree of perceptual confusability between these two vowels for the Tyneside listeners, and provide further evidence of the merging of GOAT and THOUGHT in the speech community.

Comparing the results of the production and perception analyses, the overall findings indicated that GOAT and THOUGHT are merging in Tyneside. However, some evidence of GOAT-fronting is also found in both the production and perception data, which may prevent further progression toward a full phonemic GOAT-THOUGHT merger in Tyneside English.

Acknowledgements

This research was made possible with the help of my supervisory team. I'd first like to thank Danielle Turton for her advice and support over the past 6 years, throughout both my undergraduate and postgraduate degrees. Danielle kept me on the right track with her encouragement and feedback, and helped immeasurably with funding and conference applications. Danielle also encouraged me to develop my analysis skills in R, and to learn new methods and techniques, all of which unquestionably made me a better researcher. Thanks also go to Jalal Al-Tamimi for all of his guidance on speech perception methods, and for his help with statistics. I'd also like to thank Adam Mearns for all of his help with DECTE related matters.

Special thanks go to the Economic and Social Research Council (ESRC) for funding this project, without who this research would not have been possible. I'm also grateful to all of the Tyneside speakers and listeners who took part in this project.

I have met some wonderful fellow PhD researchers over the last few years. I'd like to thank Caitlin for sharing all of her R and LaTeX tips with me, and for engaging in many long discussions about the GOAT vowel. Thanks also go to Dan for sharing his endless knowledge of phonetics with me, and for his advice on research and life in general. I'm also indebted to my writing group pals, Joaquin, Andreas, Sheradan, and Bella, for keeping me productive, sane, and social over the past year or so.

I'd like to thank my family for their unwavering support and encouragement throughout my time at university. I'm not sure I could have done this without my mam, Nicola, and my dad, Ian. Thanks for always being there when I needed to de-stress, and for always believing that I could achieve anything I wanted to achieve. I'm also grateful to my grandparents, John and Audrey, for their constant support.

Finally I'd like to thank my partner, Andrew, who has been my biggest supporter throughout this whole process. Thanks for always listening to me talk about my data and coding problems, and for, more often than not, helping me fix them. More importantly, though less directly helpful to the PhD, thank you for always helping me to put any work difficulties into perspective, for providing food and distractions when necessary, and for always believing in my abilities. You're the best!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Merger, as a sound change, is defined by Labov (1994: 331) as a 'reduction in phonemic inventory' due to the convergence of two or more phonemes. Such a change to phonemic categorisation has implications for both speech production and speech perception, as Gordon (2013: 205) explicitly acknowledges that phonemic merger results in the inability of individuals to either produce or perceive a difference between previously distinct sounds. While phonemic mergers are frequently investigated in sociolinguistic research, many studies have focused predominantly on evidence from speech production. Thomas (2002: 115) comments that perception has historically been sidelined in sociolinguistic studies, with research on speech perception largely restricted to the fields of experimental phonology and psychology. However, there are a number of examples of recent studies on mergers which examine both production and perception, indicating an increased interest in – and emphasis on – speech perception in contemporary sociolinguistic research (Thomas and Hay 2005; Hay *et al.* 2006; Drager 2010; Mok *et al.* 2013; Arnold 2015; Johnson and Nycz 2015; Kendall and Fridland 2017; Freeman 2019).

One contemporary case of potential merger in Tyneside English involves the vowels referred to as GOAT and THOUGHT in Wells' (1982) system of standard lexical sets. First noted in Watt's (1998: 161) study of the Tyneside vowel system, several speakers were found to demonstrate 'indistinguishable' realisations of the GOAT and THOUGHT vowels. This finding leads to a brief speculation on whether there may be a potential GOAT-THOUGHT merger-in-progress in Tyneside English (Watt 1998). However, Watt's (1998) study does not provide any acoustic analysis of these converging vowels, nor any information on how these vowels are perceived in the local speech community. Furthermore, there has been very little subsequent research into this possible vowel merger in recent accounts of Tyneside speech. As

such, there are still many unanswered questions regarding the status of the GOAT and THOUGHT vowels in contemporary Tyneside English, and whether there is any evidence to support the idea of a GOAT-THOUGHT merger in the speech community.

This thesis presents a sociolinguistic investigation into the potential merging of the GOAT and THOUGHT vowels in present-day Tyneside English. The main study is divided into two components; the first examines these two vowels in speech production, while the second focuses on the perception of GOAT and THOUGHT. Analysis of both speech production and perception will allow for a more comprehensive, more detailed examination of whether these vowels show any evidence of phonemic as well as phonetic convergence. The present study is the first to conduct a sociophonetic analysis of GOAT-THOUGHT merging on Tyneside, and the first to investigate how Tynesiders perceive these potentially converging vowels.

1.1 Goals of the Thesis

The overarching aims of the sociophonetic investigation presented in this thesis are stated below in Section 1.1.1, and the research questions that this thesis will aim to answer are detailed in Section 1.1.2.

1.1.1 Primary Thesis Goals

The main goals of this thesis are as follows:

1. To assess the extent of overlap between the GOAT and THOUGHT vowels in the Tyneside speech community

2. To utilise multiple innovative methods of analysis and advanced statistical measures to investigate the merging of GOAT and THOUGHT in both speech production and speech perception

3. To determine whether the overlap between the GOAT and THOUGHT vowels is indicative of a current or ongoing phonemic GOAT-THOUGHT merger in Tyneside

1.1.2 Research Questions

Following on from the primary research goals, the research questions which will be addressed in this thesis are are as follows:

1. To what extent does the evidence, from the production and perception analyses, support the idea of a GOAT-THOUGHT merger-in-progress in contemporary Tyneside English?

   i) Do the results obtained from the production data indicate convergence of the GOAT and THOUGHT vowels in phonetic space?

   ii) Do the results of the perception experiments point to a collapse of a phonemic GOAT-THOUGHT distinction?

   iii) Working within an apparent time framework, do the results show that GOAT and THOUGHT are becoming more merged over time?

2. If the findings of the main study are indicative of a merging of the GOAT and THOUGHT vowels, then:

   i) How do the results suggest that this merger developed (e.g. what type of merger is occurring)?

   ii) In which direction is the convergence of the merger (e.g. does GOAT move to coalesce with THOUGHT, or vice versa)?

   Else, if the results of the main study indicate that GOAT and THOUGHT are not merging, then:

   i) What may be preventing these two vowels from undergoing phonemic merger?

3. What is the nature of the relationship between the production and perception of the GOAT and THOUGHT vowels in Tyneside English?

   i) Is the amount of overlap between GOAT and THOUGHT in production reflected in perception?

   ii) Are the results consistent with previous accounts of the link between production and perception in situations of phonemic merger?

1.2 Structure of the Thesis

This thesis is organised as follows. Chapter 2 gives an overview of previous research on the phenomenon of mergers. It discusses definitions of 'merger', different types of merger, and how these may develop, as well as providing a number of historical and present-day examples of phonemic mergers. Such background information is critical for understanding the process of merger, and

therefore necessary in order to determine whether the GOAT and THOUGHT vowels are merging in Tyneside. Chapter 2 also considers the ways in which mergers may manifest in speech perception, and the nature of the production-perception relationship during mergers-in-progress. The chapter then concludes with a comparison of different methods for assessing whether individuals perceive certain phonetically merging sounds as belonging to the same or different phonemes.

Chapter 3 centres around Tyneside English. First, some necessary information is given on how 'Tyneside English' may be defined, and on the geographical boundaries between which this language variety is spoken. Some distinctive features of the Tyneside accent are then discussed, followed by an overview of the research into attitudes toward Tyneside English. The remainder of Chapter 3 then focuses on the GOAT and THOUGHT vowels in Tyneside. As previous accounts of Tyneside English show, there is a history of variability in the realisation of the GOAT vowel. This will influence the selection of appropriate methods with which to measure a GOAT-THOUGHT merger in speech production, and the presence of multiple GOAT variants in the speech community must be addressed in assessing the evidence of a phonemic merger between GOAT and THOUGHT in Tyneside English. Chapter 3 also briefly examines research conducted on with Tyneside speech data from the 1970s and 1990s to assess the status of GOAT and THOUGHT in older Tyneside English, and to discuss any previous evidence of GOAT-THOUGHT merging in the speech community.

Chapter 4 outlines the methodology for the examination of a potential merger between GOAT and THOUGHT in speech production. It first details how the data were obtained from the *Diachronic Electronic Corpus of Tyneside English*, and discusses which social variables were taken into consideration during participant selection. The chapter then goes on to describe the transcription and measurement processes, summarising how automatic alignment and formant extraction of the target vowels was achieved through the use of the *Forced Alignment and Vowel Extraction* suite (Rosenfelder *et al.* 2011). Multiple analysis methods were employed throughout the production analysis, all of which are discussed in Chapter 4. The use of certain methods was necessitated by the extensive variation previously reported for the Tyneside GOAT vowel, and the chapter details how an auditory analysis of GOAT and methods of merger using dynamic vowel trajectory data were used to account for this variability in the examination of a GOAT-THOUGHT merger. Finally, Chapter 4 explains the statistical methods, such as mixed-effects regression models and generalised additive mixed-models (GAMMs), which were used to aid in the assessment of vowel merger.

Chapter 5 presents the results of the investigation into GOAT-THOUGHT

merging in the speech production data. Formant values and Pillai scores reveal that many of the Tyneside speakers exhibit considerable phonetic overlap of the GOAT and THOUGHT vowels. Then, the findings of an auditory analysis of GOAT are presented, showing that several variants of the vowel are still present – although very infrequent – in contemporary Tyneside English. Chapter 4 continues on to discuss the outcomes of an analysis of vowel trajectory data using GAMMs. The outputs of these statistical models suggest a decrease in the distance between GOAT and THOUGHT in apparent time. However, results are complicated by the exhibition, among some young male speakers, of fronted GOAT vowels. This chapter concludes with a brief focus on vowel duration, with results indicating a lack of any significant durational difference distinguishing GOAT from THOUGHT in Tyneside.

Chapter 6 shifts focus to the speech perception component of the main study. This chapter discusses the methodology of perception tasks, beginning with an explanation of the experimental design and the stimuli creation process. Borrowing elements from speech perception studies more rooted in the subfields of laboratory phonology and psycholinguistics than sociolinguistics, the present study utilised two different types of experiment – both a discrimination task and an identification task – to investigate the perception of the GOAT and THOUGHT vowels among Tyneside listeners. Chapter 6 also describes how participants were recruited, and details the procedure of conducting the experiments. Also explained in the chapter, the data collected from the perception tasks were analysed in terms of accuracy, reaction time, and sensitivity. A subset of the participants in the perception tasks additionally provided speech samples via recordings of minimal pair lists. The chapter describes this recording process, explaining that collecting production and perception data from the same participants will allow for a direct comparison of how individuals' produce and perceive the GOAT and THOUGHT vowels.

Chapter 7 presents the results from the perception experiments. Measures of accuracy and reaction time indicated that, for both the discrimination and identification tasks, listeners were significantly less accurate and significantly slower in their responses to GOAT-THOUGHT vowel stimuli compared with other vowel pairs included in the experiments. Sensitivity measures also suggested that listeners had more difficulty in distinguishing the GOAT and THOUGHT vowels than the other vowel pairs presented to them. The chapter also looks for evidence of perceptual differences among different listener groups (i.e. sex and age groups), ending with a comparison of the production and perception of GOAT and

THOUGHT for the listeners who performed least and most accurately in the perception tasks.

Chapter 8 discusses the results of the production and perception analyses in greater detail, both individually and together, to assess the overall evidence for a GOAT-THOUGHT merger-in-progress in contemporary Tyneside speech. In addition, this chapter explores the evidence suggesting that the GOAT and THOUGHT vowels are converging in Tyneside via a merger-by-approximation situation, and also comments on the probability of the GOAT and THOUGHT vowels undergoing a complete phonemic merger in the speech community.

Finally, Chapter 9 summarises the strengths and limitations of the main study, with particular reference to the main aims and research questions of the thesis. Some suggestions for future research into GOAT-THOUGHT merging are also presented, before the chapter closes with a conclusion of the main findings of the thesis.

Chapter 2

Understanding Phonemic Mergers

This chapter presents a review of the literature on phonological mergers. First, in order to understand what is meant by 'merger', some definitions of the term are explored in Section 2.1, and the different mechanisms by which mergers arise are presented in Section 2.2. A number of example cases of mergers are then presented; Section 2.3 details several completed historical mergings, and Section 2.4 examines more contemporary examples of conditioned and unconditioned mergers. The unidirectionality of mergers is then discussed in Section 2.5, where arguments of mergers as irreversible sound changes are evaluated in light of cases which appear to provide evidence of 'unmergings'. Section 2.6 looks at the phenomenon of 'near mergers', investigating why their existence is controversial and discussing a well known example of near merger. Then, Section 2.7 discusses the relationship between speech production and speech perception in cases of merger and merger-in-progress. Finally, the chapter concludes with an exploration in Section 2.8 of how the perception of phonemic mergers may be tested, reviewing a number of different methods used in sociolinguistic studies of merger perception.

2.1 What is a Merger?

In discussing the phenomenon of mergers, Labov (1994: 310) notes the common definition of merger as when 'two vowels fell together'. Such a definition, however, gives little information regarding the process of merger, nor of the phonetic and phonological consequences of the loss of a phonemic distinction. Additionally, the phrase 'fell together' suggests movement of both vowels to an intermediate position, which, as discussed in further detail in the following sections, is not always the case. Labov (1994: 331) goes on to further describe mergers as a 'reduction in phonemic inventory' due to the coalescence of multiple phonemes: 'two phonemes become one'

(Labov 1994: 321).

If, as Labov proposes, 'merging' can be understood as a process which causes a reduction in one's phonemic inventory, the question arises as to how this reduction manifests in speech production and perception. Gordon (2013: 205) explains that the outcome of a merger is that individuals are no longer able to produce a distinction between two sounds in their own speech, nor are they able to perceive a difference between these sounds. Those not participating in the change, naturally, will continue to maintain a contrast between these sounds in both production and perception. Hickey (2004: 125) also highlights that for speakers who have come to merge two previously distinct sounds, the merged sounds will shift as one in future sound changes.

Wells (1982: 99) describes mergers in a similar way to Labov, in that 'previously contrasting phonemes now cease to contrast'. This definition again highlights that mergers involve the coming together of two phonemes, while emphasising that, in order to identify a merger, we must have knowledge of a previous distinction between two or more converging phonemes. The importance of diachrony in research of mergers is further highlighted by Trudgill (1982: 88), who states that evidence of mergers is most often found by comparing earlier stages of language to later, more contemporary speech. Trudgill (1982: 88) continues that, while usually able to identify stages of pre- and post-merger, linguists are somewhat 'ignorant' as to what goes on during the actual process of merger. The implication here is that the historical development of mergers is generally well known and easily identified, whereas the synchronic study of merger falls behind. However, in more recent years, studies of mergers-in-progress are very common, (examples include: Majors 2005; Hay *et al.* 2006; Labov *et al.* 2006; Baranowski 2013; Hall-Lew 2013; Arnold 2015; Freeman 2019), and thus there has been an increase in research investigating what happens during the merging of two sounds, and how this affects speech production and speech perception.

Maguire *et al.* (2013: 231) propose that 'merger' may actually refer to two separate phenomena: 'merger as change' and 'merger as a synchronic state'. In the sense of phonological change, Maguire *et al.* explain that 'merger' refers to the 'loss of a phonetic differentiation' causing two or more phonemic categories to come together as one. This is similar to the descriptions of 'merger' given in Labov (1994), Wells (1982) and in much of the literature on phonological mergers. However, Maguire *et al.* (2013: 230) posit that 'merger' may also be understood as descriptor for a lack of a phonemic distinction which is 'characteristic of earlier forms' of certain language varieties. Using the example of the FOOT-GOOSE merger

in Standard Scottish English, Maguire *et al.* (2013: 230) highlight that the absence of a distinction is not always the result of 'merging' of two phonemes, but may be due to other factors, such as the imperfect learning of the /ʊ/-/uː/ contrast by Scottish speakers learning English in the 1600s. Therefore, discussions of a 'merger' of FOOT-GOOSE in Standard Scottish English are only accurate in the synchronic sense i.e. when comparing this variety to other varieties which distinguish these vowels, but not in the sense of a historical merging of two phonemic categories. While Maguire *et al.* (2013: 230) argue for a separation of the diachronic and synchronic definitions of 'merger', they also propose that mergers are better understood when looked at from both diachronic and synchronic perspectives.

2.2 How do Mergers Develop?

If a merger, in the sense of phonological change, can be broadly defined as the loss of a distinction between two or more previously separate phonemes, one primary question to ask is how these phonemes come to coalesce with one another. Labov (1994) describes three main mechanisms via which mergers may arise; by approximation (or drift), by transfer, and by expansion.

2.2.1 Merger-by-Approximation

Merger-by-approximation, sometimes termed 'merger-by-drift', refers to the 'gradual approximation' of two or more phonemes in the vowel space until they are no longer distinct (Harris 1985: 308; Labov 1994: 321). Figure 2.1 shows an illustration of the process of merger-by-approximation, presented in Harris (1985). Within the figure, uppercase A and B refer to lexical sets, and the lowercase x, y and z represent phonemes.

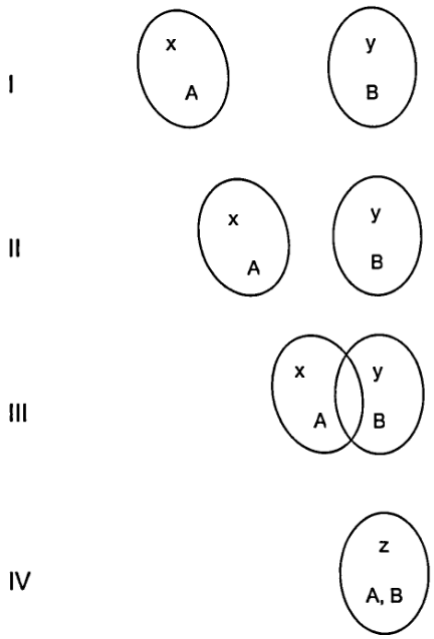


Figure 2.1: Merger-by-Approximation (*source*: Harris 1985: 310)

As Figure 2.1 shows, at Stage I the phonemes x and y are completely distinct, as are the lexical sets of A and B. At this initial stage, phoneme x occurs in lexical set A, while phoneme y occurs in lexical set B. The phonemes remain distinct in Stage II, although they are closer together in phonetic space than before. By Stage III however, x and y, are shown as overlapping. Harris' (1985) diagram here indicates that at this stage in the merger-by-approximation process, the two phonemes x and y may be variably produced as phonetically identical while remaining phonemically distinct. As the figure shows phonemic overlap of x and y, this model of merger-by-approximation also indicates that there may be variable perceptual difficulty in distinguishing the two phonemes at Stage III. Finally, once the merger is completed at Stage IV, x and y are merged together as phoneme z, which will occur in both lexical sets A and B.

Labov (1994: 321) notes that, as merger-by-approximation is a phonetically gradual process, this allows for merged forms to converge in a phonetically intermediate position. Figure 2.1 depicts a situation in which x slowly converges on y so that, in speech production, the merged vowel will have a phonetic realisation of the y phoneme. Another possible merger would be the converse, whereby y drifts phonetically closer to x before eventually converging. Also possible, both x and y may pull toward each other, finally merging so that the quality of the merged vowel is intermediate between x and y. Figure 2.1 therefore presents only one of a number of possibilities in which a merger-by-approximation may occur.

One question that arises from the process of merger-by-approximation, in

reference to Figure 2.1, is whether Stage III always leads to Stage IV; does a phonemic overlap between x and y always necessarily lead to a complete merging of the two phonemes? Harris (1985: 313) suggests that, in such situations of variable phonemic overlap, only knowledge of 'the subsequent history of the vowels' can answer the question of whether the phonemes have undergone a complete merger-by-approximation. Therefore, while overlap in the phonetic distributions of two phonemes, or some perceptual ambiguity between x and y, is a good sign of a merger-by-approximation at work, progress toward a complete phonemic merger from this stage is not a certainty. Dinkin (2016: 165) argues that in possible approximation cases, researchers describing a sound change as a 'merger-in-progress' must provide evidence that the overlap of two phonemes is 'likely' to continue on to full phonological merger.

There are many examples of merger-by-approximation in the linguistic literature. Labov (2010) reports on the approximation of the CURE vowel to the position of FORCE, resulting in the merging of the two vowels in New York City. In New Zealand, Gordon and Maclagan (2001) found evidence of a merger-by-approximation caused by the gradual raising of the SQUARE vowel, so that it now occupies the same space as the NEAR vowel. Perhaps the most notable example is the merging of the low-back vowels, as a number of studies in several different speech communities in the US have described a progressing merger-by-approximation of the LOT and THOUGHT vowels (see: Bigham (2010) on Southern Illinois; Baranowski (2013) on Charleston, South Carolina; and Kendall and Fridland (2017) on speakers from Nevada and Oregon). Not all of these studies find the same route of approximation, however, as Bigham (2010) finds evidence that LOT is raising and backing toward THOUGHT in Southern Illinois, while Baranowski (2013) finds evidence of both LOT-backing and THOUGHT-lowering in Charleston.

2.2.2 Merger-by-Transfer

Labov (1994: 321) defines a merger-by-transfer as the gradual transference of lexical items from one phonological class to another. First writing about this type of merger in their study of vowel mergers in East Anglia, Trudgill and Foxcroft (1978: 73) state that merger-by-transfer allows for speakers to conceivably 'one by one, transfer lexical items from one lexical set to another'. Figure 2.2 below presents Harris' (1985) diagram showing the process of a merger-by-transfer. As before, A and B represent lexical sets, and x and y represent phonemes.

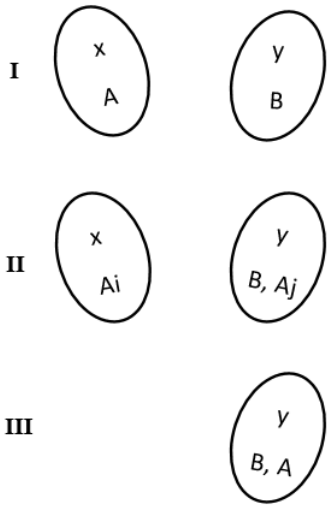


Figure 2.2: Merger-by-Transfer (*source*: Harris 1985: 300)

As depicted in Figure 2.2, at Stage I of the sound change, phonemes x and y and lexical sets A and B are distinct. At Stage II, lexical set A has split into two, with a subset of lexical items transferring to phoneme y while the rest remain associated with phoneme x. As shown in the figure, neither x nor y are moving toward the other, highlighting the phonetically abrupt and lexically gradual nature of this type of merger. Upon completion of a merger-by-transfer, as represented in Stage III in the diagram, all of the lexical items traditionally associated with phoneme x have moved to phoneme y. Of course, another example of merger-by-transfer may instead involve a lexical set associated with phoneme y transferring to phoneme x, although, unlike a merger-by-approximation, mergers-by-transfer are too phonetically abrupt to result in any intermediate phonetic forms. Harris (1985: 301) also points out that the transfer of lexical items from one phoneme to another does not necessarily lead to a phonological merger; e.g a new set of words may transfer to phoneme x from another phoneme, maintaining a phonemic separation of x and y.

Mergers-by-transfer, Labov (1994: 321) claims, are more frequent in situations where a certain linguistic variable is subject to social evaluation, and are therefore often motivated by 'change from above'. Trudgill and Foxcroft (1978: 78) suggest that southern East Anglians began transferring words traditionally belonging to the regionally marked /ʊu/ class to the /ʌu/ class due to an awareness of how differently they pronounced these words to neighbouring London speakers, resulting in merger-by-transfer. More recently, Baxter (2010: 18) hypothesises that the *merry-marry* merger found in many varieties of contemporary Quebec English may be the result of merger-by-transfer. She notes that, while there is no apparent social awareness of the merging of these two vowels among speakers today, historical evidence indicates that

the use of /a/ in certain words (e.g. *marry, barrel, wheelbarrow*) was at one point a stigmatised feature of Canadian English This may have prompted the transference of these words from /a/ to /æ/, resulting in the merging of *merry* and *marry*.

While merger-by-transfer is typically understood as individual lexical items transferring from one phonemic class to another, both Herold (1990: 53) and Trudgill and Foxcroft (1978: 73) suggest that there may be examples of this type of merger in which transference is not purely lexical. Dinkin (2016) finds evidence of this in his study of the low-back merger in Upstate New York, whereby lexical items transfer from one phoneme to another based on their phonological context. Lexical items containing LOT vowels followed by an /l/ and a labiovelar consonant (e.g. *golf, revolve*) were found to typically be produced with a THOUGHT vowel by speakers who otherwise maintained a distinction between these two vowels. This mechanism of merger, termed 'phonological transfer', is defined by Dinkin (2016: 183) as the 'replacement of one phoneme in a given phonological environment with a different phoneme'. A similar situation was recently discovered by Mellesmoen (2018) in British Columbia, where speakers were found to produce [eɪ] vowels followed by a [g] (e.g. *plague, bagel*) as [æ]-like, while /eɪ/ and /æ/ otherwise remained distinct.

Phonological transfer is similar to traditional definitions of merger-by-transfer, which Dinkin (2016: 164) dubs 'lexical transfer', in that it is a phonetically abrupt process. Just as in lexical transfer, words which meet a certain phonological condition transfer to another phoneme 'without passing through the phonetically intermediate space' (Dinkin 2016: 164). Thus, neither lexical or phonological transfer allow for realisations between phonemes x and y. Due to Dinkin finding words of only one phonological condition (i.e. LOT followed by /l/ + labiovelar) having transferred from one phoneme to another, there is therefore no evidence of phonological transfer leading to a complete phonemic merger. However, the implication of phonological transfer as a mechanism of merger is that, one at a time, groups of words containing the merging vowel in different phonological environments would transfer from one lexical set to another.

2.2.3 Merger-by-Expansion

The third mechanism by which a merger may occur, merger-by-expansion, refers to a loss of a phonemic distinction without a reduction in phonetic space. This type of merger situation was first reported by Herold (1990) in her study of the low-back vowel merger in Tamaqua, Pennsylvania. Herold (1990), as recounted in Labov (1994), notices this particular phenomenon while examining the formant values of

the LOT and THOUGHT vowels of a father and his son. Although the father was found to exhibit a vowel distinction, producing an /ɒ/ vowel that was significantly higher and fronter than /ɔ/[1], the son, on the other hand, displayed no significant distinction between pronunciations of /ɒ/ and /ɔ/. Despite this, Herold (1990) found that the phonetic space for these vowels was almost identical for both father and son. Labov (1994: 322) summarises merger-by-expansion as when the phonetic space of the new merged phoneme 'is roughly equivalent to the union of the range of the two phonemes that merged'. A visual representation of merger-by-expansion is displayed in Figure 2.3.

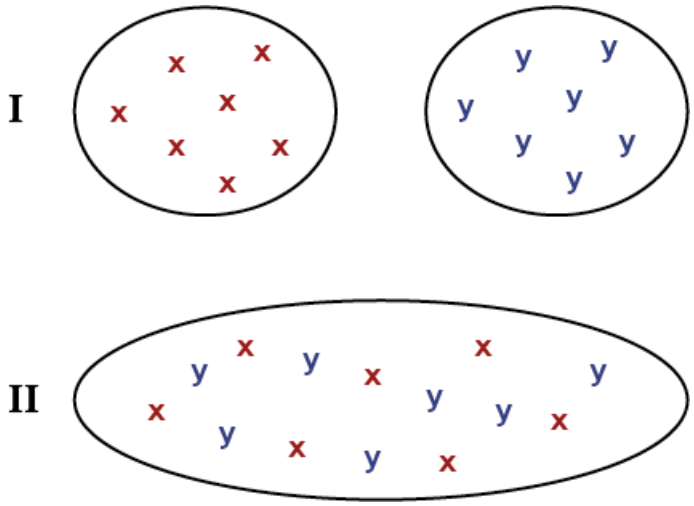


Figure 2.3: Merger-by-Expansion (*source:* adapted from Herold 1990)

As indicated in Figure 2.3, mergers-by-expansion occur very quickly, with no intermediate stage between distinct and merged phonemes. As Herold (1990) finds a LOT-THOUGHT vowel distinction in the father that is absent in the son, Labov (1994: 323) notes that merger-by-expansion is the quickest type of merger, often undergoing completion within one generation. Mergers-by-expansion resemble transfer mergers in that they do not allow for phonetically intermediate forms, however, they differ from mergers-by-transfer as all forms associated with phoneme *x* merge at once with phoneme *y*. Therefore, mergers-by-expansion can be categorised as phonetically *and* lexically abrupt. The primary difference between merger-by-expansion and other forms of merger is the resulting effect on phonetic space. Both merger-by-approximation and merger-by-transfer result in a gap in phonetic space due to one phoneme coalescing with another, while

[1]Many researchers writing on the low-back merger use the Labovian transcriptions of /o/ for the LOT or /ɒ/ vowel, and /oh/ for the THOUGHT or /ɔ/ vowel. Here, I use IPA symbols and Wells' (1982) lexical sets to refer to these vowels to maintain a consistent transcription system throughout the thesis.

merger-by-expansion does not leave a vacant space, but rather the merged phoneme shares the position of phoneme *x* and phoneme *y*. Due to this effect on the vowel space, it is probable that merger-by-expansion can only occur in vowels already adjacent to one another.

Herold (1990; 1997) proposes that mergers-by-expansion occur in high contact situations between speakers who merge two vowels and those who have a phonemic distinction. She posits that the phonetic constraints which separated /ɒ/ from /ɔ/ in speakers from eastern Pennsylvania disappeared due to contact with speakers for whom these constraints did not apply. Without these phonetic constraints, LOT and THOUGHT became indistinguishable. Johnson (2010) also finds examples of merger-by-expansion in his research on /ɒ/-/ɔ/ merging in southeastern New England. As in Herold (1990; 1997), Johnson discovers evidence of children merging LOT and THOUGHT despite their parents distinction of these vowels, and, furthermore, notes cases in which young children are found to merge these vowels where their older siblings maintain a phonemic separation.

2.2.4 Merger-by-Glide Loss

Irons (2007) proposes another mechanism of merger which he terms 'merger-by-glide loss'. Researching speakers of Kentucky English, Irons (2007) reports that the older speakers in the region exhibited a diphthongal quality to their LOT vowels due to the presence of a back upglide, thus providing resistance to a merger with THOUGHT. Conversely, younger speakers in Kentucky were found to show a merger of these two vowels, resulting from an absence of a back upglide on THOUGHT in the youngest members of the speech community. Irons (2007: 166-7) hypothesises that the loss of this back upglide in Kentucky was socially motivated, with younger speakers monophthongising THOUGHT as a conscious rejection of the locally marked diphthongal pronunciation.

This vowel merger in Kentucky English, Irons (2007: 137) argues, 'cannot be explained by current theories of merger'. However, Johnson (2007: 18) highlights the similarity between this case of LOT-THOUGHT merging in Kentucky and examples of merger-by-expansion. As the only factor distinguishing these vowels was the presence of an upglide on THOUGHT, Johnson (2007: 18) argues that the sudden disappearance of this upglide, which resulted in a merging of the low back vowels, is not too dissimilar from Herold's (1990) description of the sudden removal of phonetic constraints resulting in a phonemic merger in expansion situations. As relatively little has been written on the concept of merger-by-glide loss, evidence for it as a distinct mechanism of merger is slight, and it may be the case that low back

vowel merging in Kentucky English is a variant on a merger-by-expansion.

2.2.5 Multiple Mechanisms of Merger

Although three primary types of merger have been identified, there are examples where more than one mechanism of merger appears to be at play within one speech community. Trudgill and Foxcroft (1978) find evidence of both merger-by-approximation and merger-by-transfer in their research of East Anglian /ʊu/ and /ʌu/. In this region, the two types of merger appear to be independent from one another, with a number of speakers exhibiting evidence of the transference of lexical items from /ʊu/ to /ʌu/, while others produce intermediate realisations between /ʊu/ and /ʌu/ as in a case of vowel approximation. Therefore, while the whole region is found to participate in the merging of /ʊu/ and /ʌu/, Trudgill and Foxcroft (1978: 77) argue that different speakers are using different 'strategies' to converge these vowels.

As discussed, in Dinkin's (2016) study of merger by 'phonological transfer' in Upstate New York, words containing a LOT vowel in a following /l/ + labiovelar environment moved toward the /ɔ/ vowel. However, Dinkin (2016: 182) also observes that both phonological transfer and approximation are working in tandem in the progression of this merger. He finds evidence of LOT vowels backing slightly toward THOUGHT before being transferred over, i.e. *golf, revolve, solve* first approximated THOUGHT before their transference from one phoneme to the other. Dinkin (2016) also notes that LOT vowels before /l/ (e.g. *doll, solid*) now appear to be backing, indicating that this may be the next group of LOT vowels to phonologically transfer to THOUGHT. These findings suggest that it would be possible for the merging of two phonemes in a speech community to be the result of two different mechanisms of merger working together.

2.3 Historical Mergers

Much about the nature and development of mergers can be learned from studying cases of completed, historical mergings. The examples discussed below detail all of the sound changes that lead to the merging of previously distinct phonemes, showing the completion of the mergers from beginning to end. Such comprehensiveness is often only possible when looking at past mergers, as accounts of contemporary mergers may not yet be fully aware of all of the changes involved in a recent or ongoing merger.

2.3.1 The Long-Mid Mergers

In the majority of contemporary varieties of English, words such as *pane-pain* and *toe-tow* have identical pronunciations. However, in the Middle English period, these words were not homophonous, with contrasting monophthongal and diphthongal pronunciations. The coalescence of these vowels, which Wells (1982: 192-3) terms the 'Long Mid Mergers', led to the development of the the FACE and GOAT lexical sets, as shown below in Figure 2.4 below.

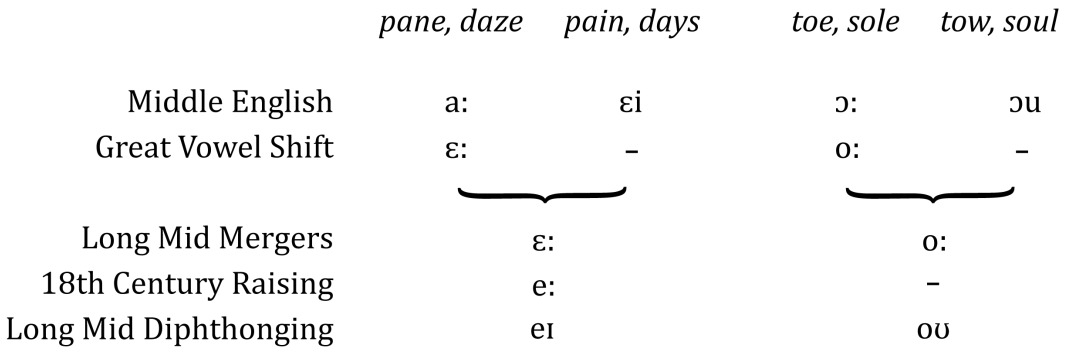


Figure 2.4: Development of the Long-Mid Mergers (*source:* Wells 1982: 193)

The FACE lexical set originally had two subsets; words such as *pane, raze, daze* which were pronounced with an [aː], and words such as *pain, raise, days* which were realized with [ɛi]. During the Great Vowel Shift, the subset with the [aː] vowel underwent raising to [ɛː]. Wells (1982: 193) proposes that the [ɛi] vowel used in the second subset became monophthongal around the 16th century, before merging with the [ɛː] vowel of the first subset. Later shifts affected the previously distinct subsets together, with the merged vowel raising to [eː] and then becoming diphthongized as [eɪ] in the 18th century. According to Wells (1982: 193), the GOAT lexical sets shows a strikingly similar development. In Middle English a subset of *toe, sole, nose* words were realized with [ɔː], while *tow, soul, know* had a diphthongal [ɔu] pronunciation. The Great Vowel Shift saw the raising of [ɔː] to [oː], and both subsets are thought to have merged in the 17th century. Later, the merged vowel was diphthongized to the [ou] pronunciation found in many present day English varieties. The Long-Mid mergers therefore neatly demonstrate Hickey's (2004: 125) point that, in cases of completed phonemic merger, the previously distinct sounds will shift together as one phoneme in any further sound changes.

The long-mid mergers of the FACE and GOAT lexical sets have not undergone completion in all accents of English. Mostly notably, speakers of Norwich English and speakers of several Welsh varieties have been found to maintain a distinction

between the two subsets of FACE and GOAT. While Trudgill and Foxcroft (1978) note the presence of this Middle English split in the GOAT lexical set in East Anglia, they find that many speakers are in fact merging the two subsets of GOAT. Wells (1982: 337) similarly reports that, although a distinction of the two GOAT subsets may still be heard in East Anglian speech, a distinction between the two subsets of the FACE vowel is 'sharply recessive' for younger speakers in the region. This indicates that East Anglian English is undergoing the process of Long Mid Merging, just much later than the majority of other accents of English. In a number of Welsh English varieties, Wells (1982: 384) notes that *pane, daze, made* are realized with a monophthongal [eː] whereas *pain, days, maid* have the diphthongal [eɪ] vowel. Such a distinction of the GOAT subsets, however, is reportedly rare. Wells (1982: 384) raises the question of whether the contemporary Welsh English distinction between the two subsets of the FACE vowel can be said to be a preservation of the historical, Middle English distinction between the two subsets. He proposes that the absence of the long-mid merger in Welsh English is not due to the variety's non-participation in the historical merger, but rather as a feature of Welsh speakers learning English pronunciations via spelling: e.g. words spelled with 'i' or 'y' *pain, day* are diphthongal, while all other words have monophthongal [eː]. This again highlights the importance, as advocated by Maguire *et al.* (2013), of looking at sound changes from both a diachronic and a synchronic perspective.

2.3.2 The *Which-Witch* Merger

The previous examples discussed in this section have all been cases of vocalic merger. One consonantal example of a historical merger is the coalescence of the labial-velar fricative /ʍ/ with the labial-velar approximant /w/ in initial position. In Old English, words written with an initial <hw> were pronounced with the glide cluster [hw], as in 'hwæt' *what* and 'hwelc' *which*. These words later came to be spelled with <wh>. Minkova (2014: 109) explains that, starting in the 11th century, /hw/ was being simplified to /ʍ/, and that, by the Middle English period, pronunciations of [w] were common in interrogatives such as *what* and *where* when weakly stressed. Wells (1982: 228) also highlights that [w] was typical in educated speakers of English by 1800. Today, many varieties of English no longer distinguish /ʍ/ and /w/, resulting in a lack of contrast between word pairs such as *which-witch*, and *whine-wine*.

The merging of /ʍ/ and /w/ occurred much earlier in the south of England than in the north. Strang (1970: 45) notes that these two phonemes have been merged for southern speakers since the Norman Conquest, and Minkova (2014: 111) remarks

that it was not until relatively recently, i.e. the past century, that the absence of a /ʍ/-/w/ distinction was characteristic of most British English varieties. Those from Northumberland and Scotland are often given as examples of speakers who continue to maintain a separation of /ʍ/ and /w/. However, Lass (2016), in looking at the data from the Survey of English Dialects (collected in between 1950-1961), finds that the majority of speakers from northern England produce only [w], and that speakers from areas such as Northumberland, Cumberland (now Cumbria), Durham, and the Isle of Man variably produced /ʍ∼w/ in words with initial orthographic <wh>. In Scotland, Stuart-Smith (1999: 210) reports that young working-class speakers of contemporary Glasgow English exhibit signs of a /ʍ/-/w/ merger-in-progress. These results mirror those of Chirrey (1999: 227), who finds frequent use of [w] among young Edinburgh speakers, signalling that a loss of a distinction between /ʍ/ and /w/ is spreading in urban Scottish English varieties. Given the amount of recent research on the weakening of this distinction, Minkova (2014: 111) comments that the /ʍ/ phoneme is generally merged with /w/ 'even in those historical strongholds of *which-witch* contrast'.

Although the distinction of /ʍ/ and /w/ in the south of England was lost around the Middle English period, RP speakers have often been found to produce [ʍ] in initial <wh> lexical items. Due to the geographical development of the merger, Strang (1970: 45) proclaims that a separation of /ʍ/ and /w/ 'hardly, therefore, belongs in RP'. The adoption of [ʍ] pronunciations in RP is was the result of the stigmatisation of [w] in <wh> words in the 18th century. Beal (1999: 177) proposes that [w] in this context was subject to stigmatisation because of its association with Cockney speech, and, echoed in Strang (1970: 45), due to the apparent deviation of pronunciation from its orthographic representation. The appearance of [ʍ] in RP could point to a reversal of the previous merger between /ʍ/ and /w/ in southern England. However, on RP speakers who use [ʍ] pronunciations of <wh> words, Wells (1982: 228-9) argues that there is no distinction between /ʍ/ and /w/ in their phonemic inventory, and that their use of [ʍ] is a 'conscious decision' due to the perception of this realisation as 'correct, careful, and beautiful'. Even so, the negative perception of [w] has since largely receded, with Strang (1970: 45), some 50 years ago, noting the decline of [ʍ] in the majority of British English varieties.

Outside of the UK, Wells (1982: 229) reports that a /ʍ/-/w/ distinction is widespread in North America, but that the loss of this distinction, which is spreading in the US, is most notable in urban speech. Labov *et al.* (2006: 49) comment on the rapid diffusion of this merger across the US, explaining that, found only in a few coastal areas in the mid-20th century, the merged /ʍ/-/w/ consonant was dominant

throughout the country by the end of the century. They add that, while a high proportion of speakers from southern states continue to distinguish /ʍ/ from /w/ in both production and perception, the distinction between these phonemes 'will soon be extinct' in all other varieties of US English. A similar situation is also found in Canada. Chambers (2013) observes speakers from four different regions in Canada, finding that /ʍ/ and /w/ are merging in apparent time in multiple Canadian English varieties. From this, he suggests that the merger is not geographically conditioned, but rather that it is a sound change affecting Canadian English as a whole.

2.4 Contemporary Mergers

Contemporary mergers, and mergers-in-progress, are frequently researched topics in sociophonetic studies. In order to further understand the phenomenon of mergers, and to gain insight into how they are studied in present-day research, a number of examples of more contemporary phonemic mergers are discussed in this section. Additionally, as so little has previously been written about the potential GOAT-THOUGHT merger in Tyneside English, there are many unknown details surrounding, for example, the quality of the merger, how it developed, and who in the speech community may exhibit signs of merger. Thus, surveying some recent sociolinguistic studies of merger will help to reveal how other researchers answer these kinds of questions on mergers and mergers-in-progress.

This section presents a number of examples of 'conditioned' and 'unconditioned' mergers. According to Hoenigswald (1960: 90) vowel mergers fall into two broad categories: conditioned and unconditioned. The primary difference between these two types of merger is the effect on phonemic inventory. Unconditioned mergers involve a collapsing of two or more phonemes, whereas Gordon (2013: 204) notes that in conditioned mergers, the phonemes involved remain distinct, but 'the contrast between them is neutralized in certain environments'. In conditioned mergers, therefore, phoneme x and phoneme y are realised as phonetically identical only in specific phonological contexts. Thomas (2001: 15) posits that conditioned vowel mergers are most likely to occur word-finally, or when followed by (1) liquid consonants, (2) nasal consonants, or (3) palatal and velar consonants, namely /g/ and /ʃ/. Hickey (2004: 126) also proposes that mergers tend to occur in pre-sonorant environments, as, when in coda position, these consonants tend to become 'absorbed' into the vowel. As a result, vowels which are coloured by a following sonorant may become difficult to distinguish from one another, thus these particular phonetic environments are more vulnerable to merger.

2.4.1 Conditioned Mergers

2.4.1.1 The PIN-PEN Merger

In certain US varieties of English, speakers are found to produce /ɪ/ and /ɛ/ identically when preceding a nasal consonant (/n, m, ŋ/) (Wells 1982; Labov *et al.* 2006). Usually referred to as the PIN-PEN merger, this sound change results in homophonous pairs such as *pin-pen*, *him-hem* and *since-sense*. Due to the environmental restrictions on this vowel merger, speakers retain a phonemic distinction between /ɪ/ and /ɛ/. For example, speakers will produce the same merged vowel in *bin* and *Ben*, while also maintaining a phonetic distinction in their production of the words *bid* and *bed*. The PIN-PEN merger is most closely associated with southern US varieties, although it may also be found in southern California (Thomas 2001; Warren and Fulop 2014), southern Illinois (Bigham 2004) and also in the Canadian province of Newfoundland (Colbourne 1982). The PIN-PEN Merger is also a widely reported feature of African American Vernacular English (AAVE) (Wells 1982; Labov *et al.* 2006; Rickford and King 2019), however, some AAVE speakers, such as those from Brooklyn, New York City, have been found to maintain a distinction between [ɪ] and [ɛ] in pre-nasal environments (Blake *et al.* 2015).

The quality of the merged PIN-PEN vowel appears to vary in different regions of the US. Some accounts of the merger, such as that of Thomas (2001), generally attribute this merger to the raising of the [ɛ] vowel to the position of [ɪ] in southern speech. Thus, the merged vowel has a quality closer to /ɪ/, e.g. '*pen*' will sound like '*pin*'. Other analyses of the PIN-PEN merger find that, although it may sound more similar to /ɪ/ perceptually, the quality of the merged vowel is actually phonetically intermediate between /ɪ/ and /ɛ/ (Koops *et al.* 2008; Baranowski 2013; Warren and Fulop 2014). Bigham (2004) finds a great deal of variation in the quality of this merged vowel among his southern Illinois speakers; some raised /ɛ/, some lowered /ɪ/, and others did both. Such findings of a form in a phonetically intermediate position between the two vowels indicate that, at least in these US regions, the converging of PIN-PEN is progressing via a merger-by-approximation. Based on this research into PIN-PEN, it appears possible that speakers of different varieties – or even speakers in the same speech community – may use varying articulatory strategies to achieve the same phonemic merger. This echoes Trudgill and Foxcroft's (1978: 77) findings of both merger-by-transfer and merger-by-approximation in the merging of /ʊu/ and /ʌu/ in East Anglia, suggesting that different speakers may have different routes toward the same phonemic merger.

Most accounts of the PIN-PEN merger discuss only the merging of the /ɪ/ and /ɛ/ vowels. Bigham (2004), however, links this sound change to the raising of /æ/ before nasal consonants in southern speech. Within his dataset, Bigham identifies several different merging systems; (1) those who merge all three vowels in nasal environments, (2) those who have a merged /ɛ/ and /æ/ but a distinct /ɪ/, and, (3) those who merge /ɪ/ with /æ/. This could suggest that, in southern Illinois, there is a sound change in progress toward a three-way PIN-PEN-PAN merger, although little further research has since been conducted on this potential merger.

2.4.1.2 The PULL-POOL Merger

A phonemic merger of the FOOT and GOOSE vowels, so that words such as '*pull*' and '*pool*' are homophones, is a noted feature of Scottish English. In Scottish English, the absence of a distinction between these lax and tense vowels is unconditional, however, for the majority of English varieties exhibiting identical pronunciations of FOOT and GOOSE, the merger is conditioned to appear only in pre-/l/ environments.

Several studies conducted in the US report findings of a merger of FOOT and GOOSE before /l/. As a conditioned merger, this is often referred to as the PULL-POOL merger (sometimes the FULL-FOOL merger). Labov, Yaeger, and Steiner (1972) first noted the merging of PULL and POOL among young Mexican-American speakers in western Salt Lake City, Utah. More recently, in the *Atlas of North American English*, Labov *et al.* (2006) show that this vowel merger is most prevalent in the western Pennsylvania area, though speakers across the country may lack a PULL-POOL distinction. Labov *et al.* (2006: 70) state that this merger is found 'consistently' in western Pennsylvania, though Eberhardt (2009) finds that, while white speakers do typically exhibit PULL-POOL merging, African American speakers in the region typically retain a distinction between these two vowels. Elsewhere, Labov (1972; 1991) notes that PULL-POOL merging is found in the area of Albuquerque, New Mexico. The pronunciation of the merged vowel is generally described as closer to PULL [ʊ] as a result of the laxing of POOL (Labov *et al.* 2006; Eberhardt 2009).

Bowie's (2000) study on the speech community of Waldorf, Maryland finds that PULL and POOL may also merge with POLE. He finds a linear progression of this merger through time, with PULL-POOL merging first, then PULL-POLE, before finally leading to a merger of POOL-POLE. Research carried out in Youngstown, Ohio (Arnold 2015; Wade 2017) has found a similar situation to that which Bowie describes in Waldorf. Arnold (2015: 9) notes the variable merging of PULL, POOL and POLE, proposing that different mergers are 'competing' with each other in this

region; she finds evidence of a PULL-POOL overlap predominantly in older speakers, while younger speakers show more of a merger between the PULL and POLE vowels. Additionally, some younger speakers in Arnold's study showed a merger of PULL, POOL and POLE, though she notes that a three-way merger is relatively rare in the speech community (2015: 6). Given the variability in the overlap of these three vowels in Youngstown – in both production and perception – Arnold (2015) expresses that is is difficult to predict how these mergers will progress in the future.

FOOT-GOOSE merging is also found to co-occur with /l/-vocalisation in a number of accents, most notably in New Zealand and Cockney Englishes. In his discussion of the Cockney accent, Wells (1982: 315) describes the merging of PULL and POOL, noting that the merger only occurs before dark (coda) /l/. As such, although '*pull*' and '*pool*' are pronounced identically, '*pulling*' and '*pooling*', and '*pulled*' and '*pooled*' remain distinct. More recently, Lindsey (2019: 28) reports that young speakers of Standard Southern British English (SSBE) may exhibit a three-way merger of '*pull, pool, Paul*', all of which are realised with a tense [oː] or [uː]-like vowel. Due to the frequency of /l/-vocalisation in Cockney speech, merged PULL-POOL vowels are typically followed by a vocalised /l/. However, both Wells (1982: 315) and Lindsey (2019: 28) state that this vowel merger, in Cockney or SSBE, may occur either when followed by a vocalised or a dark-/l/. A similar case is found in New Zealand, as Bauer and Warren (2004: 584) detail that many speakers merge PULL with POOL. The authors do not provide a description of the quality of this merged vowel, however, noting the considerable variation in vowels which undergo neutralisation before /l/ in this speech variety. Additionally, Bauer and Warren (2004: 584) continue that some New Zealand speakers may exhibit a neutralisation of the FOOT, GOOSE, GOAT and THOUGHT vowels before coda-/l/, leading to the possible four-way homonphony of '*pull, pool, Paul, pole*'.

2.4.2 Unconditioned Mergers

2.4.2.1 The Low-Back Merger

The low-back merger is perhaps the most extensively studied vowel merger in contemporary sociolinguistic research. Affecting a large number of North American English varieties, it involves the loss of a distinction between /ɔ/ and /ɒ/, or rather, between the LOT and THOUGHT classes. This creates homophonous word pairs of '*cot-caught*', '*collar-caller*', and '*knotty-naughty*'. Gordon (2008: 134) notes that, while the quality of the merged low-back vowel is subject to much variation,

the pronunciation is most frequently found to be 'unrounded, low and quite back'.

Western Pennsylvania is sometimes cited as the region of origin of the low-back merger (Wells 1982; Labov 2001). As early as the 1960s, Kurath and McDavid (1961) find that speakers from western Pennsylvania typically merge LOT and THOUGHT to an [ɒ]-like pronunciation. With speakers in eastern Pennsylvania maintaining a /ɔ/-/ɒ/ distinction, Kurath and McDavid (1961: 17) declare this geographical difference in the realisation of the low-back vowels as the 'most important isogloss within Pennsylvania'. Now found across North America, it does not appear to be the case that western Pennsylvania is the only area from which the low-back merger has spread. Labov *et al.* (2006) map the geographic distribution of the LOT-THOUGHT merger, showing that there are many areas surrounding western Pennsylvania in which speakers still distinguish these vowels in both production and perception. Rather, Labov (2001: 173) posits that the merging of the low-back vowels occurred independently in – and is now diffusing from – several different regions: from western Pennsylvania down to West Virginia and Kentucky, from east to west New England, from Canada down to Minnesota, and from the southwest US area eastward to Texas.

Labov *et al.* (2006) explore the historical development of the LOT and THOUGHT classes in their discussion of the low-back merger. First, they highlight the complicated formation of the THOUGHT class – derived from a number of different Old English vowel classes – which resulted in a lack of easily identifiable orthographic cues for which words belong to /ɔ/. Labov *et al.* (2006: 57) go on to explain that, in North American English varieties, certain words with a traditional short /ɒ/ vowel have transferred over to the long /ɔ/ class. Phonetically restricted, the vowel in LOT words came to be realised with an [ɔ] vowel when followed by a voiceless fricative or a back nasal , e.g. *loss, dog, song.* These words make up what Wells (1982) defines as the CLOTH lexical set. Labov *et al.* (2006: 57) continue that the CLOTH set is further restricted by prosody, in that most polysyllabic words are still associated with the /ɒ/ class, e.g. *hospital, soggy.* Additionally, word frequency is also a factor, with common words such as *dog, song* transferring over to the /ɔ/ class, while less frequent words, like *hog, pong*, variably remain in /ɒ/. Thus, Labov *et al.* (2006: 57) propose that the development of these vowel classes, in addition to the changes within the /ɒ/ and /ɔ/ classes in North American Englishes, have resulted in an 'unstable relation' between the low-back vowels. This suggests some kind of inevitability of a low-back merger, which provides an explanation for why the LOT-THOUGHT merger is present in so many varieties of North American English. However, Labov *et al.* (2006) also point to a

number of US varieties which have instead undergone changes that have led to an increase or strengthening of the distinction between /ɒ/ and /ɔ/.

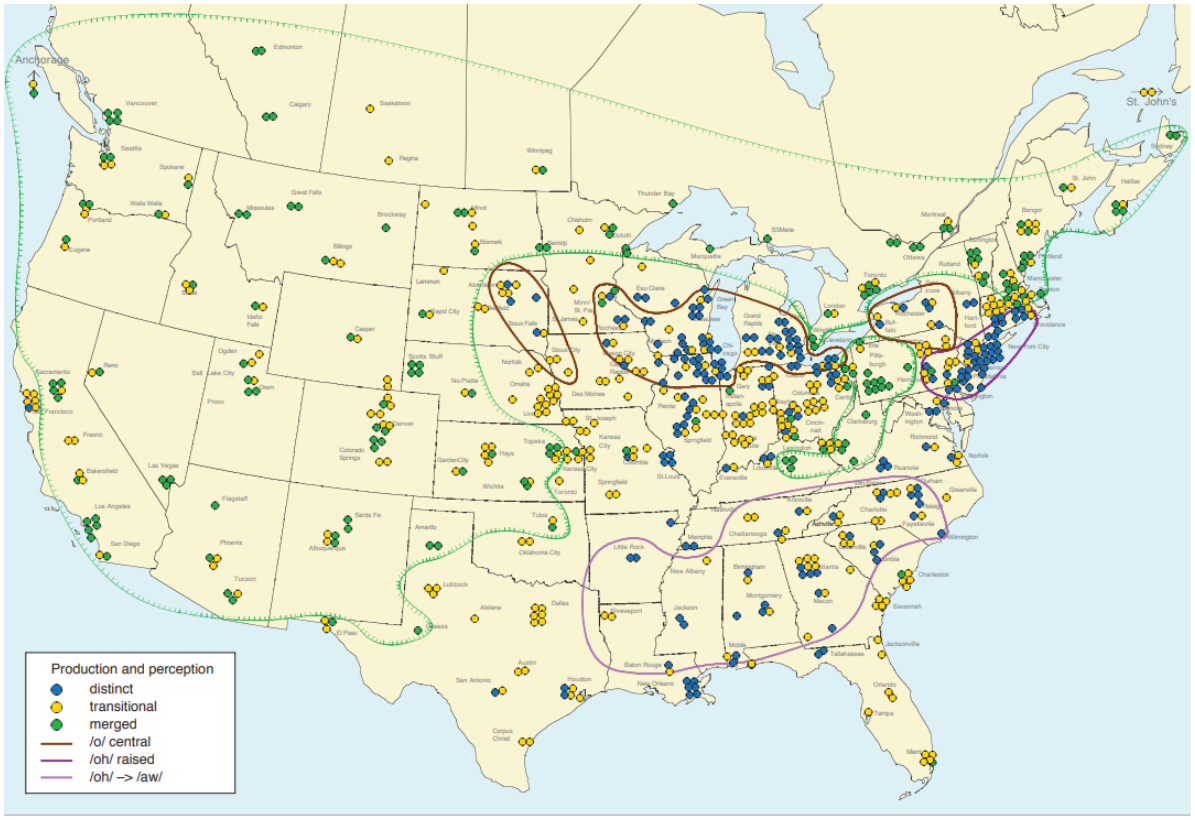


Figure 2.5: Areas resistant to the LOT-THOUGHT merger (*source:* Labov *et al.* 2006: 60)

Despite its vast geographical spread, the low-back merger is largely absent from three large US regions: the Inland North, the Mid-Atlantic, and the South. Labov *et al.* (2006: 58) argue that these areas show 'resistance' to the low-back merger due to previous changes in the LOT and THOUGHT vowel classes. This is illustrated in Figure 2.5 above. Speakers in the Inland North maintain a distinction between /ɒ/ and /ɔ/ due to the fronting of /ɒ/, and the low-back merger is prevented in the Mid-Atlantic due to the raising of /ɔ/. In the South of the US, /ɔ/ is diphthongized with a back upglide, thus blocking a merger of LOT and THOUGHT. However, more recently, Dinkin (2011) has looked for signs of progression of the low-back merger in Upstate New York (an area in the Inland North) and Hudson Valley (an area in the Mid-Atlantic). In Upstate New York, Dinkin (2011: 342) finds that the LOT-THOUGHT merger is spreading in the region, concluding that /ɒ/-fronting is an easily reversed change which does not provide stable resistance toward a merger of the low-back vowels. This finding provides evidence against Labov's (2010: 164) proposition that the LOT-THOUGHT merger cannot expand into the Inland North

due to /ɒ/-fronting, which, due to its position in the Northern Cities chain shift, is a vowel change which is not readily reversible. For the Hudson Valley speakers, Dinkin (2011: 342) reports that a raised /ɔ/ does appear to prevent the merging of the low-back vowels in the Mid-Atlantic region. Unlike /ɒ/-fronting, Dinkin (2011: 342) posits that /ɔ/-raising is a 'unidirectional change' which cannot be easily undone, providing strong resistance to the LOT-THOUGHT merger in the Mid-Atlantic region of the US.

Research conducted by Di Paolo (1992: 289) in Salt Lake City, Utah finds that the low-back vowel merger is perceived as a prestigious sound change. In a matched-guise task, listeners responded more favourably to speech samples containing merged LOT-THOUGHT vowels, judging these speakers to be more successful, more likeable, and more well-spoken. Di Paolo (1992: 289) also reports that many of the participants from Utah displayed hypercorrection toward the merger. That is, in conversational speech, the speakers would usually make a clear distinction between /ɒ/ and /ɔ/, yet would produce a merged vowel in minimal pair reading tasks. These findings indicate that the low-back merger is viewed as prestigious in the Salt Lake City area, though it may not be the case that such attitudes are echoed throughout the rest of the US. There are suggestions, however, that the low-back merger is coming to be viewed as a feature of standard American English pronunciation. Baković (2010) cites an example of a national radio panel show which used a limerick-style quiz question involving the LOT and THOUGHT vowels. On the show, a contestant from New York failed to identify that the correct answer to the question – which must rhyme with the words '*distraught*' and '*brought*' – was in fact the word'*hot*'. While this incident does not directly show that the merger is thought of as prestigious, it does demonstrate that the low-back merger is viewed as a standard, widespread feature of contemporary American English.

2.4.2.2 The NURSE-SQUARE Merger

In a number of varieties of English the vowels of NURSE and SQUARE are merged, creating homophones of *purr-pair* and *furry-fairy*. The merging of these lexical sets is exhibited by speakers from various areas, including East Yorkshire, Liverpool and Wigan in North-West England, Belfast, and Dublin (Wells 1982; Barras *et al.* 2007; Ferragne and Pellegrino 2010; Cruttenden 2014). Present in several varieties, the quality of the merged NURSE-SQUARE vowel is variable. Wells (1982: 361) notes that the typical realisation of this vowel in North West England is that of a centralized [ɜː], thus '*nurse*' [nɜːs] and '*square*' [skwɜː]. Looking specifically at Merseyside, Watson and Clark (2012) find variation within the region: St Helens' speakers were

found to use the mid-central [ɜː] pronunciation, while those from Liverpool itself merged to a mid-front [ɛː]. Noting the stigmatisation of the Scouse accent as a whole, Watson and Clark (2012: 320) discuss how, although it is not clear which specific features of the accent are subject to negative stereotyping, [ɛː] realisations of the merged NURSE-SQUARE are perceptually salient. As in Liverpool, fronted [ɛː]-like NURSE vowels are also produced by speakers from Middlesbrough, though this vowel is typically not found to be merged with SQUARE in Teesside (Kerswill and Williams 2002; Beal *et al.* 2012).

A full merger between NURSE and SQUARE is reported for speakers from Belfast and the city's adjacent areas, and is noted in certain varieties of Dublin English. Hickey writes (1999: 268) that NURSE and SQUARE are both realised as [əː] in 'fashionable Dublin English', which is spoken by a group that 'vigorously rejects a confining association with low-prestige Dublin English'. It is further noted by Wells (1982: 421) that, in contrast to its negative perception in Merseyside, the NURSE-SQUARE merger is generally viewed as prestigious in Dublin. In more conservative Dublin speech, NURSE and SQUARE are not fully merged due the absence of that which Wells (1982: 200) terms the 'First NURSE Merger' in many Irish English varieties. As Wells (1982: 200) details, in the Middle English period the short lax vowels /ɪ, ɛ, ʊ/ became merged and centralized to [ə] when preceding a pre-consonantal or final /r/, such as in *bird, fern, church.* The contemporary NURSE lexical set therefore contains words with orthographic <*i*>, <*e*> and <*u*>. As a result of the First NURSE Merger, word pairs such as *kerb-curb* and *earn-urn* became homophonous for the majority of English speakers, typically taking a mid-central [ɜː] pronunciation. However, a contrast between these word pairs is still frequently heard among Irish English and Scottish English speakers. Wells (1982: 421) outlines that in 'Typical Dublin' English, the NURSE vowel has two subsets, leading to a distinction between '*purr*' [pʌr] and '*per*' [peːr]. It is only the latter subset of NURSE which has come to merge with the SQUARE vowel in traditional Dublin speech, leading to '*per-pair*' having identical [eːr] realisations.

In terms of stability, Ferragne and Pellegrino (2010) find that a merged NURSE-SQUARE vowel is 'the norm' for their speakers from East Yorkshire, Liverpool and Dublin. However, there is evidence that the NURSE and SQUARE vowels are becoming distinct in the North-West. Barras *et al.* (2007) find evidence to suggest that younger Lancashire speakers exhibit an ongoing separation of these two vowels, increasingly using [ɛː] in SQUARE words. They posit that this apparent unmerging of NURSE-SQUARE is due to speakers' proximity to, and contact with, speakers from Central Manchester who maintain a distinction between these

vowels. In Liverpool, Watson and Clark (2012) find that although often merged in production, some speakers occasionally differentiate NURSE and SQUARE in perception. Differentiation of these vowels perceptually may indicate that this merged vowel is too becoming distinct in Merseyside. The relationship of – and the asymmetry between – production and perception in situations of merger will be discussed in further detail in Section 2.7.

Examples of contemporary mergers and mergers-in-progress, such as the cases discussed in this section, demonstrate the different types of merger that can occur. For example, some mergers are unconditional while some are phonologically restricted, and some mergers are viewed as prestigious or as the 'standard' while others are stigmatised. These examples also highlight that a particular merger may be present in multiple separate speech communities, as in the case of PULL-POOL merging found in a number of US Englishes in addition to New Zealand English, Cockney, and SSBE speech. Furthermore, examining these studies of phonemic merger has provided an understanding of how mergers are being measured and analysed in recent research, and helps to establish what type of findings and results are interpreted by other researchers as showing evidence of merger. Such information will be beneficial when it comes to understanding and determining whether the GOAT and THOUGHT vowels are merging in contemporary Tyneside English.

2.5 Merger Reversibility

While mergers may differ in terms of their development and their conditioning, one general principle of mergers is that they are irreversible sound changes. Garde's Principle, first proposed by Garde (1961) then summarised by Labov (1994: 311), submits that 'mergers are irreversible by linguistic means'. As Labov (1994: 311-312) explains, word class membership is entirely arbitrary, thus, the reversal of a merger would entail the learning of these unpredictable categories on a word-by-word basis. For example, for a speaker with a TRAP-BATH merger to acquire a distinction between these vowels they would need to learn that '*tap, hand, anthem*' belonged to TRAP and not BATH, and conversely, that '*staff, ask, sample*' are members of the BATH set only.

Crucially, Labov (1994: 312) notes that Garde's Principle does not propose the theoretical impossibility of merger reversal. Rather, he suggests that the proposition of the irreversibility of mergers stems from (1) the obvious difficulty of learning arbitrary word class categories, and (2) a lack of reports of people or communities

restoring a distinction between two previously merged vowels. Also of importance is that Garde's Principle posits specifically that sounds cannot be unmerged 'by linguistic means'. Therefore, Garde's Principle does not exclude the possibility that mergers may be reversed due to external, non-linguistic influences. Hickey (2004) argues that explanations of unmergings due to external motivations (e.g. dialect contact) are more robust than explanations which invoke internal factors, and further notes that only externally motivated merger reversals have been attested in the literature. Thus, it may be more accurate to present the general viewpoint on the reversibility of mergers as highly improbable, but not impossible.

There are a few accounts which provide some evidence that individual speakers are able to learn a distinction, particularly in cases of mobile speakers moving to new dialect areas. One of which is Sankoff's (2004) study of language change across the lifespan using interview data from the documentary series *'Seven Up!'*. Two of the subjects from Northern England were shown to have a merger of the FOOT-STRUT vowels in their early interviews, but after relocation and exposure to other non-merged dialects, had seemingly acquired a distinction between these vowels in later life. More recently, Johnson and Nycz (2015) have looked at the acquisition of local phonemic systems by adults and children who have moved to a new city. Interested in both mergers and distinctions of the LOT and THOUGHT vowels, they focused on individuals who have moved from a merged dialect area to a distinct area, and vice versa. The results indicated that, despite the difficulty of learning a distinction, many of the adults and the children began to distinguish LOT and THOUGHT in their own speech production. This leads Johnson and Nycz (2015: 116) to conclude that the majority of the relocating adults were able to 'accommodate long-term to a second dialect', regardless of whether it required the loss of or the acquisition of a LOT-THOUGHT distinction. In reviewing evidence of individual cases of merger reversal, Labov (2010: 121-122) writes that, while it is 'clearly possible' for speakers in dialect contact situations to acquire distinctions, the more pertinent question is whether a whole speech community has the ability to reverse a phonemic merger.

2.5.1 Merger Reversals in Speech Communities

2.5.1.1 /w/ and /v/ in England

Trudgill *et al.* (2004) investigate reports of a previous consonantal merger of /w/ and /v/ in south-east England. Compiling evidence of the merger, they find that late 19th and early 20th century descriptions of south-eastern English speech note the use of /w/ for /v/ in prevocalic position, as in *'village'* and *'view'*. Also recorded in

the Survey of English dialects as a feature present in Norfolk, Trudgill *et al.* (2004: 213) note that, while a stereotype of the local accent, such pronunciations have been absent from Norfolk speech since the early 20th century. Recognising that no English speakers in England today exhibit a /w-v/ merger, the authors propose that either (a) /w/ and /v/ were never truly merged, or (b) a merger of /w/ and /v/ did exist in south-eastern England, and that this merger has since been reversed.

In an effort to answer the question of whether there was a completed /w-v/ merger in England, Trudgill *et al.* (2004: 216) look for evidence of this merger in 'lesser-known colonial varieties of English' in the North-South Altantic and Pacific regions. A phonetic analysis of six different varieties across these regions found that /w/ and /v/ are indeed merged, produced with an intermediate realisation of the bilabial fricative [β]. Commenting on the rarity of the [β] sound in the world's languages, Trudgill *et al.* (2004) hypothesise that, as these geographically disparate colonial varieties all share a rare sound, it is likely that the [β] pronunciation of the merged /w-v/ consonant originated in England. This, they argue, further supports the idea that /w/ and /v/ were once fully merged in south-east England.

As with the other cases of unmergings so far discussed, Trudgill *et al.* (2004: 229) suggest dialect contact as an explanation for the reversal of the /w-v/ merger in England. The authors posit that /v/ and /w/ were merged and realised as [β] in south-eastern speech until, due to contact with non-merged middle-class accents from the north and west of England, distinct [w] and [v] pronunciations began to emerge in the region. Prior to a complete phonemic split, it is suggested that [w] and [v] were first allophones of the merged /w-v/ phoneme, before splitting into two separate phonemes. This case therefore lends support for the argument that non-linguistic external forces – such as language contact – are capable of driving community wide merger reversal.

2.5.1.2 The NEAR-SQUARE Vowels in Charleston

In Charleston, South Carolina, Baranowski (2007) investigated the merging of the NEAR and SQUARE vowels. This leads to homophony in the minimal pairs of *fear-fair, beer-bear, hear-hair.* Baranowski (2007: 104) notes that accounts of the Charleston accent from the early 20th century report a merged vowel with a realisation close to SQUARE, while later descriptions from the 1950-60s indicate that the majority of speakers produce a distinction between these two vowels. It would therefore appear that, based on accounts of the accent in the literature, this merger of NEAR and SQUARE in South Carolina has been reversed.

Surveying a large, socially stratified sample of Charleston speakers, Baranowski

(2007) found that the oldest participants exhibited a merged NEAR-SQUARE vowel. Additionally, some of these speakers showed lexical variability, for example, Baranowski (2007: 108) highlights one speaker who did not distinguish between *beer* and *bear* in either production or perception, yet produced and perceived the pair *here-hair* as distinct. The results of acoustic measurements and minimal pair judgements showed that the majority of Charleston speakers today have distinct NEAR and SQUARE vowels; although the merger was found in the speech of many over the age of 70yrs, all those aged 50yrs or under displayed a clear phonemic separation of these vowels. Such a finding lends support to the case of merger reversal in the Charleston speech community. Baranowski (2007: 112) also points out that there are cases in the data of parents and children with different systems, in that the parents merge NEAR and SQUARE in production and perception where their children have a distinction. This emphasises the abruptness of this change, which appears to have completed within only one generation.

This sound change in Charleston is another case which, on the surface, seems to violate Garde's Principle of the irreversibility of mergers. One possible explanation for this apparent phonemic split is that the NEAR and SQUARE vowels were never fully merged in Charleston. Baranowski (2007: 122) argues that, while there is not sufficient evidence to exclude this possibility, the evidence of merger exhibited by the older speakers in both production and perception would indicate that a complete NEAR-SQUARE merger was present in the traditional Charleston accent. If it is indeed the case that NEAR and SQUARE were fully merged in the region, then Baranowski (2007: 122) proposes that the unmerging of these vowels may have occurred due to 'extralinguistic' factors. Two theories are then put forward: (1) that many NEAR-SQUARE distinct speakers from other southern states migrated to Charleston in the early 20th century, resulting in local children acquiring a distinction, or, alternatively, (2) that the association of the merger with the traditional, local dialect led to speakers feeling pressured to acquire the more prestigious NEAR-SQUARE distinction.

Research of community wide merger reversals, in the cases of both /w/ and /v/ in England (Trudgill *et al.* 2004) and the NEAR-SQUARE vowels in South Carolina (Baranowski 2007), has provided evidence contrary to the notion of irreversibility across an entire speech community. These cases indicate that mergers can be reversed if the driving force of the sound change is extralinguistic. Therefore, in the strictest sense, Garde's Principle has not been refuted, as none of these cases involve a merger reversal that can be explained as occurring via internal or linguistic means. In light of these apparent counterexamples to the principle of

merger irreversibility, Labov (2010: 138) notes that, while it may be possible for sounds to become unmerged in situations of dialect contact or social pressure, we should not 'underestimate the difficulty of separating a merged word class into two components, once their historical identity has been lost.'

2.6 Near Mergers

There are some reports of merger reversal in the literature which are now generally thought to have been cases of near merger rather than completed phonemic mergers which were reversed, such as in the examples of the MEET and MATE vowels (Milroy and Harris 1980; Labov 1994) and the LINE and LOIN vowels (Nunberg 1980; Labov *et al.* 1991; Labov 1994). Individuals who produce a slight phonetic difference between two contrasting vowels while exhibiting an absence of a perceptual distinction are said to have a 'near merger'. Accounts of near mergers therefore provide evidence that some speakers are able to pronounce two words differently while perceiving them to be identical. As such, the very existence of near mergers is a topic of some controversy in the literature.

Minimal pair tests have four possible outcomes, as displayed in Table 2.1. Labov *et al.* (1991: 37) discuss how cases represented by cells (a) and (d) in Table 2.1 are uncontentious; it is expected that sounds which are pronounced as the same will also be perceived as the same, and, similarly, distinctions in production will be upheld in perception. Situations in which participants pronounce two words as the same yet perceive them to be different, as represented in cell (c), are also quite expected. This, Labov *et al.* (1991: 38) explain, is due to the influence of orthographic differences on judgements of perceptual similarity. Cell (b) in the table represents sounds which are pronounced as distinct but judged as the same, i.e. near mergers. As the notion of a near merger appears to confute several principles of that which Labov *et al.* (1991) term the 'categorical view' of phonology, results represented by cell (b) were long assumed not to be possible.

		Spoken	
		Same	*Different*
Judged	*Same*	a	b
	Different	c	d

Table 2.1: Possible outcomes of minimal pair test (*source:* Labov *et al.* 1991: 37)

That 'linguistic categories are separated into mutually exclusive sets' is a

fundamental principle of the categorical view of phonology (Labov 1994: 349). Therefore, from a categorical standpoint, a sound either belongs to a phonemic category or it does not. Labov *et al.* (1991: 46) argue that such a binary view of phonemic categories cannot account for near mergers, as sounds which are near-merged – the same in perception but different in production – must belong to some intermediate category between phoneme *x* and phoneme *y*. Additionally, as only distinctive features are deemed important to category membership in the categorical framework, other phonetic details are thought to be irrelevant to phonology (Labov *et al.* 1991: 36). With near mergers, however, a slight phonetic difference is enough to maintain a distinction between these sounds in speech production. Again, this demonstrates that near mergers are not easily accounted for in a 'categorical view' of phonology. Labov *et al.* (1991: 33) discuss how accounts of near merger have long been 'rejected in favor of theoretical assumptions that are inconsistent with them'. The authors go on to argue that evidence of near mergers must not be discounted on the basis of being difficult to account for theoretically.

One important question that often arises in research of near mergers is the following: how can it be that some individuals produce a distinction which they cannot hear? In contemplating this question, Harris (1985) points to Nunberg's (1980) theory of how near mergers are able to occur. Defining the terms 'limits of production' and 'limits of confusability', Nunberg's theory is illustrated in Figure 2.6. Here, the solid ellipses represent the 'limits of production', with pronunciation of the phonemes *x* and *y* lying within these boundaries. As shown in the diagram, the boundaries of pronunciation for the phonemes *x* and *y* are clearly distinct. The larger, dashed ellipses in Figure 2.6 show the 'limits of confusability'. This represents the boundary of perception of *x* and *y*, beyond which listeners will judge tokens to be distinctly different. Nunberg (1980: 230) therefore suggests that near mergers occur when two phonemes become phonetically close enough to result in an overlap of their 'limits of confusability'. Notably, as phonemes are required to be close in phonetic space to allow for an overlap in the 'limits of confusability', it may be possible that near mergers can only occur in situations of approximation between two or more phonemes.

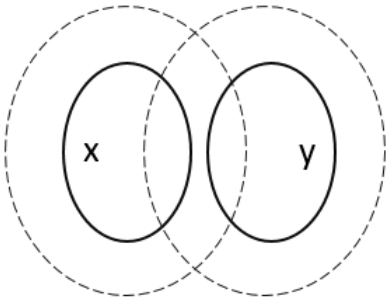


Figure 2.6: Perceptual overlap between phonetically adjacent phonemes (*source:* Nunberg 1980: 227)

The existence of near mergers also proves problematic for theories on language change, particularly regarding notions of how sound changes are transmitted. How does an individual learn to produce a phonetic distinction between two forms if they cannot themselves perceive that the forms are different? Hickey (2004: 131) highlights the difficulty of explaining how near mergers are passed on from one generation to another, pointing to the following paradoxical situation: a child must learn a phonetic distinction which they cannot perceive, yet they are only able to acquire this distinction through hearing it. As such, Hickey (2004: 132) remains sceptical about the existence of near mergers, arguing that, if they do indeed exist, they are short-lived stages in a sound change.

Hickey's (2004) approach to near mergers would indicate that these phenomena are a transitional stage in a sound change involving a complete merger of two phonemes. A similar viewpoint is voiced by Janson and Schulman (1983: 335), who argue that situations of near merger 'seldom remain stable'. Of course, in some cases it may be possible that evidence of near mergers simply captures a stage of a merger-by-approximation in progress. For example, Kontra (1993: 229) concluded that a near merger in bilingual Hungarian-American speakers was simply a transitory stage in the change toward a full merger of /ɛ-e/. Nunberg (1980: 224), however, argues that near mergers can indeed be passed on through several generations, citing the longevity of the LINE-LOIN near merger in Essex as evidence of the potential stability of near mergers. In examining historical sources, Nunberg (1980) concludes that reports of a merger of the LINE and LOIN vowels in England in the 17th and 18th centuries illustrate a case of near merger, whereby the vowels came into close approximation before becoming peripherally distinctive for the majority of English speakers. Labov (1994) also studied these vowels in three speakers from Essex in the early 1970s, finding that the speakers produced LINE and LOIN as distinct in speech production tasks, but that all three were

unable to distinguish these two vowels in a commutation test. In both Labov (1994) and Labov *et al.* (1991), this result is interpreted as further evidence that LINE and LOIN were only nearly merged in the Early Modern English period, and that near mergers can survive across generations. However, it is not completely clear whether the 17th and 18th century sources and Labov's 1970s data can be understood as showing the same near merger, and, additionally, Labov (1994: 384) notes that interviews with adolescent Essex speakers reveal that younger speakers are typically able to distinguish LINE and LOIN in perception tasks. As such, the stability and longevity of near mergers is still of some debate.

2.6.1 The Near Merger of MEAT and MATE

Perhaps the best known case of near merger involves the MEAT and MATE vowels. Prior to the eventual merger of MEET and MEAT in Standard English, some linguists have argued that there existed a merger of the MEAT and MATE vowels. In a review of sources of Early Modern English pronunciation, from Shakespeare's rhyming poetry to the works of 17th century grammarians, Labov *et al.* (1972: 278) find contrasting accounts of the pronunciation of these three vowels. Figure 2.7 shows the three possible systems of MEET, MEAT and MATE that have been attested in the literature.

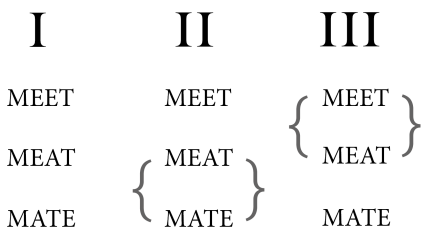


Figure 2.7: Systems of MEET, MEAT, MATE classes (*source:* Labov *et al.* 1972: 278)

System I shows the three vowels as distinct, as they were in Middle English, while System III represents the system of the majority of Present Day English varieties, in which the merged MEET-MEAT vowel is distinct from MATE. Existence of System II, however, is more contentious. As discussed in Labov *et al.* (1972), Labov (1994), and Harris (1985), Wyld (1936) and Kökeritz (1953) are some of the most vocal proponents of a complete MEAT-MATE merger in the 16th century, citing spelling errors such as '*St. Jeamsis Park*' and '*have her bed mead*' in letters written around the late 1600s as confirmation that speakers were unsure whether certain words belonged to the MEAT or the MATE vowel class. Based on this evidence, Wyld (1936) and Kökeritz (1953) propose that MEAT and MATE were at one point merged, before MEAT split from MATE to eventually merge with the MEET class.

As detailed in Section 2.5, many linguists hold the belief that complete phonemic mergers are impossible to reverse. Labov *et al.* (1972: 278) note that the 'traditional view' is that a merger of MEAT and MATE did not take place due to the impossibility of merger reversal. The authors go on to propose that MEAT and MATE in the 16th century were involved in a near merger situation; while the vowels 'were in close approximation, and could not be relied on – for a time – to distinguish words', it does not necessarily follow that these vowels were 'the same' (Labov *et al.* 1972: 296). This indicates that, while they may have been produced or perceived as being similar, the vowels of MEAT and MATE were not fully merged. Furthermore, this helps to explain how the vowels became unmerged in the 17th century, as, because the phonemic categories of MEAT and MATE remained distinct, English speakers would not be required to relearn which lexical items belonged to which word class in order to undo the MEAT-MATE merger.

Milroy and Harris (1980) discuss how both historical evidence and 20th century linguistic accounts point to a MEAT-MATE near merger in Irish English. This suggests that a near merger of MEAT and MATE was not a transitory stage in the development of these word classes in Irish English, but rather that it has been passed down through multiple generations. In their study of Belfast English speakers, Milroy and Harris (1980) find that, although speakers may report that the two vowels sound identical, there are clear phonetic differences which distinguish MEAT and MATE in speech production. Thus, it seems, speakers are producing a distinction yet perceiving the vowels as the same. Although Milroy and Harris (1980) do not themselves use the term 'near merger', their discussion of these vowels in Belfast English is consistent with Labov's (1991) definition of near merger. Additionally, Labov often refers to the case of MEAT and MATE in Belfast in his discussions of near mergers (Labov *et al.* 1991; Labov 1994).

In their quantitative phonetic analysis of MEAT and MATE in Belfast, Milroy and Harris (1980: 203) find that the vowels differ in terms of vowel height. Excluding tokens in which MEAT was realised with an [i] vowel (as in the MEET class), MATE was typically found to be pronounced with a higher vowel than that of MEAT. Milroy and Harris (1980: 203) do note that there is some overlap in the phonetic realisation of these two vowels; the highest MEAT tokens often overlap with the lowest MATE tokens. This pattern is manifest in both the data for the whole sample and for individual speakers. However, given that the majority of MEAT and MATE tokens differed in terms of vowel height, Milroy and Harris (1980: 204) 'confidently' conclude that this is not a case of 'true merger'.

Discussing their results, Milroy and Harris (1980: 207) argue that the Belfast

English speakers, who may claim that MEAT and MATE are perceptually identical, must 'in some deep sense' have an awareness that these two vowel classes are different. Labov (1991: 387) expands on this, positing that speakers must have the ability to use information regarding the phonetic distribution of these vowels (i.e. that higher vowels tend to belong to the MATE class) to preserve an underlying distinction. Following this notion, then, speakers are not required to hear an audible difference between two vowel classes in order to distinguish them. This may explain how the near merger of MEAT and MATE was able to survive for so long in Belfast, providing some evidence contrary to Hickey's (2004: 131) claim that near mergers cannot be stable because new generations of speakers must perceive a distinction that they cannot hear. Furthermore, this finding has consequences for how researchers may conduct data on speech perception, with Milroy and Harris (1980: 207) arguing, in complex cases such as that in Belfast, techniques such as minimal pair tests may not be sufficient for explorations of speakers' perceptual organisation.

2.7 The Production and Perception of Mergers

In cases of two vowels converging toward a complete phonemic merger, linguistic research acknowledges that there must have been a change in the production *and* the perception of these vowels in an individual, or in the wider speech community. However, Thomas (2002: 115) notes that, for a long time, research into perception had been largely 'neglected' by sociophoneticians, and that research into speech perception was confined to more experimental phonetic research. As such, there are more studies of merger which focus on production than perception. Notably, Nycz and Hall-Lew's (2013) paper, which discusses optimal methods for examining vowel mergers, notably only discusses measures of assessing how these merging vowels are produced. There have been, however, an increasing number of studies in the last couple of decades which investigate both production and perception – and the relationship between them – in situations of vowel merger.

In completed phonemic mergers, there is an expected symmetrical relationship between production and perception; two sounds are pronounced and perceived as identical. Conversely, in cases of phonemic distinction, the sounds will be separate in both production and perception. Symmetry between production and perception is one of the central concepts of the categorical view of phonology, which, as Labov (1994: 352) puts it, means that there is a general assumption that people are 'listeners as we speak, and speakers as we listen'. This leads to the assumption

that results from analyses of speech production data will reflect findings from speech perception tasks. However, as discussed in the previous section, this is not always the case, as sounds are found to be phonetically distinct while perceptually indistinguishable in cases of near merger. It is precisely this asymmetrical relationship between production and perception which makes the existence of near mergers so controversial. While symmetry between production and perception may be expected in cases of full, completed phonological mergers, the question arises as to the relationship between these two factors in cases of merger-in-progress. That is, in the midst of sound change toward phonemic merger, do speech production and speech perception mirror each other, or do they change at different rates?

Some studies have found evidence to suggest that perception is more advanced than production during the progression toward phonemic merger. In Bowie's (2000) investigation into pre-lateral mergers in Maryland, results indicated that participants often struggled to distinguish these vowels perceptually, yet few produced them as the same. While this appears to be a similar situation as reported in cases of near merger, Bowie (2000: 142) takes his findings as evidence that perception precedes production in mergers. One factor of note in Bowie's (2000) study is that perception of the mergers are measured via a commutation style test, whereby a single misperception of one vowel for another is taken as evidence that the listener has a perceptual merger of the two vowels. Commutation tests are discussed in further detail in Section 2.8 below, but as Bowie provides no information on the perception results beyond the labels of 'merged' and 'distinct', it is possible that this measurement of merger perception may exaggerate just how perceptually merged these vowels are among the speech community.

More recently, Freeman (2019) examined the link between production and perception in merging of the /ɛ/ and /e/ vowels before /g/ in Pacific Northwest English. The results of acoustic analysis and vowel categorisation tasks showed that the merger was more advanced in perception compared with production in this speech community. This was particularly notable when looking at the young speaker age group, in which half were merged in both production and perception, yet the rest exhibited merger only in perception. These findings could indicate that perception leads production in situations of merger. However, considering the sharp distinction among the young speakers as either exhibiting a complete merger or a complete distinction in production, Freeman (2019: 460) postulates that such results may alternatively be explained by a negative perception of the merger among certain young speakers. Thus, while all young speakers perceive a merger of these two vowels, some may be consciously avoiding pronouncing these words as

identical due to recent negative evaluations of the merger.

Evidence for perception leading production in cases of merger is not, however, found in all studies. One example is that of Thomas and Hay's (2005) research into the conditioned merger of /e/ and /æ/ before /l/ in New Zealand English. The findings suggested asymmetry between how individuals pronounced and perceived these vowels, as in the studies of Bowie (2000) and Freeman (2019), but rather that production of the merger was more advanced than perception. Young New Zealanders largely exhibited merged /e/ and /æ/ vowels during the reading tasks, while also performing well in the vowel identification perception task. Thomas and Hay's (2005: 23) propose that, because distinct /e/ and /æ/ are still heard in the wider speech community, they are still able to 'exploit this distinction' for the perception test. This is reminiscent of Jansen's (1983: 31) proposition that perception will 'lag behind' production in sound changes, as listeners are still required to perceive more traditional variants that remain in use – typically in the speech of older generations – in the local community.

Arnold's (2015) study of pre-/l/ mergers in Ohio finds little correlation between production and perception. On an individual level, Arnold (2015: 8) highlights one particular speaker who appears to merge the three affected vowels in production, yet produced no errors in perceiving the vowels as distinct in the vowel discrimination task. As with the other cases mentioned, this finding suggests that one's production does not necessarily reflect their perception of certain sounds. Interestingly, Arnold (2015) finds asymmetry between production and perception not only in individuals but across the whole speech community. The results from the production analysis suggested that /ol/ and /ʊl/ are converging in apparent-time, with young speakers displaying significantly more overlap in their realisations of these vowels than the older speakers. However, the perception results did not reveal the same pattern, as the young listeners were found to be no more or less accurate than the other listener age groups in discriminating these vowels perceptually. Such a finding suggests not only that individuals' production and perception may be mismatched, but also that sociolinguistic patterns found in production data may not be borne out in results from perception experiments.

Given the relatively few sociophonetic studies of the production-perception relationship in situations of merger, compounded by the varying results and conclusions of many of these studies, there is little consensus on how production and perception relate to one another during mergers-in-progress. Different studies have found evidence that could suggest that perception precedes production (Bowie 2000; Freeman 2019), or that production is more advanced than perception

(Thomas and Hay 2005; Arnold 2015). The studies discussed in this section do, however, share one common aspect; the results from production and perception analyses do not always neatly correlate with one another. This research therefore indicates that production and perception may not be symmetrical in mergers-in-progress, although there appears to be no overwhelming evidence to suggest which, if either, will be advance of the other.

2.8 Methods of Examining Mergers in Perception

The majority of contemporary sociophonetic studies of vowel mergers in speech production use acoustically analysed speech data to investigate overlap between two or more vowels. Once formant values have been obtained, several methods – such as Euclidean distances, Pillai scores and mixed-effects regression models – can be utilised in assessing how merged vowels are in a speaker's productions (Nycz and Hall-Lew 2013). As the current study makes use of formant analysis, as well as methods such as Pillai scores, a more detailed discussion of some methods of analysing mergers in speech production is presented in Chapter 4. In terms of speech perception, a variety of different methods have been used in sociolinguistic research to examine vowel mergers. The following techniques and tests are some of those which are most frequently used or discussed in the literature on phonemic mergers.

2.8.1 Minimal Pair Tests

Studies of the perception of vowel mergers have often used minimal pair tasks to examine whether or not participants perceive a distinction between two sounds (Boberg and Strassel 1995; Labov *et al.* 2006; Doernberger and Cerny 2008; Johnson and Nycz 2015). In a typical minimal pair test, production data is elicited as participants are asked to read aloud a list of minimal pairs, which are sets of words distinguished by one sound only e.g. *bad-bed, car-bar, pen-pet.* Researchers then typically ask participants to comment on whether they think the two words sound the 'same' or 'different'. Labov (1994: 354) comments that, from a successful minimal pair test, the researcher has gained two essential, independent pieces of information regarding (a) whether the subject pronounces these words identically, and (b) whether they are able to recognise any differences between the words. Speaker judgements in minimal pair tasks are often taken as indicators of how they perceive these vowels. For example, if a speaker categorises a minimal

pair such as *bad-bed* as sounding the 'same', this is taken to signal that the participant may not perceive /æ/ and /e/ as separate phonemes.

The extent to which speaker judgements of minimal pairs can be taken as a reliable signifier of their perception of a merger is of some debate. Labov (1994: 356) deems perception data collected from minimal pair tasks as 'limited and uncertain', arguing that factors other than the subjects' perception of speech sounds can affect their categorization of a minimal pair as 'same' or 'different'. For example, both he and Gordon (2014: 184-5) note the issue of orthography in minimal pair tests, speakers are often likely to claim two words which are spelled differently are also pronounced differently, despite that they themselves do not make a distinction in their own pronunciation.

Aside from orthography, another potential issue with using minimal pairs to examine perception is that this method relies on speakers' self-reports on how they perceive sounds. This may pose an issue if the sounds of interest in a particular study are subject to social evaluation. Inaccuracy in self-reported linguistic usage has been shown in Trudgill (1972), who found that speakers either over- or under-reported their use of vernacular variants in relation to notions of gender identity and prestige. In a similar vein, listeners in Niedzieleski's (1999) study of nationality stereotypes and vowel perception in Detroit were found to judge the exact same vowel tokens differently depending on whether they were told that the speaker was from Canada or from Detroit. These results indicated that, not only were listeners largely unaware that innovative variants associated with Canadian stereotypes were present within their own speech community, but also that notions of standard and marked forms could influence vowel perception. Thomas (2011: 26) calls for a separation of the idea of speech perception and that which he refers to as 'speaker judgement', arguing that the two are not equivalent. While speaker judgements in minimal pair tasks may be interesting data to collect, more sophisticated or experimental methods may be required in order to examine a speaker's perceptual categorisation.

2.8.2 Commutation Tests

To avoid the reliability issues sometimes associated with minimal pair tests, participants may be asked to complete what is often called a 'commutation test' in sociolinguistic literature, also known as an 'identification task' in research on speech perception. This method again requires the use of minimal pairs, but, unlike traditional minimal pair tasks, the participants do not read the words aloud themselves. Rather, in commutation tests, a speaker (or set of speakers) is recorded reading minimal pairs, and these recordings are then presented to the

listeners. After hearing a word, e.g. '*bad*', the listener is asked to report whether they heard the word '*bad*' or the word '*bed*'. Perception of the merger is then typically calculated based on the accuracy of a participant's recognition of the words they heard. An accuracy rate of 100% indicates that the listener has a clear phonemic distinction, whereas results closer to 50% recognition suggest that the participant has struggled somewhat in identifying the sounds correctly (Labov 1994: 356). The results of a commutation test, then, can be used to gain information regarding a listener's perceptual categorisation of certain sounds.

Commutation tests are said to be a more reliable method than minimal pair tests in research on the perception of speech because they assess 'perceptual responses to actual speech signals' (Gordon 2014: 185). The use of spoken minimal pairs naturally eliminates the issue of orthographic influence on perceptual responses. Additionally, commutation tests do not require participants to overtly report on their own language use, increasing the reliability of this particular method. Gordon's (2014: 185) emphasis on the use of 'actual speech signals' in commutation tests indicates a belief that this is a more appropriate method of examining perception, particularly in sociolinguistic studies, because results will more closely reflect how perception works in everyday, natural speech. Commutation tests, however, may not be applicable in all cases of perceptual investigation. Labov *et al.* (1991: 43) argue that neither minimal pair nor commutations tests could be used to investigate perception in Milroy and Harris' (1980) study of the near merged MEAT and MATE vowels in vernacular Belfast English as this accent feature is too stigmatised in the local speech community. Thus, the results of commutation tests, along with minimal pair judgements, may be influenced or skewed by stereotypes and stigmatisation.

A number of studies looking at mergers and near-mergers have used commutation tests to examine perception. In their study of the *ferry-furry* merger in Philadelphia, Labov *et al.* (1991) played listeners a tape of randomised *ferry-furry* and *merry-Murray* tokens, recorded by different speakers, and asked them to select which word they heard. In a twist on the traditional methodology, Labov *et al.* (1991: 62) presented participants not with a choice of words but with a choice of 'semantic labels'. For example, listeners were not simply circling whether they heard the word '*ferry*' or '*furry*', but whether they associate the word they just heard with a 'boat' or an 'animal'. This further tests to what extent a phonemic distinction carries meaningful, semantic information. Commutation tests have also been used to investigate vowel mergers in pre-/l/ positions, such as in Di Paolo and Faber's (1990) study of Utah English, and more recently, Arnold's (2015) research of this same merger in Ohio. However, Arnold (2015: 6) highlights a possible issue with

using commutation tests to examine perception. In accordance with Labov's (1994: 356) outline of commutation tests, any score below 100% accuracy is labelled as a failure to distinguish two sounds. Although some of her participants scored 95% and others 50%, Arnold (2015: 6) found that all listeners 'failed' the test, which indicates that they all show evidence of a perceptual vowel merger. Accuracy rate results from commutation tests may therefore be difficult to interpret, particularly with regards to variation in accuracy rates among speakers, given the pass-fail categorisation of results laid out in Labov (1994).

2.8.3 Coach Tests

Another method of examining a listener's perception of a phonemic distinction is through the use of what are often termed 'coach tests'. This method, devised by Labov *et al.* (1991), sees listeners presented with a story containing a sentence involving the merger under investigation. The example used by Labov *et al.* (1991) involves a narrative surrounding the coach of a sports team choosing to play a girl, *Merion*, or a boy, *Murray* in a crucial game. Two versions of the narrative were recorded by a speaker, containing one of the following sentences:

(a) "I gotta play *Merion* there"
(b) "I gotta play *Murray in* there"

Listeners were then asked to comment on the coach's decision. If the listener believed the coach chose the girl, they heard *Merion*, whereas if they commented that the coach chose the boy, they instead heard the phrase *Murray in*. After the participants had given their opinions the narrative was played to them again, this time containing the opposite (a) or (b) sentence. The assumption here is that listeners with a phonemic distinction will believe they had misinterpreted the story the first time it was played and alter their answer, while those who perceive a merger between *merry-Murray* will not change their interpretation of the narrative.

The main advantage of a coach test is that participants are completely unaware that the perception of a phonemic distinction is being tested. Unlike minimal pair and commutation tests which demand that the listener consciously thinks about differences and similarities between two words, coach tests were designed so that linguistic knowledge could be tested without the need for 'reflection and introspection' on the listener's part (Labov *et al.* 1991: 58). That is, participants are asked to semantically interpret the story, focusing on the content of the narrative rather than the vowel distinction itself. This eliminates certain issues found with other perception tasks, such as unreliability from self-reports, and

greatly reduces the possibility of an awareness of – or a negative attitude toward – a particular merger influencing the listener's responses.

Labov *et al.* (1991: 59) themselves note that the coach test is a somewhat 'elaborate' method. Requiring a narrative in which the distinction under investigation can be seamlessly inserted, coach tests can be a tricky and time-consuming method. Firstly, it may be difficult to find a minimal pair in which two items are interchangeable in the story; the pair would usually have to be the same part of speech, e.g. both nouns or names as in the Labov *et al.* (1991) example of *Murray* vs. *Merion*. This may be more difficult for some mergers than others. Perhaps the biggest disadvantage to the coach test method is its lack of repeatability within one experiment. Some mergers may be more advanced in certain lexical items or phonological environments than in others, and this can be tested for in minimal pair and commutation tests. However, with the coach test, narratives cannot be created for each possible occurrence of the merger in different phonological contexts. Similarly, the success of the coach test rests on the fact that the listener is unaware that a distinction is being tested, and that they don't realise that different versions of the story have been played to them. This is unlikely to work multiple times in one experiment session, meaning that the researcher can only obtain a limited amount of information on a participant's perception of the relevant sounds.

2.8.4 Vowel Continuum Tests

Perception tasks which make use of vowel continua involve the creation of a synthetic vowel continuum from one vowel to another. This type of method was used by Janson and Schulman (1983) in their study of the Swedish vowel merger between /e/ and /ɛ/. To create the vowel continuum Janson and Schulman (1983: 323) researched the formant values of typical Stockholm realizations of [ɪ], [e] and [a], using these values as a guide for their continuum. Formant values were converted to the mels scale – a measurement which is argued to give a more accurate representation of perceptual distances than formant values (Janson and Schulman 1983: 323) – then manipulated so as to create intermediate vowel stimuli 30 mels apart from each other. This produced in a continuum from [ɪ], through [e], and finally to [a], which contained in total 23 stimuli. These synthetic vowels were then inserted into natural stimuli of a Stockholm male speaker producing the word *sett*. Listeners were then presented with a tape of the stimuli, and were asked to indicate whether they heard the word *sitt, sett, sätt* or *satt*. Based on the listeners' vowel identifications, Janson and Schulman (1983) hypothesised that it would be possible to locate each listener's

phonemic boundary between /e/ and /ɛ/.

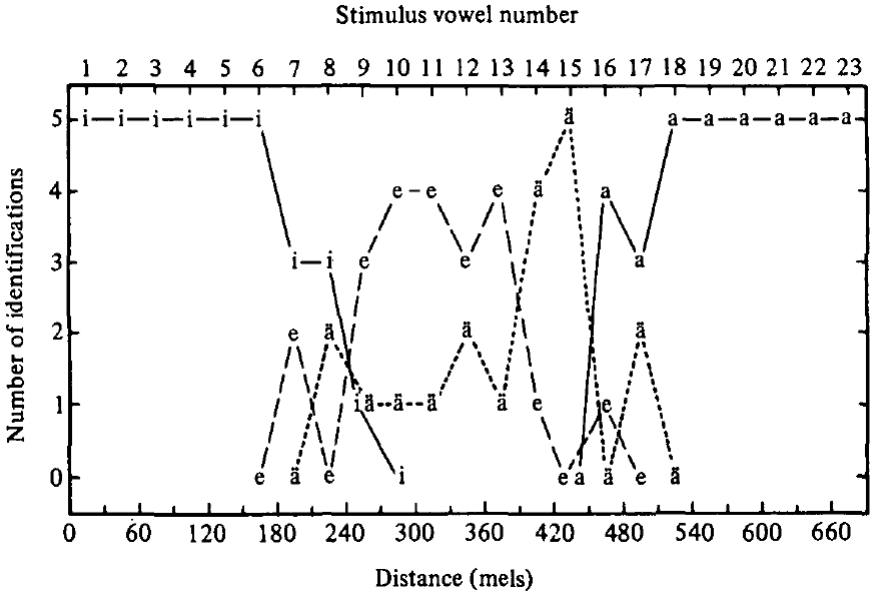


Figure 2.8: Perception results of one Lycksele speaker (*source:* Janson and Schulman 1983: 328)

The main advantage of the vowel continuum method is that it produces detailed results regarding the merger in perceptual space. Unlike the other methods discussed above, using a continuum of equidistant vowel stimuli allows the researcher to more closely pinpoint where the phonemic boundaries cross in merger situations, and can identify narrow regions in the perceptual vowel space where the merger occurs. As shown in Figure 2.8, Janson and Schulman (1983) were able to visualise the phonemic boundaries of their speakers, and to examine in detail where the phonemic boundaries of the merged /e/ and /ɛ/ vowels crossed. Such information is particularly useful in looking at the direction and the advancement of the merger in perception. Other methods such as the minimal pair, commutation, and coach tests merely tell the researcher whether the listener distinguishes certain sounds perceptually; these methods do not indicate where and to what extent sounds are merged in perceptual space.

There are, however, a number of disadvantages to the vowel continuum method. Firstly, in using synthetic stimuli, Janson and Schulman (1983: 329) could not be sure that the unnaturalness of the stimuli did not in some way influence listeners' perception. As Labov *et al.* (1991: 50) point out, perhaps participants in this study were not hearing the same speech cues they would hear in natural speech which may be important in the differentiation of the /e/ and /ɛ/ vowels. Thomas (2005; 2011) agrees, arguing that tasks involving the identification of isolated, synthetic

stimuli does not accurately reflect the process of perception in casual speech. Janson and Schulman (1983) were required to run a second experiment, this time using natural speech tokens, in order to verify the validity of the first experiment. Another more practical disadvantage to this method is the time and care needed to produce 'natural' sounding resynthesised stimuli. Not only do the formants of the vowel need to be manipulated in a systematic way to create a continuum, but the formant transitions to and from the vowel also need to be altered to prevent the stimuli from sounding too unnatural. This can be a lengthy and difficult process.

2.9 Summary

This chapter has presented an overview of the literature on phonemic mergers. As the aim of this thesis is to investigate the potential merging of the GOAT and THOUGHT vowels in contemporary Tyneside English, there are a number of findings in this chapter which will help to shape the method, the analysis techniques, and the interpretation of results in the main study.

From the discussion of different mechanisms of merger, and the investigation into several historical and contemporary cases of merger, this chapter has highlighted that not all cases of merger look the same. In terms of the mechanisms of merger, upon completion – whether via a process of approximation, transfer, or expansion – these sound changes all result in a reduced phonemic inventory due to the converging of two or more phonemes. However, while the merger is in progress, cases of mergers which take different development routes may look very different from one another. In order to determine by which mechanism the Tyneside GOAT-THOUGHT merger is progressing, if indeed these vowels are coalescing, it will be necessary to look for clues, manifest in speech production, that point to certain mechanisms. For example, if intermediate realisations between the GOAT and THOUGHT vowels are found in Tyneside, this would suggest that a merger-by-approximation situation is taking place. Alternatively, if there are no intermediate realisations, and a degree of lexical variability within the production of these vowels is found, this may be indicative of a merger-by-transfer situation.

This chapter has also highlighted a few key points to consider in examining the relationship between speech production and speech perception. As laid out in Labov (1994) it is generally expected that, in phonemic mergers and distinctions, production and perception are symmetrical. The assumption that production and perception reflect one another is also important in the 'categorical' view of phonology (Labov 1994: 352). The necessity of looking at both production and perception in

studies of potential mergers is also demonstrated in the discussion of near mergers, as it is only through a comparison of production and perception that can reveal whether two or more sounds are near merged or completely phonemically merged. As detailed in Section 2.7, there are relatively few studies investigating the relationship between production and perception in cases of mergers-in-progress, with some finding that production is ahead of perception (Thomas and Hay 2005; Arnold 2015), and others finding that perception seems to lead production (Bowie 2000; Freeman 2019) in the sound change. Based on this evidence, it is unclear what to expect of relationship between the production and perception of the GOAT and THOUGHT vowels if they are found to be merging in contemporary Tyneside English.

A review of different methods in examining mergers in speech perception has highlighted several methodological considerations. In selecting a particular method of measuring the perception of the GOAT and THOUGHT vowels in Tyneside, the goals of this thesis must be taken into account. As, first and foremost, a sociolinguistic study into Tyneside GOAT-THOUGHT merging, the most appropriate methods are those which measure perception in contexts that most closely resemble casual, natural speech. Methods which make use of synthesised stimuli, such as the vowel continuum tasks utilised by Janson and Schulman (1983), may therefore not be suitable for this study. Additionally, simple tasks such as minimal pair judgements may not be sophisticated enough to provide access to a listener's perceptual knowledge, thus yielding insufficient information on how listeners are able to distinguish and categorise these vowels perceptually. To investigate the perception of GOAT and THOUGHT among Tyneside listeners, then, it will be necessary to find a balance between methods that may be too 'blunt' (Gordon 2014: 184) to examine perception of these merging vowels, and methods which test perception in settings and situations that are too far removed from the perception of these vowels as heard in the speech community.

Following this general discussion of the phenomena of phonemic mergers, Chapter 3 will provide a background of Tyneside English before introducing evidence for a GOAT-THOUGH merger found in previous research on Tyneside.

Chapter 3

GOAT and THOUGHT in Tyneside English

This chapter explores Tyneside English, first outlining the geographical boundaries of the Tyneside region, and detailing how the term 'Tyneside' can be understood and defined. Section 3.3 presents past descriptions of Tyneside English in the literature, highlighting some notable features of the accent, while Section 3.4 discusses attitudes toward Tyneside English from both those inside and outside of the region. Sections 3.5 and 3.6 focus specifically on the quality of and variation in the THOUGHT and GOAT lexical sets in Tyneside English. Finally, Section 3.7 presents previous evidence of GOAT-THOUGHT merging in Tyneside English.

3.1 The Geography of North-East England

Throughout this dissertation, the term 'North-East' is used to refer to the northernmost region of England. Situated directly south of the Scottish border, the North-East comprises the counties of Northumberland, Tyne and Wear, and County Durham, in addition to the Tees Valley area of North Yorkshire. The majority of the North-East population live in the three conurbations of Tyneside, Wearside and Teesside, which surround the urban centres of Newcastle-upon-Tyne, Sunderland and Middlesbrough respectively.

Figure 3.1: Map of the North-East region (*source*: Beal *et al.* 2012: 25)

Often referred to as the 'Far North' (Wells 1982; Hickey 2015), the North-East region is commonly thought of as being isolated from the rest of England. Pearson (1994: 136-7) (as cited in Beal *et al.* 2012) writes that this sense of remoteness is due to the geography of the region, as the North-East is separated from its surrounding counties by the Pennines in the west, the North Yorkshire Moors to the south, and the Scottish border in the north. Given the region's proximity to Scotland, along with its geographical isolation from other parts of England, Watt (2002: 54) notes that many North-Easteners report a 'closer affinity' with Scotland than with England in terms of national identity. Much has also been written about North-Easteners' strong local identity (Beal 1999; 2009, Watt 2002, Beal *et al.* 2012), which is perhaps a result of this perception — held both by outsiders and residents of the region — that the North-East is distant and detached from other regions in England in terms of both geography and culture.

3.2 Defining 'Tyneside'

The term 'Tyneside' is often broadly thought of as encompassing the towns along the north and south banks of the River Tyne. However, Beal *et al.* (2012: 2) note that previous changes to county and local authority boundaries have led to a certain level of ambiguity in defining the Tyneside area, particularly the 'controversial' 1974 change which placed towns such as Newcastle, Gateshead, Wallsend and South Shields alongside Sunderland in a new county named 'Tyne and Wear'. The 2011 Census (Office for National Statistics) uses the term

'Tyneside' to refer to the city of Newcastle-upon-Tyne in addition to riverside towns such as South Shields, Tynemouth, and Whitley Bay. However, other places collapsed under the 'Tyneside' header in the census include the towns of Sunniside and Longbenton, which are located a couple of miles inland from the banks of the Tyne.

In Watt's (1998: 109) study of the Tyneside English vowel system, he defines Tyneside as containing the aforementioned riverside towns, in addition to Ponteland, a town in Northumberland located some 8 miles north of Newcastle-upon-Tyne. As different researchers and sources tend to differ slightly in their definitions of Tyneside, and as to which towns and villages can be labelled as 'Tyneside', it is difficult to draw strict geographical boundaries in defining this region. Watt (1998) also highlights how the attitudes of those within the region may impact definitions of 'Tyneside English'. He notes that speakers from north of the Tyne may assert that their speech is distinct from speakers born and raised south of the river, whose accent, they claim, more closely resembles that of Sunderland than that of Newcastle. It may therefore be useful to take into account attitudes and perceptions, alongside geographical information, in definitions of 'Tyneside English'.

Pearce's (2009) work on perceptual dialect mapping in North-East England is a valuable source in helping to define a geographical boundary by which to include or exclude participants for the present study of Tyneside English. Based on the responses from a dialect questionnaire, Pearce divided the North-East into three perceptually distinct dialect zones: Northern, Central, and Southern. The Northern sector is made up of South-East Northumberland (Ashington, Blyth, Cramlington) in addition to Tyneside[1], while the Central sector comprises of Sunderland and County Durham. This indicates that there is a perceptual difference between the accents of Newcastle and Sunderland speakers, while there is less of a distinction between the accents of speakers from the Tyneside area and speakers from Northumberland. Other researchers also acknowledge this, as Burbano-Elizondo (2015: 185), in citing Beal's (1999: 34) assertion that the 'important boundary is in the south', states that the difference in accent and identity are 'not as salient' between Newcastle and Northumberland as they are between Newcastle and Sunderland. These accounts of North-Eastern Englishes suggest that, north of the river Tyne, the boundaries within which Tyneside English is spoken are somewhat ambiguous, both perceptually and geographically.

[1]While Pearce does not specifically define what he means by 'Tyneside', the analysis makes clear that he is referring to the towns along the riverbank (e.g. Newcastle, Gateshead, Whitley Bay) as well as other more 'inland' towns in the area (e.g. Gosforth, Longbenton, Whickham).

3.3 Descriptions of the Tyneside Accent

Tyneside English, often otherwise referred to as the 'Geordie' accent, is the language variety spoken by those born and raised in the Tyneside region. It shares several features with other varieties in the North of England, namely the lack of a STRUT-FOOT distinction e.g. *put* and *putt* are both realised with an [ʊ] vowel, and the use of the lax [a] vowel in words belonging to both the TRAP and BATH lexical sets e.g. *bat* [bat] and *bath* [baθ][2]. However, Trudgill (1999: 70) notes that the accent is 'very distinctive', while Beal *et al.* (2012: 19) propose that it is one of the 'most easily identifiable' accents to non-linguists. Descriptions in the literature often detail a number of features which distinguish Tyneside English from other Northern varieties (Wells 1982; Beal 2008; Hughes *et al.* 2012; Hickey 2015). These include the pronunciation of a tense [i] vowel word-finally in *happy, very*, where other Northern varieties typically exhibit a lax [ɪ], and the retention of word-initial [h], which serves to distinguish Tyneside English from the majority of varieties in England, including the nearby Sunderland, Durham, and Middlesborough accents. Some of the most salient features of Tyneside English are briefly outlined below

3.3.1 Consonants

Highlighted as one of the most distinctive features of Tyneside English (Wells 1982; Docherty and Foulkes 1999a; Watt and Allen 2003), glottal reinforcement of the voiceless stops /p t k/ may occur between sonorants, as in words such as *butter* [bʊ͡ʔtə] and *jumper* [dʒʊmʔ͡pə]. Beal *et al.* (2012: 38-40) note that glottal reinforcement is found across the North-East, but that intra-regional differences are found: /k/ is most prone to glottal reinforcement in Middlesbrough English, while /p/ is least likely to be glottalized by Sunderland speakers.

Another consonantal feature often described in accounts of Tyneside English is the use of 'light' or 'clear' /l/ in all phonetic environments, e.g. *lamp* [lamp], *hill* [hɪl], *fully* [fʊli] (Wells 1982; Beal 2008; Hickey 2015). However, although often cited as a notable feature of the accent, Watt and Allen (2003: 268) find evidence of 'dark' /l/ in coda position in their analysis of a Tyneside speaker, indicating that the use of velarized [ɬ] is more common in contemporary Tyneside speech. More recently, in Turton's (2017) analysis of /l/-darkening in several different accents of British English, one speaker from Newcastle was shown to exhibit a difference in velarisation between /l/ in onset and coda position. This finding provides further

[2]Noted lexical exceptions to this rule include the words *master*, *plaster*, and occasionally *disaster*, which in Tyneside and Northumberland take a long [ɑː] vowel (Beal 1985; 2008).

evidence to suggest that, in contemporary Tyneside English, /l/ is no longer 'light' in all contexts.

Previous accounts of Tyneside English have suggested that linking and intrusive-/r/ occur less frequently in than in other British varieties, with Foulkes (1999: 264) reporting that the overall intrusive-/r/ occurrence rate within his sample of Tyneside speakers was just 8.5%. While Foulkes posited that linking-/r/ would eventually fall out of use in the region, more recent research has found an overall slight increase in the use of linking-/r/ among today's Tyneside speakers (Warburton and Turton 2017).

3.3.2 Vowels

Some of the most studied features of Tyneside English, and North-East accents more generally, are the FACE and GOAT vowels. Most accounts of the Tyneside accent report that, traditionally, the FACE and GOAT vowels are realised with a centring diphthong, usually transcribed as [ɪə] and [ʊə] respectively. However, Wells (1982: 375) expresses that these diphthongal variants are 'rather old-fashioned', and more recent research indicates that North-East speakers primarily use the more broadly northern monophthongal variants of [eː] and [oː] (Kerswill 1984; Watt 2002). A more detailed description of variation within and changes to the Tyneside GOAT vowel is presented later in this chapter.

Another oft mentioned feature of the Tyneside vowel system is the PRICE vowel, which is often pronounced with a high mid nucleus, as in [ɛɪ], where many other accents in England have a more open [aɪ] vowel. Both [ɛɪ] and [aɪ] are reported as present in speech community, and the usage of these variants on Tyneside is thought to be phonetically conditioned. [ɛɪ] does not occur word finally, in pre-/r/ position, or before a voiced fricative, i.e. /v ð z/, so that the vowel in *knife* [nɛɪf] differs from that in *knives* [naɪvz] (Beal *et al.* 2012; Hughes *et al.* 2012).

The MOUTH vowel is frequently discussed in descriptions of Tyneside English. Traditionally, this vowel is realised as the long monophthong [uː], as in *house* [huːs] and *about* [əbuːt]. Although this is perhaps one of the most stereotypical features of the Tyneside accent, [uː] pronunciations of MOUTH are found to be restricted to the speech of older working-class males in present-day Tyneside English (Beal 2008). The traditional, local [uː] pronunciation has become lexicalised in certain words, namely in *town* [tuːn], orthographically represented as *<toon>*, which is often used to refer to the city centre of Newcastle, or alternatively, in reference to Newcastle United, the city's beloved football club (Beal *et al.* 2012; Mearns 2015).

The unstressed vowels of the lettER and commA lexical sets are often noted for

their open back quality in Tyneside speech e.g.*scooter* [sku:tɐ] and *pasta* [pæstɐ] (Beal *et al.* 2012; Hughes *et al.* 2012). Usually transcribed as [ɐ] in the literature, this realisation is distinctly more open than the [ə] pronunciation present in the majority of British English varieties, and is also less backed than the [ɒ] found in Manchester and Sheffield Englishes (Beal 2008).

3.4 Attitudes to Tyneside English

3.4.1 Outsiders' Attitudes

Trudgill (1999: 70) writes that Tyneside English is difficult for people from outside of the region to understand, specifying that listeners from the south of England in particular struggle with the accent. This notion of Tyneside speech as incomprehensible is noted in Watt (1998: 123) as a trope frequently used in entertainment media, typically played for comic effect. Despite this stereotype of the unintelligible Geordie, more recently, Beal (1999: 38) highlights the favourable perception of Tyneside English by outsiders, noting in particular that Tyneside English is often rated as being one of Britain's most 'friendly' and 'sexy' accents in attitudinal surveys, and pointing to the growth of the call centre industry in the North-East as an indicator of the perception of Tyneside speech as friendly and approachable. Such attitudes are further evidenced in Montgomery's (2007: 79) perceptual study of northern Englishes, in which participants scored the Tyneside accent most highly in terms of 'pleasantness' when compared with other northern varieties.

3.4.2 Tynesiders' Attitudes

Tynesiders are often noted for the pride they display toward their own accent and their home region. Beal (1999: 37) comments that speakers of Tyneside English are aware of the distinctiveness of their own speech variety, and that they embrace this uniqueness in the form of their regional 'Geordie' identity. Further evidence of Tynesiders' pride in their own accent is noted in Beal's (2009:) study of enregisterment in Newcastle and Sheffield. She finds examples of the 'commodification' of Tyneside English in the plethora of merchandise adorned with Geordie phrases and orthographic representations of Tyneside accent features. Interestingly, Beal (2009: 147) gives an example of a novelty Geordie 'borth sortificat' (*birth certificate*), which uses an orthographic representation of the NURSE-NORTH merger. As research suggests this feature has all but vanished from

the region (Maguire 2007), this perhaps indicates that there is an idealized, more commercial version of the Geordie accent, different from contemporary Tyneside speech, which retains traditional, regional accent features that have since been lost.

Beal (1999: 37) notes that, based on the results of several newspaper and magazine surveys, Tynesiders have the highest level of 'accent loyalty' within England. However, based on anecdotal evidence and attitudes exhibited by his Tyneside participants, Watt (1998; 2002) writes that Tynesiders often hold negative views of their own accent. Despite research showing that Geordie is typically viewed favourably by outsiders (Beal 1999b; Montgomery 2007), Watt (1998: 124) mentions that the speakers in his study often displayed an awareness of local forms which may be viewed negatively by those from outside the region, avoiding these features and forms in their recordings. In studying the FACE and GOAT vowels, Watt (1998; 2002) finds that the majority of speakers favour supralocal, more generally Northern pronunciations, i.e. monophthongal [eː] and [oː], as opposed to the more traditional local centering diphthongs [ɪə] and [ʊə]. This, Watt posits, allows speakers to distance themselves from perceptions of Tyneside as industrial, isolated, and old-fashioned, while still retaining a 'northern' identity. Given the evidence of dialect levelling of the Tyneside vowel system in Watt's (1998; 2002) research, it would be interesting to trace whether the 'Geordie' accent and identity becomes less embraced in favour of a 'north-eastern' or a 'northern' identity, or less commercialised due to the rising usage of supralocal over local features.

3.4.3 North-Eastern Attitudes

As previously mentioned, Newcastle is the North-East's largest urban centre, with Beal (1999: 35) stating that there is no other English region where 'a single centre has so great an influence over such a wide area'. Pearce (2009: 164) highlights that media depictions of the North-East often focus exclusively on Newcastle or feature Tyneside English speakers, noting the rarity of representations of other accents in the region. Intra-regional accent differences have been reported in the literature (Pearce 2009; Beal *et al.* 2012), although, Trudgill (1999: 77) notes that, while a North-Eastener would never mistake a Tynesider for a Teessider, outsiders to the region may incorrectly assume that a Middlesbrough speaker is from Newcastle. This can often lead to the broad labelling of any speaker of a North-Eastern variety as a 'Geordie'. A number of studies which have investigated language and identity in Sunderland, Middlesbrough, and County Durham have found that speakers from these areas often take offence to being referred to as 'Geordie' (Burbano-Elizondo

2006; Llamas 2007; Pearce 2009).

Burbano-Elizondo (2006) writes of the strong rivalry between Sunderland and Newcastle, finding that many of her Sunderland participants resent the dominance which Newcastle has over the entire North-East region. Although often insistent that their speech variety is distinct from the Tyneside accent, Burbano-Elizondo (2006: 123) notes that many interviewees found it difficult to describe the differences between Sunderland English and Tyneside English. Similar viewpoints were also reported in Llamas' (2007: 599) study of language and identity in Middlesbrough, with many participants declaring their 'strong aversion' to Newcastle and the Tyneside accent.

Attitudes toward Tyneside English are more mixed in other areas of the North-East. In the south Durham town of Newton Aycliffe, West (2009: 101) finds that, while some participants fiercely rebuff the 'Geordie' label, many were 'un-offended' by the term. Moreover, several informants were found to express their dislike of Middlesbrough and Sunderland in particular, despite being geographically closer to these two urban centres than to Newcastle. In his study of Darlington, another south Durham town, Atkinson (2011: 210-211) finds further evidence of hostility towards Middlesbrough English, with participants describing the accent as 'harsh', and 'a poor man's Geordie'. Results from this study showed that Darlington speakers would rather be mistaken for a Tyneside English speaker than a Middlesbrough English speaker, with several speakers commenting on the 'flatness' of their own accent compared to Tyneside, and on how Geordie is a more 'North-Eastern' accent than their own language variety. Additionally, Kerswill (2003: 238) finds overwhelmingly positive attitudes to Tyneside English from his young Durham interviewees, who report that there exists 'no rivalry' between Durham and Newcastle. Furthermore, Kerswill (2003: 238) notes that these Durham speakers do not consider themselves to have a distinct identity from Tyneside, and, where linguistic and identity differences are acknowledged, they lament that they are not in fact 'real Geordies'.

3.5 THOUGHT in Tyneside English

3.5.1 The THOUGHT Vowel

The lexical set of THOUGHT is defined by Wells (1982: 144) as representing words realised with the vowel /ɔ(ː)/ in Received Pronunciation (RP) and General American (GenAm) accents, such as *caught, crawl, haul* and *walk*. In terms of vowel quality, Wells (1982: 145) describes RP /ɔː/ as a mid-back 'closely-rounded'

vowel intermediate between cardinal vowels 6 and 7, while GenAm /ɔ/ is more open, lying between cardinal vowels 5 and 6. Cruttenden (2014: 120) remarks that there is little dialectal variation in the pronunciation of THOUGHT, with the exception of the diphthongal and triphthongal variants found in Cockney speech, e.g. *board* [bɔʊd] and *bored* [bɔwəd].

Excluded from the THOUGHT lexical set are cases of /ɔ(ː)/ followed by /r/, which are subsumed under the NORTH lexical set. Wells (1982: 159) states that NORTH comprises of words with orthographic <or> and <ar> (e.g. *order, corn, wardrobe*), which are pronounced as /ɔr/ in GenAm. Additionally, words with orthographic <or> may fall into the FORCE lexical set. Words belonging to this set are defined as having /or/ pronunciations in most varieties of GenAm, and are usually spelled with one of the following: <or>, <ore>, <oar>, <oor>, <our> (e.g. *before, door, flour*) (Wells 1982: 161).

In many present-day British Englishes, a loss of rhoticity has led to a lack of a distinction between the THOUGHT, NORTH and FORCE lexical sets. For RP, Wells (1982) gives /ɔː/ pronunciations for all three lexical sets, indicating a merger of THOUGHT, NORTH and FORCE[3]. Most northern varieties are also described in Wells (1982: 360) as increasingly using /ɔː/ pronunciations for THOUGHT, NORTH and FORCE, though more traditional diphthongal pronunciations such as [ʊə] and [oə] may be found in some FORCE words, as in *door*.

3.5.2 THOUGHT Realisations on Tyneside

Like most non-rhotic varieties in England, Tyneside English does not retain a distinction between the THOUGHT-NORTH-FORCE vowels. Previous research has suggested that rhoticity survived in areas of North-East England longer than in many other English varieties, with the *Survey of English Dialects* (Orton and Dieth 1962-1971) mapping rhoticity to Northumberland in the 1950s and 1960s. Wells (1982: 368) also mentions that rhoticity can be found in the speech of some Northumbrians, however, more recent research finds little evidence of rhoticity in the region. In an investigation into rhoticity around the Scottish border, Watt *et al.* (2013: 90) note that 'derhoticisation has gone almost to completion' in the Northumberland towns on the English side of the border. While Beal (2008: 140) suggests that coda /r/ may still be heard today in the northernmost parts of Northumberland, she asserts that 'it would certainly not be found in Newcastle'. Furthermore, Beal *et al.* (2012: 90) state – in their book on urban varieties in

[3]Wells (1982: 161) mentions the possible exception of conservative RP speakers who retain /ɔə/ in some FORCE words.

North-East England – that all contemporary North-Eastern varieties can be described as 'non-rhotic'. As most recent research on North-East accents indicates a lack of rhoticity, particularly in the Tyneside area, it would not be expected that speakers of contemporary Tyneside English would distinguish THOUGHT from the NORTH and FORCE lexical sets. Therefore, for the remainder of this thesis, references to the THOUGHT vowel in Tyneside English will refer to the merged lexical sets of THOUGHT-NORTH-FORCE. It is however important to note that phonological differences may remain between the THOUGHT and NORTH-FORCE lexical sets. For example, while *'saw'* and *'soar'* may be produced and perceived identically in Tyneside English, *'sawing'* will typically be distinguished from *'soaring'* due to the retention of /r/ in cases where it is followed by a vowel, either word internally or across word boundaries.

The THOUGHT vowel in contemporary Tyneside speech is most often transcribed as a back open-mid [ɔː] vowel like that found in RP and most other British English accents (Watt 1998; Watt and Milroy 1999). However, more locally marked pronunciations of certain THOUGHT words can also be found within the region. For words with an orthographic <*a*>, Beal *et al.* (2012: 35) note that realisations of a low central [aː] vowel are common in the areas of Northumberland and Tyneside, such as in *talk* [taːk], *draw* [dɹaːɹ], and *all* [aːl]. This phenomenon is also reported to occur further south in the North-East, in County Durham, only the quality of the vowel is said to more closely resemble a low back [ɑː] vowel (Kerswill 1984: 1). Pronunciations of [aː] in THOUGHT words in contemporary Tyneside speech is reported to be somewhat rare, with Watt and Milroy (1999: 28) describing its usage as 'sporadic' and largely restricted to working-class speech. In other orthographic environments, such as before <ght> consonant clusters, Beal *et al.* (2012: 35) report that THOUGHT may be found to have a diphthongal [aʊ] pronunciation, as in *thought* [θaʊt] and *daughter* [daʊtə].

In their survey of monophthongal vowels in several varieties of British English, Ferragne and Pellegrino (2010: 28) find that the average formant values of THOUGHT in the Newcastle region are an F1 value of 489 Hz and an F2 value of 837 Hz. This indicates that contemporary Tyneside THOUGHT vowels are pronounced with an mid-back realisation. One factor of note with Ferragne and Pellegrino's study, however, is that these results are based off of only two male speakers from the Tyneside area. As such, the results may not be entirely representative of the larger speech community. Nevertheless, given that most literature on Tyneside English report the use of back open-mid [ɔː] vowels, and that the formant values for the Tyneside speakers do not differ noticeably from those reported for the majority of

other varieties surveyed in Ferragne and Pellegrino (2010), these F1-F2 values likely do provide a good indication of the typical quality and position of THOUGHT for the speakers of contemporary Tyneside English.

3.5.3 The NURSE-NORTH Merger

One of the more stereotyped features of the Tyneside accent is the apparent merger between the lexical sets of NURSE and NORTH. This results in homophony between words like *shirt-short* and *bird-board*. Wells (1982: 374) notes that the 'broadest Geordie' accents exhibit a merger of the NURSE and NORTH lexical sets, with NURSE being realised as the low-mid back vowel [ɔː]. The retraction of the NURSE vowel is often attributed to the influence of the 'Northumbrian Burr', which refers to uvular [ʁ] pronunciations found in the region prior to the loss of rhoticity (Påhlsson 1972; Wells 1982; Trudgill 1999; Beal *et al.* 2012). The merging of the NURSE and NORTH vowels is perhaps best illustrated by this oft cited 'Geordie joke', as transcribed in Viereck (1966: 95):

A man with a leg injury goes to the doctor
Doctor: "Can you walk [wɔːk]"
Patient: "Work [wɔːk]? I can hardly walk [waːk]"

Although termed 'The NURSE-NORTH Merger', Maguire (2007: 69) includes the FORCE set in his analysis of the phenomenon due to the absence of a NORTH-FORCE distinction in Tyneside English. However, Maguire (2007: 216) excludes THOUGHT in discussions of the NURSE-NORTH merger due to the continuing phonological differences between THOUGHT and NORTH-FORCE in cases such as *'sawing'* vs. *'soaring'*, as outlined above in Section 3.5.2. Maguire (2007: 216) also highlights the use of regional [aː] in certain THOUGHT words (*walk, talk, all*) as further justification of excluding this lexical set from analysis. Despite this, one could make a case for including THOUGHT in discussions of NURSE-NORTH merging. Firstly, the Viereck (1966: 95) transcribed joke – referenced above, and in much of the literature surrounding the NURSE-NORTH merger – illustrates a case of perceptual ambiguity between the NURSE word '*work*' and the THOUGHT word '*walk*'. Variation in a speaker's production, such as [aː] pronunciations in some THOUGHT words, may not necessarily indicate that the speaker has two distinct perceptual categories for THOUGHT and NURSE. Furthermore, the lack of a THOUGHT-NORTH-FORCE distinction in present-day Tyneside English, coupled with accounts which give [ɔː] as the typical pronunciation for all three lexical sets (Watt 1998; Watt and Milroy 1999), indicate that, while THOUGHT is perhaps not

included in the historical development of the NURSE-NORTH merger, today, backed [ɔː] pronunciations of NURSE would be homophonous with THOUGHT words as well as those in the NORTH-FORCE sets. Watt and Foulkes (2017) do include THOUGHT words in their analysis of NURSE-NORTH merging in World War 1 recordings of Tyneside English, excluding only pre-lateral THOUGHT tokens due the high occurrence of [aː] pronunciations. They state that if NURSE had merged with NORTH, then, due to the similar [ɔː] pronunciations of all three lexical sets, it must, by extension, also have merged with the FORCE and THOUGHT sets (Watt and Foulkes 2017: 159).

Despite being a well-known feature of the accent, merged NURSE-NORTH vowels are somewhat rare in contemporary Tyneside English. Looking at speech data collected from Tynesiders in the 1960-70s, Maguire (2007: 293) finds that, while some speakers exhibited some phonetic overlap of NURSE and NORTH, these vowels were distinct for the majority of speakers. The results also show that, even 50-60 years ago, backed realisations of NURSE were disfavoured by certain speaker groups, particularly young middle-class females. Watt and Milroy (1999: 39) also find NURSE-NORTH merging to be rare in their research into Tyneside English in the 1990s. The authors find that backed [ɔː] NURSE tokens made up only 7% of their data, with Tyneside speakers overwhelmingly favouring the centralised [ɜː] or the the fronted [øː]. Regardless of whether these vowels were ever completely phonemically merged in Tyneside (see: Maguire 2007), the findings of previous research indicates that centralised and fronted pronunciations of NURSE are those most common in the region, and that contemporary Tyneside speakers are unlikely to demonstrate any overlap between the NURSE and NORTH vowels.

3.6 GOAT in Tyneside English

3.6.1 The GOAT Vowel

Wells (1982: 146) defines the GOAT lexical set as comprising of words containing /əʊ/ in RP accents and /o ∼ oʊ/ in GenAm. The most frequent spellings of GOAT words are <o>, <o_e>, and <oe>, as in *so, home, toe* (Cruttenden 2014: 134). Wells (1982: 146) notes that the quality of the GOAT vowel is 'particularly variable both regionally and socially', and may be realised with one of a number of monophthongal or diphthongal variants.

The GOAT lexical set is derived from the Middle English (ME) /ɔː/ (*toe, sole, nose*) and /ɔu/ (*tow, soul, knows*) vowels. As a result of the Great Vowel Shift,

Wells (1982: 193) explains that ME /ɔː/ became raised to /oː/, while /ɔu/ became monophthongised to merge with /oː/. During the 18th and 19th centuries, diphthongisation of this vowel occurred, resulting in /oʊ/ pronunciations (Wells 1982; Cruttenden 2014). The development of GOAT, and the parallel development of the Present Day English FACE lexical set were discussed in further detail in Section 2.3.1. It is important to note that many accents of English have not undergone diphthongisation of the long-mid GOAT and FACE vowels. Wells (1982: 210) details that monophthongal variants are still common in northern England, in Wales, Scotland and Ireland, as well as in the US. Reduction of GOAT to schwa may also occur in word final position in regional speech, so that, for example, *window* is pronounced [wɪndə].

3.6.2 GOAT Realisations on Tyneside

A variety of different pronunciations of the GOAT vowel can be found throughout the North-East of England. Table 3.1 gives the four primary variants of GOAT which have previously been found to occur in Tyneside speech (Watt 1998; Pearce 2009; Beal *et al.* 2012).

Vowel	Description
oː	peripheral monophthong
ʊə	centering diphthong
oʊ	closing diphthong
ɵː	fronted/centralised monophthong

Table 3.1: Variants of the GOAT Vowel in Tyneside English

In contemporary Tyneside English, the most frequent variant of GOAT is reported to be the back monophthong [oː] (Watt 1998; Watt and Tillotson 2001; Watt and Milroy 1999). Not exclusive to the North-East, Watt (2002: 47) describes [oː] as a 'supralocal' variant, due to its occurrence in Scottish English and a number of other varieties in northern England. This monophthongal [oː] variant of GOAT is typically found, in both Tyneside speech and other varieties of British English, to occupy a higher position than THOUGHT in the vowel space (Watt and Tillotson 2001; Watt and Allen 2003). Also noted in Table 3.1, the centering diphthong [ʊə] and the non-peripheral [ɵː] monophthong are pronunciations which are more localised to the North-East. Usage of these traditionally regional variants has been found to greatly depend on social factors. Evidence suggests that [ʊə] and [ɵː] are used frequently by older working-class males, but are strongly disfavoured by female speakers in the

region (Watt 2002: 48). Realisations of the RP-like [oʊ] closing diphthong are also found in the region, usually in the speech of middle-class and young female speakers (Watt 1998; Watt 2002).

There are a number of exceptions to the pronunciations of GOAT outlined in Table 3.1. Like the regional pronunciations of THOUGHT, Watt and Milroy (1999: 28) note that some GOAT words can take an [aː] pronunciation in Tyneside English. Moody (2007: 2) indicates that [aː] pronunciations of GOAT are only permitted by certain spellings: <old> as in *cold* [kaːd], <ong> as in *long* [laːŋ], and <ow> as in *snow* [snaː]. In addition to these lexical restrictions, [aː] is reported to be largely confined to the speech of older speakers in the region (Watt and Allen 2003: 269; Corrigan *et al.* 2014: 119). A possible exception to this pattern is the realisation of *know* as [naː], which Crinson and Williamson (2004) find frequently in the speech of teenage Tynesiders. However, within the study, [naː], often written as *knaa*, is categorised as an example of non-standard lexis as opposed to a phonetic feature of the dialect; Crinson and Williamson (2004: 214) claim that the [aː] pronunciation in *knaa* does not 'fit[s] into patterns of the phonology of Tyneside'.

Diphthongal [aʊ] pronunciations are reported for certain GOAT words, such as in *soldier* [saʊldʒə] and *cold* [kaʊld] (Watt 2002; Moody 2007). Again, there are parallels with THOUGHT, as words belonging to both lexical sets may have regional [aʊ] pronunciations. Watt and Allen (2003: 269) highlight further possible variation in the Tyneside GOAT vowel, presenting examples of [iː] pronunciations, as in *no* [niː], and the diphthongal realisation [ɪə] in *stone* [stɪən].

3.7 Evidence of GOAT-THOUGHT Merging

In the late 1990s, Watt (1998: 161) first commented on a possible merger between the GOAT and THOUGHT vowels in Tyneside, stating the following:

> '[T]he quality of NORTH very frequently overlaps, to the point of homophony, with the quality of GOAT. For many speakers [...] the adjacent wordlist items *bought* and *boat* are perceptually identical, or at least so close as to be indistinguishable by the phonetician's ear.'

Since this first report, a few sources (Watt and Tillotson 2001; Watt 2002) have noted the indistinguishability of Tyneside GOAT and THOUGHT vowels. Despite these acknowledgements, no study of Tyneside English has yet provided an analysis of a possible GOAT-THOUGHT merger. Little is known of the quality of the homophonous GOAT-THOUGHT realisations, and the prevalence of this alleged merger in the speech

community is largely unknown. Furthermore, little consideration is given in the literature to the origins and development of this possible merger.

The current study presents an analysis of Tyneside speech taken from interviews obtained for the *Diachronic Electronic Corpus of Tyneide English*, hereafter referred to as DECTE (Corrigan *et al.* 2012). DECTE comprises of three subcorpora of Tyneside sociolinguistic interviews conducted in different decades throughout the late-20th and early-21st centuries. Data used in this thesis is taken from the most recent component of DECTE, the *Newcastle Electronic Corpus of Tyneside English* (NECTE) corpus, which includes interviews from 2007 to the present day. Watt's (1998) analysis of the Tyneside vowel system – in which he found initial evidence of a GOAT-THOUGHT merger – utilised data from the *Phonological Variation and Change in Contemporary Spoken English* (PVC) corpus, another component of DECTE consisting of interviews from the 1990s. The oldest subcorpus of DECTE, the 1960s-1970s *Tyneside Linguistic Survey*, has also been used to study the vowel system of Tynesiders, such as in Jones-Sargent (1983). Given the history of DECTE, an examination of the results of research carried out using these previous subcorpora of DECTE may shed some light on the timeline of GOAT-THOUGHT merging in Tyneside English.

3.7.1 1970s: The TLS Corpus

Conducted in the 1960s-1970s, the *Tyneside Linguistic Survey* (TLS) (Strang 1968) comprises of interviews with speakers from the Newcastle and Gateshead areas of Tyneside, and therefore includes speakers both north and south of the River Tyne. The general aim of the TLS, as detailed in Pellowe *et al.* (1972: 1), was to identify the speech varieties used in the Tyneside region, and to discover which dialectal features were used by which social groups within the speech community. Pellowe *et al.* (1972) also note the long-term aims of the TLS, which were to extend the scope of the survey to include other geographical areas in the North-East, and to carry out follow-up surveys to allow for analyses of dialectal change. Now a component of DECTE (Corrigan *et al.* 2012), the longitudinal aims of the TLS have been fulfilled.

A major concern of the TLS methodology was to avoid making assumptions regarding which linguistic variables were relevant in the differentiation of speech varieties and social groups on Tyneside (Jones-Sargent 1983; Jones-Sargent 1985). Critical of the 'Labovian model', Jones-Sargent (1983: 19) argues that an approach which pre-determines the linguistic and social features of interest will inevitably lead to the exclusion of important social and linguistic factors. As such, the TLS researchers worked toward a less restrictive method in which relevant features and

variables were not selected in advance, but rather that they were 'empirically' determined (Jones-Sargent 1983: 29). This involved the creation of an extensive questionnaire which asked participants to fill in their biographical details (age, sex, education level etc.) and to give details relating to factors such as hobbies, voting behaviour, and television watching preferences. Survey responses were then entered into a cluster analysis model which determined relevant social groupings. Thus, social factors of interest were determined empirically by the model.

Following the initial survey, one-to-one interviews were conducted with speakers of Tyneside English. The TLS originally aimed to conduct interviews with 250 speakers in the region. Of the audio recordings made during the 1970s, 37 TLS interviews are available via DECTE. All of the interviews included in the corpus are with participants from the Gateshead area.

3.7.1.1 GOAT and THOUGHT

Phonetic analyses of the 1970s TLS interview data can be found in Pellowe *et al.* (1972) and Jones-Sargent (1983). In these studies, the coding schema used for transcribing phonetic realisations found in the TLS interviews is unique. A hierarchical, three-tiered model is used to transcribe vowels in the data. First, and most broadly, the 'Overall Unit', or OU, is a 'categorical label' which closely correspond to Wells' (1982) lexical sets (Pellowe *et al.* 1972: 21). The principal difference between OUs and lexical sets is that, whereas the latter uses keywords to represent categories, 'Overall Units' make use of IPA symbols. For example, the FLEECE lexical set is represented by an OU of /i:/. The next tier is referred to as 'Putative Diasystemic Variables' or PDVs, which represent a broad phonetic transcription of variants of a particular OU. Each 'Overall Unit' has several PDVs. The bottom tier of the transcription model used for the TLS analysis is referred to as 'states'. These 'states' correspond to a fine phonetic transcription, and at least two states are given for every PDV in the data. Table 3.2. gives Jones-Sargent's (1983: 296-7) transcriptions of the GOAT and THOUGHT pronunciations used by TLS interviewees.

This TLS transcription scheme provides evidence of Tyneside pronunciations of the GOAT and THOUGHT vowels in the 1960s and 1970s. Table 3.2 shows that both THOUGHT and GOAT have a number of different realizations in the TLS data. The table highlights the high amount of variation within these vowels on Tyneside, while showing that this variation is often phonetically conditioned and lexically restricted. For example, as previously discussed, THOUGHT is often realised as [a] in pre-/l/ environments, while a diphthongal [ɪə] vowel may be heard only in certain GOAT

words, such as in *stone, home*. Jones-Sargent (1983) does not explicitly indicate the frequency of use of each of these variants in the TLS data, nor which speakers in the community use which particular variants. However, the adoption of the symbols /ɔ/ and /əʊ/ to represent the Overall Unit of THOUGHT and GOAT would indicate that these were the two most common variants of each vowel in 1970s Tyneside speech.

Lexical Set	**Overall Unit**	**PDV**	**Lexical Examples**
THOUGHT	ɔ	a	*all, talk*
		ɔ	*or, four*
		ɒ	*auction*
		ɛ	*more, sore*
		aʊ	*four, more*
		ʊə	*door, course*
GOAT	əʊ	əʊ	*so, phone, nose*
		əɪ	*so, no*
		ɔː	*so, smoke*
		uː	*go, nose*
		ɑː	*old, know, no, cold*
		ɪə	*stone, home*
		ɛʊ	*bolt, hope*
		ə	*pillow, yellow*

Table 3.2: Realisations of THOUGHT and GOAT in the TLS data (*source:* Jones-Sargent 1983: 296-7)

THOUGHT, represented in the table as /ɔ/, has six possible variants, both monophthongal and diphthongal. Some of the listed pronunciations are noted as regional THOUGHT variants in the literature, such as the pre-lateral [a] as in *all, talk*, along with the diphthongal realisations found in certain lexical items, such as [aʊ] in *four* and [ʊə] in *door* (Wells 1982; Beal *et al.* 2012). Perhaps unsurprisingly, [ɔ] pronunciations are also found in the TLS interviews. Notably, the transcription used by Jones-Sargent (1983: 296) does not include length marks, using [ɔ] where, at least for British varieties of English, [ɔː] is typically used. This could indicate that the Tyneside realisation of [ɔ] is shorter than that found in other varieties. However, other researchers studying Tyneside English typically transcribe [ɔː] as being the most frequent pronunciation of THOUGHT in the speech community (Watt 1998; Watt and Milroy 1999). As shown in Table 3.2. Jones-Sargent (1983) also uses [a] where others use [aː] (Beal *et al.* 2012),

indicating that the difference is a transcription discrepancy rather than signalling a lengthening of the [ɔː] vowel over time.

Turning to the GOAT vowel, Jones-Sargent (1983) gives [ɔː] as a possible realisation of GOAT on Tyneside. This differs from the transcription of the [ɔ] THOUGHT variant only by the presence of a length mark, indicating that some variants of GOAT and THOUGHT may be similar in terms of vowel quality. This brings into question the importance of vowel duration in phonemic distinctions; it may be the case that some Tyneside GOAT and THOUGHT pronunciations differ only in terms of duration. However, this overlap in the transcription of vowel quality for GOAT and THOUGHT cannot be taken as reliable evidence of a merger in the 1970s data. Due to the range of variants used in the community for these two vowels, it is not clear from Jones-Sargent's findings whether any speakers are found to use [ɔ] variants for THOUGHT in addition to [ɔː] pronunciations of GOAT. Thus, there is little evidence within Jones-Sargent's findings to suggest that any of the TLS speakers exhibit a phonetic overlap of the GOAT and THOUGHT vowels.

It it important to note that the use of the IPA symbol [ɔː] to transcribe certain realisations of the Tyneside GOAT vowels is somewhat perplexing. As shown in Table 3.2, [ɔː] is used by Jones-Sargent (1983) to describe a number of realisations of GOAT among the Tyneside speakers, notably appearing in words such as *so, smoke*. Moisl and Maguire (2008: 64), in their study of the TLS interviews, note that Jones-Sargent's (1983) [ɔː] transcription corresponds to the long back monophthong most frequently written as [oː] in research on Tyneside vowels. This would suggest that the monophthongal GOAT realisations are less open than the TLS transcription may at first suggest. However, things are further complicated by evidence found in Corrigan (2012: 3), which lists both [oː] and [ɔː] as possible variants of the GOAT vowel in Tyneside English. As such, there is a lack of clarity in research which provides only phonetic transcription (e.g. no auditory descriptions or acoustic measurements) as to whether GOAT may be realised with a THOUGHT-like [ɔː] vowel on Tyneside.

3.7.2 1990s: The PVC Corpus

The *Phonological Variation and Change in Contemporary Spoken English* (PVC) corpus emerged from a project in the 1990s which aimed to investigate sound variation and change in a number of urban varieties of British English (Milroy *et al.* 1999). In terms of its focus on Tyneside English, the PVC includes 18 dyadic sociolinguistic interviews with speakers from the Newcastle-upon-Tyne area. Unlike the TLS, the researchers involved in the PVC project set out with a clear

aim to investigate language use in different pre-determined social groupings (e.g. male and female speakers, younger and older speakers, working-class and middle-class speakers). In addition to interview data, the PVC also contains word list recordings for all of the speakers.

3.7.2.1 GOAT and THOUGHT

As referenced above, Watt's (1998) thesis on the Tyneside English vowel system reports a possible GOAT-THOUGHT merger in the speech community. Analysing speech data from the PVC corpus, Watt noticed instances of homophony between the GOAT and THOUGHT vowels for some speakers in the region. In his dataset of 33 speakers, Watt (1998: 161) states that roughly one third of the speakers in the sample exhibit a phonetic GOAT-THOUGHT overlap in the word list recordings. However, little detail is given regarding the quality of this 'merged' GOAT-THOUGHT vowel, nor is it discussed whether the realisation is closer to GOAT, closer to THOUGHT, or rather some intermediate pronunciation.

Among the PVC participants, Watt (1998: 161) finds that identical realisations of GOAT and THOUGHT in the word list data are largely produced by female speakers. As younger females are typically found to lead in cases of language change (Labov 1990; Labov 2001), this finding could be taken as evidence in support of the idea that a GOAT-THOUGHT merger is underway in Tyneside. However, Watt (1998: 161) adds that it is predominantly the older and working-class women who exhibit an overlap of GOAT and THOUGHT. He explains this by noting that a merging of these vowels in production requires that the speaker produces a back monopthongal [oː] GOAT vowel, whereas younger and middle-class female speakers generally produced the closing diphthong [oʊ] in the reading task. Given that Watt (1998: 233) finds the back, peripheral monophthong [oː] to be the most frequently used GOAT variant among young and middle-class females in the data taken from the PVC sociolinguistic interviews, it may be possible that there is evidence of GOAT-THOUGHT merging among these speakers in more naturalistic speaking styles. Nevertheless, as this potential merger was not the main focus of Watt's (1998) study, there is no discussion of GOAT-THOUGHT merging in the sociolinguistic interview data.

The only evidence of GOAT-THOUGHT merging presented in Watt (1998) is in the context of the words '*boat*' and '*bought*'. As such, questions remain unanswered on the contexts in which homophony may occur. It could be the case that only certain words show overlap between GOAT and THOUGHT, as in potential merger-by-transfer situation. Or, perhaps GOAT and THOUGHT are participating in a conditioned merger, in which the vowels only surface as identical in specific

phonological contexts. Due to the lack of literature on this GOAT-THOUGHT overlap, and the absence of any existing detailed analysis of this variable, very little is currently known about this potential vowel merger in Tyneside English.

3.7.3 GOAT-THOUGHT Merging in Other Varieties

There are a number of studies which find evidence of GOAT-THOUGHT merging in varieties other than Tyneside English. Jansen (2015: 213) gives an overview of the vowel system of speakers from Carlisle, a city approximately 50 miles west of Newcastle-upon-Tyne. She reports the variable pronunciations of THOUGHT as [o̞ː], [ɔː], or [ɔ̟ː], while GOAT is typically realised as either [oː] or [ɔ̝ː]. This indicates that the quality of GOAT and THOUGHT are very similar for Carlisle English speakers, and that there is a possibility of vowel overlap. Due to the similarity of the Carlisle and Tyneside accents (Hughes *et al.* 2012: 124), and the relative close proximity of the two cities, it would not be too far-fetched to hypothesise that Jansen's transcriptions of Carlisle GOAT and THOUGHT may also reflect the quality of these vowels in contemporary Tyneside speech. Elsewhere in Cumbria, Jansen's (2018) recent study of speakers in the town of Maryport explores the merger of the GOAT and THOUGHT vowels, though findings indicate that these vowels are becoming more distinct in younger speakers.

Evidence of GOAT-THOUGHT merging has also been found in other Northern Englishes which have monophthongal realisations of the GOAT vowel. Watt and Tillotson (2001) find an overlap of GOAT and THOUGHT for some speakers of Bradford English, the quality of both vowels being close to the low-mid back [ɔː]. However, due to the advance of GOAT-fronting in the region, the results of the study indicate that older speakers exhibit the most overlap, while younger speakers are increasingly distinguishing the distance between the two vowels. In what Wells 1982: 360) terms the 'Middle North' – i.e. the area encompassing Greater Manchester and South and West Yorkshire – THOUGHT is often realised as [ɒː]. He attributes this lowered pronunciation to the 'pressure to preserve a clear distinction between GOAT and THOUGHT, as [lo̞ːʷ] '*low*' vs.[lɒː] '*law*'. This is consistent with the description of the Lancashire accent in Hughes *et al.* (2012: 150); although the GOAT vowel has a monophthongal [ɔː] realisation, open [ɑ̟ː] pronunciations of THOUGHT help to preserve a distinction between these two vowels in the speech community. However, in investigating the vowel system of speakers from Burnley, Lancashire, Ferragne and Pellegrino (2010: 17) note that several speakers exhibit 'perfect homophony' between the GOAT and THOUGHT vowels, as exemplified in the similar formant frequencies for the vowels in the

words *'hoed'* and *'hoard'*. It may therefore be the case that in contemporary Lancashire English, some speakers are raising the once lowered THOUGHT vowel to the position of GOAT, resulting in an overlap of the two vowels in phonetic space. While the majority of evidence for GOAT-THOUGHT merging comes from accents in the North of England, Wells (1982: 382) notes that speakers from many areas of Wales, typically those with less of an English influence, will pronounce the GOAT vowel with a monophthongal [oː]. In these regions, Wells states that speakers often lack a distinction between the GOAT and the THOUGHT-NORTH-FORCE lexical sets.

3.8 Summary

The focus of this chapter has been on Tyneside English, providing definitions and descriptions of the accent alongside findings from perceptual research into attitudes toward North-East English varieties. This chapter has also presented an overview of the status of – and variation within – the GOAT and THOUGHT vowels on Tyneside. Finally, previous evidence of a potential Tyneside GOAT-THOUGHT merger was discussed, with a particular focus on studies using older subcorpora contained within the larger DECTE corpus. While there is evidence to suggest a merging of these two vowels in Watt's (1998) analysis of the Tyneside vowel system, there is a distinct absence of information in the literature regarding the production and/or perception of these vowels in the Tyneside speech community.

The remainder of the thesis is dedicated to investigating a possible merger of the GOAT and THOUGHT vowels in contemporary Tyneside English. The following two chapters, Chapters 4 and 5 will focus on examining these vowels in speech production, using sociolinguistic interview obtained from DECTE. Then, Chapters 6 and 7 will detail the methodology of and results from an investigation into the perception of GOAT and THOUGHT among Tyneside listeners.

Chapter 4

Production: Method

This chapter discusses the methodology of the investigation into the production of the GOAT and THOUGHT vowels in contemporary Tyneside English. In Sections 4.1-4.4, the data collection process is described, detailing the nature of the corpus used in the study, the social variables chosen for analysis, and the participant selection process. Then, Section 4.5 presents the experimental procedure, providing an overview of how the data was transcribed, and how the target vowels were automatically aligned and extracted to allow for an acoustic analysis. Multiple methods and measures are used in the analysis of the production data, such as Pillai scores and Generalised Additive Mixed Models (GAMMs), each of which are explained in Section 4.6. Finally, the techniques used for a quantitative analysis of the data are discussed in Section 4.7.

4.1 The Diachronic Electronic Corpus of Tyneside English

The data used in this study were obtained from the *Diachronic Electronic Corpus of Tyneside English* (DECTE) (Corrigan *et al.* 2012), a large corpus of speech data taken from North-Eastern English speakers. As briefly discussed in the previous chapter, DECTE comprises of three components or sub-corpora collected in three separate decades: the *Tyneside Linguistic Survey* (TLS) created in the 1960s (Strang 1968); the *Phonological Variation and Change in Contemporary Spoken English* (PVC) project collected in the 1990s; and most recently, the ongoing *Newcastle Corpus of Tyneside English 2* (NECTE2) which has been collecting data at Newcastle University since 2007. As the present study is concerned primarily with contemporary Tyneside speech, only data from NECTE2 (2007-onwards) is

used in the following analysis.

DECTE, across all three of its components, is a database of hundreds of sociolinguistic interviews. The 1960-70s TLS interviews each feature one interviewer and one speaker (randomly selected from the electoral register) and typically follow a conventional question-answer interview format. By contrast, the PVC and NECTE2 projects feature dyadic interviews with a pair of speakers, often friends or relatives, who were recruited via snowball type sampling. This results in a more informal, conversational style of sociolinguistic interview. Given the size, the contemporariness, and the naturalness of its interview data, DECTE is a valuable source for examining linguistic variation in today's Tyneside speakers.

Recent research on Tyneside English often uses DECTE as a data source, from Buchstaller's (2011) work on variation and change in Tyneside quotative usage, to Beal and Burbano-Elizondo's (2012) research into lexical variation in North-East England, to Fehringer and Corrigan's (2015b) look at changes to the Tyneside modal system (see also: Barnfield and Buchstaller 2010; Fehringer and Corrigan 2015a; Buchstaller and Mearns 2018). Various researchers have also used DECTE as a data source in their work on the Tyneside accent, including Foulkes's (1997) work on [r]-sandhi usage in Newcastle, Watt's (1998; 2000; 2002) look at variation and change in Tyneside vowels, and Moisl and Maguire's (2008) research into sociophonetic variation in North-East England.

4.2 Participant Selection

The lack of previous, in-depth research on GOAT-THOUGHT merging in the North-East means that there is no indication as to whether this feature is limited to Tyneside speech, or as to how geographically widespread this potential merger may be in the region. Additionally, as noted in the previous chapter, there is a certain ambiguity in delimiting the borders of Tyneside English. Researchers typically find that those in the region consider the Tyneside accent to be distinct from the variety spoken in the neighbouring city of Sunderland (Watt 1998; Beal 1999b; Pearce 2009), and a number of phonetic differences have been found between the two accents (Beal *et al.* 2012; Burbano-Elizondo 2015). The northern boundary of Tyneside English, however, is more difficult to define, as Beal (2000: 355) notes that many Northumbrians consider themselves as 'Geordies', and as speakers of Tyneside English.

In light of Pearce's (2009) study of perceptions of North-East varieties, in which he divides the region into three different accent zones (North, Central, South), it

was decided that all speakers in the current study must have been born and raised in the Northern zone. Therefore, participants who were selected for analysis were those who had categorised their own regionality as either Tyneside, South Tyneside, North Tyneside, or (South-East) Northumberland.

4.3 Social Variables

4.3.1 Sex

The social variable of speaker sex is included in the majority of studies on language variation and change. Labov (1990: 205-6) summarises the differences between men and women linguistically in two principles:

> **Principle I:** In stable sociolinguistic stratification, men use a higher frequency of non-standard forms than women.
> **Principle II:** In the majority of linguistic changes, women use a higher frequency of the incoming forms than men.

Labov's (1990) Principles I and II posit that the linguistic differences between men and women also differ depending on the variable of study. Due to the amount of evidence supporting these principles, Chambers (2013: 302) notes that, in each speech community, there are 'partly predictable linguistic correlations with sex'. If a stable variable is being investigated, men would be expected to exhibit higher rates of vernacular variants than women. In cases of a linguistic change in progress, women are expected to be the leaders. However, Eckert (2013: 245) warns researchers against simplifying the linguistic differences between the sexes as being 'more or less conservative'. While Labov (1972: 243) suggests that men use more stigmatized variants and women use more prestige forms, Milroy and Gordon (2003: 103) find that women use more 'supra-local' forms which 'may or may not be identifiable as prestigious'. While Principle II stresses that women lead language change in the 'majority' of cases, Labov (2001: 284) does note that there are a 'small minority' of cases where men lead linguistic change. Still, Labov (2001: 280) asserts that, within sociolinguistic research, women are overwhelmingly found to 'be in advance of men in most linguistic changes in progress'.

Speaker sex has previously been found to affect the GOAT and THOUGHT-NORTH-FORCE vowels in Tyneside English. In Maguire's (2007) research

into the reported merger of NURSE-NORTH[1] on Tyneside, which looked at speech data from the 1970s, sex was found to be the most important social predictor in the pronunciation of the NURSE vowel. Female speakers were found to prefer fronted realizations of NURSE, while male speakers were more likely to exhibit backed variants which showed greater overlap with NORTH. In terms of GOAT variants, Watt (1998; 2002) finds that male speakers are significantly more likely than females to use regionally marked pronunciations such as [ʊə] and [ɵː], with females preferring the unmarked and prestige forms of [oː] and [oʊ]. Additionally, in his brief discussion of GOAT-THOUGHT homophony in Tyneside English, Watt (1998: 161) notes that it is primarily female speakers for whom the GOAT and THOUGHT vowels sound 'indistinguishable'. Given this evidence of how greatly speaker sex has been found to affect the pronunciation of both GOAT and THOUGHT on Tyneside, it will be important to include sex as a social factor in the following analysis of GOAT-THOUGHT merging in the Tyneside speech community.

4.3.2 Age

Speaker age is a vital social variable to include in apparent-time studies – such as that presented in this thesis – as it acts as a proxy for real-time linguistic change, showing which linguistic variants may be entering, leaving, or changing in the speech community. Research using apparent-time studies works on the assumption that the linguistic differences manifest in speakers of different ages reflect changes in the way people speak over time. As Milroy and Gordon (2003: 35) explain, the speech of a 75 year-old reflects the linguistic norms of a particular speech community at an earlier point in time than the speech of a 50-year old or a 25 year-old, and so on. Due to the comparative ease of collecting data for an apparent-time study as opposed to a real-time study, the apparent-time framework is the most frequently used in research into language change (Cuckor-Avila and Bailey 2013: 258).

Chronological age is perhaps the easiest social variable to define, and therefore collecting data from speakers of different ages is relatively uncomplicated. However, in terms of analysis, Milroy and Gordon (2003: 38-9) note that speakers must be grouped in 'meaningful ways' in order to allow for a comparison across age groups. They continue that arbitrary groupings of age, such as into decades (e.g. 20-29, 30-39 etc.), will make it difficult for researchers to explain any linguistic differences between age groups that might bear out in any data analysis.

[1]Maguire (2007) uses the notation NORTH to refer to both the NORTH and FORCE lexical set, while THOUGHT is used in this thesis to refer to the THOUGHT, NORTH and FORCE lexical sets. See section 3.5.3 for further discussion.

Age is included as a variable in apparent time studies of GOAT and THOUGHT in Tyneside English, notably in Watt (1998; 2002), Watt and Milroy (1999) and Maguire (2007). Researching the NURSE-NORTH merger using data from the 1970s, Maguire (2007: 281) found age to play a small role in determining whether speakers used fronted or backed variants of NURSE, determining that age was a less important factor than speaker sex and social class. However, examining 1990s Tyneside speech data, Watt and Milroy (1999: 38) do find age to be an important predictor of the pronunciation of NURSE. The authors find that the backed [ɔː] variant of NURSE is relatively rare among the participants – making up only 7% of all NURSE tokens – and that its usage is largely restricted to older speakers. The use of fronted NURSE realizations by younger speakers in Tyneside therefore increases the distance between NURSE and the THOUGHT-NORTH-FORCE vowels, and reduces the possibility of overlap. Similarly, Watt (2002: 47-48) reports that younger speakers are increasingly using fewer localised variants of the GOAT vowel, concluding that the Tyneside vowel system is undergoing levelling due to the rise in usage of 'supralocal' and national forms of the GOAT and FACE vowels.

Regarding a potential Tyneside GOAT-THOUGHT merger, Watt (1998: 161) suggests that middle-aged and older female speakers are more likely to exhibit homophony between the two vowels, as younger females were more likely to use the non-regional closing diphthong [oʊ] variant of the GOAT vowel. As such, it is important to include age as a predictor in an analysis of this potential merger in Tyneside English. Speakers in the current study range from 19-74 years of age, split into three age groups: young (19-25 years); middle (34-55 years); and older (60+ years). These groups are defined to an extent by the data available in the DECTE corpus e.g. few speakers under 19 years of age, few speakers aged 25-35 years. However, it serves to compare speakers at different approximate life stages e.g. comparing students and individuals early in their career with speakers who may have more work-related and familial responsibilities, and to those close to and post-retirement age. These age groups should therefore provide a good overall representation of the adult Tyneside speech community.

4.3.3 Social Class

A speaker's socio-economic status, or social class, is often included as a social variable in sociolinguistic studies. Ash (2013: 350) comments that the inclusion of social class as a variable 'regularly produces valuable insights into the nature of linguistic variation and change'. In stable situations, Labov (1990: 220) states that speakers categorised as being of a higher social class use more prestige forms, while

the speakers categorised as having a lower socio-economic status use more stigmatised variants. Furthermore, social class is often found to interact with speaker sex, with Labov (1990: 220) continuing that it is lower middle class women who most disfavour stigmatised forms, and most frequently use prestigious forms. In cases of language changes that occur below the level of conscious awareness (often termed a 'change from below'), Labov (1990: 226) notes that incoming and innovative forms are used more frequently by the 'intermediate groups' (i.e upper working and lower middle classes) than by those speakers at either end of the social class hierarchy.

Social class is another social variable which has been found to affect the use of different GOAT and THOUGHT variants in Tyneside English. Comparing working and middle-class speakers, Watt and Milroy (1999: 39) find that backed variants of the NURSE vowel are primarily used by working-class older male speakers, and Maguire (2007: 287) notes that middle-class speakers avoid backed [ɔː] pronunciations of NURSE. This suggests that the 'merging' of NURSE with NORTH was limited to working-class speakers, predominantly the older males. Turning to the GOAT vowel, Watt (2002: 47-48) reports that middle-class speakers in Tyneside are most likely to use a 'prestige' diphthongal [oʊ] realisation. Watt (1998: 161) also highlights that social class may determine whether a Tyneside speaker is likely to merge GOAT and THOUGHT; because middle-class speakers in the region prefer to use a diphthongal [oʊ], working-class speakers, who primarily use monophthongal [oː] pronunciations of GOAT, are more likely to exhibit GOAT-THOUGHT homophony.

4.3.3.1 Measuring Social Class

To include socio-economic status as a variable in a study, researchers must first work out how to measure social class. Labov (1990: 220) refers to social class as being 'a generally recognized hierarchical organization of the speech community'. This vague definition reflects that different researchers take different approaches to defining and measuring social class, with Ash (2013: 350) commenting on the 'lack of consensus' among researchers as to which variables are important in determining a speaker's socio-economic status. Trudgill (1974: 31) notes that it is 'necessary', in every study which includes it as a social variable, to define 'exactly what is meant on each occasion by "social class"'. Nevertheless, despite the number of different approaches to measuring social class, Labov (1990: 220) observes the 'remarkable uniformity' of evidence showing the effects of socio-economic status on linguistic variation and change.

A number of researchers (Labov 1966; Trudgill 1974) have conducted studies

in which they create their own 'social index' to measure social class. These social indices combine multiple factors – such as occupation, level of education, and income – which are rated on a point scale (e.g. the higher participant's income, the higher they score on the point scale). Such a method is advantageous given that researchers have often struggled to pinpoint precisely which variables give the best measure of social class. However, Milroy and Gordon (2003: 46-7) point out that social indices are infrequently used in contemporary research.

Some researchers propose that the occupation of a participant may be the most important factor in determining their social class. Ash (2013: 351) explains that 'the most simplistic' categorisation of social class, used across the social sciences, places speakers on a scale whereby non-manual occupations are given a higher position in the social class hierarchy than speakers with more manual occupations. Labov (1990), in revisiting the data from his earlier 1980 study in Philadelphia, found occupation to have a greater effect on the sound changes taking place in the speech community than the combined social index comprising of the factors of occupation, education level, and residence value. Similarly, Baranowski and Turton's (2018) recent investigation into the best indicators of social class finds occupation to be the best predictor of social class for some linguistic variables, while combining occupation with education level works best for others.

Further complications to including social class as a variable arise due to an individual's ability to move up and down the social hierarchy. As social class is commonly measured using factors such as occupation, education level and residence value, one's social class is subject to change if they start a new job, spend another year or two in education, or possibly if they move to a different house or neighbourhood. Trudgill (1974: 33) states that due to social mobility and the fluidity of socio-economic status, there are often 'anomalous individuals' who cannot be confidently assigned to a specific social class. Milroy and Gordon (2003: 44) also highlight that the social class hierarchy is geographically variable. They argue that two speakers from different areas of the country may share an occupation but be perceived differently in terms of prestige, e.g. a lawyer living in a non-wealthy area will enjoy more prestige than a lawyer living in a region in which a great percentage of the population are classified as being of a high socio-economic status. This suggests that different speech communities cannot be so readily compared using the same definitions and measures of social class.

4.3.3.2 Social Class in DECTE

Each subcorpus of DECTE has its own experimental design and methodology, and thus social class is measured differently in each. The oldest component of DECTE, the 1960s-70s TLS interviews, asked participants to fill in an extensive questionnaire (discussed in more detail in section 3.5.3), and social class groupings were subsequently determined by a cluster analysis model which was used on the questionnaire data (Jones-Sargent 1983). In the 1990s PVC subcorpus, social class was defined according to neighbourhood and social networks, as interviews were carried out with participants from pre-selected housing estates in Newcastle; one generally defined as 'working class' and the other as 'middle class' (Watt 1998; Docherty and Foulkes 1999b). Unlike the PVC, the more recent NECTE2 subcorpus did not pre-define social class, nor was it experimentally designed to collect data from equal numbers of working-class and middle-class speakers. This would be impossible given the nature of the corpus, as each year university students carry out new interviews by selecting members of their own family and friendship groups. NECTE2 does however collect biographical data from participants which could be used to define and measure socio-economic status, such as self-reports of social class, occupation, level of education, and parents' occupations.

Self-Reports of Social Class

In NECTE2, speakers are asked to fill in a short biographical questionnaire prior to the interview. This questionnaire asks participants to fill in information about their own 'social class'. The most common responses in the corpus are 'working-class' followed by 'middle-class', with a small minority of participants categorising themselves as being either 'lower middle-class' or 'upper middle-class'. The lack of a consistent scale is an obvious problem here, as it is difficult to place self-identifying 'middle-class' speakers on a scale which also includes 'lower middle-class' and 'upper middle-class' boundaries. Further complications arise in speakers who struggle to categorise themselves, with some speakers reporting that they are on the 'working-class to lower middle-class' boundary. Some responses to the social class question avoid using this traditional working/middle-class hierarchy altogether, instead using the classes as laid out by the Office for National Statistics (ONS) (2010) e.g. a class of L15 is given for full-time students. Again, it would be difficult to confidently place these participants using ONS categories to describe their own social class in any social class hierarchy alongside those who self-identify using terms such as 'working-class' or 'middle-class'. The NECTE2 database also shows that

social class information is missing for a large number of speakers. This could be for a number of reasons, such as participants being unwilling or unable to categorise themselves as belonging to a particular socio-economic group. Notably, social class information is missing for the majority of participants who were interviewed between the time period of 2012-2013, perhaps indicating that interviewees were not required to provide self-reports of their own social class on the biographical questionnaire during these years.

Self-reports of social class are therefore not a viable measure of social class in the present study due to (1) the difficulty of comparing self-reports of class, as the definition of and boundaries between socio-economic classes may differ from person to person, (2) the lack of a complete and consistent social class scale within the NECTE2 data, and (3) the extent of missing self-reports of social class for interviewees.

Education

The education levels of the NECTE2 speakers fall into three general categories. Many of the interviewees, primarily the younger speakers, are categorised in the corpus as having had a 'higher education'. This means that they have obtained, or are in the process of obtaining, some type of university degree. Other participants are described as having completed 'further education', which refers to any post-secondary education at a school or college e.g. A-levels or equivalent qualifications. Remaining speakers can be clustered into a 'secondary education' group. The majority of these speakers, where specified, left education at 15 or 16 years of age. While some participants detail which type of exams they sat (e.g. GCSEs, O-levels), the majority provide no further information. As such, it would be difficult to refine this group further, given the lack of information for most participants about (1) the age they left school, and (2) which, if any, exams they completed. Furthermore, within NECTE2, education level correlates closely with age: a high proportion of younger speakers hold a degree or are currently at university, while the older speakers typically did not remain in education post secondary school. Education level is therefore unlikely to be the fairest measure of social class in the DECTE data.

Occupation

Early sociolinguistic studies often utilised occupation-based social class scales used in official surveys and censuses. Labov's (1966) New York City study categorised speakers into class groups using classifications from the Bureau of the Census,

while several British studies (Trudgill 1974; Macaulay 1978) created occupation groupings based on the Registrar General's Classification. As there is a precedent in sociolinguistic research to use official occupation-based categorisation schemes, the National Statistics Socio-economic Classification (NS-SEC) is used in the present study to determine the social class of the NECTE2 participants. According to the NS-SEC guidelines, information regarding participants' occupations was used to categorise speakers into four social class groups: Lower-Working Class (LWC), Upper-Working Class (UWC), Lower-Middle Class (LMC) and Upper-Middle Class (UMC).

One complication that arises in using occupation as a predictor for social class within the DECTE data is that the majority of young speakers in the corpus are university students. Given that data collection for the corpus is continued each year via students at Newcastle University, it is perhaps inevitable that the corpus would skew towards collecting data from (1) younger speakers (i.e. aged 18-21 years), and (2) speakers who are studying at the university. The National Statistics Socio-economic Classification (NS-SEC) allocates the class of L15 to full-time students. However, anyone within the L15 boundary is excluded from the occupation-based social class hierarchy proposed by the NS-SEC. The Office for National Statistics (2010) instead propose that for any data analysis including full-time students, other measures such as parental occupation should be used in order to allow for this group to be better defined in terms of their socio-economic status. As information regarding parents' occupations is available for the majority of speakers in DECTE, the social class groupings of participants listing their occupation as 'student' are here determined by their parental occupation[2].

4.3.4 Ethnicity

There has been little research on the Tyneside accent which includes ethnicity as a social variable, and thus no previous studies investigating how ethnicity may affect realizations of GOAT and THOUGHT in Tyneside English. According to the results of the 2011 UK census, the North-East region has the highest percentage (93.6%) of residents identifying as 'White British' in the whole of England and Wales (GOV.UK 2018). As DECTE is overwhelmingly made up of speakers identifying as 'White British', the following analysis focuses only on these speakers, and therefore excludes ethnicity as a social variable.

[2]The occupation of both parents were checked against the NS-SEC guide, and students were allocated a social class based on the occupation of the parent who was determined to be of a higher social class.

4.4 Participants

The present study features 28 speakers, all of whom were born in and reside in the North-East of England. Each interview obtained from DECTE features two interviewees, typically friends or relatives, of approximately the same age. The majority of the selected interviews feature a same-sex pair of speakers, although the interviews with the oldest speakers (70+ years of age) feature mixed-sex married couples. In each interview, both interviewees are of a similar age (the largest age gap being 10 years).

The map displayed in Figure 4.1 below shows the hometowns of the participants in the present study. As detailed in section 4.2, this study makes use of speech data from speakers from the Tyneside, North Tyneside, South Tyneside, and the (South-East) Northumberland regions.

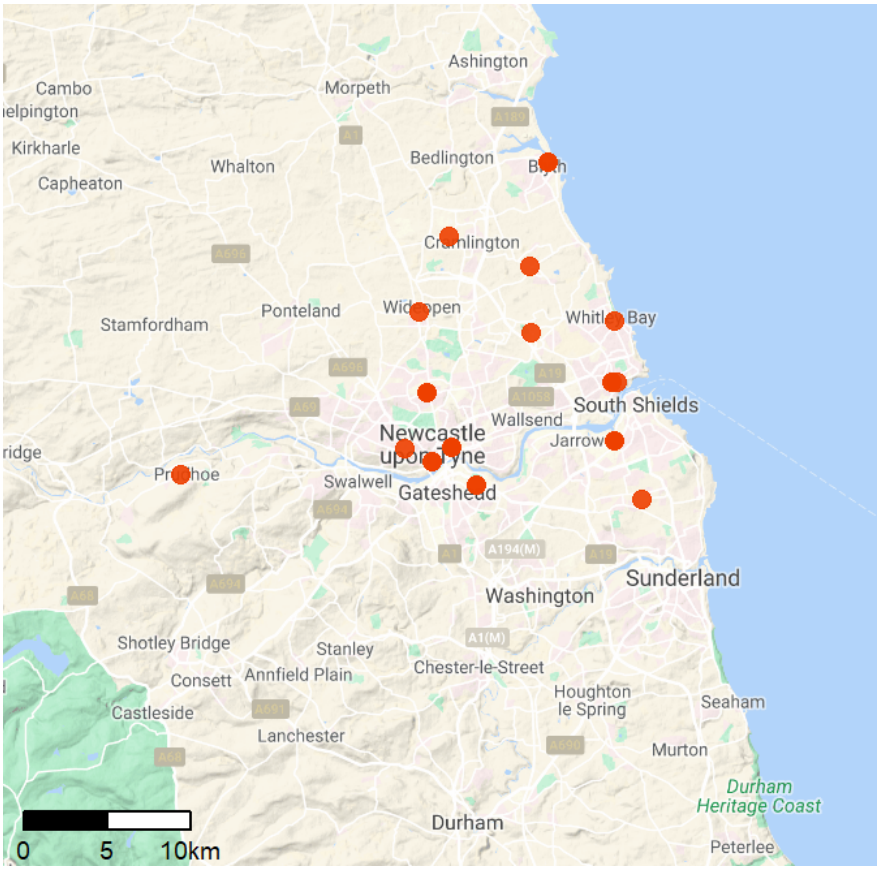


Figure 4.1: Map of Speaker Locations

4.4.1 Breakdown of Speaker Sample

Table 4.1 presents the amount of speakers in each speaker group, according to sex, age and social class.

	LWC		UWC		LMC		UMC	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Young		I		I	IIII		II	II
Middle	I	III	II	II	II	I		I
Older	I	I			I	III		

Table 4.1: Breakdown of speakers by social group

Firstly, the table shows that there is a range of age groups and social classes represented within the sample. However, given the low numbers in some cells and the presence of some empty cells in the table, the sample is not well represented across all three social variables of sex, age group, and gender. This lack of representation across social variables somewhat mirrors the nature of the DECTE corpus. There is a much higher percentage of young speakers to old speakers in DECTE, which is reflected in the sample. Similarly, Table 4.1 shows that the most well represented socioeconomic group is lower-middle class. As outlined above in Section 4.3.3.2, there are a number of complications of assessing social class in DECTE. Additionally, the corpus is not structured with the aim of obtaining speech data from speakers of different social backgrounds, and since the participant selection process is carried out by students of Newcastle University, it is likely that speakers from the lowest and highest groups on the socioeconomic scale are underrepresented in the corpus. This means that any findings relating to social class in the following analysis of speech production may provide tentative evidence of socioeconomic patterns in realisations of the GOAT and THOUGHT vowels, however these results cannot be taken to be conclusive. Given the speaker sample, it will not be possible to look at social class alongside speaker sex and age in the analysis, meaning that, speakers cannot be categorised into groups based on all three social factors, e.g. young lower-middle class females, older upper-working class males. Rather, in the following analysis, speakers are grouped based on their sex and age, and social class is discussed separately.

Table 4.2 gives a more detailed overview of speakers sample of the current study[3]. Along with information relating to the social variables which will be included in the following analysis, the table details the broad regionality or each speaker.

[3]Participants in DECTE either provide their own pseudonym, or one is given to them by the interviewer

Speaker	Sex	Age	Region	Residence	Social Class
April	F	19	Northumberland	Seaton Sluice	UMC
Chloe	F	19	Tyneside	Newcastle	LWC
Kate	F	19	Northumberland	Blyth	UWC
Rebecca	F	19	Tyneside	Newcastle	UMC
Samantha	F	37	Tyneside	Gateshead	LWC
Mary	F	39	Tyneside	Gateshead	LWC
Cheryl	F	46	Tyneside	Newcastle	LMC
Joanna	F	47	Northumberland	Prudhoe	UWC
Nicola	F	48	Northumberland	Seaton Delaval	UWC
Claire	F	52	North Tyneside	Backworth	LWC
Tracy	F	54	North Tyneside	Shiremoor	UMC
Andrea	F	63	South Tyneside	South Shields	LMC
Sarah	F	69	South Tyneside	South Shields	LMC
Beatrice	F	71	South Tyneside	South Shields	LWC
Pam	F	73	Tyneside	Newcastle	LMC
Josh	M	19	Tyneside	Newcastle	LMC
Tom	M	19	Tyneside	Newcastle	LMC
Evan	M	21	South Tyneside	East Boldon	LMC
Rupert	M	21	Northumberland	Cramlington	LMC
Shaun	M	21	Tyneside	Newcastle	UMC
John	M	22	Northumberland	Blyth	LMC
Monty	M	34	Tyneside	Newcastle	LMC
Raymond	M	34	Tyneside	Newcastle	UWC
Phil	M	43	Tyneside	Newcastle	LWC
Daniel	M	48	Tyneside	Newcastle	LMC
Stephen	M	53	North Tyneside	North Shields	UWC
Wallace	M	70	South Tyneside	South Shields	LWC
Ian	M	71	Tyneside	Newcastle	LMC

Table 4.2: Biographical information of selected speakers from DECTE

4.5 Procedure

4.5.1 Interviews

The interviews selected for this study, obtained from the NECTE2 component of DECTE, were conducted and recorded by students studying English Language or Linguistics courses at Newcastle University. Students recruit their own participants for the dyadic interview, usually two friends or relatives, and collect biographical details from each interviewee. The interviews typically take the form of semi-structured, conversational interviews, and last for approximately 60-90 minutes. Topics and themes vary from interview to interview, with popular interview topics including childhood and teenage memories, holidays and hobbies, and thoughts on the North-East region and accent. The majority of the selected interviews took place in the home of the interviewees, which helps to create a relaxed, informal environment in which to collect naturalistic speech data.

Sound quality was a factor in selecting which interviews to use in this study. The majority of recent interviews are recorded in a quiet room with high quality audio equipment. However, some interviews could not be used in this study due to background noise e.g. household appliances, voices from non-interviewees, or issues with the recording e.g. participants sound too quiet, noisy interference throughout recording. Additionally, while the majority of recent DECTE interviews are in the WAV format, interviews saved in the MP3 format were excluded due to the loss of information incurred following compression.

4.5.2 Transcription

In addition to interview audio, the NECTE2 component of DECTE contains transcription files. Initially, the student interviewers orthographically transcribe a 30 minute segment of their interview. For the present study, these student transcribed files were checked for errors before alignment took place. The interview audio and text transcription were time-aligned into breath groups using ELAN (2018, v5.2). This outputs a transcription file which gives the specific timestamp of each speech segment in the audio file.

4.5.3 Forced Alignment and Automatic Extraction

The *Forced Alignment and Vowel Extraction* (FAVE) program suite (Rosenfelder *et al.* 2011) is a tool which allows for automatic alignment and vowel formant extraction. Taking both a sound file and a transcription file as input, alignment

using FAVE converts a time-aligned orthographic transcription to a time-aligned phonemic transcription. The output of FAVE alignment is a Praat TextGrid file, which can then be used with the sound file, again via FAVE, for the automatic extraction of vowel formant values.

The present study uses FAVE to obtain both phonemic transcriptions of the DECTE interviews and vowel formant measurements for each speaker. Recent sociophonetic research often makes use of FAVE for automatic alignment and/or extraction of formant measurements (Hall-Lew *et al.* 2015; Sneller 2015; Strelluf 2016; Thiel and Dinkin. 2017). There are obvious advantages to automatic alignment and extraction which make FAVE an attractive tool for conducting sociophonetic analysis, namely, (1) it is quicker than phonemically transcribing and measuring formants by hand, thereby, (2) it allows for more data to be analysed e.g. more speakers, more tokens, more variables. The majority of research utilising FAVE focuses on varieties of American English. FAVE's acoustic models were trained on a corpus of American English speech data, and alignment relies upon an American pronouncing dictionary to convert orthographic to phonemic transcriptions. Despite this, MacKenzie and Turton (2020) find that FAVE performs well in accurately aligning speech data from several different varieties of British English. Thus, FAVE is a viable tool for use on British Englishes, and data from Tyneside speakers should not pose a problem for FAVE's automatic alignment and formant extraction.

4.5.3.1 Alignment

To allow for successful forced alignment of the Tyneside interviews, modifications were made to the FAVE pronouncing dictionary. Dialect words, place names, and words with British English spelling variants were phonemically transcribed and added to the dictionary. Given that FAVE uses an American English pronouncing dictionary, further checks and edits were necessary following alignment e.g. segment boundaries recoded to reflect lack of coda /r/ in Tyneside English. As this thesis is primarily concerned with GOAT and THOUGHT, the TextGrid files were double-checked to make sure words containing these vowels were forced-aligned correctly. As GenAm varieties typically lack a distinction between the CLOTH-THOUGHT lexical sets, FAVE automatically labels them as the same vowel. CLOTH words such as *off, across* were recoded by hand so as to be separated from THOUGHT words and excluded from further analysis.

4.5.3.2 Extraction

FAVE automatically extracts formant values for each speaker's vowels. Before extraction, biographical details for each speaker were input into FAVE. It is necessary to tell FAVE the sex of the speaker to ensure accurate formant measurements; the maximum formant frequency settings differ depending on whether the speaker is male (5000 Hz) or female (5500 Hz). Most vowels, including THOUGHT, are measured at 1/3 of the duration of the vowel, while GOAT is measured halfway between the vowel onset and the point of maximum F1. The FAVE extract output includes the raw formant values and Lobanov normalized formant values, the latter of which will be used in the following analysis to allow for reliable inter-speaker comparisons. In addition to formant values (F1, F2 and F3), the FAVE extract output gives measurements for vowel duration, as well as providing information on the phonological environment of the vowel e.g. word position, preceding and following segments.

In total, over 4,600 vowel tokens (2,704 tokens of GOAT, 1,954 tokens of THOUGHT) were measured via FAVE. The mean number of GOAT tokens per speaker is 85, and the mean number of THOUGHT tokens per speaker is 61. Tables 4.3 and 4.4 below show the speakers with the highest and lowest number of GOAT-THOUGHT tokens. While Milroy and Gordon (2003: 164) claim that 30 tokens per variable per speaker is the ideal number to aim for, they add that any amount over 10 tokens is a 'sensible goal'. The majority of speakers in the current data exceed this 30 tokens per variable goal, and all exhibit at least 15 tokens per vowel. Therefore, although Table 4.4 shows that two speakers (Kate and Sarah) used few THOUGHT tokens in their interviews, these numbers should still be sufficient to allow for a comparison between their GOAT and THOUGHT vowels.

Speaker	GOAT	THOUGHT	**Total**
Jack	181	155	336
Samantha	157	77	234
Patricia	128	101	229
Cheryl	122	106	228
Evan	98	109	207

Table 4.3: Speakers with the highest usage of GOAT and THOUGHT tokens

Speaker	GOAT	THOUGHT	**Total**
Kate	37	15	52
Sarah	49	19	68
Shaun	45	30	75
Mary	58	28	86
Beatrice	71	123	94

Table 4.4: Speakers with the lowest usage of GOAT and THOUGHT tokens

4.6 Methods and Measures

The following examination of the potential merging of the GOAT and THOUGHT vowels in Tyneside speech uses a combination of several different methods of analysis, including Pillai scores, auditory analysis, and analysis using vowel trajectory data.

4.6.1 Pillai Scores

According to Nycz and Hall-Lew (2013), an ideal method for examining vowel mergers would allow for measures of both the distance and overlap between two or more word classes while also allowing for inter-speaker comparisons. First used by Hay *et al.* (2006) in their study of the NEAR-SQUARE merger in New Zealand, Pillai scores, which are an output of Multivariate Analysis of Variance (MANOVA) models, are one possible method of measuring vowel merger. Pillai scores, otherwise known as Pillai-Bartlett trace statistics, give a measure of *difference* between the F1-F2 distributions of different word classes by taking into account the distance and overlap. As Pillai scores are generated by MANOVA models, Nycz and Hall-Lew (2013: 13) argue that this measure is appropriate for inter-speaker comparisons because phonological and lexical predictors are easily incorporated into the statistical model, therefore reducing potential imbalances between the data for each speaker. Recent research often uses Pillai scores, either alone or, typically, in combination with other methods and measures, to examine vowel mergers, distinctions, and splits (Hall-Lew 2010; Wong 2012; Ebabel *et al.* 2013; Sloos 2013; Kendall and Fridland 2017).

To calculate a Pillai score for each of the 28 speakers in the present study, MANOVA models were run on the data. These statistical models included the Lobanov-normalised F1 and F2 values of each vowel token as the dependent variables, and also included predictors for vowel type (i.e. GOAT or THOUGHT) and phonological environment (preceding and following segments). Previous research on Tyneside English has provided little information on how phonological environment may affect GOAT-THOUGHT merging. In the absence of any prior known effects of phonological environment, this predictor was therefore defined in the model simply as the segments immediately preceding and following the vowel. Pillai scores range from 0 to 1: a score of 0 indicates *no difference* between the distributions of two word classes, and a score of 1 indicates *no similarity* between word classes. Therefore, speakers with a higher Pillai score can be said to have a greater vowel distinction, while speakers exhibiting a lower Pillai score show more evidence of vowel merging.

4.6.2 Auditory Analysis

As discussed previously, GOAT has been found to be highly variable in Tyneside English (Watt 1998; Pearce 2009; Beal *et al.* 2012). Most research on Tyneside English in the past few decades has found the monophthongal [oː] realisation of GOAT to be the most frequently used in the region, supplanting more local pronunciations such as the centering diphthong [ʊə] and the non-peripheral monophthong [ɵː] (Watt 2002; Beal *et al.* 2012). Watt (2002) also notes that a closing diphthongal [oʊ] realisation, used in a number of varieties across Britain, may be found in the speech of middle-class Tynesiders. If multiple GOAT variants are still used on Tyneside today, this poses some complications in assessing the evidence of a possible GOAT-THOUGHT merger. First, a theoretical question arises of whether a merger can take place if multiple variants of a potentially 'merging' vowel still exist in the speech community. Second, a more methodological complication emerges, as, if it is possible that both monophthongal and diphthongal variants of GOAT are still being used in contemporary Tyneside English, any analysis which takes one formant measurement per vowel token (such as with Pillai scores) may not be sufficient. Analysis using dynamic rather than static data may therefore also be necessary in order to more accurately measure the GOAT vowel, and its relationship to THOUGHT, in Tyneside English. This is discussed further in Section 5.3 below. In order to investigate the current status of GOAT variation in Tyneside, an auditory analysis of the DECTE interview data was conducted.

For an auditory analysis, each speaker's GOAT vowels were analysed by ear and coded using the Praat Handcoder (Fruehwald 2011). Initially, each token of GOAT was categorised as being one of the four variants, displayed in the table below, previously found in the Tyneside region. Given that previous research has found locally marked variants like [ʊə] and [ɵː] to be largely restricted to the speech of older working-class males (Watt 1998; Corrigan *et al.* 2014), such pronunciations are expected to be infrequent – or even absent – in this data from contemporary Tyneside speech.

Vowel	Description
oː	peripheral monophthong
ʊə	centering diphthong
oʊ	closing diphthong
ɵː	fronted/centralised monophthong

Table 4.5: Previously recorded variants of the GOAT vowel in Tyneside English

A number of GOAT tokens were excluded from the handcoding process. Firstly, tokens in which the vowel was difficult to hear due to e.g. background noise, laughter, overlapping speech were rejected during coding. Additionally, examples of unstressed GOAT vowels (e.g. often found in words such as *'so'*, *'don't'* in rapid, conversational speech) were excluded from analysis, as were words in which word-final GOAT had been reduced to schwa (e.g. *'follow'*, *'tomorrow'*). Finally, tokens of the word *'know'* pronounced with an [a:] were omitted from analysis, as this is thought to be a lexically restricted pronunciation rather than a variant of GOAT in the Tyneside phonological system (Crinson and Williamson 2004: 214). In total, 2,544 GOAT tokens were auditorily coded.

4.6.3 Dynamic Vowel Analysis

Several methods are available for examining cases of potential vowel merger. Possible measures include Pillai scores (as discussed in Section 4.6.1), the Spectral Overlap Assessment Metric (SOAM), and mixed-effects regression adjusted Euclidean distances (Hay *et al.* 2006; Wassink 2006; Nycz 2011; Nycz and Hall-Lew 2013). These methods take in static data as input e.g. one measure for F1 and one measure for F2 per vowel token. However, an increasing number of researchers are taking a more dynamic approach to investigating vowel variation, taking two or more formant measurements per vowel token (Irons 2007; Strelluf 2016).

Di Paolo *et al.* (2011) recommend taking measurements at multiple points for all vowels, as even monophthongs show formant movement across their duration. The number of measurement points to take, however, is of some debate. Labov *et al.* (2006: 38) show concern over the use of too many measurements which could 'obscure' any patterns in the data, while additionally making comparisons between speakers and quantitative analysis difficult to carry out. On the other hand, Wieling *et al.* (2018: 86) propose that oversimplification of dynamic data may lead to patterns in the data being left 'undiscovered'. Baranowski (2013: 7) argues that two measurements, one at the vowel nucleus and one at the vowel glide, are typically sufficient for identifying patterns of social and regional variation. However, Baranowski (2013: 7) also highlights that in certain situations, namely in cases of vowel merger, a more dynamic analysis using formant trajectories may be preferable. Formant trajectories allow for an examination of 'the total formant movement' within a vowel (Thomas 2011: 152). Measurement points for vowel trajectories are typically obtained either by taking measurements at specified time-points during the vowel duration (e.g. every 20 milliseconds), or by taking measurements at percentages of the vowel duration (e.g. 25%, 50%, 75%). The

advantage of the latter is that all vowels, regardless of duration, will have the same number of measurement points, which allows for an easy comparison across tokens of all vowels (Di Paolo *et al.* 2011; Thomas 2011b).

As mentioned above in Section 4.6.2, a more dynamic analysis of the GOAT and THOUGHT vowels may be necessary in the current study given the possibility of monophthongal and diphthongal variants being present in the data. Therefore, multiple formant measurements taken at different points throughout the vowel trajectories were carried out. The output of FAVE's (Rosenfelder *et al.* 2011) automatic vowel extraction gives measurements for every token at 20%, 35%, 50%, 65% and 80% of the vowel's duration. While this would be sufficient for a dynamic analysis, measurements taken at shorter, more regular intervals will enable a more detailed analysis of vowel trajectories. As such, the FAVE extract script was altered so as to collect formant measurements at every 10% of the vowel's duration. This gives 11 F1 and F2 measurements per vowel token rather than the 5 measurements available when using the unaltered FAVE script. Again, the Lobanov normalised formant values automatically output by FAVE are used for the following analysis.

4.7 Quantitative Analysis

4.7.1 Mixed-Effects Regression Models

Mixed-effects regression models are used to assess the effects of social and linguistic variables on the realisation of the GOAT and THOUGHT vowels. These statistical models were run using the `lmer()` function of the lme4 R package (Bates *et al.* 2015). The output of mixed-effects regression models run with the lme4 package do not provide p-values, instead presenting t-values. For a large dataset (i.e. containing hundreds/thousands of observations) such as that used in the current study, Baayen (2008: 398) states that significance can be 'gauged informally' by looking at the t-values: a predictor can be estimated to be significant if its t-value is greater than +/-2. Thus, in the following analysis, a t-value of higher than 2 or -2 is interpreted as statistically significant.

4.7.1.1 Model Predictors

Models were initially maximally fit to include the social variables of interest (i.e. sex, age group, social class) alongside the following linguistic predictors: the place of the preceding segment, the place, manner, and voicing of the following segment,

context (e.g. word initial, medial, or final), and word frequency. Information regarding the articulation of the preceding and following segments was taken from the FAVE-extract output[4]. The word frequency values included in the model come from SUBTLEX (Van Heuven *et al.* 2014), a corpus containing over 200 million words used in subtitles for BBC film and television broadcasts. SUBTLEX uses log transformed data to categorise words on a 7-point frequency scale, with scores between 1-3 labelled as low frequency, while a score of between 4-7 signals a high frequency word. The advantages of using SUBTLEX over other word frequency databases are (1) the large size of the corpus, (2) the contemporariness of the data in the corpus (collected between 2010-2012), and (3) that the word frequency values are already log transformed and easily interpretable. Many recent studies have used SUBTLEX to explore the effects of word frequency on phonetic variation and change, including Fruehwald (2016), Reubold and Harrington (2017), Tamminga (2018), and Schleef and Turton (2018).

The mixed-effects regression models used in this study also include the random effects of speaker and word. Following Baayen (2008: 390), the inclusion of speaker as a random effect allows for the model to incorporate individual differences between the speakers. Other than the social variables of sex, age group, and social class that are included as fixed effects in the regression models, any other possible factors differentiating the individual speakers are deemed unpredictable, therefore the variable of speaker must be included as a random effect. Similarly, as the word choice of speakers in their sociolinguistic interviews is a random sample of all possible words, word is an unpredictable factor, and therefore must be included in the model as a random effect. If these random effects are not built into the regression models, it could lead to inaccurate modelling of the fixed effects.

To select the optimal model to use for analysis, ANOVA model comparisons were performed in R using the `anova()` function. Maximal models which included all of the social and linguistic variables mentioned above were compared to smaller nested models which removed non-significant predictors. In order to ensure that the optimal model is parsimonious (e.g. best explains the data with as few predictors as possible), if an ANOVA comparison does not find the the maximal model to be significantly better at explaining the variance in the data (i.e. p-value >0.05), then the smaller, reduced model is used for analysis.

[4]The FAVE-extract descriptions of the preceding segment's place of articulation were collapsed into two categories: those which have a fronting effect on the vowel (e.g. coronal consonants), and those which do not (e.g. bilabial, velar and glottal consonants).

4.7.2 Multinomial Logistic Regression Models

For a statistical analysis of the auditory coding of the different GOAT variants used by the DECTE speakers, multinomial logistic regression was performed on the data in R using the `multinom()` function from the 'nnet' package (Venables and Ripley 2002). As the dependent variable in the model is the pronunciation of GOAT — which contains four different variants — an analysis using multinomial regression was necessary due to its ability to take input from data in which the dependent variable contains two or more categories. The multinomial regression models output p-values to indicate whether a predictor has a significant effect on the usage of a particular GOAT variant (significance is taken at the <0.05 level).

4.7.2.1 Model Predictors

The predictors included in these models are the same social and linguistic variables as those outlined in Section 4.7.1.1: sex, age group, social class, place of preceding segment, place, manner, and voicing of the following segment, and word frequency. However, unlike in mixed-effects regression, multinomial regression models do not include random effects.

Model selection was conducted by comparing the Akaike Information Criterion (AIC) values of different models. When selecting the ideal statistical model, a lower AIC value indicates the model which best fits the data. A similar model selection process to that detailed above in Section 4.7.1.1 was carried out, with a preference for parsimonious models. Thus, the model with both the smallest amount of predictors and the lowest AIC value was chosen as the model which best fits the data.

4.7.3 Generalised Additive Mixed Models (GAMMs)

Generalised Additive Mixed Models (GAMMs) can be used for analysis of dynamic data in linguistic research. Weiling (2018: 87) notes that linguistic analysis of dynamic data was 'computationally prohibitive' until the last couple of years. Recently, however, researchers in linguistics have been using GAMMs to perform analyses on vastly different types of dynamic data. Baayen *et al.* (2018) illustrate that GAMMs can be used on a multitude of dynamic data types: pitch contours of speaker productions of English three-word compounds, time-series data from a lexical decision type task, and data recording the amplitude of participants' electrophysiological responses to ungrammatical English compounds.

The use of Generalised Additive Mixed Models is becoming more common in phonetic and sociophonetic research. Coretta (2017) uses GAMMs on tongue

contour data to compare how voiced and voiceless segments affect tongue root advancement in Polish and Italian speakers. Another example of recent GAMM usage is in Wieling's (2018) study of differences between L1 and L2 speaker productions of /θ/, using articulatory trajectory data to compare Dutch speakers of English with speakers of Standard Southern British English (SSBE). While GAMMs have been used in more articulatory phonetics studies, Sóskuthy (2017) notes that these models are also useful for analysing formant trajectories. In Sóskuthy's (2017) tutorial on how to use GAMMs for a linguistic analysis of dynamic data, he uses data – initially presented in Stuart-Smith *et al.* (2015) – of F3 trajectories of word-final /r/ + preceding vowel tokens. Here, a GAMM is used to investigate diachronic change in the /r/ pronunciations of Glasweigan speakers by comparing the formant trajectories of speakers born in the 1970s, 1980s, 1990s and 2000s. More recently, Kirkham *et al.* (2019) have conducted an analysis using GAMMs to investigate dialectal variation, using formant trajectory data, in the pronunciations of /l/ + vowel(s) sequences among speakers from Manchester and Liverpool.

Elsewhere, researchers are beginning to use GAMMs to assess overlaps and distinctions between sounds across their trajectories. Renwick (2017) uses GAMMs to compare the F1-F2 trajectories of tense vs. lax vowels in Southern American English. Looking at the formant trajectories of the KIT and FLEECE vowels, Renwick finds that these two vowels are becoming less distinct, with younger speakers raising KIT while lowering FLEECE. More recently, Renwick and Stanley (2020) have also investigated overlap in the trajectories of the FACE and DRESS vowels in the American South, using GAMMs to illustrate that young female European Americans in particular show overlap in the trajectories of these two vowels. Another recent example is Cole and Strycharczuk's (2019) use of GAMMs to examine changes in the trajectories of the PRICE and MOUTH vowels in Cockney speech. Traditionally, speakers of Cockney English show an overlap in the onsets of these two vowels, but the authors find that the younger speakers typically exhibited less of a crossover between the onsets of PRICE and MOUTH than the older speakers, suggesting that the vowel system for these young speakers is changing in the direction of modern RP.

GAMMs are a suitable method of statistically analysing cases of vowel merger. In using dynamic data, the formant trajectories of tokens from different word classes can be compared in terms of trajectory shape and slope. This type of analysis is not available when using data which measures formant values at only one or two time-points. GAMMs provide a method of statistical comparison of vowel trajectories,

showing the overlap and distance between vowels across their duration. As Nycz and Hall-Lew (2013) note that measures of overlap and distance are favourable in any method of measuring of merger, GAMMs appear to be a good candidate for merger assessment. GAMMs have further advantages for research into vowel mergers, such as the ability to test for significant differences between two word classes. Statistical significance testing is possible for static data with methods such as Pillai scores and adjusted Euclidean distances, but GAMMs provide a way of performing similar statistical analysis on dynamic speech data. Additionally, GAMMs are able to predict the formant trajectories of vowels according to predictors input into the model e.g. trajectory predictions for certain words, speakers, or speaker groups.

For a dynamic analysis of a potential Tyneside GOAT-THOUGHT merger, GAMMs were run on the data output from the altered FAVE script which gave 11 measurements per vowel token. The GAMMs were run in R using the `bam()` function from the 'mgcv' package (Wood 2017), and the model predictions were plotted using the `plot_smooth()` function from the 'itsadug' (Rij *et al.* 2017) package.

4.8 Summary

This chapter has detailed the methodology of the current study's investigation into the production of the GOAT and THOUGHT vowels in the Tyneside speech community. The data collection process was outlined, highlighting the advantages and disadvantages of using the DECTE corpus. Additionally, the range of methods and quantitative approaches used in an analysis of the speech production data has been explained and evaluated. The following chapter will present the results of the production analysis, with the aim of assessing whether there is evidence that a GOAT-THOUGHT merger-in-progress is taking place in present day Tyneside English.

Chapter 5

Production: Analysis

This chapter aims to investigate whether a potential GOAT-THOUGHT merger is present in contemporary Tyneside English. The results of a range of analysis methods and quantitative measures, as outlined in the previous chapter, are presented here. First, findings from analyses using static data (one-measurement) are discussed, as mixed-effects regression models give evidence of the positions of GOAT and THOUGHT in the vowel space of these Tyneside speakers, while Pillai scores are used to assess the overlap between these vowels. Then, results from an auditory analysis of GOAT variation are presented, with the aim of providing evidence of whether multiple variants of the GOAT vowel are still found in present-day Tyneside speech. Findings from analyses using dynamic data (vowel trajectories) are then discussed, as results from Generalised Additive Mixed-Models (GAMMs) are used to assess the similarity of the vowel trajectories of GOAT and THOUGHT. Finally, an analysis of duration is carried out in order to examine whether vowel length differences between these two vowels may prevent a GOAT-THOUGHT merger on Tyneside.

5.1 Acoustic Analysis of Static Data

5.1.1 GOAT-THOUGHT Overlap

In order to get an overview of the position of GOAT and THOUGHT in the vowel space, the formant values (F1-F2) of each speakers' GOAT and THOUGHT tokens were plotted in R. These plots are based on the one-point FAVE-extract measures, which were taken at 1/3 of the vowel duration for THOUGHT, and midway between vowel onset and the point of maximum F1 for GOAT. Figures 5.1 (female speakers) and 5.2 (male speakers) below show the position and overlap of GOAT and THOUGHT

for each individual speaker. Each point in these plots represents a token of a GOAT or THOUGHT vowel, while the ellipses represent the 95% confidence interval of where each word class is positioned in the vowel space.

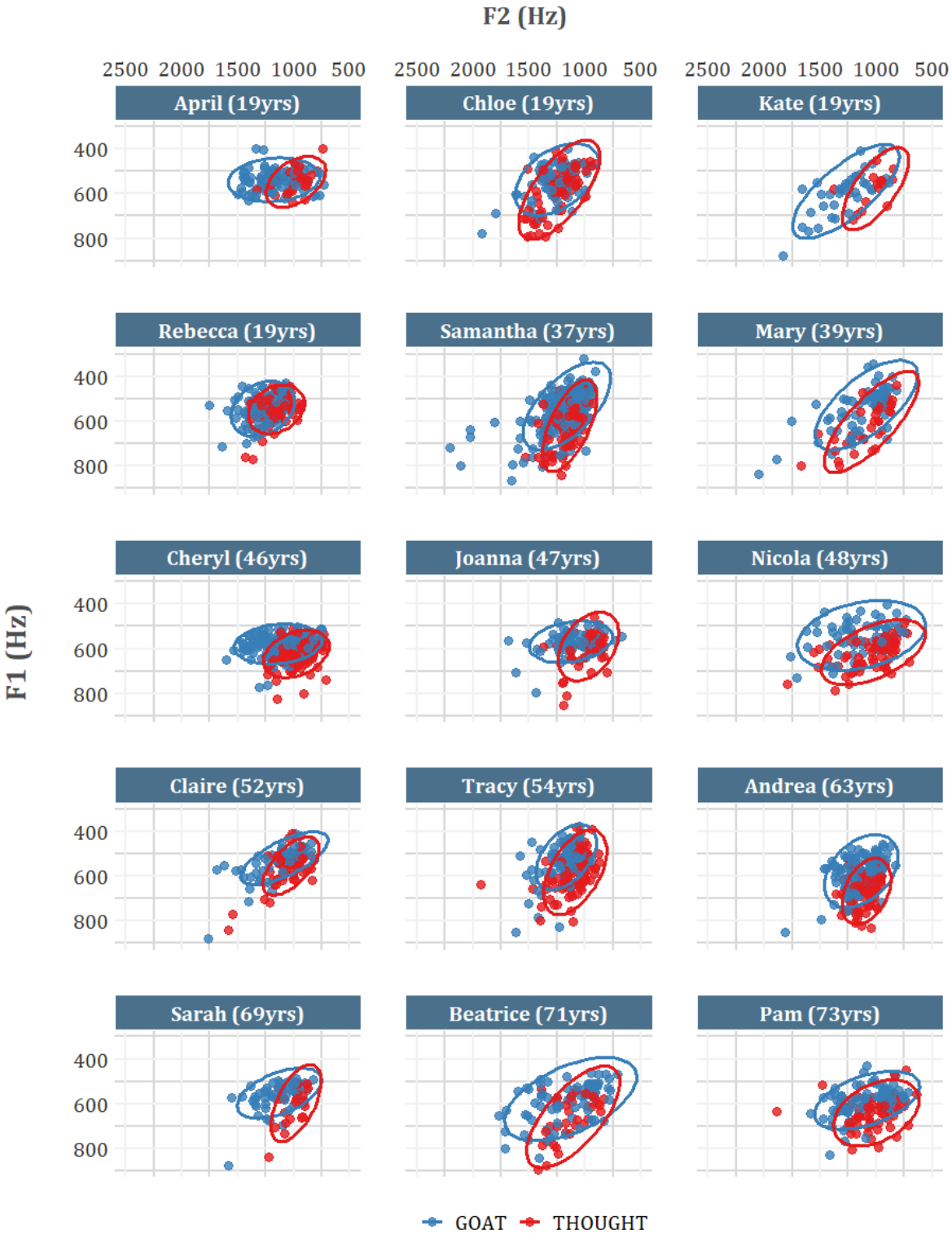


Figure 5.1: F1-F2 values of the GOAT and THOUGHT vowels of individual female speakers

Figure 5.2: F1-F2 values of the GOAT and THOUGHT vowels of individual male speakers

The plots displayed in Figures 5.1 and 5.2 give some indication of the position of GOAT and THOUGHT in Tyneside speech. As THOUGHT is generally transcribed as an open-mid back [ɔː] vowel and contemporary Tyneside GOAT vowels are typically

realised as a close-mid back [oː] monophthong, the expected difference between these vowels would be in the F1 measure, with GOAT in a higher position than THOUGHT at the back of the vowel space. Yet, for many of the speaker plots shown in Figures 5.1 and 5.2, the GOAT and THOUGHT vowel ellipses have similar F1 values, indicating little to no height difference between the two vowels. Although both GOAT and THOUGHT are traditionally classed as back vowels, Figures 5.1 and 5.2 show that the majority of speakers exhibit a GOAT vowel which is more fronted than THOUGHT. Additionally, for a number of speakers, the GOAT vowel ellipses are larger than the THOUGHT ellipses, with several stretching from back to mid position on the F2 plane (e.g. April, Patricia, Ian, Phil). Such results could highlight that the GOAT vowel is still highly variable in Tyneside English.

In terms of vowel overlap, Figures 5.1 and 5.2 show that many speakers – particularly the female speakers – exhibit a considerable overlap of the GOAT and THOUGHT vowels (e.g. Beatrice, Claire, April, Rebecca). The close proximity of these vowels in phonetic space may suggest preliminary evidence of a GOAT-THOUGHT merger-in-progress on Tyneside. Also of note is that several speakers appear to have a THOUGHT vowel which is located almost entirely within the boundaries of the GOAT vowel ellipsis (e.g. April, Rebecca, Daniel, Wallace). This indicates that, for these speakers, THOUGHT is very similar in quality to the most backed realisations of GOAT. However, given the large amounts of variation which can be seen in Figures 5.1 and 5.2, it is difficult to get a clear picture of vowel mergedness from these plots alone.

Some differences between different speaker groups are apparent when looking at Figures 5.1 and 5.2. Most of the plots showing the biggest overlap of GOAT and THOUGHT are those of the young female speakers (notably April, Chloe, and Rebecca). As a group, the female speakers appear to exhibit more of an overlap of the GOAT and THOUGHT vowels than the males. A few of the young male speakers (namely Shaun, Evan, and to a lesser extent, Josh) display a notable lack of GOAT-THOUGHT mergedness. For these individuals, the GOAT vowel is much fronter than THOUGHT. Based on these initial findings, these results present a complex picture of the GOAT-THOUGHT vowels in Tyneside English, whereby young females appear to exhibit the largest overlap between these vowels, while young males appear to show a larger GOAT-THOUGHT distinction than both females and older male speakers.

For the majority of speakers, Figures 5.1 and 5.2 display the boundaries of the THOUGHT vowel as lying largely between 500-700 Hz in the F1 space. As previous literature generally describes the Tyneisde THOUGHT vowel as an open-mid vowel, and previous acoustic analysis has found that the average height of THOUGHT in

Tyneside is just below the 500 Hz mark (Ferragne and Pellegrino 2010), the results presented in the formant plots above do not suggest that any raising of the THOUGHT vowel has taken place. It is more difficult to assess any changes to the position of the GOAT vowel on Tyneside, due to factors such as (a) the number of GOAT vowel variants used in the region, and (b) the relative lack of acoustic analyses of monophthongal GOAT vowels in Tyneside speech. However, as some descriptions of Tyneside English show monophthongal GOAT vowels as occupying a higher position than THOUGHT in the vowel space (Watt and Allen 2003: 268), and as analyses of other British varieties which use monophthongal GOAT realisations find that GOAT is generally higher than THOUGHT (Watt and Tillotson 2001), it would be expected for GOAT tokens to show higher F1 values than THOUGHT for the Tynesiders in the current study. Therefore, due to the lack of a height distinction between GOAT and THOUGHT in Figures 5.1 and 5.2 – coupled with the fact that THOUGHT does not seem to have raised from its expected position – these results would point to a lowering of the GOAT vowel as the instigator for the overlap between these vowels in phonetic space.

5.1.2 GOAT and THOUGHT: Mixed-effects Regression

To look for differences in the positions of GOAT and THOUGHT for different speaker groups, mixed-effects regression models were run on the data using the F1-F2 measurement points of each vowel. Details of the predictors included in the model, along with the model selection process, are detailed in Section 4.7.1.1. The output from the selected models is shown in Tables 5.1 and 5.2 below.

	F1			**F2**		
	Estimate	Std. Error	t-value	Estimate	Std. Error	t-value
(intercept)	554.29	19.36	28.62	1281.53	49.17	26.06
sex: male	31.86	13.31	2.40	118.91	33.24	3.58
age group: mid	-27.84	15.74	-1.77	-106.49	39.25	-2.71
age group: old	0.30	19.80	0.02	-124.50	49.43	-2.52
class: UWC	13.85	19.51	0.71	14.25	46.31	0.31
class: LMC	19.34	17.25	1.12	16.21	43.06	0.38
class: UMC	0.11	21.13	0.01	18.93	47.88	0.40
context: initial	-	-	-	-68.63	36.64	-1.87
context: medial	-	-	-	-85.46	22.36	-3.82

Table 5.1: Mixed-Effects regression results for F1-F2 values of GOAT

Table 5.1 shows the results of the mixed-effects regression model which took F1-F2 measurements of GOAT as the dependent variables. The first result of note is that the model indicates that male speakers differ significantly from the female speakers with respect to both the height and frontness of the GOAT vowel. The model output suggests that the males typically demonstrate lower and fronter GOAT vowels than the females. In terms of the effects of speaker age group on the realisation of GOAT, the model finds that the middle-aged and older speakers typically have lower F2 values, and therefore more backed pronunciations of GOAT than the younger speakers. Compared with the younger speakers, the model finds the difference in GOAT frontness to be significant for both the middle and older speaker groups. Finally, no significant effect of social class on the height or the frontness of the GOAT vowel was found in the model.

An ANOVA model comparison found that including a predictor for context significantly improved the model looking at the frontness of GOAT ($p = <0.001$). As displayed in Table 5.1, the model output indicates that, among the DECTE speakers, GOAT vowels are significantly more backed in word medial position when compared to word final position. The results also show that GOAT is considerably more backed in initial than in final position, although this does not quite reach statistical significance. This therefore suggests that context has a significant effect on the frontness of GOAT in Tyneside, and that GOAT vowels are subject to greater fronting when they are not immediately followed by another segment within the same word.

	F1			**F2**		
	Estimate	Std. Error	t-value	Estimate	Std. Error	t-value
(intercept)	586.16	12.89	45.46	1110.60	33.82	32.84
sex: male	15.99	8.54	1.87	19.40	23.03	0.84
age group: mid	15.58	10.12	1.54	-32.60	27.23	-1.20
age group: old	36.72	12.67	2.90	-15.60	34.15	-0.46
class: UWC	-4.27	12.53	-0.34	-38.24	33.70	-1.13
class: LMC	11.63	10.95	1.06	-72.50	29.64	-2.45
class: UMC	-9.38	13.40	-0.70	-46.55	36.28	-1.28

Table 5.2: Mixed-Effects regression results for F1-F2 values of THOUGHT

Table 5.2 presents the output from the mixed-effects regression model for the THOUGHT vowel. Firstly, although not statistically significant, the model indicates that male speakers have a lower THOUGHT vowel than female speakers. The model

output also appears to show an effect of age group on vowel height, with older speakers exhibiting significantly lower THOUGHT pronunciations than younger speakers, while the middle-aged speakers also show lowered THOUGHT realisations when compared with the young speakers. In terms of socioeconomic class, the model output suggests that lower-working class speakers have fronter pronunciations of THOUGHT, while the lower-middle class speakers were found to exhibit more backed realisations of the THOUGHT vowels.

Looking at the output of the regression models, both Tables 5.1 and 5.2 show similar estimated F1 values for GOAT and THOUGHT. This echoes the results of the above analysis of GOAT-THOUGHT overlap, which, as displayed in Figures 5.1 and 5.2, indicated a lack of a height distinction between the two vowels. While the findings of the overlap analysis suggested that many speakers exhibited lowered GOAT vowels, which overlapped to some extent with the boundaries of the THOUGHT vowel, this is not so clearly reflected in the regression model output. As previously discussed, Figures 5.1 and 5.2 suggested that the female speakers showed a greater GOAT-THOUGHT overlap than the male speakers. It may therefore be expected that females would show a significantly lowered GOAT vowel, signalling their lead in this merger-in-progress on Tyneside. However, as Table 5.1 shows, the model indicates that male speakers show significantly lower GOAT vowels when compared with the female speakers. Given the result that males have significantly fronter GOAT vowels than females, the model output appears to be suggesting that the male speakers, in general, are lowering GOAT while also fronting the vowel. As such, for the male speakers, it may be the case that GOAT is converging with THOUGHT in terms of height, but also diverging from THOUGHT in terms of frontness.

Another result of note, in discussing the direction of the potential phonetic merging of GOAT and THOUGHT, is the effect of age group on the height of THOUGHT. The model finds that the middle age group have lower THOUGHT vowels than younger speakers, and that the older speakers exhibit significantly lower THOUGHT vowels than the younger speakers in the sample. If it is the case that Tyneside English is seeing a merger-in-progress between the GOAT and THOUGHT vowels, then younger speakers would be expected to show increased signs of merger. Given this, the results shown in Table 5.2 may suggest that the THOUGHT vowel is subject so some raising on Tyneside. Therefore, it may be the case that both GOAT and THOUGHT are undergoing sound change, coalescing in an intermediate position between the two vowels. A clearer picture of how these vowels have come to overlap may yet emerge with continued analysis using different measures of vowel merger.

5.1.3 Pillai Scores

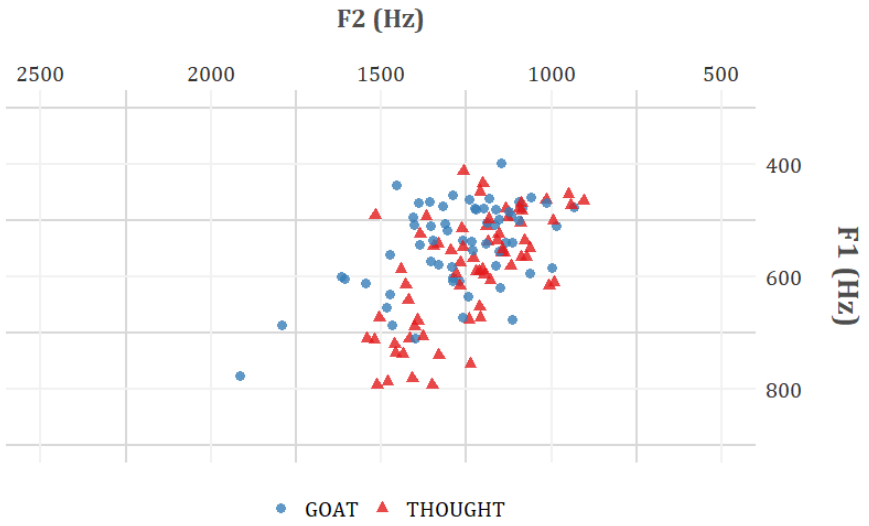
Pillai scores are a way of calculating the difference between the distributions of two word classes, and are therefore a good measure to assess vowel mergedness in speech data. The Pillai scores presented here are outputs from MANOVA models which take the F1-F2 measurements of GOAT and THOUGHT as the dependent variable in addition to a predictor of word class. Pillai scores range from 0 to 1, with scores closer to 0 indicating no difference between two vowels, while scores closer to 1 suggest no similarity between vowels. Pillai scores were calculated for each speaker separately, and these results are presented below in 5.3.

Speaker	**Sex**	**Age**	**Pillai Score**	**Speaker**	**Sex**	**Age**	**Pillai Score**
Chloe	F	19	0.164	Andrea	F	63	0.478
Beatrice	F	71	0.225	Tom	M	19	0.492
Phil	M	43	0.232	Wallace	M	70	0.496
Samantha	F	37	0.278	Tracy	F	54	0.500
Stephen	M	53	0.316	Cheryl	F	46	0.529
Claire	F	52	0.316	Mary	F	39	0.530
Rebecca	F	19	0.318	Daniel	M	48	0.556
Joanna	F	47	0.346	John	M	22	0.571
April	F	19	0.408	Monty	M	34	0.633
Pam	F	73	0.414	Kate	F	19	0.656
Ian	M	71	0.443	Josh	M	19	0.683
Nicola	F	48	0.448	Rupert	M	21	0.707
Raymond	M	34	0.455	Shaun	M	21	0.852
Sarah	F	69	0.474	Evan	M	21	0.934

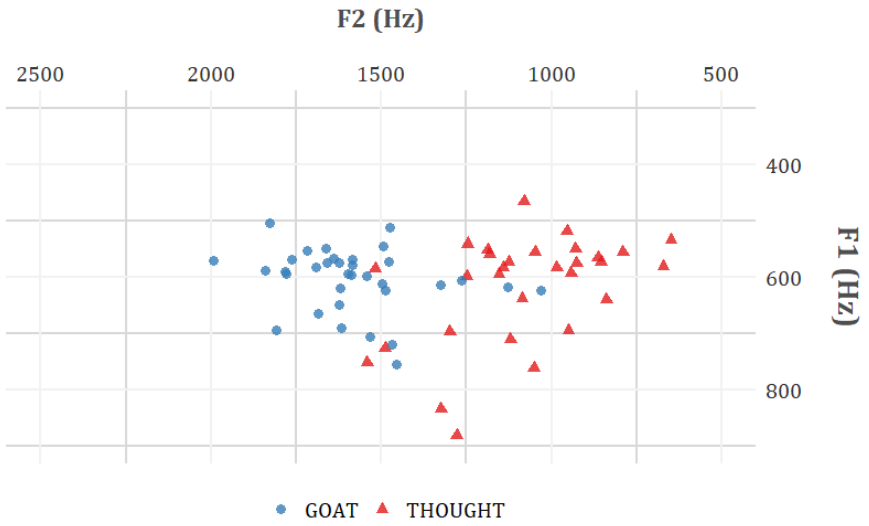
Table 5.3: Pillai scores of each speaker

Table 5.3 shows a wide range of Pillai scores, suggesting that the difference between GOAT and THOUGHT varies considerably from speaker to speaker. Given that a Pillai score of 0 indicates no difference between word classes, several speakers, notably Chloe, Beatrice, Phil, and Samantha, appear to have very little distinction between their GOAT and THOUGHT vowels. The majority of speakers have a Pillai value of somewhere between 0.4 - 0.7, indicating that there is considerable phonetic overlap between GOAT and THOUGHT in the Tyneside speech community. Table 5.3 also shows that some speakers have a clear distinction between their GOAT and THOUGHT vowels, most apparent in the speakers Evan

and Shaun. To check that the Pillai scores for the speakers in the sample are reasonable measures of mergedness, Figure 5.3 shows the formant plots for the speakers with the lowest and highest value Pillai scores. Here, it is clear that the Pillai scores reflect well the results of the formant measurements: Chloe, the speaker with the lowest Pillai score shows a large amount of overlap between the two vowel classes, while Evan, the speaker with the highest Pillai score exhibits a clear GOAT-THOUGHT distinction in the vowel space.



(a) Chloe, 19yrs (Pillai = 0.164)



(b) Evan, 21yrs (Pillai = 0.934)

Figure 5.3: Formant plots for the speakers with the lowest and highest Pillai scores

Figure 5.4 displays Pillai scores by speaker sex and age. Each point on the plot represents a speaker, and trend lines are plotted to show the patterns for the male and female speaker groups. This plot clearly shows different trends for the male

and female speakers. Within the speakers aged 40 and over, little difference can be seen between the Pillai scores of males and females. This indicates a lack of a sex effect on the mergedness of the GOAT and THOUGHT vowels in the older members of the speech community. However, when looking at the youngest speakers, there is a clear difference between males and females. The trend line shows that, as a group, younger females have the lowest Pillai scores. A different pattern is manifest in the male speaker data, as Figure 5.4 shows that younger male speakers have much higher Pillai scores than both the female and the older male speakers. Therefore, while older male and female speakers are at a similar mergedness level (many with Pillai scores of around 0.5), the younger males and females are diverging. This result may indicate that young male and female speakers in the same speech community are leading sound change in opposite directions.

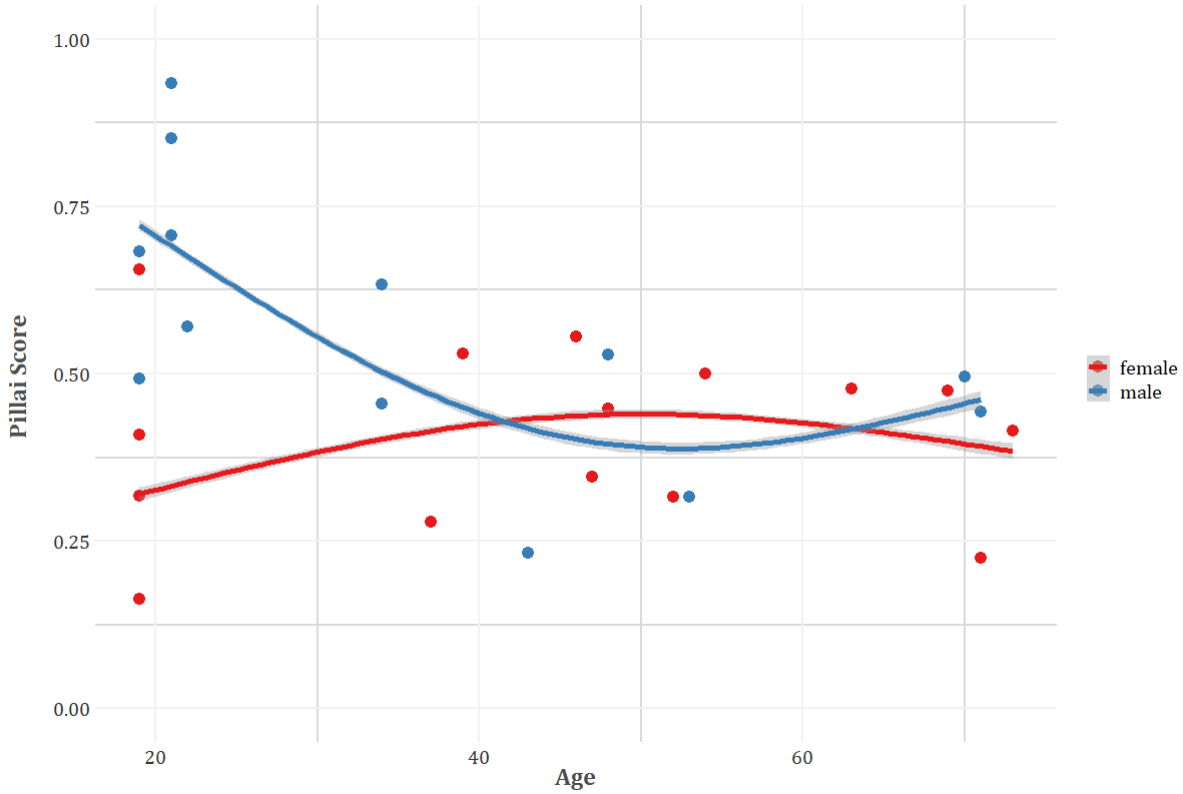


Figure 5.4: Pillai scores by speaker sex and age

The trend highlighted in Figure 5.4 broadly reflects the results of the F1-F2 plots displayed in Figures 5.1 and 5.2. Both show that younger female speakers, as a group, demonstrate the largest similarity between the GOAT and THOUGHT vowels, while young male speakers show the greatest difference between the two vowels. However, as analysis up to this point has focused on using static one-point vowel measurements, Pillai scores may not be able to fully represent the differences in the

overlap of GOAT and THOUGHT between young female and males speakers. Analyses using more dynamic data, as will be discussed in section 5.3, may shed more light on how and why these particular speaker groups appear to differ so vastly with respect to the overlap of the GOAT and THOUGHT vowels.

5.1.3.1 Pillai Scores and Social Class

Taking a brief look at social class, Figure 5.5 shows how socioeconomic class interacts with Pillai scores. The lower-working class speakers appear to have the lowest Pillai values, indicating that they have the greatest similarity in quality between the GOAT and THOUGHT vowels. At the opposite end of the scale, several of the lower-middle and upper-middle class speakers show the highest Pillai scores. The line of best fit in Figure 5.5 also shows a positive correlation between socioeconomic status and Pillai score, suggesting that speakers of a higher social class typically have more of a difference between GOAT and THOUGHT. This could be taken as evidence that middle class speakers are not participating in a GOAT-THOUGHT merger-in-progress, a change that the working class speakers appear to be leading. However, it may be the case that middle class Tynesiders are using different variants of GOAT to the working class speakers, as previous research on Tyneside English has shown increased usage of the diphthong [oʊ] among speakers of higher socioeconomic status (Watt 1998; Watt 2002). This will be discussed in further detail in Section 5.2

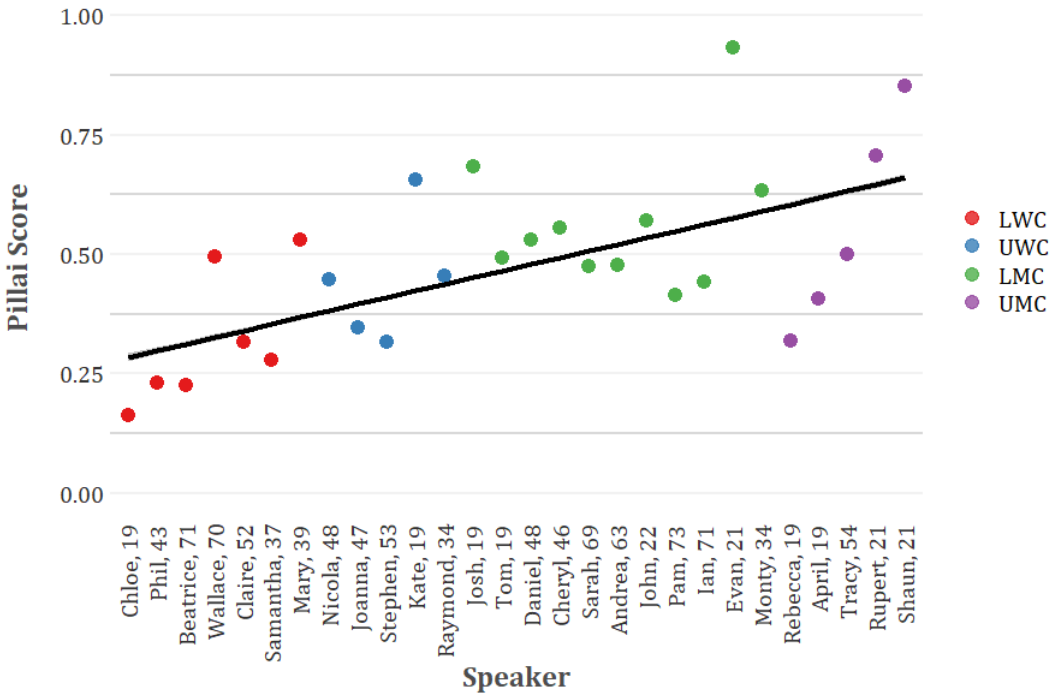


Figure 5.5: Pillai scores by socioeconomic class

Although Figure 5.5 gives some insight into how a Tynesider's social class may affect their GOAT-THOUGHT mergedness, as noted in Chapter 4, due to certain limitations of DECTE, the current sample of speakers is not ideally balanced across the social factors of sex, age and social class. Thus, while an analysis of social class cannot be ignored entirely, within the current study, it is not possible to comprehensively investigate the interactions between socioeconomic class and speaker age/sex, nor is it possible to draw concrete conclusions on the effects of social class on GOAT-THOUGHT merging in Tyneside English.

5.2 Auditory Analysis of GOAT

An auditory analysis of the data was conducted in order to survey the use of different GOAT variants in the speaker sample. Previous research on Tyneside English has found four main variants of GOAT in the speech community. As discussed in Chapter 3, monophthongal [oː] is reported to be the region's most frequently used realisation of the GOAT vowel (Watt 2002; Beal *et al.* 2012). However, to investigate whether GOAT is merging with THOUGHT in Tyneside English, it is necessary to examine which – if any – other pronunciations of the GOAT vowel are used by the speakers in the current sample. Any evidence which suggests the continued usage of multiple GOAT variants in contemporary Tyneside speech will have both theoretical and methodological implications for how GOAT-THOUGHT merging in speech production may be assessed.

5.2.1 Overall Variant Usage

For a general overview of GOAT variant usage in the present study, Table 5.4 shows the count total and the percentage of each variant in the data. While, as expected, monophthongal [oː] is by far the most prominently used variant in the DECTE interviews, some variation in the realisation of GOAT is present. The second most frequently used pronunciation is [oʊ], which is a variant that can be found in many varieties of English across the UK. As shown in the table, variants local to Tyneside such as [ʊə] and [ɵː] make up a very small percentage of the data. Such figures indicate that speakers of contemporary Tyneside English favour the supralocal and national pronunciations of the GOAT vowel.

Variant	No. of Tokens	Percentage
oː	2265	89%
ʊə	25	0.98%
oʊ	198	7.7%
ɵː	56	2.2%

Table 5.4: Usage of each GOAT variant in DECTE data

5.2.2 Results: Social Variables

Figure 5.6 shows the GOAT vowel variants used by different speaker groups according to speaker sex and age. Clearly, the most prominently used variable across all speaker groups is the monophthongal [oː]. This finding is unsurprising given previous research on North-Eastern speech. However, Figure 5.6 also shows that there is still some variation within the pronunciation of GOAT on Tyneside.

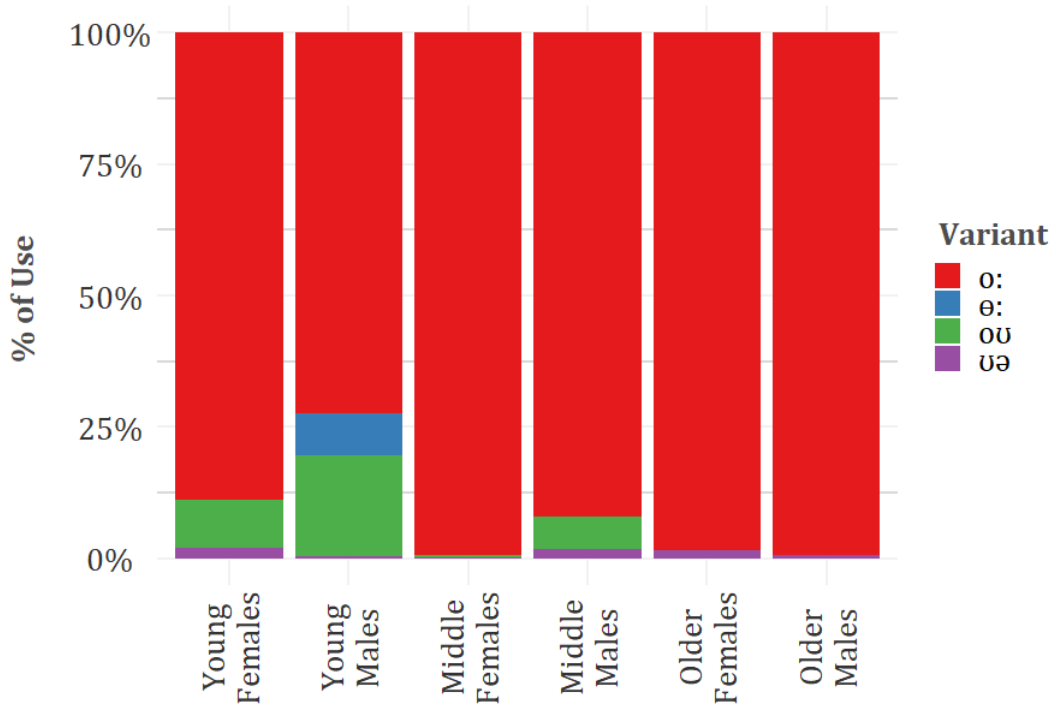


Figure 5.6: GOAT variants by speaker group

The second most frequently used variant, as Figure 5.6 shows, is the non-local closing diphthong [oʊ]. Both male and female speakers are found to use this variant, although results suggest an effect of age: younger speakers have increased usage of [oʊ], while the oldest speakers avoid it entirely. There are a few tokens (N = 25) of the regionally marked centering diphthong [ʊə]. This variant is used very

infrequently, although there are examples of its usage in all speaker groups. The figure shows that most tokens of [ʊə] are found in the speech of young females. This is a highly surprising result given that research in recent years has found it almost exclusively in the speech of males (Watt 1998; Corrigan *et al.* 2014). However, further analysis, to be discussed below in section 5.2.3, indicates that this unexpectedly high rate of [ʊə] in the female speakers may have more to do with the phonetic context of the GOAT vowel than any sex difference between male and female speakers in the use of this particular variant. Use of a fronted monophthong [ɵː] is used only by young male speakers, and within this group it is used solely by two speakers, Evan and Shaun. Figure 5.7 below shows the GOAT variants used by these two young male speakers, illustrating that Shaun predominantly uses the [ɵː] pronunciation in his own speech.Shaun is the only speaker in the current sample not to overwhelmingly favour [oː] pronunciations of GOAT.

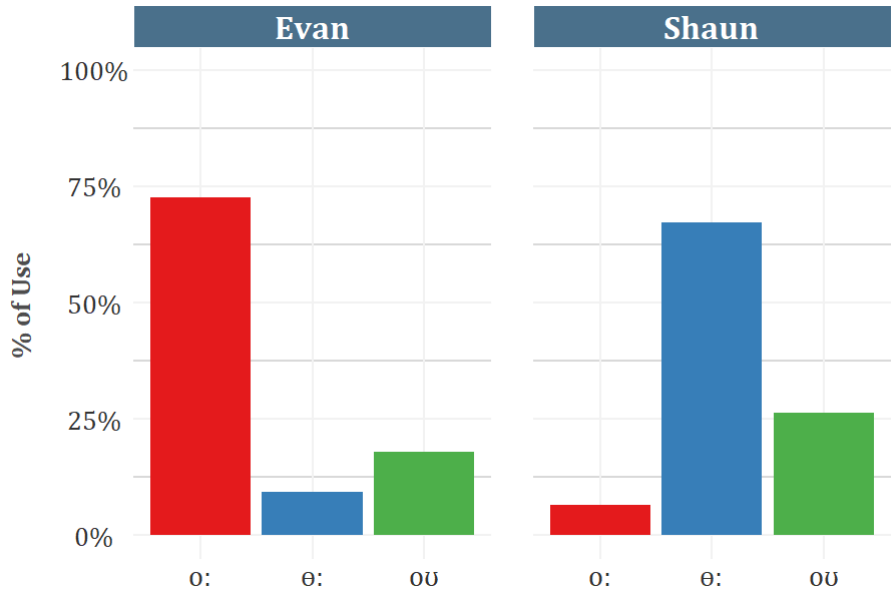


Figure 5.7: GOAT variants of two young male speakers

The quality of [ɵː], at least for the speakers in the present study, is often very close to the quality of NURSE. Similar [ɜː]-like fronted realisations of GOAT have been reported as a feature of the Hull accent (Watt and Smith 2005; Hickey 2015). Watt (1998) finds some usage of [ɜː] for GOAT among his Tyneside speakers, subsuming these realisations, along with other fronted monophthongal pronunciations, under the broad transcription of [ɵː]. While the NURSE-like pronunciations used by Evan and Shaun in the current study could be transcribed as [ɜː], the remaining analysis will continue to represent these pronunciations broadly as [ɵː] to allow for clearer comparisons with previous accounts of the Tyneside GOAT variation.

In terms of socioeconomic status, Figure 5.8 shows the use of GOAT variants by different social class groups. As with the results for the sex and age group variables, monophthongal [oː] is the predominant variant across all socioeconomic groups. Notably, [oʊ] is least present in the LWC speakers, while the UMC group show the highest percentage of [oʊ] usage. These results are unsurprising given that [oʊ] is the prestige, non-local variant of GOAT, and are broadly consistent with previous findings on contemporary Tyneside English (Watt 1998; Corrigan *et al.* 2014).

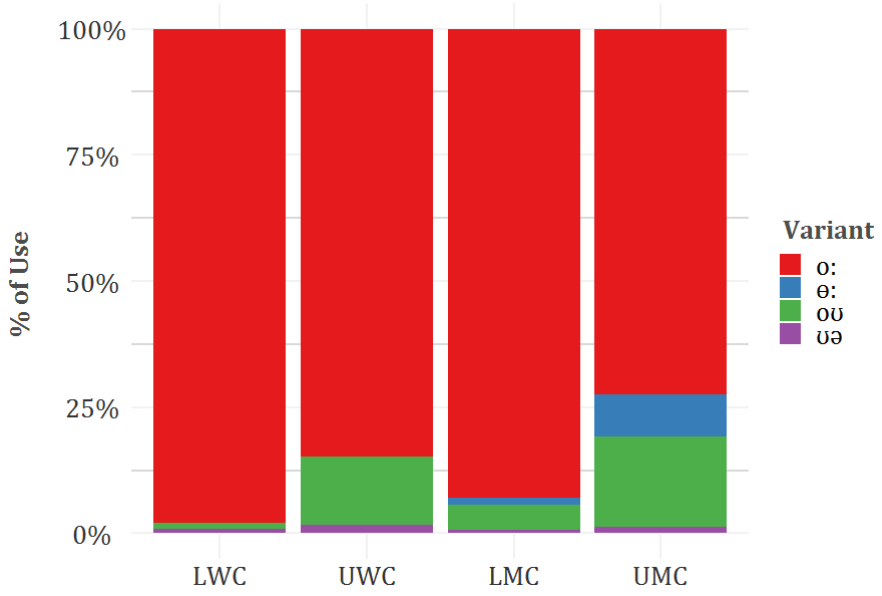


Figure 5.8: GOAT variants by socioeconomic class

As detailed above in section 5.2.2, only two young male speakers show any evidence of [ɵː] pronunciations of GOAT, with only one speaker using the fronted monophthong as their majority variant. Both of these speakers fall into middle class speaker groups (Evan in the lower-middle class group, Shaun in the upper-middle class group). Therefore, according to the results of the present study, [ɵː] is used exclusively by young, male, middle-class Tynesiders. Watt (1998: 236-7) finds a high usage of [ɵː] among his male speakers, positing that [ɵː] allows males to signal their Tyneside locality without using the 'stigmatised' regionally marked [ʊə]. This no longer seems to be the case, with the majority of male speakers favouring [oː] or even [oʊ] realisations of GOAT. It is possible that [ɵː] is being used by these two young middle-class males as a way of indicating their Tyneside identity. However, given that this variant is exhibited predominantly by one speaker (Shaun), and that the quality of this pronunciation sounds similar to the East Yorkshire [ɜː], perhaps his use of this variant is not as a traditional Tyneside pronunciation but as influence from elsewhere. While nothing in the

demographic information collected from Shaun would explain his use of [ɜː][1], given that his distribution of GOAT variants differs significantly from other speakers in his age/sex/social class group — and from the rest of the speakers in the sample — it is difficult to confidently determine whether Shaun is using a form of the local [ɵː] pronunciation of GOAT, or whether his use of this variant is due to an unrecorded outside influence.

5.2.3 Results: Effects of Phonetic Context

The effects of the following segment's manner of articulation on GOAT is shown in Figure 5.9[2]. Here, it can be seen that monophthongal [oː] is the most frequently used variant, irrespective of the following sound. Figure 5.9 shows no variation in GOAT pronunciations when followed by an affricate, with only monophthonal [oː] being used in this environment. This apparent lack of variation is more likely due to the comparatively low number of tokens of GOAT ___ [+ affricate] contexts (N = 7) than any phonetic effect of an affricate on a preceding GOAT vowel. For the fricative, nasal, and stop contexts, a similar distribution pattern can be found: [oː] is the majority variant, followed by [oʊ], then [ɵː] and finally [ʊə]. This indicates that, overall, the manner of the following segment has little effect on the use of different GOAT variants.

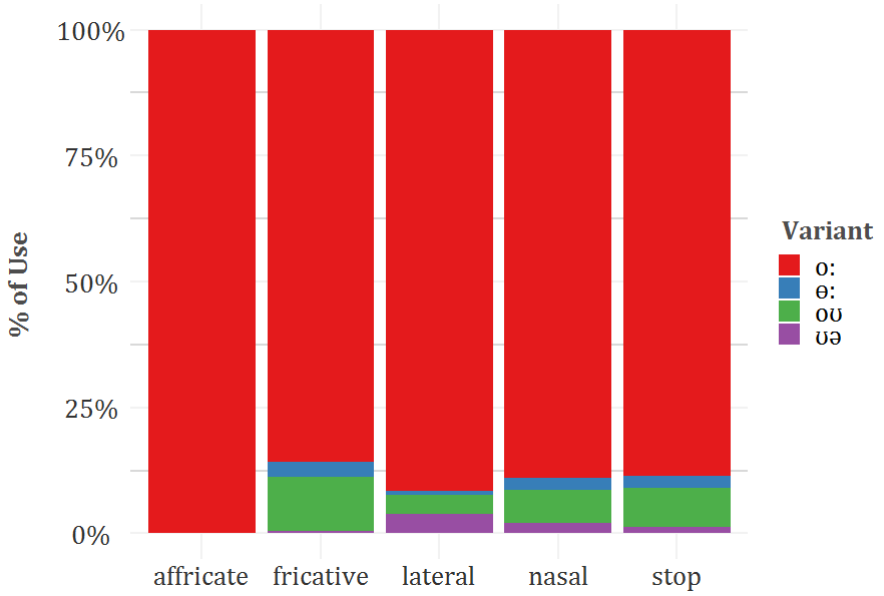


Figure 5.9: GOAT variants by manner of following segment

[1]This speaker was born and raised in Tyneside, currently resides in the region, and both parents are from the North-East

[2]Again, the FAVE default categories for phonetic environment were used in this analysis.

One notable finding from Figure 5.9 is that the [ʊə] variant is most frequently found in pre-/l/ contexts. Watt (1998: 237) finds a similar pattern in his Tyneside data, reporting that the centering diphthong appeared most frequently before lateral and nasal consonants. He also notes that female speakers only use [ʊə] in pre-/l/ and pre-/n/ environments. This is reflected in the present study, as Figure 5.10 shows how the manner of articulation of the following segment effects usage of [ʊə] differently for male and female speakers.

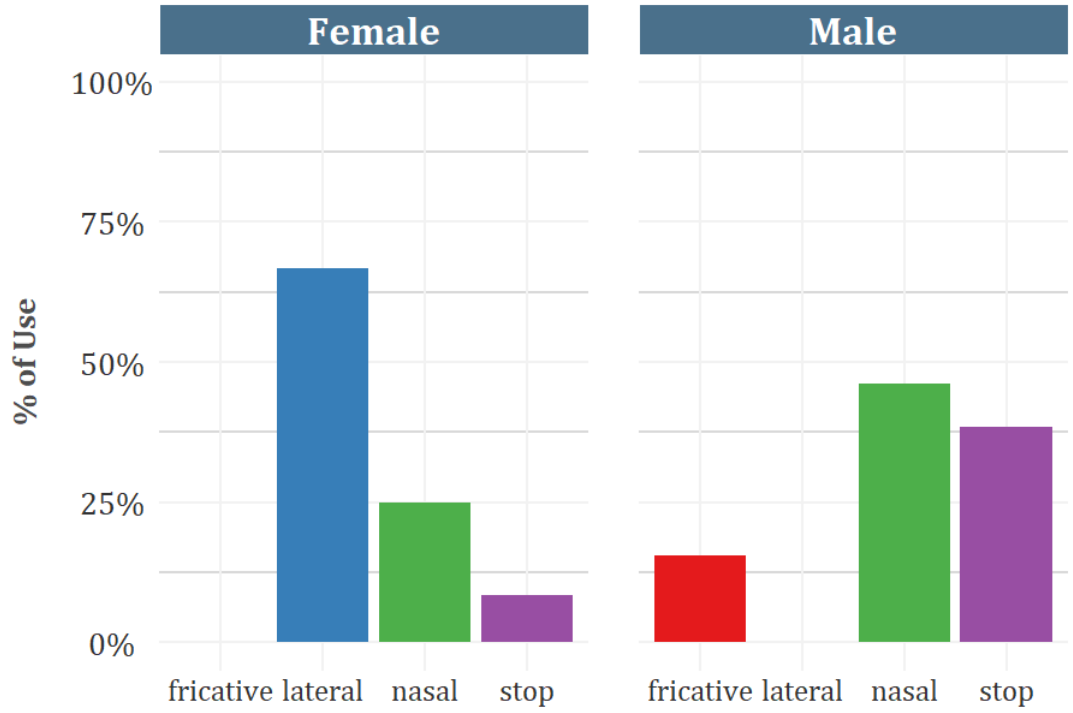


Figure 5.10: Effects of speaker sex and following segment manner on [ʊə] usage

As Figure 5.10 shows, when female speakers use the [ʊə] variant, it is most often followed by an /l/. However, use of the centering diphthong in male speech is found to be most frequently followed by a nasal or a stop consonant. Watt (1998: 237) posits that such results can be explained by the tendency of English speakers to diphthongise monophthongs in pre-/l/ position. More specifically, examples of this typically involve a monophthong being realised as a centering diphthong, as in *'real'* [ɹɪəɬ], *'feel'* [fɪəɬ] and *'cool'* [kuəɬ], *'school'* [skuəɬ]. It is therefore possible that the pre-lateral tokens of [ʊə] found in the speech of these Tyneside females is the result of occasional dipthongisation of monophthongal [oː] when preceding /l/, rather than the use of the traditional Tyneside centering diphthong. This may help to explain why, as shown in Figure 5.6, female speakers unexpectedly show the highest rates of [ʊə] usage when compared with other speaker groups.

5.2.4 Statistics

Multinomial logistic regression models were run on the results of the auditory analysis to further examine the use of different GOAT variants by different speaker groups. The model selection process was discussed in section 4.7.2, and the details of the model comparison are given below in Table 5.5.

model	AIC
Variant ~ Sex + Age Group	1820
Variant ~ Sex + Age Group + Social Class	1678
Variant ~ Sex + Age Group + Social Class + Manner	1217
Variant ~ Sex + Age Group + Social Class + Manner + Other	1210

Table 5.5: Comparison of multinomial logistic regression models for GOAT variant usage

Table 5.5 shows that models which include all social variables (i.e sex, age group, social class) have a lower AIC value and therefore better explain the data. The table also shows that adding in the manner of articulation of the following segment greatly improves the model. The model with the lowest AIC value is that which contains all social variables plus independent variables for manner, place and voicing of the following segment and the backness/frontness of the preceding segment. However, the improvement in AIC value from adding these additional independent variables is minimal, so, in order to keep the ideal model as parsimonious as possible, the favoured model here is the one with the second lowest AIC value. The output of a `multinom()` model does not include any values for significance testing, so a two-tailed z-test was performed on the model output, using the coefficients and standard error values to generate z statistics and p values. A z-test is carried out on each category of the dependent variable separately to compare it against the reference category. Here, [oː] was used as the baseline in the models as it is the most ubiquitous variant. The results of the z-tests are shown below in Tables 5.6 - 5.7. A p value of <0.05 indicates significance.

Table 5.6 shows the probability of speakers using the diphthongal [ʊə] rather than the most frequently used variant [oː]. Looking at the z statistics and p values, the model output shows that males are slightly more likely to use [ʊə] than females, and that the oldest speaker group show a higher probability of using this variant; however, the model does not find these differences to be significant. Similarly, social class does not seem to have a significant effect on [ʊə] usage for these Tyneside speakers. The model output does show that speakers are more likely to use an [ʊə]

when the GOAT vowel is followed by a lateral consonant, although the p value does not quite reach significance ($p = 0.052$). As shown above in Figure 5.10, a following /l/ appears to promote use of [ʊə] in female speech, though male speakers do not exhibit the same effect.

	Coefficient	**St. Errors**	**z statistic**	**P value**
(intercept)	-4.827	0.819		
sex: male	0.259	0.435	0.825	0.409
age group: middle	-0.088	0.543	-0.163	0.870
age group: older	0.902	0.680	1.326	0.184
social class: UWC	1.074	0.673	1.594	0.110
social class: LMC	-0.432	0.609	-0.709	0.477
social class: UMC	0.766	0.753	1.017	0.309
manner: fricative	-1.005	0.821	-1.225	0.220
manner: nasal	0.425	0.533	0.796	0.425
manner: lateral	1.071	0.553	1.935	0.052

Table 5.6: Model output showing use of [oː] vs. [ʊə]

	Coefficient	**St. Errors**	**Z statistic**	**P value**
(intercept)	-54.990	0.770		
sex: male	32.779	0.322	6.701	<0.001
age group: middle	-25.655	0.295	-5.560	<0.001
age group: older	-13.302	—	—	—
social class: UWC	-7.545	0.761	2.002	<0.045
social class: LMC	19.411	0.770	0.250	0.801
social class: UMC	21.711	0.762	2.382	0.017
manner: fricative	-0.052	0.258	0.510	0.609
manner: nasal	-0.024	0.280	-0.799	0.423
manner: lateral	-1.591	0.425	-2.776	0.005

Table 5.7: Model output showing use of [oː] vs. [oʊ]

Table 5.7 displays the model output for the use of [oʊ] compared with [oː]. Here, the table shows that male speakers are significantly more likely to use [oʊ] than female speakers, and that younger speakers are significantly more likely to use this variant than middle aged speakers. There are a few missing values for the older age group in the model output: as no examples of [oʊ] were found in the speech of

any of the older speakers, the model could not output a standard error value, and it was therefore not possible to work out the z statistic or p value for this speaker category. In terms of social class, the model output shows that the UWC and UMC speakers are significantly more likely to use [oʊ] ($p = <0.001$). The model indicates that a following lateral significantly affects whether speakers use a particular GOAT variant, with [oʊ] having a lower probability of occurring in pre-lateral position.

5.2.5 Summary

The results of the auditory analysis show that monophthongal [oː] is by far the most favoured GOAT variant in contemporary Tyneside speech. However, there are examples of other pronunciations in the data, therefore some variation in the realisation of the GOAT vowel must still exist in the region. The presence of these other variants may help to account for the high degree of variance of GOAT in the vowel space of several speakers, as shown in the formant plots in Figures 5.1 and 5.2. Auditory analysis has also shown that, overall, the second most used variant is the non-local diphthong [oʊ]. Results indicate that this variant is not limited to middle class speech, but rather that is is used in some capacity by speakers of each socioeconomic group. However, the results do suggest an effect of age, with younger speakers using [oʊ] more than the older speaker groups.

Findings also suggest that the young males in the sample, as shown in Figure 5.6, exhibit the most variation in the use of different GOAT pronunciations. This is somewhat consistent with the results of Watt's (1998; 2002) Tyneside study, which found more variability in GOAT amongst the male speaker groups. However, while Watt did note a decline in the use of the traditional Tyneside variants [ɵː] and [ʊə] in his data, the frequency of these variants in the current data is comparatively low. Watt finds [ʊə] to be used almost exclusively by (primarily older) working-class male speakers, while the current study finds very low rates of [ʊə] usage in both male speakers and working-class speakers. This could signal a further, more rapid decline of the use of the local diphthongal GOAT variant in the speech community, suggesting that it is almost completely absent in contemporary Tyneside English. Alternatively, it is possible that the difference in results between Watt's studies and the current study is due to a difference in data collection methods and the categorization of social class boundaries. Whereas Watt used the PVC corpus, which was set up to interview participants from pre-determined 'working-class' and 'middle-class' neighbourhoods, as highlighted in Chapter 4.4.1, DECTE is not structured to allow for a strict sampling of speakers from different social backgrounds. Given that the DECTE interviews are collected by university students via Newcastle University, the current

sample may not contain speakers from a similar 'working-class' background to those sampled in the PVC corpus. As such, it may be possible that [ʊə] does survive among male working-class speakers in Tyneside, but that speakers who use this variant are neither reached nor explicitly targeted by DECTE.

Auditory analysis has shown that there is still some variation in the pronunciation of GOAT on Tyneside, and that diphthongal variants are found in the data. As such, acoustic analyses which use one measurement point per vowel to retrieve the F1-F2 values are unlikely to fully capture this variation in GOAT pronunciations. Given the presence of diphthongal GOAT variants in the data, any assessment of the mergedness of GOAT and THOUGHT using static data may be inaccurate. It is therefore advantageous to carry out an additional analysis utilising more dynamic data (i.e. multiple F1-F2 measurements per vowel) in order to more accurately assess GOAT-THOUGHT merging in Tyneside English.

5.3 Analysis of Dynamic Data

The results of an analysis of GOAT-THOUGHT merging in Tyneside English using dynamic data are presented in this section. As detailed in section 4.7.3, formant measurements were taken at every 10% of the vowel duration for each token of GOAT and THOUGHT in the data. This allows for an analysis of GOAT-THOUGHT similarity and overlap across the vowel trajectory. Using this dynamic data, Generalised Additive Mixed Models (GAMMs) were performed on the data.

5.3.1 Generalised Additive Mixed Models

Figures 5.11 and 5.12 show the predicted formant trajectories of GOAT and THOUGHT for each speaker group (sex and age) in the data, with Figure 5.11 showing the results for the female speakers and Figure 5.12 showing those of the male speakers. Each plot within the figures shows separate trajectory lines for THOUGHT (red) and GOAT (blue), with the y-axis marking the formant value and the x-axis representing the time-point within in the vowel's duration. The p-values displayed on each plot indicate whether the model found the GOAT and THOUGHT vowels for that speaker group to be significantly different with respect to F1-F2. The plots showing the F1 trajectories are displayed on the bottom row of each figure, with the F2 trajectory plots shown on the top row.

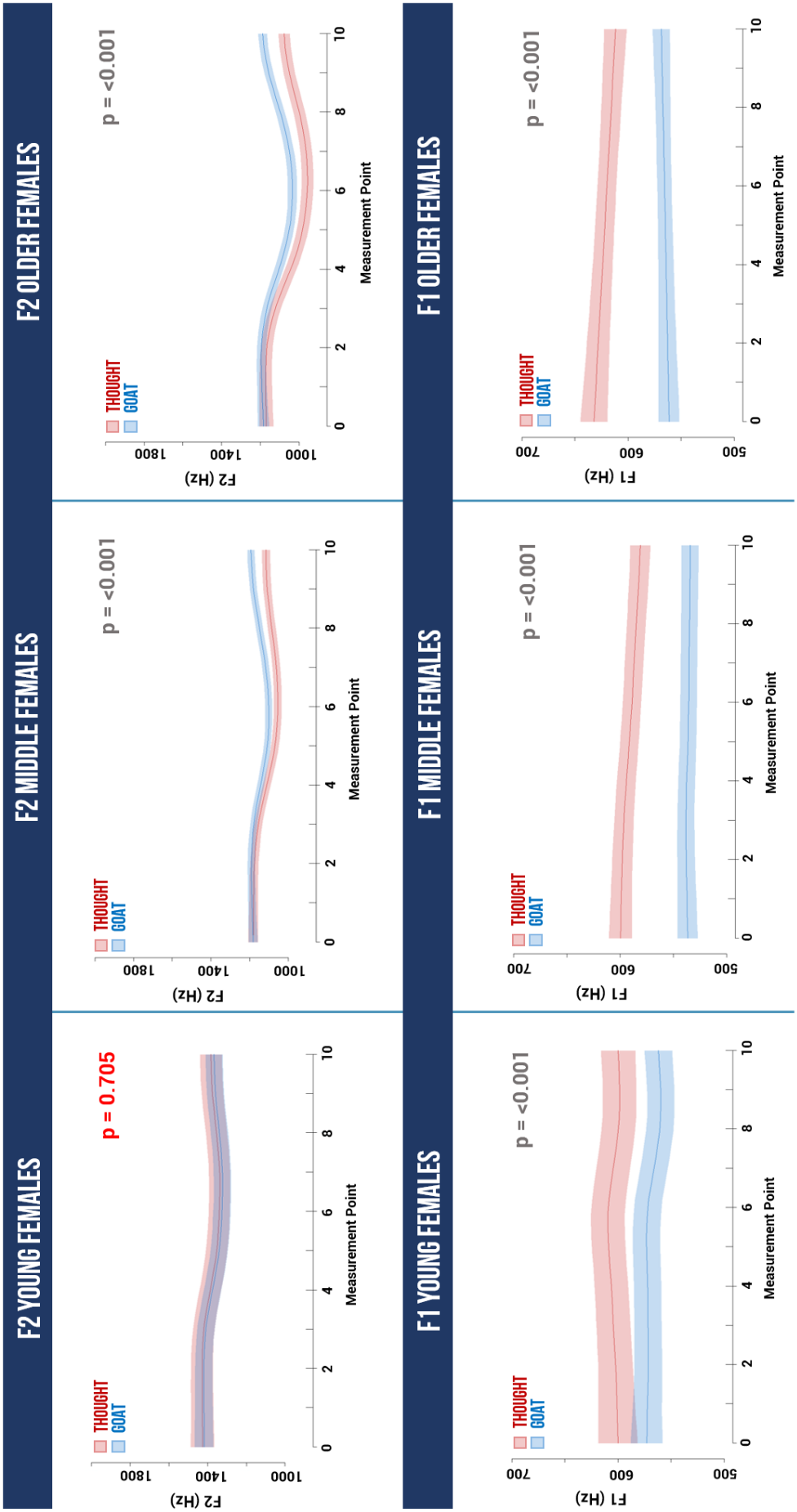


Figure 5.11: Predicted F1 and F2 trajectories of GOAT and THOUGHT for Tyneside female speaker groups

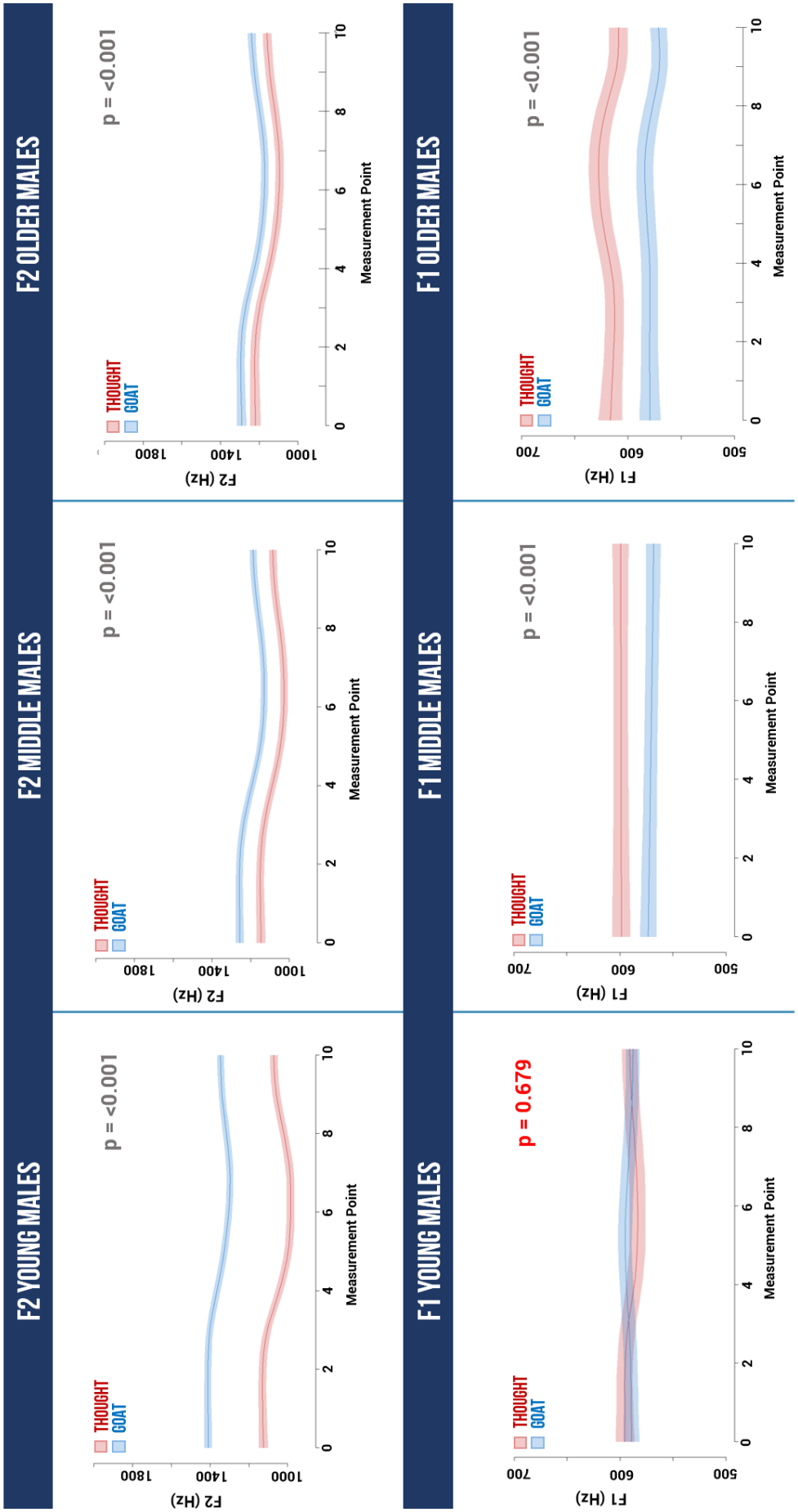


Figure 5.12: Predicted F1 and F2 trajectories of GOAT and THOUGHT for Tyneside male speaker groups

Figure 5.11 shows the predicted formant trajectories of GOAT and THOUGHT for each of the female speaker groups. The figure shows that young female speakers are expected to show the most overlap in F2, with the GOAT and THOUGHT trajectories showing considerable overlap across the vowel duration. This indicates that the two vowels are of a similar frontness, confirmed by the p-value ($p = 0.705$) which shows that GOAT and THOUGHT are not significantly different in terms of frontness for young Tyneside females. The middle and older female speakers also show GOAT-THOUGHT overlap during the first 1/3 of the vowel before diverging, with GOAT appearing fronter than THOUGHT toward the end of the vowel. This highlights two important things: (1) vowels may still show formant movement even if they sound monophthongal, and, (2) static measures of monophthongal vowels may not always capture potentially merging vowels in enough detail. Also of note is that the predicted F2 trajectories for both GOAT and THOUGHT are higher for the young than the middle and older females. This could suggest that some slight fronting of both vowels is taking place in the Tyneside speech community.

A similar pattern in the predicted F1 trajectories can be seen in Figure 5.11. The middle and older female speakers show no overlap of GOAT and THOUGHT, with a considerable height difference separating the two vowels. For these speaker groups, the GOAT and THOUGHT vowels were found to have significantly different F1 trajectories ($p = <0.001$). The plot for the young female speakers in 5.11 also shows minimal predicted overlap of the GOAT and THOUGHT F1 trajectories, with the GAMM model indicating that these two word classes are significantly different ($p = <0.001$) in terms of F1 for this speaker group. However, the figure also shows that the distance between the predicted F1 trajectories of GOAT and THOUGHT is much smaller for the young female speakers than for the middle and older females. As both vowels appear to be of a similar height and frontness, these results could suggest that the young females may be leading a sound change toward a GOAT-THOUGHT merger in Tyneside English.

Figure 5.12 displays the predicted GOAT-THOUGHT F1 and F2 trajectories for the male speakers. Looking first at the F2 plots, young males are found to have a greater distance between the predicted trajectories of GOAT and THOUGHT than the middle and older male speakers. This result reflects the previous results of the overlap plots and Pillai scores which indicate that young males have a greater distinction between these two vowels than both female and older male speakers. While the distance between GOAT and THOUGHT appears larger for the young males, all three age groups were found to have a significant difference between the two word classes ($p = <0.001$). Comparing all three age groups, the GOAT vowel is much fronter for

younger male speakers. Similar fronting of GOAT was shown in the young female speakers in Figure 5.11, again suggesting that GOAT may be undergoing fronting in Tyneside.

The predicted F1 trajectories shown in Figure 5.12 show the opposite pattern to F2. The middle and older male speaker groups show no overlap of GOAT and THOUGHT in terms of F1, though the distance between the trajectories appears quite small (between 20-50 Hz difference). Both groups are determined by the model as having a significant difference in F1 between GOAT and THOUGHT ($p = <0.001$). Conversely, the plot for the young males shows a lot of GOAT-THOUGHT overlap, with the model finding no significant height difference between the trajectories of GOAT and THOUGHT ($p = 0.679$). As such, the results of the GAMMs present a complex picture; young males are the most GOAT-THOUGHT merged group in the speech community in terms of vowel height, yet they also have the biggest distinction between GOAT and THOUGHT in terms of vowel frontness.

The results of these GAMM models indicate, for both the female and male speakers, that there is a trend in apparent-time towards the loss of a height distinction between GOAT and THOUGHT. Looking at the predicted F1 trajectories in Figures 5.11 and 5.12, the young male and female speakers exhibit GOAT vowels with higher F1 values than the other speaker groups, indicating that GOAT is undergoing a process of lowering toward the position of THOUGHT. The figures also suggest some slight raising of the THOUGHT vowel over time, which is most notable in comparing the F1 trajectories of the middle and older speaker groups. However, this difference is rather slight, and the height of the THOUGHT vowel is similar for the younger and middle age groups. The predicted trajectories output from the GAMM models would therefore indicate that the merging of the GOAT and THOUGHT vowels in the F1 plane – as exemplified by the young female and, especially, the young male speakers – is largely the result of a lowering of the GOAT vowel over time in the Tyneside speech community.

5.4 Duration

In analysing vowel mergers, Gordon (2013: 206) comments that researchers are often guilty of relying too heavily on static formant values as a measure of mergedness. He continues that vowel duration is an additional acoustic measure which should be included in any analysis of merger, as a difference in duration may preserve a distinction between vowels which overlap in F1-F2 space. In the past decade or so, studies of mergers and distinctions have often included measures of vowel duration

in their analysis (Wassink 2006; Labov and Baranowski 2006; Irons 2007; Fridland *et al.* 2014; Arnold 2015).

Vowel duration has been found to be affected by phonological environment, e.g. vowels are shorter in polysyllabic words, and when followed by a voiceless consonant (House 1961; Klatt 1976; Labov and Baranowski 2006). The question therefore arises as to how much durational flexibility is permitted before one vowel starts to sound like another, e.g. /i/ begins to sound more like /ɪ/. Klatt (1976: 1219) defines the *just noticeable difference* (JND) for vowel duration as 25 milliseconds (msec), indicating that durational changes below this threshold will not be perceptually important for listeners. More recently, Labov and Baranowski (2006) conducted a perception experiment to examine whether durational differences were enough to uphold the contrast between /ɛ/ and /ɑ/ in Northern Cities Shift speech. The authors find that, despite these two vowels showing considerable overlap in F1-F2 space, listeners were able to differentiate these vowels based on duration. They conclude that a difference of approximately 50 msec is enough to 'effectively alter the majority perception of the phoneme'.

There is little literature available detailing the approximate expected durational difference between the GOAT and THOUGHT vowels. In Williams and Escuerdo's (1999) comparison of vowels in Standard Southern British English (SSBE) and Sheffield English speakers, they find that GOAT has a shorter duration than THOUGHT on average, however, these differences were found to be slight at approximately 20 msec. It is important to note that the duration for GOAT found in SSBE and Sheffield English speakers may not be comparable to the duration of this vowel for the Tyneside speakers in the current study, as Tynesiders typically have a monophthongal [oː] realisation whereas speakers of SSBE and Sheffield English are found to use a [əʊ] diphthong. Currently, duration measurements are not provided in any description of the Tyneside vowel system.

5.4.1 Analysis

Measurements of vowel duration were obtained from the FAVE-extract output. Figure 5.13 shows the distribution of GOAT and THOUGHT vowel durations in the Tyneside dataset. Due to the FAVE-extract default settings, vowels with a duration of less than 50 msec are not measured, therefore no vowels below this duration are included in the data or the analysis. The figure also shows that there are a few tokens of each vowel within the 350-600 msec range. Given the infrequency of tokens of this length, and given that the data comes from conversational sociolinguistic interviews, these tokens are likely to be examples of

over-emphasised, elongated vowels. The dashed line in the plots marks the mean duration values for each vowel, which are 127.39 msec for GOAT and 131.55 msec for THOUGHT. As Figure 5.13 shows, both the distribution of vowel durations and the mean vowel duration are very similar for GOAT and THOUGHT, indicating that, overall, there is no difference in duration between these two vowels.

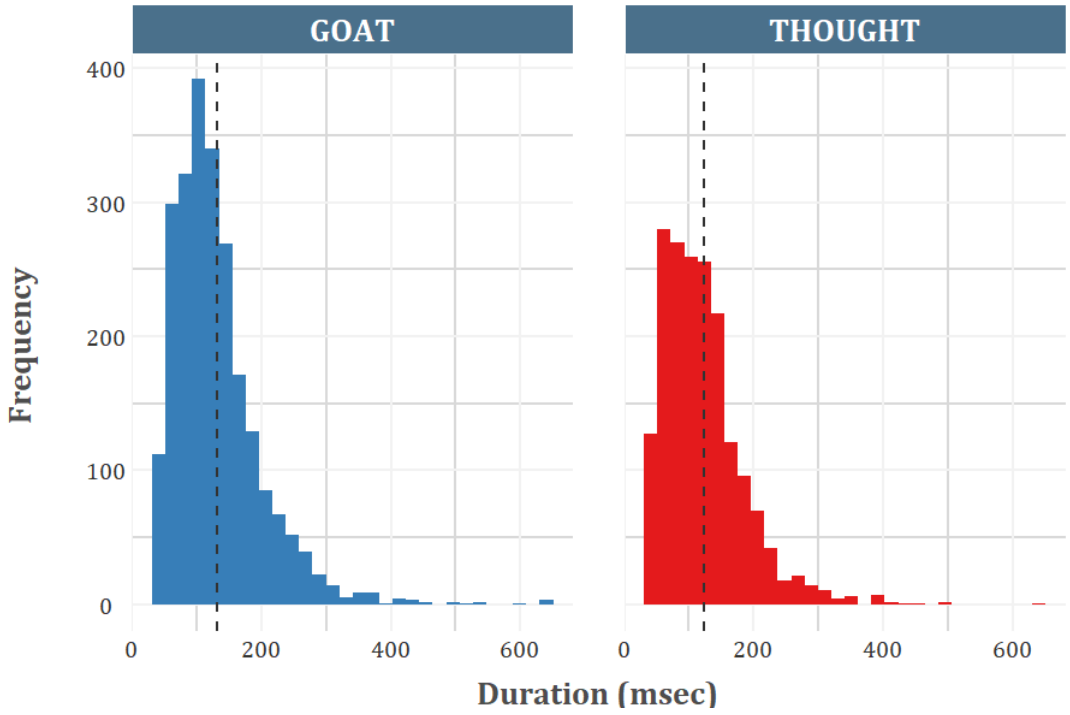


Figure 5.13: Vowel duration of GOAT and THOUGHT tokens

For further analysis of duration, the vowel length measurements in milliseconds were log-transformed (log10) in R. This log-transformation helps to make the data more normally distributed. Figure 5.14 below presents a comparison of the duration of the GOAT and THOUGHT vowels using these log-transformed duration measures. The first thing of note is that the height and position of the boxes are very similar for GOAT and THOUGHT, indicating that the interquartile range of duration measurements is similar for both vowels. The box representing the THOUGHT tokens is slightly lower than the GOAT box, but given the similarity in the position of the boxes and the median measurements, this figure suggests that there is no durational difference between these two vowels for the Tyneside speakers. The boxplot also shows several outliers at the higher end of the duration scale range. Again, these are likely to be examples of elongated or emphasised vowels which can be heard quite frequently in sociolinguistic style interviews.

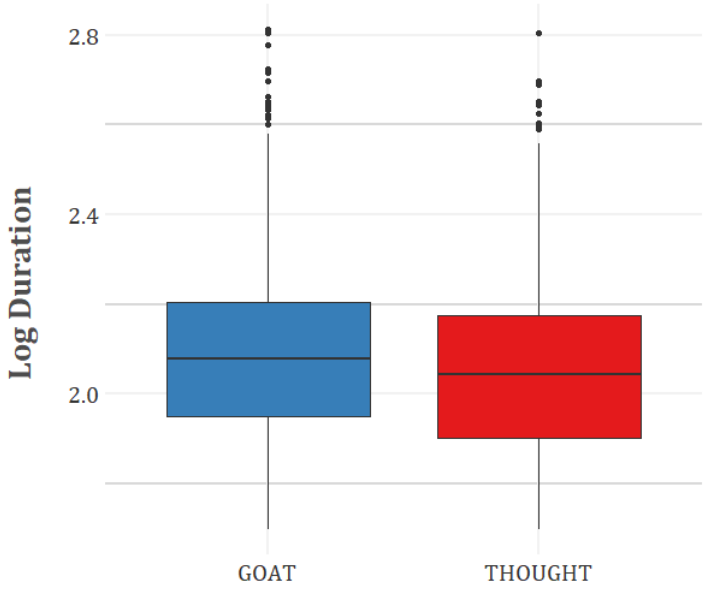


Figure 5.14: Comparison of GOAT and THOUGHT duration

A mixed-effects regression was performed on the log-transformed duration data. The results of this statistical analysis are displayed below in Table 5.8. The best model was selected using the model selection criteria outlined in Chapter 4.7.1.1, and contains interactions between the predictors of vowel type (i.e. GOAT or THOUGHT), speaker sex, and age group[3].

	Estimate	Std. Error	t-value
(intercept)	2.05	0.02	82.32
vowel: THOUGHT	-0.02	0.023	-0.72
sex: male	0.007	0.03	0.25
age group: mid	0.001	0.03	0.05
age group: old	0.09	0.03	2.63
vowel: THOUGHT * **sex:** male	0.007	0.02	0.29
vowel: THOUGHT * **age group:** middle	0.02	0.02	0.84
vowel: THOUGHT * **age group:** older	-0.02	0.03	-0.73
sex: male * **age group:** middle	0.004	0.04	-0.12
sex: male * **age group:** older	-0.08	0.05	-1.56
vowel: THOUGHT * **sex:** male * **age group:** middle	0.001	0.03	0.56
vowel: THOUGHT * **sex:** male * **age group:** middle	0.03	0.04	0.65

Table 5.8: Mixed-Effects regression results for GOAT and THOUGHT duration

[3]Social class was not included in the final model and models containing this predictor had higher AIC values, and were found not to significantly improve the model

As table 5.8 shows, there is a lack of a statistical difference between the duration of GOAT compared to the duration of THOUGHT. This again indicates that these two vowels cannot be distinguished based on vowel length. The only predictor that the model finds significant, as highlighted in Table 5.8, is that the older speakers in the sample have longer vowels. This finding could possibly be due to the reduced speech rate of older speakers leading to vowel lengthening. However, no durational effect was found for either sex or age group on vowel type, suggesting that no speaker group produces a vowel length distinction between GOAT and THOUGHT.

5.4.2 Summary

The results of an analysis of vowel duration suggest that there is no difference in vowel length between GOAT and THOUGHT. Summary statistics show that the average duration of both vowels is very similar, although GOAT is very slightly longer in length. However, as the difference between the mean duration of GOAT and THOUGHT is found to be only 4.1 msec, this figure is well below the 50 msec that Labov and Baranowski (2006) argue is enough to change phonemic perception, and under the 25 msec threshold that Klatt (1976) proposes listeners require in order to notice a difference in duration. This therefore indicates that, where GOAT and THOUGHT overlap in F1-F2 space, vowel duration would not be a reliable cue in helping listeners to distinguish these two vowels. Thus, there is not enough of a difference in duration to maintain a distinction between GOAT and THOUGHT when these two vowels converge in the vowel space of Tyneside speakers.

5.5 Discussion of Production Results

This chapter has presented an analysis of the production of the GOAT and THOUGHT vowels in contemporary Tyneside English, using both auditory and acoustic methods, and a combination of static and dynamic data. The key findings of this analysis are summarised below, and are discussed in greater detail in Chapter 8.

An analysis of formant values of each speaker's GOAT and THOUGHT tokens has shown that, for many speakers, these vowels occupy a similar position in the vowel space. Notably, many speakers appeared to show a lack of a vowel height distinction between GOAT and THOUGHT. This finding was further supported by the results of the generalised additive mixed models (GAMMs), which showed closeness and overlap in the predicted F1 trajectories of the GOAT and THOUGHT vowels for both young male and young female speakers. Mixed-effects regression models showed a

slight indication that THOUGHT is more raised in the speech of younger Tynesiders, while formant plots indicated that many speakers had GOAT vowel which was closer in height to the THOUGHT vowel. The predicted trajectories output by the GAMM models did show some slight THOUGHT-raising in the younger and middle speaker groups compared with the oldest speakers in the sample, but the most notable finding was the indication that a process of GOAT-lowering is leading to a lack of a height distinction between these vowels in the youngest members of the speech community. These results suggest a phonetically gradual sound change in which GOAT is lowering to the position of THOUGHT in apparent time.

While the results which show a lessening of a height distinction between GOAT and THOUGHT in the younger speakers could be indicative of a merger-in-progress, analysis of vowel frontness and Pillai scores complicate the picture. Formant analyses suggested that, for many speakers, GOAT has a slightly fronter realisation than THOUGHT. Additionally, the results of a mixed-effects regression model showed that younger speakers tend to have fronter GOAT vowels than older speakers, and that male speakers exhibited significantly more GOAT fronting than female speakers. According to the GAMM models, the young females showed no significant difference in the predicted F2 trajectories of GOAT and THOUGHT, signifying a lack of a difference in frontness between these vowels. However, the young males have the largest frontness difference in the predicted trajectories of GOAT and THOUGHT, with the GOAT trajectory having a much higher F2 value than THOUGHT. This result could be in part explained by the findings of the auditory investigation, which found that the young male speaker group produced the highest frequency of non-monophthongal [oː] variants of GOAT, namely the closing diphthong [oʊ], and two speakers (the only speakers in the sample to use this variant) who showed high rates of usage of the fronted monophthong [ɵː].

Analysis using the Pillai score method has indicated that many speakers showed considerable overlap of GOAT and THOUGHT. Over half of the speakers were found to have a Pillai score of <0.5, signifying a lack of a clear difference between the realisations of these vowels for many of these Tyneside speakers. Further analysis showed that, as a group, young females have the smallest Pillai scores, which, within an apparent-time framework, could be taken as evidence of a young female led GOAT-THOUGHT merger-in-progress in the speech community. However, the Pillai score results also show that young males exhibit the greatest distinction between GOAT and THOUGHT. This lack of overlap between GOAT and THOUGHT may again be explained by the continued usage of multiple GOAT variants by young males in the speech community, e.g. [ɔː] may merge with [oː]-like pronunciations of GOAT, but

other GOAT realisations such as [oʊ] and [ɵː] will remain distinct from THOUGHT. The question of whether a merger can ever take place in a speech community where several variants of one or more vowels remain in usage will be discussed in further detail in Chapter 8.

A durational analysis found that the length of the GOAT and THOUGHT vowels was very similar for this sample of Tyneside speakers. Furthermore, given the similarity in the duration measures of both vowels, it is very unlikely that durational differences could uphold a contrast between GOAT and THOUGHT if a speaker lacked a distinction in vowel quality.

Overall, the results show that GOAT and THOUGHT are very similar in the production of many Tyneside speakers, with young females showing the most similarity and overlap between the vowels. These findings provide some evidence of a sound change in Tyneside English, whereby young female speakers are decreasing the distance and distinction between the GOAT and THOUGHT vowels. The low Pillai scores for this speaker group could suggest evidence of a merger-in-progress, although the GAMM models indicate that young females still retain a significant distinction between these two vowels in terms of height. Further discussion of whether the findings of this chapter support the idea of a GOAT-THOUGHT merger in Tyneside English are reserved for Chapter 8, where the results will be discussed in conjunction with the findings of the perception study.

Chapter 6

Perception: Method

This chapter presents the methodology of the perception component of the main study, which examines how Tyneside listeners perceive the GOAT and THOUGHT vowels. Sections 6.1 and 6.2 give some background into the two main types of speech perception task – *discrimination* and *identification* – and the different types of stimuli which can be used in these experiments. Then, Section 6.3 details the process of stimuli creation, before Section 6.4 discusses the participants who were recruited to perform the perception tasks. Section 6.5 describes the procedure of carrying out the perception experiments, providing further information on the 'same-different' (discrimination) and 'labelling' (identification) tasks used in this study. Analysis of the results of the perception tasks uses measures of accuracy, reaction time, and sensitivity, all of which are discussed in Section 6.6. Finally, Section 6.7 explains the production data that was recorded for some of the listeners, and how this is used for a comparison of participants' production and perception of the GOAT and THOUGHT vowels.

6.1 Speech Perception Experiments

The experimental design of speech perception tasks typically falls into two groups: *discrimination* experiments and *identification* experiments (Pollack and Pisoni 1971; Yeni-Komshian 1998; Macmillan and Creelman 2005; McGuire 2010; Drager 2014). These types of experiment also have applications outside of linguistic research, widely used by psychologists and social scientists (Macmillan and Creelman 2005). Speech perception studies often use a combination of discrimination and identification tasks, as they ask different questions of the listener: discrimination tasks assess a listener's ability to perceive a difference between two sounds or stimuli, while identification tasks ask whether a listener is

able to recognise and categorise a particular sound or stimulus.

6.1.1 Discrimination Tasks

Discrimination experiments, when used in speech perception research, investigate whether listeners are able to distinguish two or more audio stimuli.

A frequently used variant of a discrimination experiment is a 'same-different' task, also known as an AX discrimination experiment. Such tasks typically involve the presentation of two or more audio stimuli on each trial, and the listener is then asked whether the stimuli they heard were the 'same' or 'different' to each other. This allows for four possible combinations of stimuli: the 'same' pairs *(xx, yy)* and the 'different' pairs *(xy, yx)*. One major advantage of this task type is that it is very easily explained to participants (McGuire 2010: 3). This should reduce the number of erroneous responses resulting from a listener's lack of understanding of the task.

Another type of discrimination task that can be used in studies of speech perception is known as an *oddity* experiment. In an oddity task, listeners are presented with multiple stimuli and are asked to choose the 'odd one out' (Macmillan and Creelman 2005; McGuire 2010). As explained in Iverson (2012: 579), oddity tasks in speech perception research most often play three stimuli by three different speakers to listeners, two of which contain the same phoneme and one of which contains a different phoneme e.g. /pi-pi-bi/. The listener is then asked to decide which stimulus was the 'different' one. The advantage of this type of task is that it requires listeners to rely on their own phonemic categories, as they must ignore any inter-speaker phonetic differences in the production of the stimuli in order to decide which one differs from the others. As this type of task is useful in assessing a listener's phonemic categories, it could therefore be a useful method for investigating vowel mergers in speech perception.

The discrimination experiment used in the present study is a combination of a same-different (or AX) task and an oddity experiment. Rather than presenting two stimuli on each trial as in a classic same-different task, three stimuli are presented from three different speakers (two males, one female). This ensures that the stimuli are acoustically different while remaining phonemically identical, pushing listeners to rely on phonemic rather than phonetic differences to distinguish the stimuli. As such, the discrimination task in the current study will examine whether listeners recognise that GOAT is phonemically different to THOUGHT, and vice versa. Further details of this experiment procedure are discussed in section 6.5.1. This type of discrimination experiment which utilises stimuli from several different speakers has been used in other studies of vowel mergers and contrasts in speech perception, as

in Brunellière *et al.* (2009) and Dufor *et al.* (2013), and more recently by Zhang (2019) in her study of the production and perception of English vowels by Basque and Catalan speakers.

6.1.2 Identification Tasks

Speech perception studies which make use of identification experiments are primarily interested in examining a listener's ability to categorise or label audio stimuli. Identification type tasks have been used in several previous studies investigating the perception of vowel mergers (Labov 1994; Thomas and Hay 2005; Hay *et al.* 2006; Koops *et al.* 2008; Drager 2010).

One popular type of identification experiment is a 'labelling' (sometimes termed a 'forced choice identification') task. Typically, only one stimulus is presented in each trial, which the listener is subsequently asked to label. McGuire (2010: 7) notes that main advantages to this type of task are that it is easy to explain to participants – thus can be carried out quickly with a low chance of error, and that the data collected from this type of task is easy to analyse.

In the identification task used in the present study, participants are presented with one audio stimulus and are given two labels to choose from (e.g. Word A or Word B). Other versions of this task may instead present the listener with more label options, or even have the listener supply their own label (e.g. a *'write the word you heard'* type question). For this study, the identification task is used to assess whether the Tyneside listeners can reliably categorise GOAT and THOUGHT audio stimuli. A more comprehensive explanation of the procedure of the labelling task used in this study is presented in section 6.5.2.

6.2 Natural vs. Synthetic Stimuli

The stimuli used in speech perception experiments may be either natural (i.e. produced by a real speaker), or generated using speech synthesis software, such as the Klatt synthesizer (Klatt 1980) or Praat (Boersma and Weenink 2019). Synthetic stimuli have been frequently used in research on vowel perception, as in the Bell-Berti *et al.* (1978) study of the link between production and perception of the vowel pairs /i-ɪ/ and /e-ɛ/ in American English, in research by Frieda *et al.* (2000) on the perception of /i/ in different speech styles, and in Freeman's (2019) paper on pre-velar merging of /e-ɛ/ in the Pacific Northwest. However, as purely synthetic stimuli may sound unnatural, Thomas (2002: 130) notes that sociophonetic studies which include a perception experiment component typically

prefer natural stimuli to those that are synthetically constructed. Consequently, listeners participating in the perception task will be responding to stimuli that more closely resemble the sounds that they might hear in day-to-day conversation, therefore stimuli based on the voices of real speakers are more appropriate in studies focusing on perceptual differences in multiple social groups among the same speech community.

In addition to natural and synthetic stimuli, McGuire (2010: 11) defines the term *hybrid stimuli* as when naturally produced stimuli undergo modification or resynthesis. The most common experimental design using resynthesized stimuli in speech perception involves the manipulation of stimuli to create a specified step-continuum (e.g. a 11-step continuum from /a/-/e/ is created by increasing the F2 value by a specific amount at regular intervals). During the experiment, listeners are usually asked to decide which sound was played (e.g. /a/ or /e/) after they heard each resynthesized stimulus. This type of methodology is popular in recent socioperceptual studies (Fridland *et al.* 2004; Drager 2010; Kleber *et al.* 2011; Fridland and Kendall 2012; D'Onofrio 2019).

The use of resynthesized speech in continuum-type methods allows for the drawing of perceptual boundaries in phonetic space. For example, if /a/ begins to be perceived as /e/ by a listener when the stimulus continuum hits a certain point on the F1 plane, this gives an indication of where that listener's phoneme boundaries for /a/ and /e/ lie. This type of methodology is beneficial for studies of the perception of vowel contrasts and mergers – as used in Janson and Schulman (1983) – and is only possible with the use of resynthesized stimuli. However, there are complications in using resynthesized stimuli in speech perception research.

Firstly, resynthesisng stimuli is more time consuming than using natural stimuli, requiring knowledge of and training in software which can perform speech resynthesis. Resynthesis of speech is a delicate job, as it frequently involves word list data whereby the segments preceding and following the vowel need to be removed before the vowel is resynthesised. Subsequently, the manipulated vowels are spliced back together with the naturally produced consonants. This can lead to a few problems. Modified stimuli may sound unnatural, and therefore the listener's perceptions in the task may not be comparable to their perception of certain sounds in non-experimental environments. If the sounds present in the stimuli do not sound like they were produced by real voices from the speech community, then the appropriateness of resynthesised stimuli in sociolinguistic research is brought into question. Furthermore, as Labov *et al.* (1991: 50) note, this method might lead to the omission of speech cues found in natural speech which could be

important for the perception of speech. With studies which suggest that dynamic information in the acoustic signal – such as formant transitions – are utilised in the identification of vowels (Jenkins and Strange 1999), resythesizing stimuli risks chopping up and possibly excluding important cues in the perception of the vowels under investigation. As the perception experiment in the current study is a component of the larger sociophonetic study presented throughout this thesis, given the pros and cons in the existing literature, the use of unmanipulated, natural stimuli from real speakers is the most suitable method to use in the perception tasks.

6.3 Stimuli Creation

In order to examine perceptions of the GOAT and THOUGHT vowels in the Tyneside speech community it was first necessary to create stimuli for the discrimination and identification perception experiments. Stimuli were created by recording local speakers reading a list of words and minimal pairs. Audio recordings from the minimal pair lists were used as the stimuli for the discrimination experiment, and audio from both the word and minimal pair list readings were used as stimuli for the identification task.

6.3.1 Participants

Three native North-East English speakers took part in the recording of the stimuli (JC and RB from Tyneside, AR from County Durham). The speakers, recruited via word of mouth at Newcastle University, were two males and one female aged between 23-30 years of age at the time of recording. Aside from being born and raised in the North-East of England, the main criteria for these speakers was that they exhibited monophthongal pronunciations of the GOAT vowel. This is due to Watt's (1998: 161) suggestion that a merger in production between GOAT-THOUGHT may only occur in speakers who have peripheral monophthong [oː] realisations of the GOAT vowel. The results of the production analysis supported this idea, to a certain degree, showing that the speaker group who exhibited the most variety in their use of GOAT variants (i.e. the young male speakers) were found to have the biggest distinction between the GOAT and THOUGHT vowels.

Recordings took place in the Phonetics Laboratory at Newcastle University, and lasted approximately 15 minutes. None of the speakers were aware of the specific purpose of the study, only that they were taking part in an experiment which was interested in accents and dialects of North-East England. Prior to the recording

session, participants were given an information sheet outlining the general nature of the project, and were asked to sign a consent form. Copies of the participant information sheets and consent forms can be found in Appendix A.

6.3.2 Word List

Participants were first asked to read a list of words aloud. This word list contained 40 monosyllabic words, mainly comprising words containing the GOAT, THOUGHT, and NURSE vowels. Within the list, 10 filler words were also included to help disguise which vowels were being investigated. The list contained examples of GOAT, THOUGHT and NURSE vowels in a variety of different phonetic contexts, so as to allow for an exploration of the effects of phonetic environment on the perception of the vowels of interest.

6.3.3 Minimal Pairs

Participants also read a list containing 40 minimal pairs. They were instructed to read both words presented on the screen aloud, with a short pause between the words. Of these minimal pairs, 10 contrasted the GOAT and THOUGHT vowels, 10 contrasted THOUGHT with NURSE, and filler minimal pairs made up the rest of the list. Fortunately, there are quite a number of minimal pairs available that contrast GOAT with THOUGHT. The words in the minimal pair lists were chosen so as to represent the phonetic environments in which GOAT and THOUGHT contrast most often: in open syllables (e.g. *'low-law'*), in pre-/l/ contexts (e.g. *'bowl-ball'*), in pre-nasal environments (e.g. *'bone-born'*), and before voiceless stops, both coronal (e.g. *'boat-bought'*) and non-coronal (e.g. *'folk-fork'*). The word and minimal pair lists used in the stimuli creation process are presented in full in Appendix C.

6.3.4 Procedure

Participants were seated in front of a computer screen in a sound treated booth. Stimuli were presented to participants using OPA (Chen 2007), a text presentation program created for speech production experiments. Before recording began, speakers were instructed to read aloud as clearly as possible the word(s) which appeared on the screen. A brief training segment preceded the experiment to ensure that participants understood the task. Stimuli in these training trials did not contain any GOAT, THOUGHT or NURSE vowels, and the audio of these trials was not used as stimuli in the perception experiment. During the recording, each item from the word and minimal pair lists appeared individually on the screen.

Audio was recorded using a Zoom H4nPro recorder. Participants first read the word list, followed by the minimal pair list. Items from the lists were presented to the speakers in a randomised order, and each word/minimal pair was presented to the speakers twice. In total, each participant read aloud 80 words and 80 minimal pairs.

OPA allows for the display of stimuli on dual monitors, so both participant and researcher are able to see the words on the screen. This is advantageous as it allows the researcher to ensure that the word read aloud matches the word which appears on the screen. Another benefit of OPA is that, to continue with the experiment, the researcher must push a key to first accept the speaker's reading of the word or minimal-pair. As such, during the recording of the stimuli, mispronunciation or interruptions (e.g. due to coughing) were flagged as such, and the affected stimulus was then added to the end of the experiment to be read aloud again.

6.3.5 A Note on NURSE-THOUGHT Stimuli

While the focus of this thesis is on the potential merger between the GOAT and THOUGHT vowels, the NURSE vowel is also included in the word and minimal pair lists. Previous reports on Tyneside English indicate that a merger – or at least a near merger – of the NURSE and THOUGHT vowels existed in the speech community (Wells 1982; Maguire 2007; Beal *et al.* 2012; Watt and Foulkes 2017). However, research by Maguire (2007) and Watt and Milroy (1999) indicates that a full phonemic merger of the NURSE and THOUGHT vowels is largely absent in more contemporary Tyneside speech.

Although recent findings typically point to a lack of NURSE-THOUGHT merging in Tynesider's *productions*, little research has been carried out on the *perception* of this reported merger. It may be presumptuous to conclude that this vowel merger has completely disappeared from contemporary Tyneside English based on evidence from production alone, as this presupposes that production and perception are symmetrical, and change at the same rate. It may be possible that present-day Tyneside listeners taking part in a perception experiment will struggle to discriminate NURSE and THOUGHT words, regardless of whether they produce the vowels identically or distinctly. The primary reason for including NURSE-THOUGHT stimuli in this perception experiment, and for assessing the extent to which listeners are able to distinguish them perceptually, is to examine how this interacts with a potential GOAT-THOUGHT merger. The perception experiment was designed to allow for an investigation into the following questions: (1) do any listeners demonstrate any evidence of a perceptual NURSE-THOUGHT

merger?, and (2) if so, does the presence of a perceptual NURSE-THOUGHT merger prevent a GOAT-THOUGHT merger in perception?

6.3.6 Analysis of Stimuli

Each speaker's GOAT, THOUGHT and NURSE tokens were subject to a brief acoustic analysis. This analysis was conducted to assess how distinct or merged these speakers were in their GOAT-THOUGHT and NURSE-THOUGHT productions. As audio recordings of these speakers would be used as the stimuli for the perception experiment, speakers were required to show some phonetic difference in their pronunciation of these vowels. If the speakers of the stimuli produced the vowels under investigation as phonetically identical, then the listeners in the perception task would struggle to successfully distinguish the vowels.

A Praat script was written to automatically measure the F1 and F2 values of these vowels at their midpoint. While static, one-point measurements may not be optimal for a comprehensive analysis of the Tyneside GOAT vowel (as discussed in further detail in Chapter 5), they do still allow for an overview of the position of GOAT in the vowel space of these North-Eastern speakers. Figures 6.1-6.2 show the measurements for each speaker's target vowels. Speakers JC and AR are male, and speaker RB is female[1].

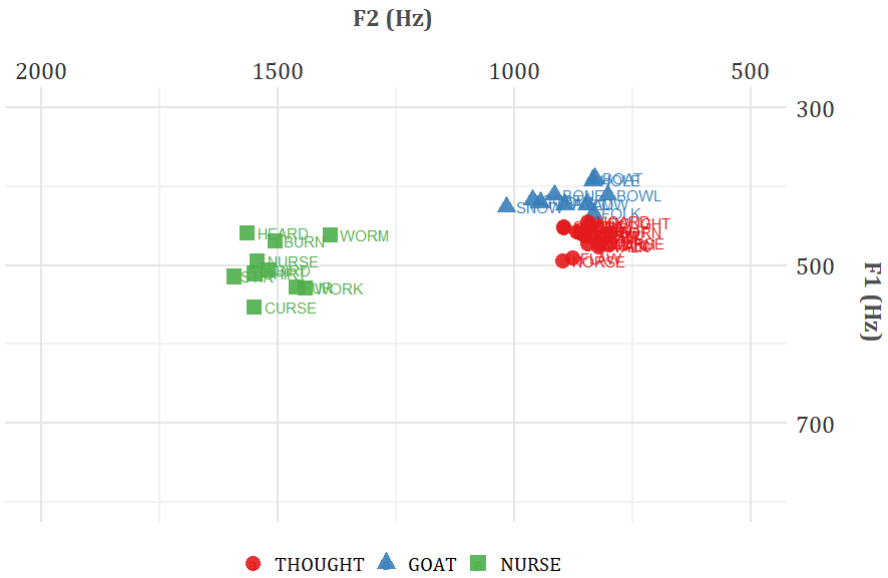


Figure 6.1: Formant values of JC's GOAT, THOUGHT and NURSE tokens

[1]The Praat script measured male speakers using a maximum formant value of 5000Hz, and the female speaker using a maximum formant value of 5500Hz.

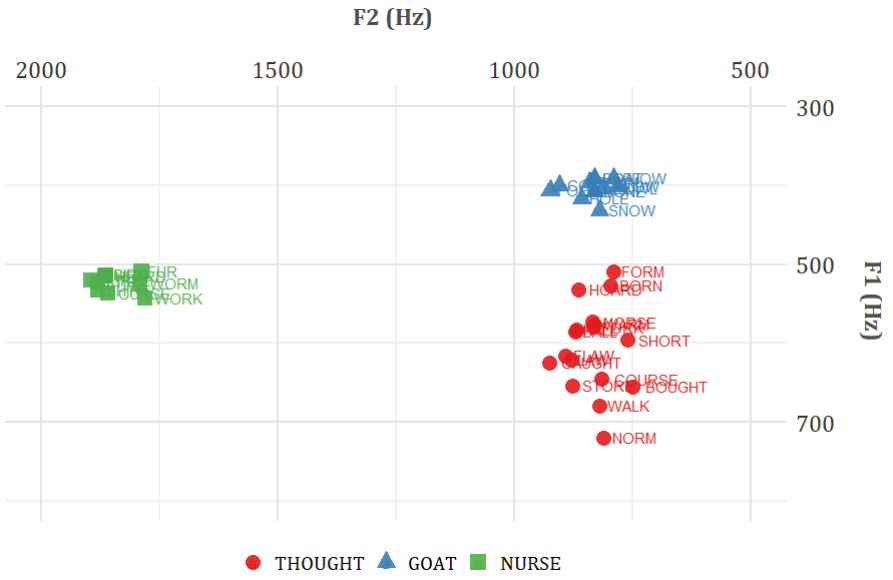


Figure 6.2: Formant values of RB's GOAT, THOUGHT and NURSE tokens

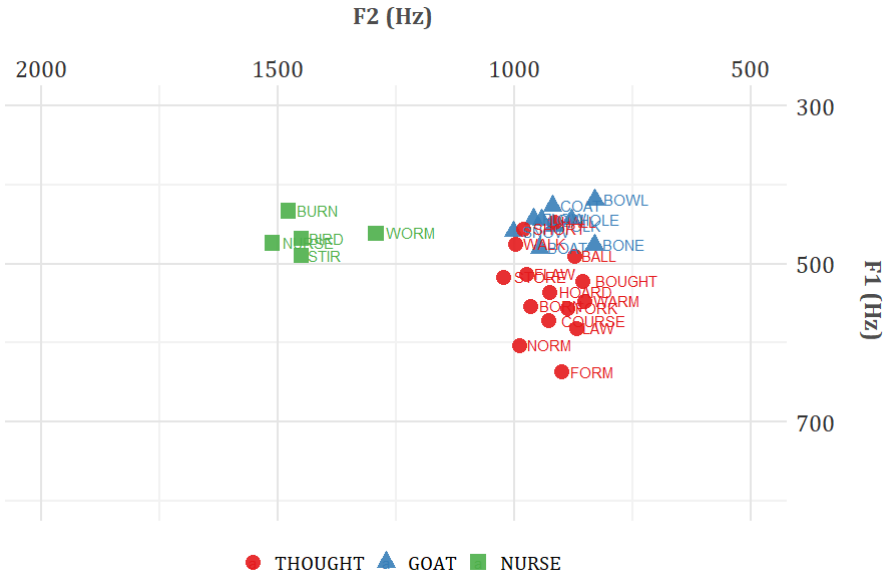


Figure 6.3: Formant values of AR's GOAT, THOUGHT and NURSE tokens

Firstly, Figures 6.1-6.3 highlight that all three speakers maintain a large phonetic distinction between the NURSE and THOUGHT vowels. This result is unsurprising given the findings in recent literature that the merging of NURSE and THOUGHT is no longer a feature of the speech of most Tynesiders (Watt and Milroy 1999). Figure 6.2 also shows that RB, the female speaker, has a more fronted NURSE vowel than both of the male speakers, and, hence, a larger distinction between NURSE and THOUGHT. This is consistent with Watt's (1998: 278) finding that NURSE-fronting can be found in in the speech of young females in Tyneside. It is clear to see that all

three speakers exhibit a much smaller phonetic distinction between the GOAT and THOUGHT vowels than between NURSE and THOUGHT. The figures show that there is little to no difference between GOAT and THOUGHT in terms of F2, suggesting a lack of difference in vowel frontness. Furthermore, the majority of GOAT tokens have a slightly lower F1 value than THOUGHT. This demonstrates that, for these three North-Eastern speakers, GOAT occupies a marginally higher position than THOUGHT in the vowel space.

Figure 6.1. shows that JC has two distinct clouds of measurement for GOAT and THOUGHT tokens. While there is no visible overlap, the two word class clouds occupy a very similar position in the vowel space. This indicates that JC does distinguish GOAT and THOUGHT in production, but, nevertheless, that these two vowels are quite similar phonetically for this speaker.

Speaker RB, as shown in Figure 6.2, also appears to maintain a distinction between GOAT and THOUGHT in production. The two word class distributions are more separated in terms of vowel height than in speaker JC, with RB exhibiting a difference of around 100Hz or more in F1 between GOAT and THOUGHT. Therefore, it seems that RB does not exhibit a phonetic GOAT-THOUGHT overlap in production. Notably, there is a lot more variation in RB's THOUGHT tokens than we see in JC's vowel plot. This could indicate that RB's THOUGHT vowel is undergoing change, perhaps raising to become closer to GOAT.

Looking at Figure 6.3, speaker AR appears to show some slight overlap between the distributions of the GOAT and THOUGHT vowels. However, the majority of AR's THOUGHT vowels occupy a lower position than GOAT in the vowel space, and, auditorily, this speaker can be heard to make a clear distinction between these two vowels. AR's GOAT vowel appears to be in a slightly lower position when compared with the other two speakers, though, in line with JC and RB, the majority of THOUGHT tokens here have higher F1 values than GOAT. As with RB, speaker AR shows much more variation in the height of THOUGHT than GOAT, which may further suggest that THOUGHT is undergoing change in the speech community.

Overall, while GOAT and THOUGHT appear to be phonetically close for these speakers, this brief analysis does indicate that these speakers largely make a distinction between these vowels in production. Thus, the word and minimal pair lists recorded by these speakers are suitable to be used as stimuli in the perception tasks.

6.4 Participants

Participants for the perception experiment were recruited via a university mailing list. Criteria for recruiting listeners was minimal, asking only (a) that the participant report no hearing problems, and (b) that the participant was born and/or raised in the Tyneside area, and has lived in the region for the majority of their life. In total, 43 listeners (19 females, 24 males) between the ages of 18-69 volunteered to take part in the perception experiment. Table 6.1 below displays the biographical details of all the listeners who took part in the perception experiments. As in the production experiment, all participants were anonymised and given pseudonyms, and were divided into three age groups: young (18-29yrs), middle (30-59yrs), and older (60yrs and over)

Name	**Sex**	**Age**	**Regionality**	**Name**	**Sex**	**Age**	**Regionality**
Henry	M	18	North Tyneside	Paul	M	25	North Tyneside
Abigail	F	19	South Tyneside	Alex	M	26	Tyneside
Emma	F	19	South Tyneside	John	M	26	South Tyneside
Joel	M	19	Northumberland	Chris	M	28	North Tyneside
Laura	F	19	North Tyneside	Hannah	F	33	Tyneside
Roy	M	19	South Tyneside	Lucy	F	33	Tyneside
Charlie	M	20	Tynesdie	Lisa	F	34	Tyneside
Clark	M	20	North Tyneside	Sally	F	34	North Tyneside
David	M	20	South Tyneside	Zara	F	34	Tyneside
Jennifer	F	20	Tyneside	Sophia	F	35	Tyneside
Natalie	F	20	South Tyneside	Allan	M	39	Tyneside
Oscar	M	20	Northumberland	Dan	M	44	Tyneside
Rhys	M	20	North Tyneside	Joseph	M	48	North Tyneside
Hayley	F	21	South Tyneside	Harriet	F	51	South Tyneside
Josh	M	21	Tyneside	Jackie	F	52	South Tyneside
Rose	F	22	South Tyneside	Amanda	F	53	South Tyneside
Simon	M	23	Tyneside	Kevin	M	56	Northumberland
Justin	M	24	Tyneside	James	M	57	Tyneside
Kaleigh	F	24	Tyneside	Ross	M	64	North Tyneside
Kathryn	F	24	Northumberland	Anthony	M	68	South Tyneside
Rob	M	24	South Tyneside	Monica	F	69	South Tyneside
Steven	M	24	South Tyneside				

Table 6.1: Biographical information of listeners in perception experiment

Figure 6.4 below shows a map of listener birthplaces. As with the speakers in the production analysis of the current study, most participants are from the areas of Tyneside, North Tyneside and South Tyneside, along with a couple of listeners from the Northumberland area. The criteria for participant inclusion/exclusion based on their regionality was the same as in the production analysis, as described in Chapter 4.2.

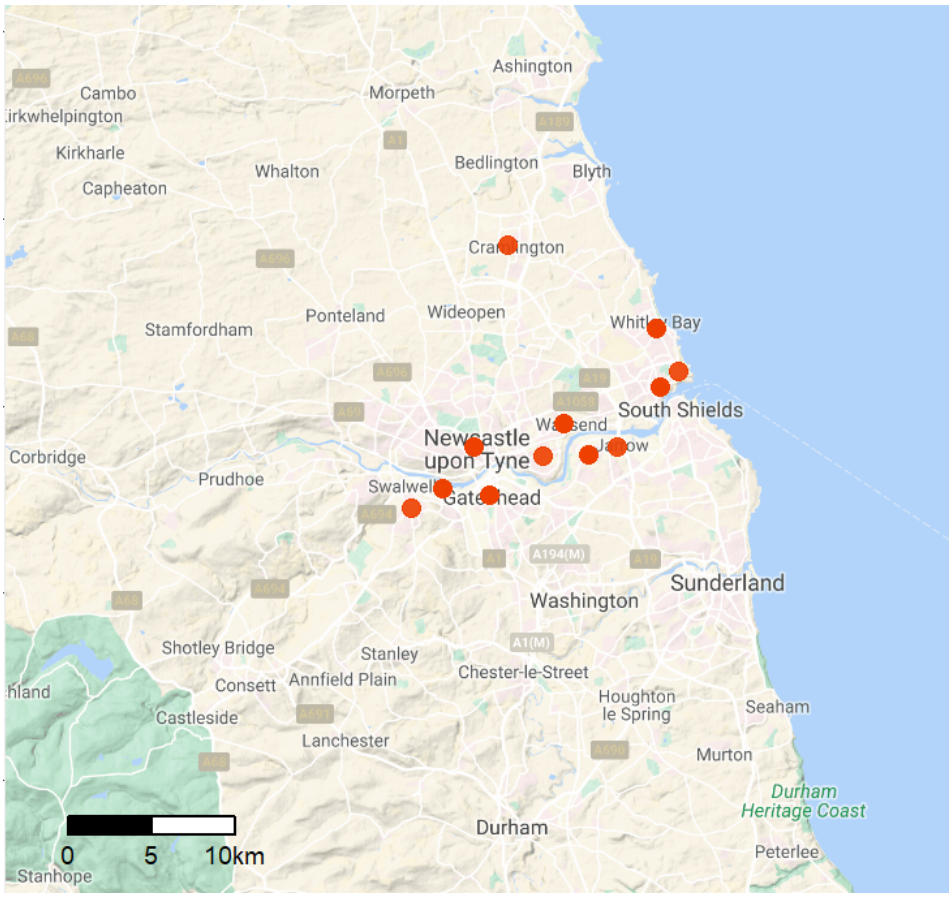


Figure 6.4: Map of Listener Locations

6.5 Procedure

The perception experiment took place in a quiet room in a building on Newcastle University's campus. While advertised as a listening task surrounding the themes of variation and change in the Tyneside accent, participants were unaware of the specific nature of the experiment. Prior to the experiment, participants were asked to sign a consent form, and to complete a short questionnaire involving questions about their biographical information and regional background. Listeners completed two tasks in the experiment session, first a discrimination task and then an identification task. Further details of these tasks are given below in sections 6.5.1 and 6.5.2.

The tasks were presented to listeners on a touchscreen Microsoft Surface Pro 4 computer using the experimental program OpenSesame (v.3.1) (Mathôt *et al.* 2012). Audio stimuli were played through a pair of Audio-Technica ATH-M50 headphones. Each of the two tasks took listeners approximately 15-20 minutes to complete, with experiment sessions lasting around a total of 40 minutes.

6.5.1 Discrimination Task

The discrimination experiment used in the present study is a variant of the same-different task with elements of an oddity task, similar to those carried out in Brunellière *et al.* (2009) and Dufor *et al.* (2013). Each trial of the discrimination task was made up of three stimuli, each separated by a pause of 1000 milliseconds. Listeners heard three different speakers producing a word: the first two words, spoken by two different male speakers, were phonemically identical, while the final word, spoken by a female speaker, would either match or differ from the previous two. After listening to the sequence of three stimuli, listeners were asked whether the third word that they heard (i.e. the word produced by the female speaker) was the 'same' or 'different' to the previous two words.

The minimal pairs used in the discrimination task are shown below in Table 6.2. Five of the pairs contrasted the GOAT and THOUGHT vowels, and five contrasted the NURSE and THOUGHT vowels. Four control pairs that the Tyneside listeners were expected to be able to easily discriminate were also included: two NEAR-SQUARE pairs and two FOOT-GOOSE pairs. These fillers were included in order to mask the main purpose of the study, and to act as a point of comparison for the vowel distinctions under investigation.

GOAT-THOUGHT	NURSE-THOUGHT	Control Pairs
boat-bought	burn-born	beer-bear
bone-born	curse-course	chair-cheer
foam-form	heard-hoard	full-fool
hole-hall	shirt-short	sun-soon
low-law	work-walk	-

Table 6.2: Minimal pairs used in the discrimination experiment

For each of the minimal pairs used in this task, four different trial sequences were possible: two 'same' sequences (e.g. *'boat', 'boat', 'boat'* and *'bought', 'bought', 'bought'*), and two 'different' sequences (e.g. *'boat', 'boat', 'bought'* and *'bought', 'bought', 'boat'*). In total, 56 trials were created, 20 testing

GOAT-THOUGHT discrimination, 20 testing NURSE-THOUGHT discrimination, along with 16 control trials. Each trial was played twice during the experiment, producing a total of 112 trials. All trials were presented in a randomised order, which differed from listener to listener.

During this task, participants were then seated in front of a touch-screen computer and asked to put on a pair of headphones. Before the experiment began, participants were presented with on-screen instructions explaining the task ahead. The task would only begin once listeners confirmed via a touch-screen response that they understood the nature of the experiment. Participants were also instructed to respond to each trial as quickly and as accurately as possible. A short training task was completed by the listeners before the main task. This was included to ensure that listeners understood how they should respond to the stimuli presented to them. The training exercise consisted of 5 trials, and contained only control stimuli.

Listeners were shown a screen after each audio stimulus. This, as shown in Figure 6.5, reminded them of the question they were answering, i.e. 'was the third word the same or different to the previous two words'. Participants were also presented with two boxes on the screen labelled as *'same'* and *'different'*, and were instructed to tap on the box which they thought corresponded to the correct answer. The responses of the listeners were recorded automatically by OpenSesame.

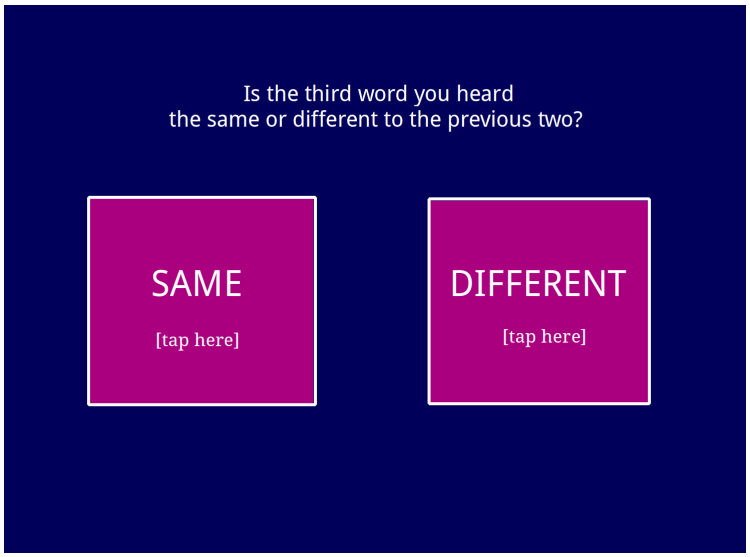


Figure 6.5: Response screen for discrimination experiment

6.5.2 Identification Task

In the identification task, listeners were presented with one-word audio stimuli and subsequently asked to identify the word that they had just heard. The stimuli for this task were taken from the word and minimal pair list recordings from one male (JC) and one female speaker (RB). As shown in Table 6.3, five GOAT-THOUGHT, five NURSE-THOUGHT and four control minimal pairs were used in the identification experiment. The use of the same stimuli in both the discrimination and identification tasks was avoided so as to try and prevent either (a) loss of attention in the second task leading to poorer performance, or (b) increased performance on the second task due to the over-repetition of certain words. Of the total of 112 trials, 40 were GOAT-THOUGHT stimuli, 40 were NURSE-THOUGHT stimuli, and the remaining 32 trials were fillers. As in the previous task, the selected filler pairs should be easily discriminable for the Tyneside listeners. Each word from every minimal pair was heard 4 times, twice spoken by the male speaker (JC), and twice spoken by the female speaker (RB). The order of the trials in the identification experiment was randomised.

GOAT-THOUGHT	NURSE-THOUGHT	Control Pairs
coat-caught	fur-for	bad-bed
bowl-ball	bird-board	had-head
folk-fork	nurse-Norse	pin-pen
flow-flaw	stir-store	pit-pet
snow-snore	worm-warm	-

Table 6.3: Minimal pairs used in the identification experiment

During the identification experiment, participants were again presented with two boxes on a touchscreen computer. This time, the boxes displayed either word A or word B of a word pair. An example trial is shown in Figure 6.6. On each trial listeners heard audio of one word being spoken, for example the word *'boat'*, and were then asked to select by tapping on one of the boxes whether the word they heard was *'boat'* or *'bought'*. Each word was played to the listeners 4 times; twice corresponding to the word A box and twice corresponding to the word B box. Again, a short training task comprising 5 control stimuli trials preceded the main experiment to ensure that listeners understood the objective of the task.

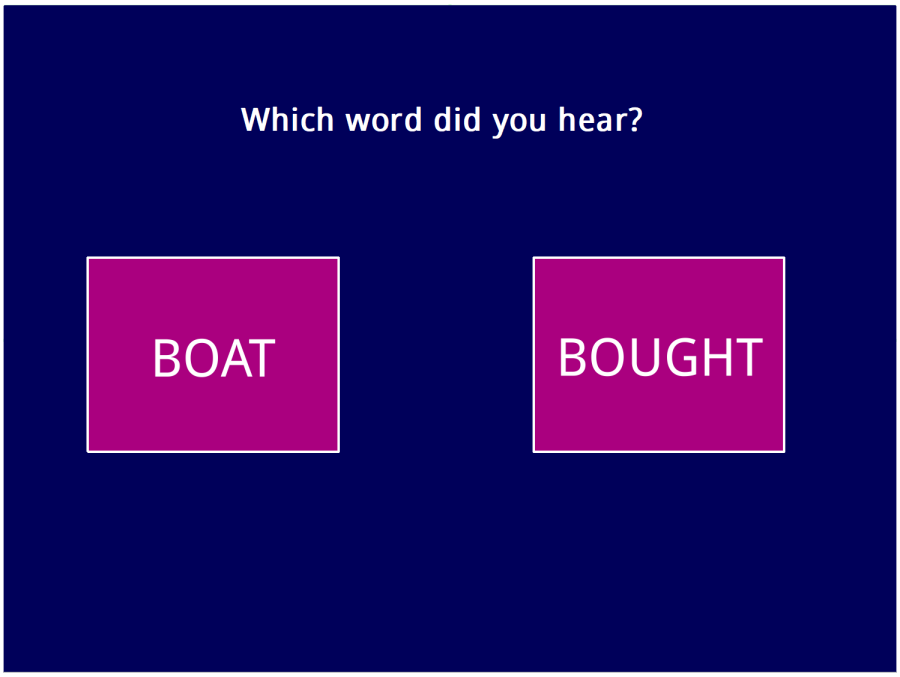


Figure 6.6: Response screen for identification experiment

6.6 Analysis of Results

To investigate whether the results of the perception experiment indicate a phonemic merging of the GOAT and THOUGHT vowels in Tyneside English, the data collected from the discrimination and identification tasks are analysed in terms of accuracy, sensitivity, and reaction time.

6.6.1 Accuracy

Given the types of discrimination and identification tasks used in this perception experiment, one method of quantifying the results is to assess how accurately each listener performed in each task. Accuracy, or percentage correct, is a frequently used analysis measure for same-different (or AX) and labelling tasks in psychology and linguistic research (Dixon 2008; McGuire 2010). In the following analysis, listener accuracy is defined as the proportion of correct responses that a particular listener gave in each task, and was calculated by dividing the number of correct responses by the total number of relevant trials. A high percentage of correct responses indicates that a listener performed well in the task, while a low percentage of correct responses suggests, in the tasks carried out in this perception experiment, that listeners had some difficulty in discriminating or identifying certain vowels. Throughout the analysis of the perception tasks, the terms 'correct' and 'incorrect' response are used. These terms refer specifically to

whether the participant's given response matched the expected response for a particular trial, and are not intended as a commentary on the 'correctness' of listeners' ability to discriminate certain sounds.

In terms of statistical analysis, logistic regression is reported to be suitable for modelling accuracy in binomial data (Dixon 2008; McGuire 2010). For the present study, a statistical analysis of the accuracy of listeners' perceptions of vowel pairs is conducted using mixed-effects logistic regression models. These are run using the `glmer()` function from the lme4 R package (Bates *et al.* 2015). As the dependent variable in these models is *response* (0 = incorrect, 1 = correct), it is specified in the model that the data follows a binomial distribution. Selection of the optimal model for accuracy followed a similar procedure to that outlined in 4.7.1.1: models are compared by their Akaike Information Criterion (AIC) values and by ANOVA comparisons, with a preference for model parsimony.

6.6.2 Sensitivity: *d'*

Measures of sensitivity can be used on data from discrimination experiments to provide additional assessments of a participant's performance. Analysis of accuracy using percentage of correct responses gives an insight into how accurately a listener can distinguish stimuli, but it does not account for any listener response bias (e.g. participants who may favour 'same' responses to 'different' responses, or vice versa). Using the results of the discrimination experiment performed by the Tyneside listeners, sensitivity measures will allow for an examination of how well listeners are able to differentiate two vowels, while taking into account any individual listener biases in response choices which could affect performance in the task.

To calculate sensitivity, a *stimulus-response matrix* must first be created by sorting the trials into four categories. As shown below in Figure 6.7, for the discrimination experiment in the current study, there are four possible permutations of stimulus-response: *Hit* (a correctly selected 'same' trial), *False Alarm* (a trial that is incorrectly selected as 'different'), *Miss* (a trial that is incorrectly selected as 'same'), and finally, *Correct Rejection* (a correctly selected 'different' trial). Then, for each participant, the *Hit Rate* (the proportion of 'same' responses in to 'same' stimuli) and the *False Alarm Rate* (proportion of 'same' responses to 'different' stimuli) is calculated. As Macmillan and Creelman (2005: 5) explain, a participant who performs perfectly on a discrimination task, and is therefore a 'perfectly sensitive' listener, would have a *Hit Rate* of 1 and a *False Alarm Rate* of 0.

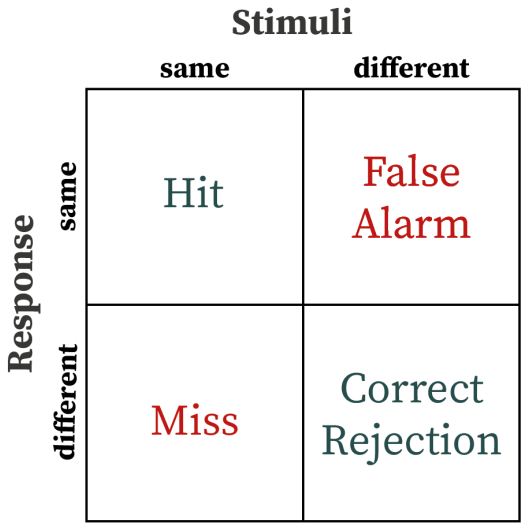


Figure 6.7: Stimulus-Response Matrix

Macmillan and Creelman (2005: 7) state that *d'* (sometimes written as d-prime) is the 'most widely used' measure of sensitivity, and there are several examples of its use in speech perception research (Sussman and Lauckner-Morano 1995; De Jong and Hao 2017). *d'* is calculated by Z-transforming the values of *Hit Rate* and *False Alarm Rate*, and subtracting the latter from the former:

$$d' = \text{z}(Hit\ Rate) - \text{z}(False\ Alarm\ Rate)$$

A higher *d'* value indicates increased sensitivity. Therefore, in the present study, listeners who exhibit the highest *d'* value are more successful in discriminating between certain vowel pairs (e.g. GOAT-THOUGHT, NURSE-THOUGHT, or control pairs). *d'* sensitivity values were calculated for each of three vowel pairs for each participant using the `dprime()` function from the 'psycho' r package (Makowski 2018).

For a statistical analysis of sensitivity, one-factor ANOVA tests may be run on the data to test whether a certain predictor significantly affects *d'*. This is the method used by Bradlow *et al.* (1999) to determine whether speaker, speech rate, and amplitude had a significant effect on the listener's ability to recognise new and repeated lexical items in a memory task. The authors then conducted further post-hoc statistical comparisons to assess the differences in the effects of these factors on *d'*. The following analysis of sensitivity in the discrimination experiment will take a similar approach, first performing a one-factor ANOVA to determine whether the vowel pair presented in a particular audio stimulus (e.g. whether GOAT-THOUGHT, NURSE-THOUGHT, or control vowels) has a significant effect on *d'*. This will indicate whether listeners are more sensitive to the differences between certain vowel pairs.

If vowel pair is found to effect *d'* significantly, then a Tukey's Honest Significant Difference comparison (Tukey's HSD) will be conducted, using the `TukeyHSD()` function in R, in order to compare the differences in *d'* values for each vowel pair. The results from these statistical tests will indicate whether listeners are less accurate in their responses to GOAT-THOUGHT trials than to trials containing other vowel pairs.

6.6.3 Reaction Time

McGuire (2010) notes that reaction time (or response time) is another popular method of analysing speech perception data collected from same-different and identification experiments. Slow reaction times can indicate that a listener has some perceptual difficulty in discriminating between or identifying certain sounds. Reaction time data does not usually conform to a normal distribution, rather it is often right-skewed (or positive-skewed) in its distribution. To reduce this skew, reaction time data in the present study was log-transformed in R using the `log10()` function prior to analysis.

In order to ensure that reaction time results are comparable across different participants, all listeners completed the perception tasks on the same laptop computer. This helps to eliminate the possibility of different devices having an influence on reaction time results, such as OpenSesame running slower on certain computers, or participants responding more quickly when using a particular device. Furthermore, while both the discrimination and identification experiments contain only monosyllabic words as stimuli, listeners heard stimuli from three different speakers. This means that, although the stimuli used in the experiment should not differ greatly in terms of their duration, it is possible that different speech rates lead to words from one speaker being shorter or longer in duration to that of another speaker. In light of this, the reaction time measurement in OpenSesame was set to begin immediately after the stimulus had played, therefore removing the influence of durational differences in words – due to different speakers, or varying phonetic contexts – on reaction time results.

Statistical analysis of the reaction time data takes the form of mixed-effects linear regression models, with the log-transformed reaction time data as the dependent variable. These models, which were also used in the analysis of formant values of GOAT and THOUGHT in the production data (as outlined in Section 4.7), are run using the `lmer()` function from the lme4 R package (Bates *et al.* 2015). As before, the optimal model is selected by comparing the AIC values of the models, and by an ANOVA model comparison.

6.7 Production Evidence

The initial intention of the main study was to only compare the results from the perception tasks with the production data collected from DECTE. This would allow for a comparison of the production and perception of the GOAT and THOUGHT vowels in the Tyneside speech community. However, as the same participants did not take part in both the production and perception components of the main study, the study design would not allow for a comparison between the production and perception of these two vowels at the individual level. Midway through running the perception tasks, it was decided that it would be beneficial to collect production data for a subset of the listeners taking part in the perception tasks. This would allow for a direct comparison of how each participant produces and perceives the GOAT and THOUGHT vowels. Of the 43 participants who completed the perception tasks, 24 were also recorded reading a list of minimal pairs aloud.

6.7.1 Procedure

As with the recording of the stimuli for the perception tasks, the minimal pair recordings were conducted using the OPA (Chen 2007) text presentation software, and carried out in a sound treated room in the Phonetics Laboratory at Newcastle University, using a Zoom H4nPro recorder. The minimal pair list consisted of the same monosyllabic word pairs used as stimuli in the perception tasks. In an effort to combat any effects of task order (e.g. improved performance in the perception tasks due to prior reading of minimal pairs, or, hyperawareness of speech production in reading task following the perception tasks), 12 participants completed the reading task first while the remaining 12 completed the perception tasks first.

As with the analysis of the DECTE interviews outlined in Chapter 4, the recorded minimal pair lists were first time-aligned in ELAN (ELAN. 2018). The Forced Alignment and Vowel Extraction (FAVE) (Rosenfelder *et al.* 2011) suite was subsequently used to automatically align the sound and transcription files, and to extract vowel formant values. Again, the FAVE defaults for measuring GOAT and THOUGHT vowels were used. Lobanov normalised F1-F2 values are used during formant analysis.

The analysis of the DECTE sociolinguistic interviews, presented in Chapter 4, utilised a number of different methods and measurement types to examine the status of the GOAT and THOUGHT vowels in contemporary Tyneside speech. However, an investigation of the production data collected during the perception tasks has a different goal. The recordings come from minimal pairs rather than sociolinguistic

interviews. As this is a more careful speech style, the production data discussed in this and the following chapter were not elicited with the aim of capturing or studying variation in everyday speech in the Tyneside speech community. Rather, the minimal pair reading task aimed to study participants' phonetic distinction between GOAT and THOUGHT in a more controlled environment (i.e. by pre-selecting phonetic contexts, repeating words etc.). As the main aim of the minimal pair data is to look generally for any overlap between GOAT and THOUGHT in the vowel space of these participants, and to compare this to the perception results, this can be achieved with static, one-point formant measures.

6.7.2 Analysis

An analysis of the realisations of GOAT and THOUGHT, as produced by the participants taking part in the perception experiments, primarily sets out to compare the difference – or lack thereof – between these two vowels. Collecting both production and perception data from the same participants is advantageous as it allows for an investigation into how closely their production of GOAT and THOUGHT reflects their perception of the same vowels. For this comparison, the main information required from a production perspective is an analysis of the distance and overlap between GOAT and THOUGHT in the vowel space. This can then be compared with the same participant's performance on the discrimination and identification tasks to look for evidence of a GOAT-THOUGHT merger in both production and perception.

6.8 Summary

The methodology of the perception study of this thesis has been laid out in this chapter. This chapter has covered the predominant components of the methodology, including stimuli creation, participant recruitment, experiment procedure, and analysis measures. The following chapter will present the results of the perception tasks, examining whether there is any evidence for a phonemic merger of the GOAT and THOUGHT vowels in the Tyneside speech community.

Chapter 7

Perception: Results

The goal of this chapter is to assess whether there is evidence of a perceptual merger of the GOAT and THOUGHT vowels in Tyneside English. Analysis of the two perception tasks, a discrimination and an identification task, is presented in this chapter. First, the results of the discrimination task are examined. Sections 7.1.1 to 7.1.3 give a discussion of listeners' accuracy in correctly discriminating between GOAT and THOUGHT, using mixed-effects logistic regression models for statistical analysis of accuracy. Section 7.1.4 then focuses on sensitivity, as *d'* values are used to assess accuracy while taking into account any listener response biases. Next, Section 7.1.5 investigates listener reaction times in the discrimination task, and 7.1.6 shows the output of mixed-effects linear regression models which were used to model the reaction time data. Following a similar structure, the results from the identification experiment are then discussed, with Sections 7.2.1 to 7.2.2 exploring listener accuracy, and Sections 7.2.3 to 7.2.4 focusing on reaction time. Finally, this chapter presents a brief comparison of participants' performance in the perception tasks with production data obtained from minimal pair readings in Section 7.5.

7.1 Discrimination Task

7.1.1 Overall Accuracy

Each listener's overall accuracy (i.e. proportion correct) in the discrimination task is displayed in Figure 7.1 below. The percentage of correct responses given by each participant is shown on the y-axis, and the x-axis shows the listeners, sorted by overall performance (l-r lowest to highest accuracy). As emphasised in the previous chapter, the terms 'correct' and 'incorrect' response refer specifically to whether the

listener's given response matched the expected response to a specific trial (e.g. on a trial which presents three phonemically identical sounds, the expected response would be that the listener selects that the sounds they heard were the 'same').

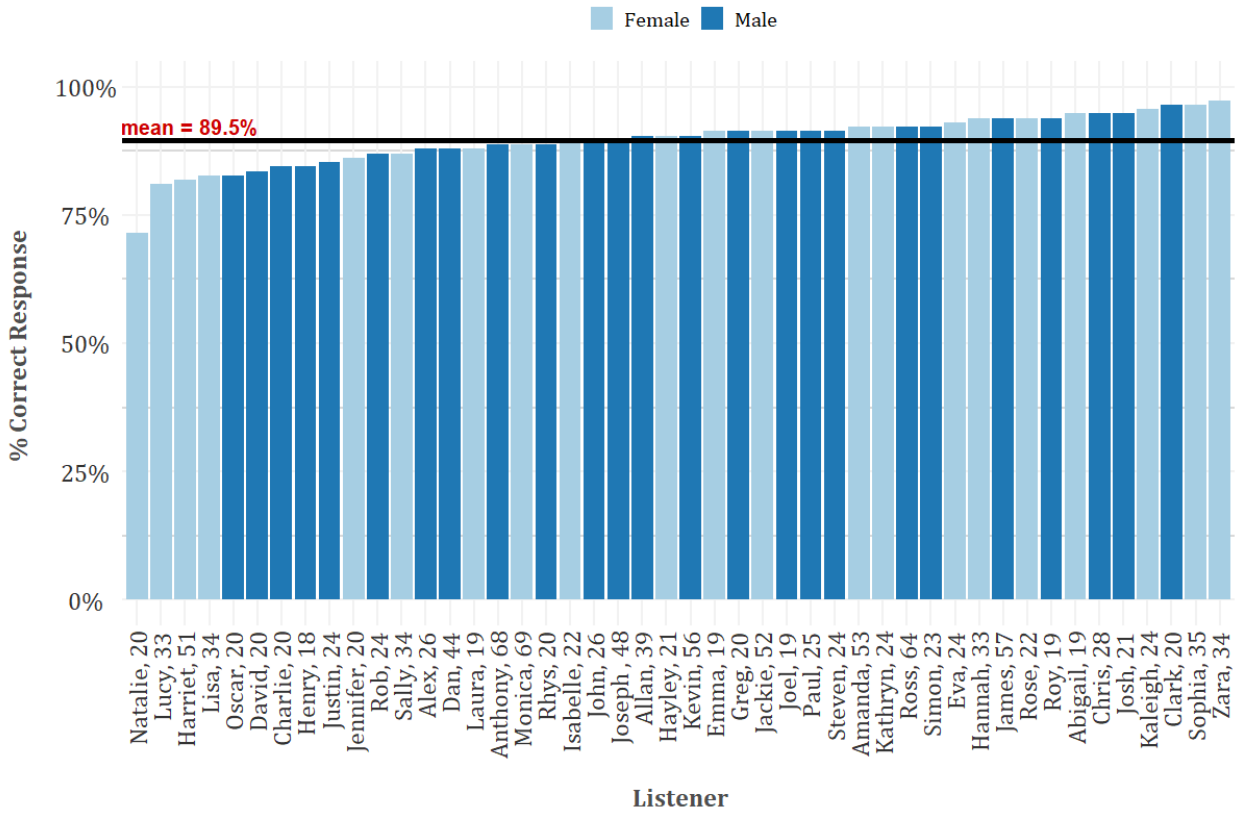


Figure 7.1: Overall accuracy percentage of each listener

Looking at Figure 7.1, the first thing to note is that the overall accuracy percentages are quite high for the majority of listeners. The mean accuracy rate is 89.5%, as represented by the black horizontal line in the figure. As the lowest accuracy percentage is 71.5% and the highest accuracy percentage is 97.4%, this indicates that no listener performed very poorly in the discrimination task overall. It can therefore be assumed that none of the listeners misunderstood the task, or had any other issues in completing the experiment. Figure 7.1 does not appear to show any obvious pattern between accuracy and either of the social factors of age or sex. This is somewhat unsurprising, however, as this figure shows the correctness percentage for all trials in the discrimination task (GOAT-THOUGHT, NURSE-THOUGHT, and fillers), and most participants performed quite well in the overall task. Notably, female listeners aged between 20-35 years of age occupy both ends of the overall accuracy spectrum, exemplified by the two least accurate listeners (Natalie and Lucy) and the two most accurate listeners (Zara and Sophia). This immediately suggests a wide range of variation in the performance

of young-middle aged females in the discrimination task.

7.1.2 GOAT-THOUGHT Accuracy

Figure 7.2 shows how well each each listener performed specifically on the GOAT-THOUGHT trials in the discrimination experiment. The mean accuracy rate for the GOAT-THOUGHT trials is 76.5%, which is considerably lower than the mean of the overall accuracy rate. This immediately suggests that participants experienced more difficulty in deciding whether an audio stimulus belonged to the GOAT or THOUGHT lexical set than when discriminating between other vowel pairs. As shown in Figure 7.2, there is substantial variation between the participants, with accuracy rates ranging from 45% to 95%. Such variation between the listeners, coupled with the result that two of the listeners (Natalie and Lucy) have accuracy rates of 50% and below, indicates that some of these Tyneside listeners struggle to distinguish the GOAT and THOUGHT vowels perceptually.

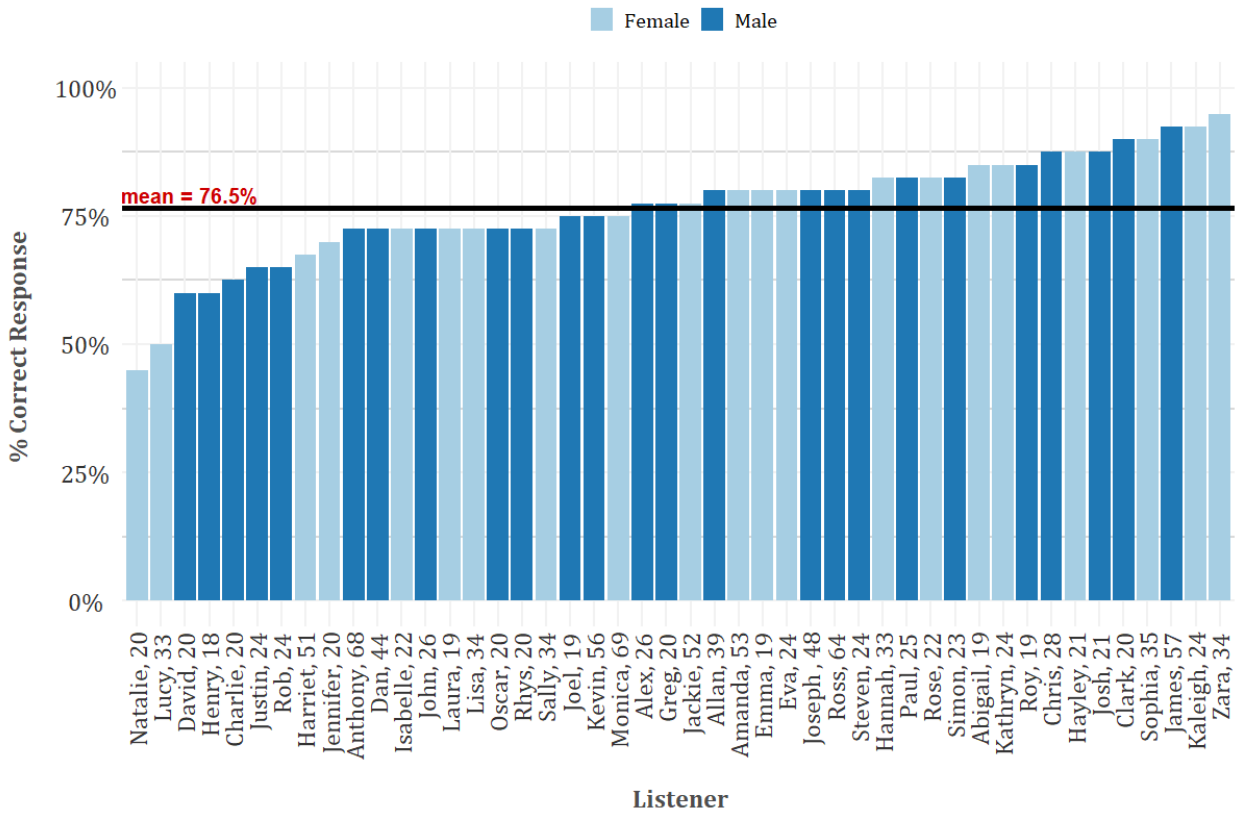


Figure 7.2: Accuracy rates for GOAT-THOUGHT discrimination

Table 7.1 below shows the number of times listeners incorrectly responded to the stimuli pairs which contained the GOAT and THOUGHT vowels. The table shows that participants were most accurate in discriminating the GOAT and THOUGHT vowels

when present in the context of the words *'low'* and *'law'*. Conversely, listeners performed worse on trials including the words *'foam'* and *'form'* as stimuli, as these trials make up 27.4% of incorrect responses to GOAT-THOUGHT stimuli.

Stimuli Pair	n	%
foam-form	118	27.4%
bone-born	96	22.3%
hole-hall	78	18.1%
boat-bought	76	17.7%
low-law	62	14.4%

Table 7.1: Incorrect Responses to GOAT-THOUGHT Stimuli

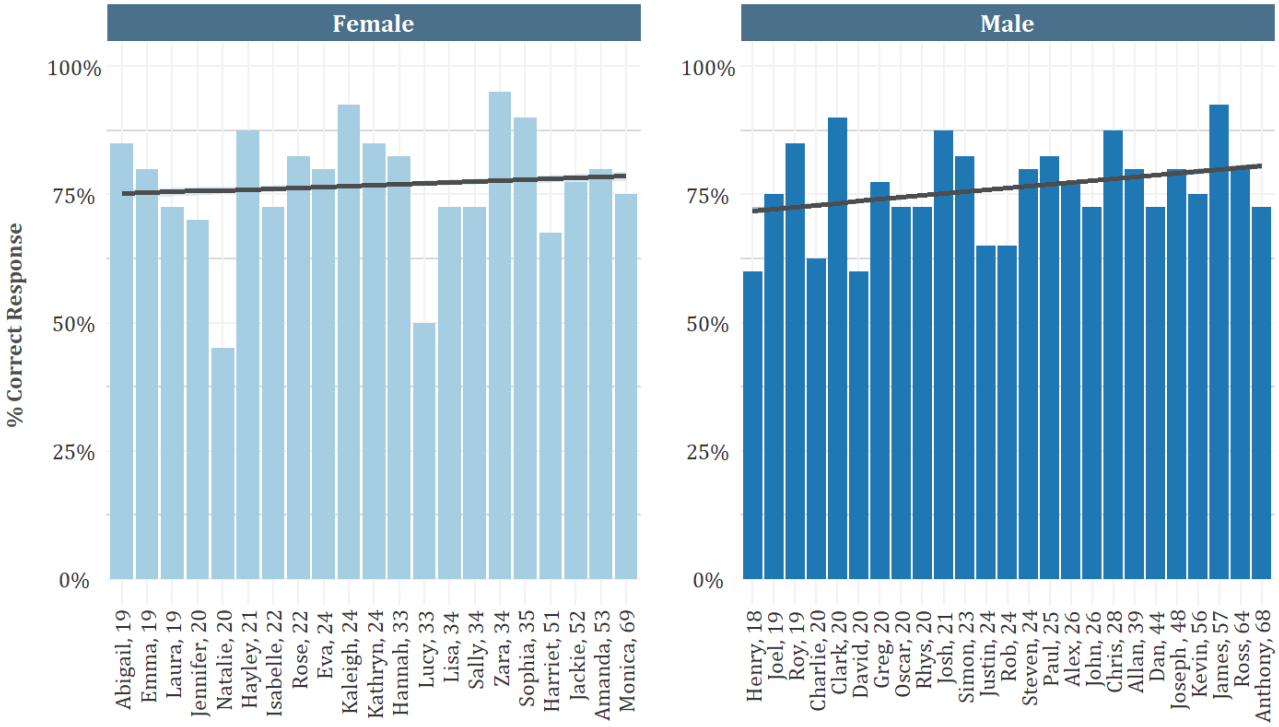


Figure 7.3: GOAT-THOUGHT discrimination by sex and age
(L-R youngest to oldest listener)

Figure 7.3 presents a comparison of the accuracy rates on the GOAT-THOUGHT trials of the discrimination experiment for female and male listeners. Firstly, the plots for the females and males are not too dissimilar, with both showing fluctuation in accuracy rates across the age range. Additionally, the figure highlights the variation found within the listener groups. This is particularly noticeable when looking at the female listeners aged between 33-35, who display

some of the lowest (Lucy) as well as some of the highest (Zara, Sophia) accuracy rates in the whole sample. Figure 7.3 also includes trend lines which do suggest some marginal effect of age on accuracy rate. Both the female and male plots show trend lines which indicate that older listeners typically show marginally higher accuracy rates, and thus performed better in the discrimination experiment. This factor of age impacting the ability to discriminate the GOAT and THOUGHT vowels appears to have a slighty more pronounced effect on the male rather than the female listeners.

7.1.3 Modelling Accuracy

7.1.3.1 Overall Accuracy

To compare listener accuracy in the GOAT-THOUGHT trials to trials containing other vowel pairs, a mixed-effects logistic regression model was run on the data from the discrimination experiment. Whether a listener's response on a particular trial was correct or incorrect (i.e. 1 or 0) was taken as the dependent variable, and the model included vowel pair as a predictor (with 3 levels for GOAT-THOUGHT, NURSE-THOUGHT, and control pairs). Listener and word pair (the pair of words being compared on a given trial e.g. *'boat-bought'*, *'short-shirt'*) were also added into the model as random effects. The output of the regression model is shown below in Table 7.2. Statistical significance is indicated by a p-value of <0.05.

fixed effects	**Estimate**	**St. Error**	**z value**	**Pr(>|z|)**
(intercept)	3.843	0.262	14.667	<0.001
vowel pair: GOAT-THOUGHT	-2.570	0.307	-8.383	<0.001
vowel pair: NURSE-THOUGHT	-0.605	0.327	-1.853	0.064
random effects	**Variance**	**St. Deviation**		
listener	0.259	0.509		
word pair	0.143	0.378		

Table 7.2: Response ~ Vowel Pair + (1| Listener) + (1| Word Pair) Model

As displayed in Table 7.2, the model indicates that listener accuracy differs significantly with respect to vowel pair. The estimates and p-values show that participants have increased accuracy when responding to NURSE-THOUGHT and control trials than when responding to GOAT-THOUGHT trials. This suggests that, for this group of Tyneside listeners, perceptual discrimination of the GOAT and THOUGHT vowels proved more difficult than distinguishing any of the other vowel

pairs included in the experiment. Notably, the model output also points to lower accuracy in discriminating NURSE-THOUGHT vowel pairs than control vowel pairs, though this is not found to be a statistically significant difference. This is discussed in further detail in Section 7.4.

Following the logistic regression model, the probability of responding correctly to a trial depending on the vowel pair presented was calculated. This was done using the `predict()` function in R, which takes the output of a regression model to produce a probability value for any given variable. Overall, trials which presented GOAT-THOUGHT stimuli had a 78% chance of being discriminated correctly, while the chances of correctly distinguishing NURSE-THOUGHT trials and trials containing control vowel pairs were 96% and 98% respectively. Again, this suggests that listeners performed worst when trying to differentiate GOAT from THOUGHT. Although the participants also seemed to have a slightly higher chance of correctly discriminating control trials than those containing NURSE-THOUGHT stimuli, these two values differ very little in comparison with the markedly lower likelihood of a correct response found for the GOAT-THOUGHT trials. Such a result provides further evidence to indicate that there is some degree of perceptual overlap between the GOAT and THOUGHT vowels for these Tyneside listeners.

7.1.3.2 GOAT-THOUGHT Accuracy

Further mixed-effects logistic regression models were run on the data to examine accuracy in distinguishing GOAT and THOUGHT for the different listener groups. Table 7.3 shows two different models that were tested, along with their AIC values. As before, the model includes listener and word pair as random effects. The table shows that adding an interaction between the predictors of age group and sex increases the AIC value, and an ANOVA model comparison confirms that adding this interaction term does not significantly improve the model ($p = 0.07$). Therefore, the selected model is that which includes the predictors of listener age group and sex without any interaction.

model	AIC
Response ~ sex + age group + (1| listener) + (1| word pair)	1970
Response ~ sex * age group + (1| listener) + (1| word pair)	1974

Table 7.3: Comparison of logistic regression models for accuracy in discrimination of GOAT and THOUGHT

Table 7.4 below presents the output of the optimal logistic regression model of

GOAT-THOUGHT accuracy.

fixed effects	**Estimate**	**St. Error**	**z value**	**Pr(>|z|)**
(intercept)	1.254	0.205	6.10	<0.001
sex: male	-0.034	0.185	-0.19	0.853
age group: middle	0.098	0.204	0.48	0.630
age group: older	-0.043	0.362	-0.12	0.905
random effects	**Variance**	**St. Deviation**		
listener	0.22	0.464		
word pair	0.07	0.28		

Table 7.4: Response ∼ sex + age group + (1| listener) + (1| word pair) Model Output

The results show that, in terms of listener sex, males and females are similarly accurate in the discrimination of the GOAT and THOUGHT vowels. This is indicated by the similarity in the estimates of accuracy for male and female listeners, and by the p-value showing a lack of a significant difference between the two groups. Similarly, the model output suggests that there is little effect of listener age on accuracy in GOAT-THOUGHT trials. While Figure 7.3 above demonstrated a trend line that indicated that older speakers had a very slightly higher accuracy rate, the results of the logistic regression model show no significant difference in the accuracy of distinguishing GOAT and THOUGHT for the three different age groups. The results of the model are unsurprising given the overall lack of any obvious effect of either age group or sex on accuracy, as demonstrated in 7.3 above.

A look at the incorrect responses to different GOAT-THOUGHT stimuli in the raw accuracy data suggested that listeners were more accurate on certain trials depending on which words were presented. As shown in Table 7.1, participants were most accurate on trials which required them to discriminate between the words *'low'* and *'law'*, and were least accurate in discriminating between the words *'foam'* and *'form'*. The effect of each stimulus on accuracy rate can be further investigated by looking at the random effect of word pair in the optimal logistic regression model. Again using the `predict()` function, it is possible to use the model output to calculate the probability of a correct response depending on the word pair presented in a given trial. Figure 7.4 displays the probability of a correct response for each word pair containing GOAT and THOUGHT vowels in the discrimination task.

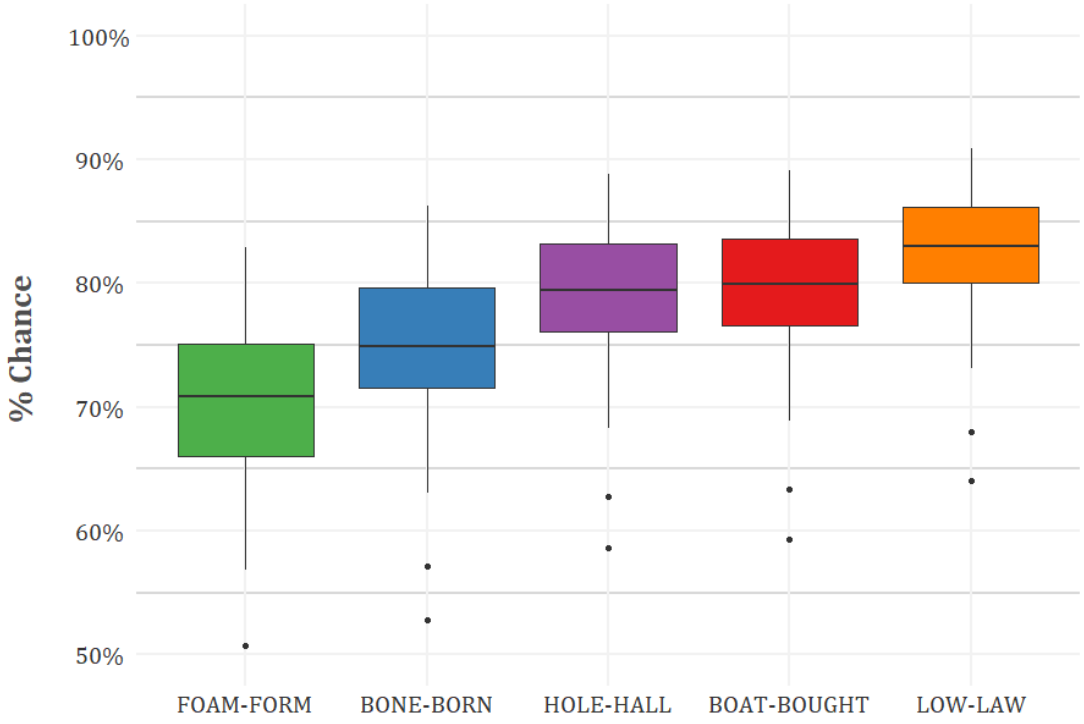


Figure 7.4: Probability of a Correct Discrimination by Word Pair

This figure further indicates that word pair has an effect on accuracy. Listeners have the highest probability of correctly discriminating between *'low'* and *'law'* when compared to the other word pairs, as the median line shows a 86% chance of correctly responding to these trials. Conversely, participants had a noticeably lower chance of correctly responding to trials containing the words *'foam'* and *'form'*, as Figure 7.4 shows the median chance of a correct response to be 72%. These results point toward an effect of phonological environment on listeners' ability to distinguish between GOAT and THOUGHT, as open syllable contexts (as in *'low'* and *'law'*) are those which are correctly discriminated most frequently, while the participants were most inaccurate in their responses to stimuli in which the vowel was followed by a nasal consonant (as in *'foam'* and *'form'*, and *'bone'* and *'born'*).

7.1.4 Sensitivity

As detailed in Chapter 6, sensitivity measures such as *d'* (d-prime) allow for an analysis of accuracy which takes into account any bias among participants for a particular response, e.g. a listener who favours 'same' responses over 'different' responses. Higher *d'* values indicate better performance in a task, and for same-different type discrimination tasks, *d'* values typically range between 0-4. Figure 7.5 below shows the *d'* values of each listener in their discrimination of GOAT-THOUGHT stimuli. The mean *d'* value is 1.596, as indicated by the horizontal black line in the

figure.

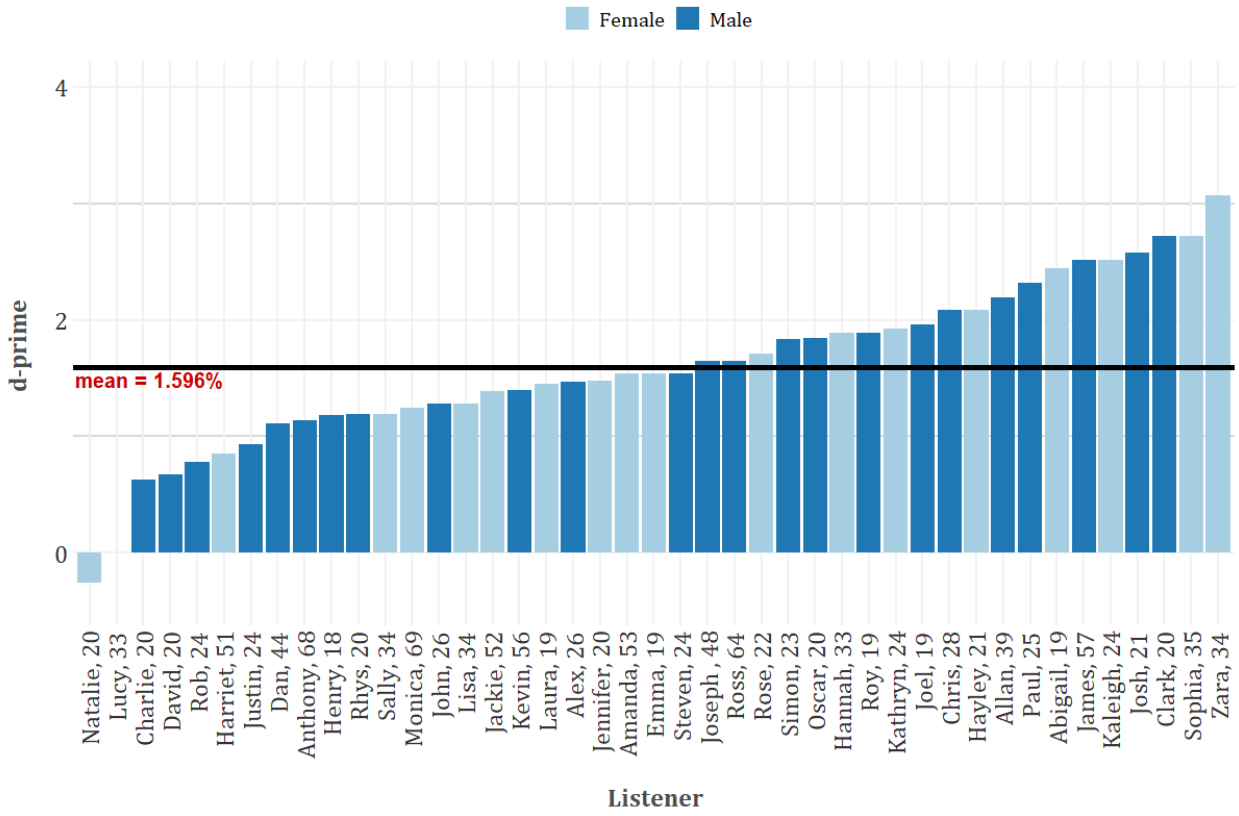


Figure 7.5: *d'* values for GOAT-THOUGHT discrimination

Here, the *d'* results shown in Figure 7.5 are overall not too dissimilar to the percentage correct results presented in Figure 7.2. This is particularly noticeable when looking at the extreme ends of the spectrum, as both *d'* and percentage correct measures of accuracy show Natalie and Lucy as the least accurate listeners (performing at or below chance level), and Zara as the listener best able to discriminate correctly between GOAT and THOUGHT.

7.1.4.1 Accuracy by Stimuli

Figure 7.6 below shows the three listeners with the lowest and the three listeners with the highest *d'* values on the GOAT-THOUGHT trials. The figure compares the *d'* values for each of these participants on GOAT-THOUGHT, NURSE-THOUGHT and control trials.

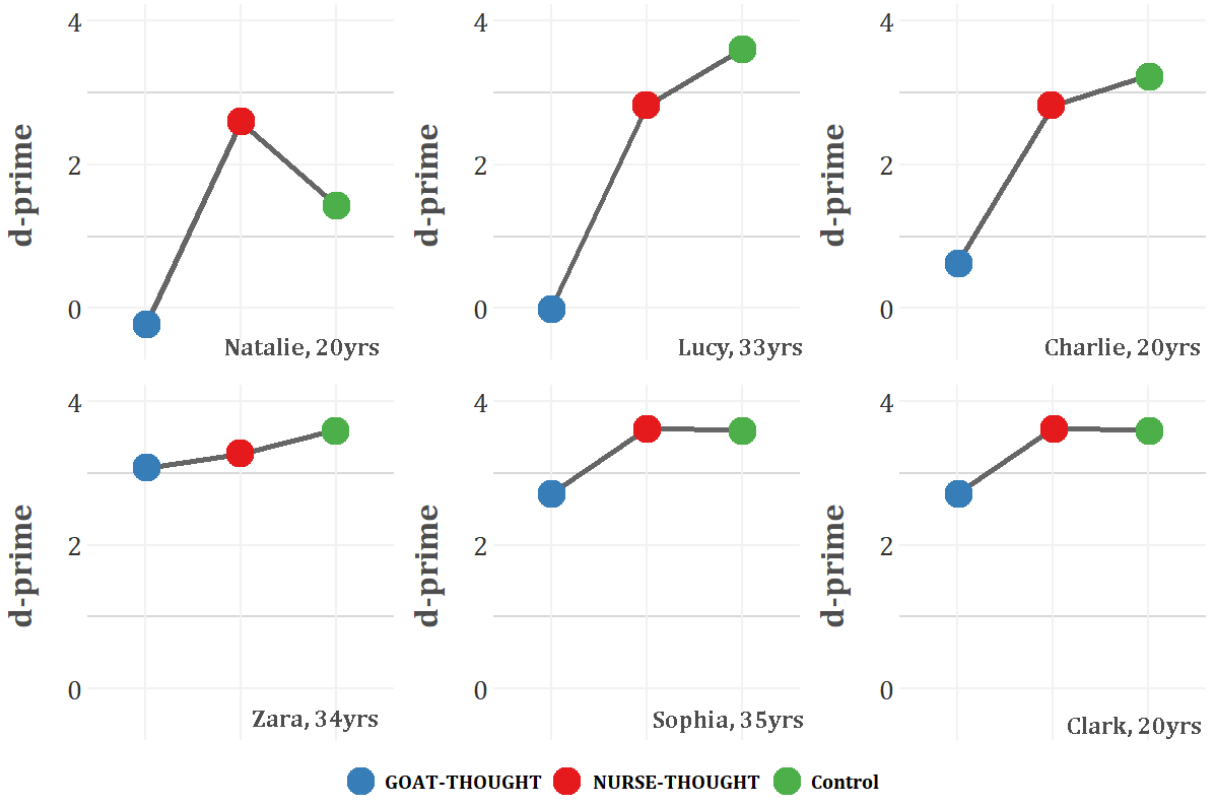


Figure 7.6: Listeners with the lowest (top row) and highest (bottom row) *d'* scores

Firstly, all listeners were found to exhibit lower *d'* scores on GOAT-THOUGHT trials compared to other trials. As displayed in Figure 7.6, even the best performing listeners showed at least slightly lower accuracy in discriminating GOAT-THOUGHT than the other vowel pairs. This suggests that all listeners struggled most to distinguish the GOAT and THOUGHT vowels. Nevertheless, the three least accurate listeners shown in Figure 7.6 exhibit a much lowered *d'* value for GOAT-THOUGHT stimuli than the other vowel pair stimuli, compared to the most accurate listeners who show a much smaller difference in the *d'* values of the three trial types. This indicates that, while all listeners were least accurate in discriminating the GOAT and THOUGHT vowels, there are listeners who appear to struggle substantially more than others in distinguishing these two vowels. There is also some variation among the listeners as to whether *d'* is higher on NURSE-THOUGHT or control trials, although for all listeners, these two trial types have more similar *d'* values than when compared with the *d'* values of the GOAT-THOUGHT trials.

One method of statistically analysing sensitivity is to perform a one-factor ANOVA to test for the significance of a certain predictor on the *d'* values, and then to use post-hoc comparisons to determine whether differences between group means are significant, as in Bradlow *et al.* (1999). In assessing the significance of

vowel pair on the *d'* values in the current study, a one-factor ANOVA found vowel pair to have a significant effect on *d'* value (p = <0.001). This indicates that listeners were considerably less accurate in their responses to certain stimuli. As vowel pair was found to be a significant predictor of *d'*, a Tukey's Honest Significant Difference (Tukey's HSD) comparison was run on the data to further investigate the effects of vowel pair. The results of this comparison are displayed below in Table 7.5.

vowel pairs	**difference**	**lower**	**upper**	**adj. p-value**
NURSE-THOUGHT and GOAT-THOUGHT	1.3909683	1.3452925	1.4366442	<0.001
Control Pairs and GOAT-THOUGHT	1.6847062	1.6387225	1.7306899	<0.001
Control Pairs and NURSE-THOUGHT	0.2937379	0.2471827	0.3402932	<0.001

Table 7.5: Tukey HSD comparison of *d'* values by Vowel Pair

The output of the Tukey HSD comparison in Table 7.5 shows that *d'* values are significantly different when comparing GOAT-THOUGHT trials to trials containing both NURSE-THOUGHT and control stimuli (p = <0.001). Such a result suggests that listeners were more accurate in responding to particular trials. One disadvantage of a Tukey's HSD test is that it cannot provide information regarding which vowel pair participants were most or least accurate in discriminating. However, when combined with the above results in Figure 7.6 which demonstrate that listeners consistently exhibit lower *d'* values for GOAT-THOUGHT trials, these results can be interpreted as showing that participants were significantly less accurate – even after accounting for any response bias – in their discrimination of GOAT and THOUGHT than of any other vowel pair included in the task. This is further supported by the result of the comparison between NURSE-THOUGHT and control trials. While the test finds these two vowel pairs to be significantly different in their effects on *d'*, the difference between these two groups, as shown in Table 7.5, is much smaller than the differences resulting from the comparisons of the other vowel pairs. Thus, this result provides further indication that it is the GOAT-THOUGHT stimuli which have the largest effect on accuracy, and that listeners are, on the whole, worse at correctly discriminating between the GOAT and THOUGHT vowels than between the other vowel pairs presented in the discrimination task.

7.1.5 Reaction Time

The reaction times for each trial were automatically recorded by the OpenSesame (Mathôt *et al.* 2012) experimental software. As previously detailed, listeners were

asked to respond to each trial as quickly as possible. If no response was selected within 10000 milliseconds, the next trial would automatically start, and the listener's response would be marked as incorrect. Figure 7.7 displays the overall distribution of reaction times in the discrimination trial. The majority of responses took place within 2000 msec. The overall mean reaction time, as represented by the black vertical intercept in the figure, lay at 1180 msec.

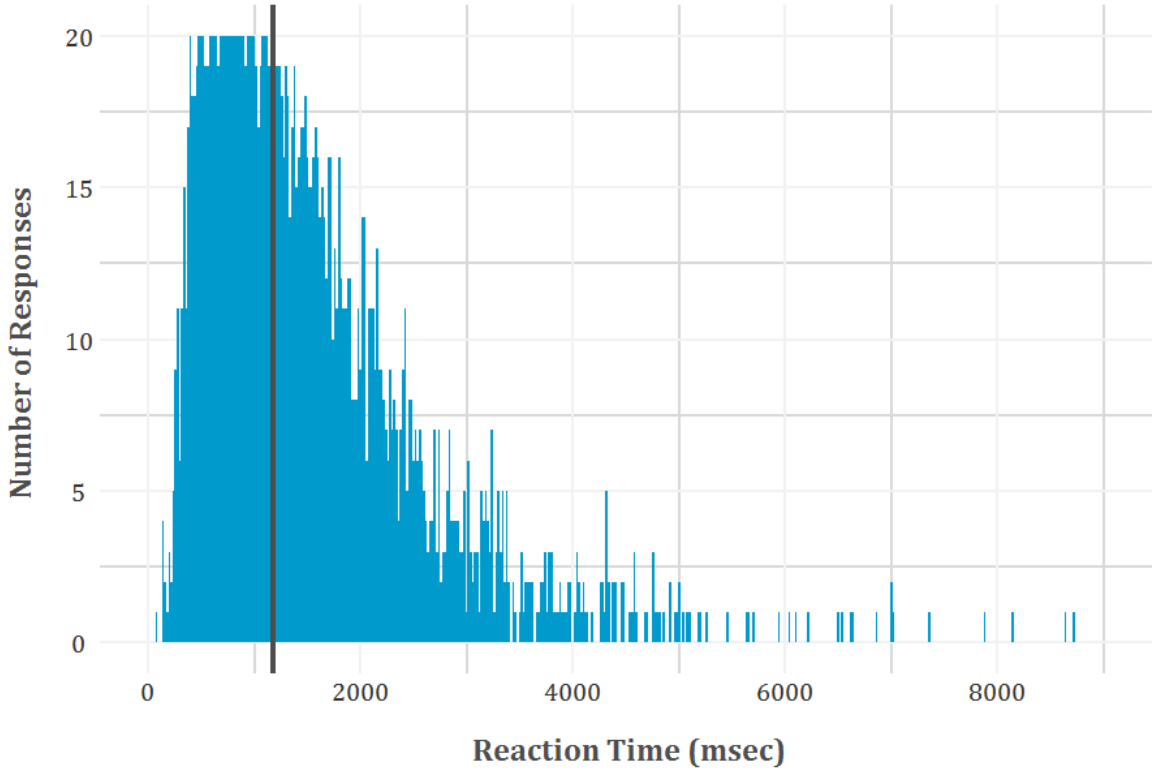


Figure 7.7: Overall distribution of vowel discrimination reaction times

The density plot in Figure 7.8 shows the distribution of reaction times recorded for each vowel pair included in the discrimination experiment. Reaction time values (milliseconds) were log transformed to reduce skew in the data. While the histogram in Figure 7.7 showed that the majority of responses in the discrimination task were rapid, Figure 7.8 highlights a number of differences between the vowel pairs in terms of response time.

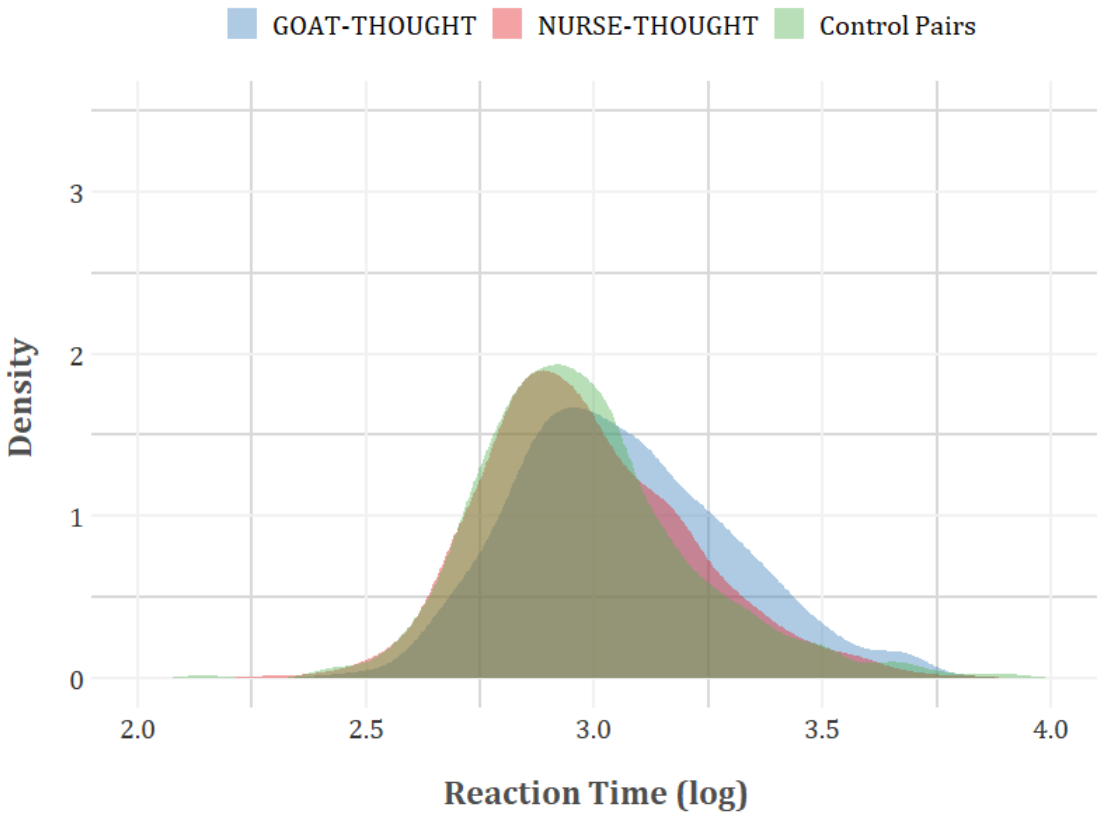


Figure 7.8: Reaction time of vowel discrimination by vowel pair

Firstly, looking at the distributional spread of response time data for each vowel pair, Figure 7.8 shows an increased number of slower responses on trials containing GOAT-THOUGHT stimuli than trials containing NURSE-THOUGHT or control pairs. The peak of the distribution of GOAT-THOUGHT reaction times also begins at a later timepoint, again demonstrating that reaction times to GOAT-THOUGHT trials were slower when compared with other trials. As listeners take longer to discriminate GOAT-THOUGHT stimuli, this result signals that listeners have greater difficulty in distinguishing these vowels perceptually. Figure 7.8 further shows little difference between the NURSE-THOUGHT and control trials in terms of reaction time, as shown by the similarity in the spreads and peaks of their distributions. Therefore, based on these reaction time results, it does not appear that listeners found distinguishing NURSE and THOUGHT more challenging than discriminating the control vowel pairs.

Figure 7.9 displays boxplots of log-transformed reaction times for GOAT-THOUGHT and NURSE-THOUGHT discrimination by different listener groups. Each boxplot compares how quickly listeners, grouped by sex and age group, responded to GOAT-THOUGHT and NURSE-THOUGHT stimuli.

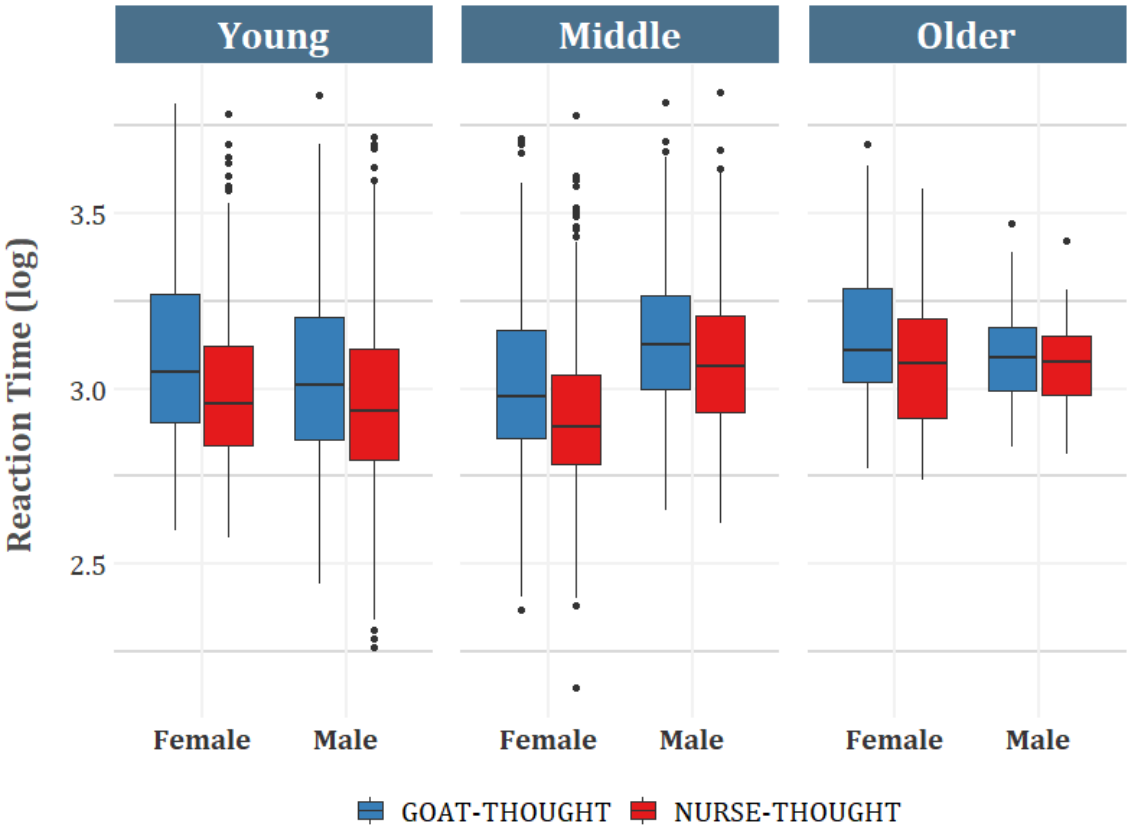


Figure 7.9: Reaction time of vowel discrimination by listener sex and age

Looking first at the results for the young speakers, Figure 7.9 shows little difference between the male and female listeners in their response time to GOAT-THOUGHT vs. NURSE-THOUGHT trials. The plot displays similar box heights and median values for the males and females, with both taking considerably longer to respond to GOAT-THOUGHT trials compared to NURSE-THOUGHT trials. This provides some evidence that the young listeners who took part in the experiment exhibit greater difficulty in discriminating between the GOAT and THOUGHT vowels than other vowel pairs, and also that listener sex does not appear to have an effect on reaction time for the youngest participants in the sample.

Figure 7.9 displays a slight difference between the reaction times of male and female listeners in the middle age group. The results for the females in the middle age group look similar to the results of both the males and females in the young listener plot; similar box heights, similar medians, and a notable difference in the response time between GOAT-THOUGHT and NURSE-THOUGHT trials. This indicates that the middle females pattern with the young listeners in finding discriminating between the GOAT and THOUGHT vowels more difficult than differentiating NURSE and THOUGHT. Figure 7.9 indicates that males in the middle age group were slightly

slower than the females in their reaction times in the task overall. Furthermore, while the middle males display slower reaction times to GOAT-THOUGHT trials, the difference in response time between the two vowel pairs appears slightly reduced compared to the middle female and young listeners.

The older listeners, as presented in Figure 7.9, also show a small sex difference in terms of their reaction times. Like the young and middle listener age groups, the older females take longer to discriminate GOAT-THOUGHT than NURSE-THOUGHT. By contrast, the older males show little difference in their response times to GOAT-THOUGHT vs. NURSE-THOUGHT stimuli.

Overall, Figure 7.9 suggests a small effect of the social factors of listener sex and age group on response time in the discrimination experiment. Young listeners demonstrate the largest difference in response time on GOAT-THOUGHT trials compared to NURSE-THOUGHT trials, while, as a group, the older speakers show the smallest reaction time difference between these trials. Additionally, male and female listeners appear to differ more in their response times at the higher end of the age spectrum. This could provide some evidence that females are overall perceptually ahead in this sound change toward perceptual GOAT-THOUGHT merging, however the differences between the listener groups shown in Figure 7.9 are small.

7.1.6 Modelling Reaction Times

7.1.6.1 Overall Reaction Times

Mixed-effects linear regression models were used for a statistical analysis of reaction time. This model takes the dependent variable of reaction time (log-transformed), and includes the predictor of vowel pair (i.e. GOAT-THOUGHT, NURSE-THOUGHT, or control pairs). As before, both listener and word pair were added to the model as a random effects. The output of this linear regression model is shown below in Table 7.6. Statistical significance is indicated by a t value of greater than +/-2.

fixed effects	**Estimate**	**St. Error**	**t value**
(intercept)	2.979	0.020	147.988
vowel pair: GOAT-THOUGHT	0.083	0.020	4.262
vowel pair: NURSE-THOUGHT	0.002	0.020	0.123
random effects	**Variance**	**St. Deviation**	
listener	0.008	0.0914	
word pair	0.0007	0.026	

Table 7.6: Reaction Time ∼ Vowel Pair + (1| Listener) + (1| Word Pair) Model

The model output presented in Table 7.6 shows that reaction time differs significantly depending on which vowels were presented in each trial. The t values indicate that, overall, listeners were significantly slower in responding to trials in which they were asked to discriminate GOAT and THOUGHT vowels when compared with other trials. Combined with the rest of the findings from the discrimination task, this result provides further evidence to suggest that the Tyneside listeners find it more difficult to distinguish GOAT from THOUGHT than the other vowel pairs presented in the experiment. The model output also suggests that there is no significant difference between the reaction times to NURSE-THOUGHT and the control trials. This indicates that listeners did not have trouble discriminating between the NURSE and THOUGHT vowels, as they did not take longer to respond to these trials than those containing control stimuli.

7.1.6.2 GOAT-THOUGHT Reaction Times

Four more mixed-effects linear regression models were run on the discrimination task data to investigate how social factors may affect reaction time. Table 7.7 displays the predictors which were included in each of the models. As discussed, models with lower AIC values are preferred, as they are a better fit for the data. The models tested contained predictors for listener age group and sex. Again, all models contained listener and word pair (5 levels: *boat-bought, bone-born, foam-form, hole-hall, low-law*) as random effects.

	model	**AIC**
Reaction Time ∼	age group + sex + (1| listener) + (1| word pair)	-146
Reaction Time ∼	age group * sex + (1| listener) + (1| word pair)	-147

Table 7.7: Comparison of linear regression models for reaction time in discrimination of vowels

Table 7.7 shows very little difference between the models in terms of AIC values. ANOVA comparisons indicate that the model is not significantly improved by the addition of word pair as a predictor, nor by the inclusion of an interaction between listener age group and sex ($p = 0.08$). Therefore, the optimal model is taken as the first one presented in Table 7.7, which contains age group and sex as fixed effects without an interaction term. The output of this model is presented below in Table 7.8.

fixed effects	**Estimate**	**St. Error**	**t value**
(intercept)	3.021	0.028	106.580
age group: middle	0.034	0.034	0.99
age group: older	0.085	0.063	1.35
sex: male	0.023	0.031	0.73
random effects	**Variance**	**St. Deviation**	
listener	0.009	0.097	
word pair	0.0002	0.016	

Table 7.8: Reaction Time ∼ Age Group + Sex + (1| Listener) + (1| word pair) Model Output

Table 7.8 shows the coefficients of the selected model. Results indicate that neither of the social factors of listener age group or sex significantly affect the reaction time of discriminating the GOAT and THOUGHT vowels. Looking at the estimates and t values in Table 7.8, the model suggests that older speakers tend to exhibit longer reaction times, but, as significance is taken as a t value of greater than +/-2, this effect is not significant.

7.1.7 Summary

Overall, the results from the discrimination experiment show evidence of Tyneside listeners struggling, to a certain extent, to distinguish the GOAT and THOUGHT vowels. This is signalled by the lower accuracy rates in GOAT-THOUGHT trials in comparison to trials which asked listeners to discriminate between other vowel pairs. Listeners are shown to be significantly less accurate in deciding whether GOAT and THOUGHT stimuli are the 'same' or 'different' than when asked the same question with NURSE-THOUGHT stimuli, despite prior literature reporting NURSE and THOUGHT as previously merged in Tyneside English. This holds true for both the proportion correct measure of accuracy, as well as the sensitivity measure which took into account any response bias among the listeners. The reaction time

results also indicate perceptual difficulty in discriminating between the GOAT and THOUGHT vowels, as listeners are found to take longer to respond to these trials than trials containing other vowel pairs. These results suggest that, overall, this group of Tyneside listeners are not always able to reliably distinguish between GOAT and THOUGHT perceptually, and it is tempting to take these results as evidence to support arguments of a merger between the GOAT and THOUGHT vowels in the Tyneside community.

Analysis of the discrimination task results suggests that neither the age nor the sex of the listener plays a significant role in their ability to distinguish GOAT from THOUGHT. Based on measures of accuracy and reaction time, the output of the mixed-effects logistic regression and linear regression models show that age and sex are not significant factors in the ability of listeners to accurately discriminate GOAT and THOUGHT. As the results from the production analysis – presented in Chapter 5 – indicated that female speakers show greater overlap of GOAT and THOUGHT in phonetic space, it may have been anticipated that female listeners would find it more difficult to distinguish these vowels in perception. Such a pattern does not emerge in the perception results for the discrimination task, as female speakers of the same age group are shown to occupy both ends of the accuracy spectrum. Rather, it appears that the ability to differentiate GOAT from THOUGHT perceptually differs on a listener-by-listener basis.

7.2 Identification Task

7.2.1 Results: Accuracy

Accuracy rates for the identification experiment were calculated using the same method as in the discrimination experiment. Figure 7.10 shows the accuracy rate (proportion of correct responses) of each listener for all trials in the identification task.

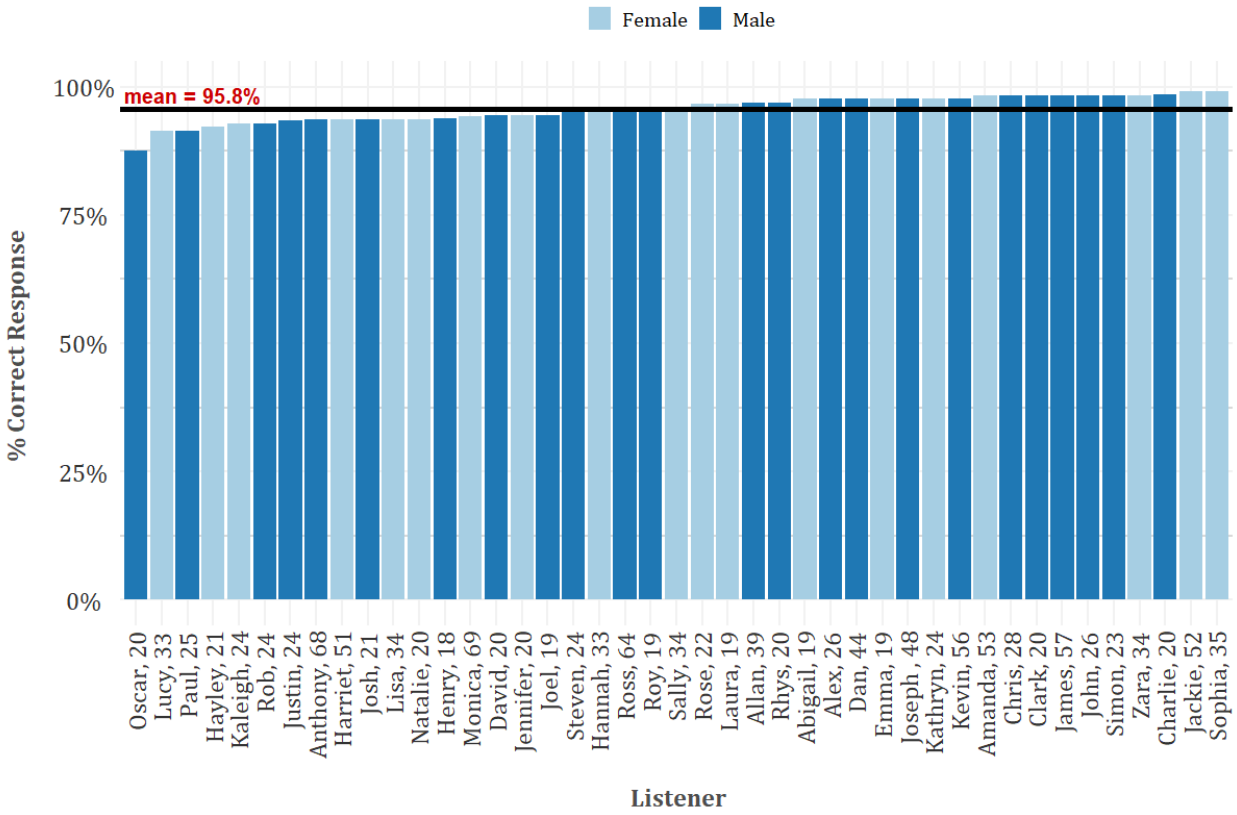


Figure 7.10: Overall accuracy percentage of each listener

Figure 7.10 shows that, as with the discrimination experiment, the overall accuracy rates among the Tyneside listeners are high. The mean accuracy rate for this experiment is 95.8%, notably higher than the mean accuracy rate in the discrimination task. The correct response rate ranges from 87.5% to 99.2% among the participants, indicating that there is little variation between listeners in terms of their performance on this task overall.

Accuracy rates for GOAT-THOUGHT identification are shown in Figure 7.10. The mean accuracy rate for the GOAT-THOUGHT trials is 89.8%, which is considerably higher than the 76.5% mean accuracy rate for distinguishing the GOAT and THOUGHT vowels in the discrimination task. There is clear variation between speakers in their ability to accurately identify the GOAT and THOUGHT vowels, with accuracy rates ranging from 70% to 100%. This result indicates that some listeners do struggle to correctly identify stimuli as belonging to either the GOAT or THOUGHT word class. As with the discrimination experiment, these results show two females listeners of a similar age (Lucy and Sophia) exhibiting some of the lowest and highest accuracy rates. Similarly, while the majority of male listeners exhibited accuracy levels above the mean, the listener with the lowest percentage of correct responses in the task is a male. This immediately suggests a lack of a

clear influence of listener age or sex on performance in the identification task.

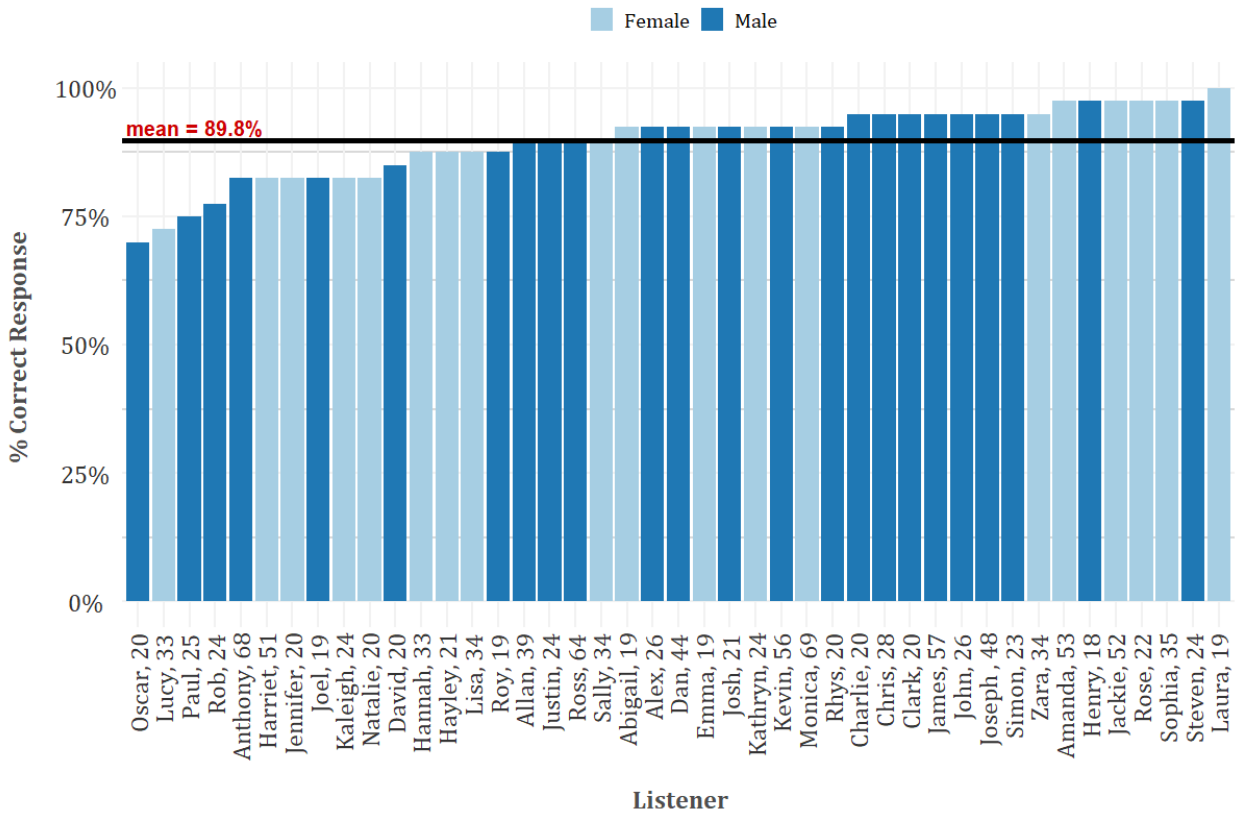


Figure 7.11: Accuracy rates for GOAT-THOUGHT identification

For the identification task, Table 7.9 shows the number of times listeners incorrectly responded to different word pairs containing GOAT and THOUGHT vowels. Participants were least accurate in identifying the GOAT and THOUGHT vowels when presented in the words *'coat'* and *'caught'*, while listeners were most accurate in their responses to the words *'snow'* and *'snore'*, which make up only 2.8% of incorrect answers on GOAT-THOUGHT trials.

Stimuli Pair	**n**	**%**
coat-caught	48	34.3%
bowl-ball	33	23.6%
folk-fork	38	27.1%
flow-flaw	17	12.1%
snow-snore	4	2.8%

Table 7.9: Incorrect Responses to GOAT-THOUGHT Stimuli

A comparison of the accuracy rates of male and female listeners in identifying GOAT and THOUGHT stimuli is displayed in Figure 7.12.

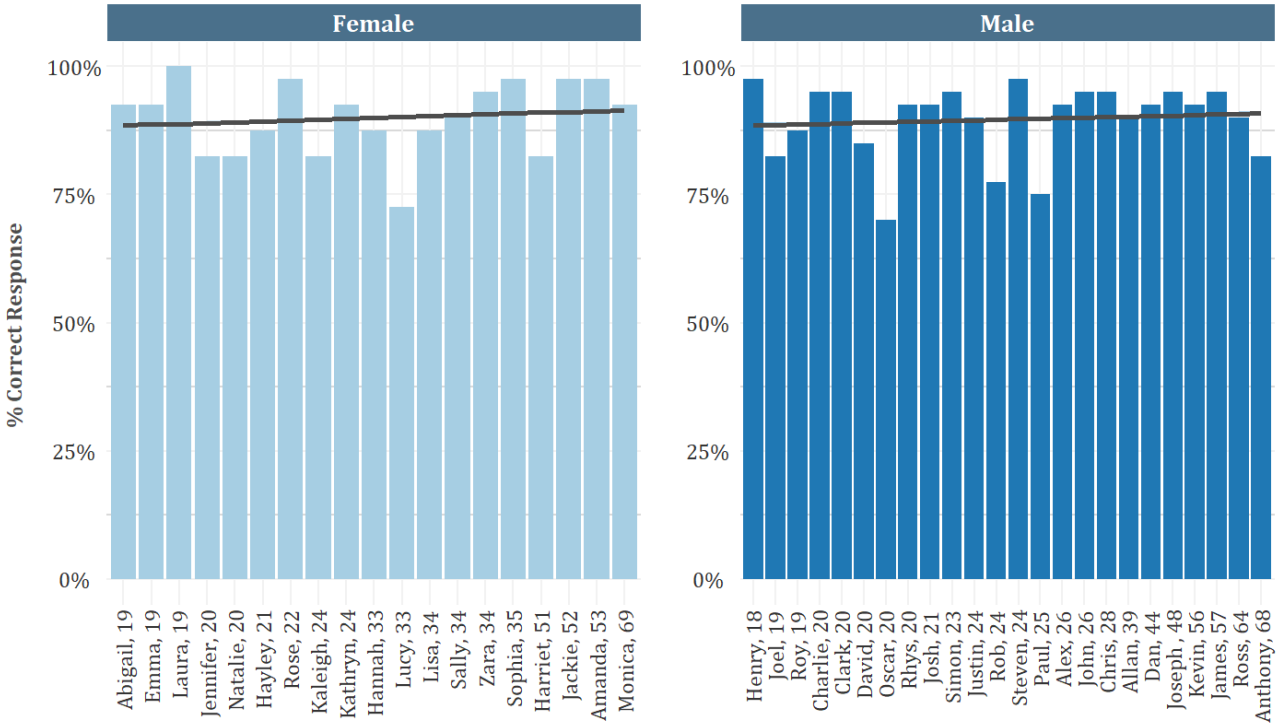


Figure 7.12: GOAT-THOUGHT identification by sex and age (L-R youngest to oldest listener)

As with the results of the discrimination task, these plots show variation in listener's accuracy rates for both male and female listeners across the age range of the sample. The trend lines shown in Figure 7.12 suggest that accuracy on the identification task improves marginally with age, although this trend is very slight and unlikely to come out as significant in statistical tests.

7.2.2 Statistical Modelling

7.2.2.1 Overall Accuracy

As in the discrimination experiment, a mixed-effects logistic regression model was run on the data, using response (i.e. whether a listener's response on a given trial was correct or incorrect) as the dependent variable. This model compares accuracy rates of the identification of the vowels in different vowel pairs. Again, the model includes the predictor of vowel pair (GOAT-THOUGHT, NURSE-THOUGHT, and control pairs) and the random effects of listener and word. Statistical significance is indicated by a p-value of <0.05. The output of the model is shown below in Table 7.10.

fixed effects	**Estimate**	**St. Error**	**z value**	**Pr(>|z|)**
(intercept)	4.685	0.001	4792.1	<0.001
vowel pair: GOAT-THOUGHT	-2.133	0.001	-2183.3	<0.001
vowel pair: NURSE-THOUGHT	-0.645	0.001	-660.2	<0.001
random effects	**Variance**	**St. Deviation**		
listener	0.296	0.544		
word	0.487	0.698		

Table 7.10: Response ∼ Vowel Pair + (1| Listener) + (1| Word) Model

The model output presented in Table 7.10 reflects the results of the same model run on the discrimination experiment data, as listener accuracy in the identification experiment appears to be significantly affected by vowel pair. The estimates and the p-values indicate that listeners are more accurate in their identification of audio stimuli when presented with a choice between NURSE-THOUGHT and control words than when asked to decide whether a stimulus belonged to the GOAT or THOUGHT lexical set. The model output also suggests that listeners performed best on control trials when compared to trials involving GOAT-THOUGHT and NURSE-THOUGHT stimuli.

Once again, the output of the model was fed into the `predict()` function in R to calculate the probability of a listener responding correctly to a trial depending on which vowel pairs were presented. Participants were very likely to be able to correctly identify the audio stimulus when presented with a forced choice between NURSE and THOUGHT words (a 98% chance) and control word pairs (a 99% chance). Additionally, listeners were overall found to be slightly less likely to correctly respond to GOAT and THOUGHT trials, with a 92% chance of selecting the correct vowel category.

7.2.2.2 GOAT-THOUGHT Accuracy

To investigate how accurately different listener groups were able to identify GOAT and THOUGHT vowels, a number of mixed-effects regression models were performed on the data. Table 7.11 shows a comparison of several models with different predictors. The optimal model, as signalled by the lowest AIC value, is that which contains the predictors of listener sex and age group without an interaction between these two predictors. Additionally, an ANOVA model comparison indicates that including an interaction term between age group and sex does not significantly improve the model ($p = 0.954$). As with the models run on the results

of the discrimination task, both listener and word pair (5 levels: *bowl-ball, coat-caught, flow-flaw, folk-fork, snow-snore*) were included in the model as random effects.

model	**AIC**
Response ~ sex + age group + (1| listener) + (1| word pair)	925
Response ~ sex * age group + (1| listener) + (1| word pair)	929

Table 7.11: Comparison of logistic regression models for accuracy in identification of GOAT and THOUGHT

Table 7.12 below shows the output from the selected logistic regression model. As with the results from the discrimination experiment, the model output shows that the social factors of listener sex and age group have little effect on accuracy in the identification task. Male and female listeners do not appear to differ significantly in their identification of GOAT and THOUGHT vowels, and accuracy rates do not differ significantly for any of the age groups.

fixed effects	**Estimate**	**St. Error**	**z value**	**Pr(>|z|)**
(intercept)	2.571	0.358	7.185	<0.001
sex: male	0.052	0.259	0.20	0.840
age group: middle	0.064	0.280	0.23	0.819
age group: older	-0.066	0.485	-0.14	0.893
random effects	**Variance**	**St. Deviation**		
listener	0.29	0.536		
word pair	0.33	0.58		

Table 7.12: Response ~ sex + age group + (1| listener) + (1| word pair) Model Output

As in the discrimination experiment, listener accuracy appears to vary depending on the word pair presented in a given trial. Indicated in Table 7.9 above, participants were able to categorise the words *'snow-snore'* and *'flow-flaw'* more frequently than the other word pairs containing GOAT and THOUGHT vowels. Again, as word pair was built into the logistic regression model as a random effect, it is possible to use the model output to calculate the probability of a correct response to a trial depending on the word pair presented. These results are presented below in Figure 7.13.

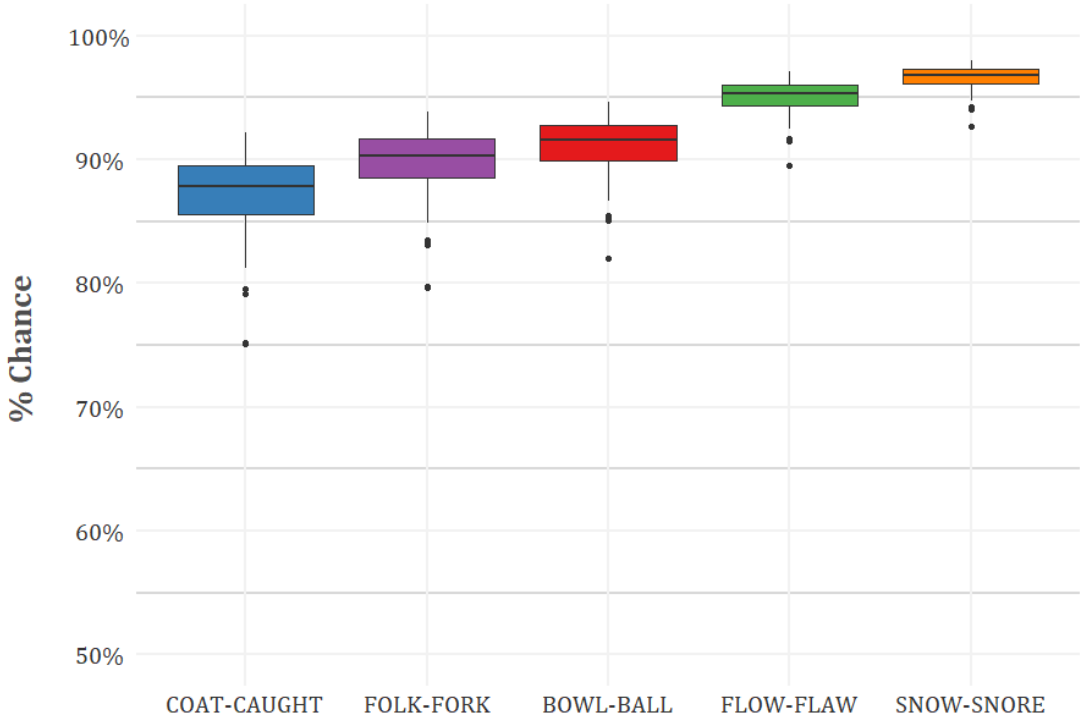


Figure 7.13: Probability of Correct Identification by Word Pair

This figure shows that, for each word pair, listeners typically had over an 85% chance of correctly identifying which word matched the audio stimulus. Once again, this is indicative of the overall good performance on this task among the participants. Figure 7.13 also illustrates that listeners had an increased ability to correctly categorise GOAT and THOUGHT when presented with the *'snow-snore'* and *'flow-flaw'* word pairs. This is signalled by the smaller amount of variance, and the higher median probabilities for these word pairs (i.e. more than a 95% chance of a correct response). Such results are consistent with the findings from the discrimination experiment, which suggested that listeners performed better on trials which presented audio stimuli with GOAT and THOUGHT in open syllable environments. The potential effects of phonological context on the perception of the GOAT and THOUGHT vowels will be discussed in further detail in Section 7.5.2.

7.2.3 Reaction Time

Figure 7.14 displays the overall distribution of reaction times in the identification task, showing that the vast majority of responses took place within 2000 milliseconds. The mean reaction time is 1042 milliseconds. Comparing the reaction times of the two task types, the narrower distribution and the lower mean value indicate that listeners were typically quicker to respond in the identification task than in the discrimination task.

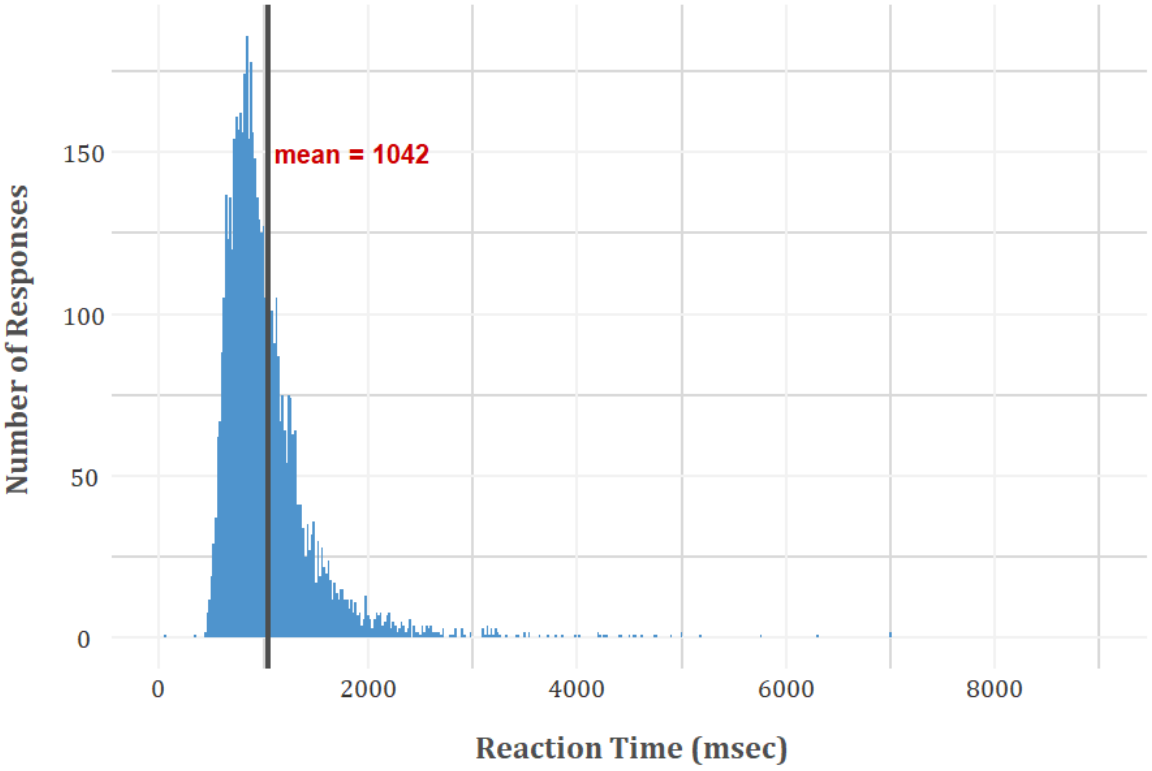


Figure 7.14: Overall distribution of vowel identification reaction times

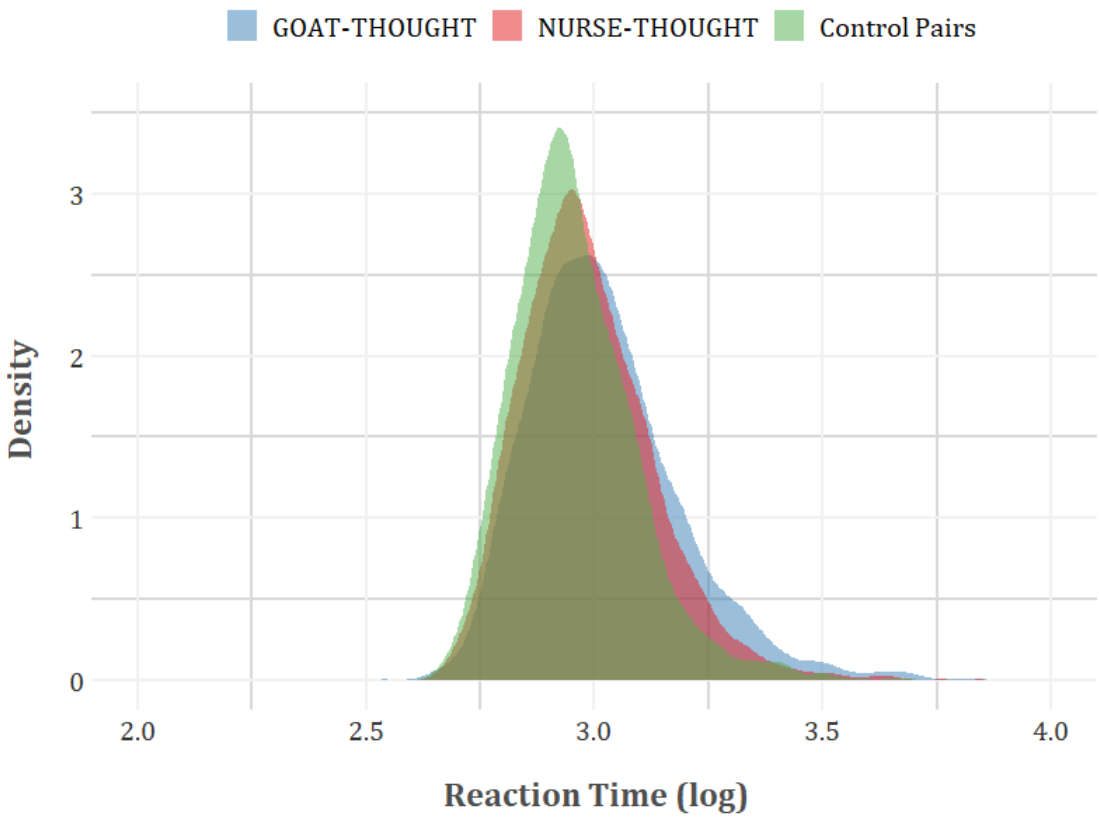


Figure 7.15: Reaction time of vowel identification by vowel pair

Figure 7.15 displays the distribution of the log-transformed reaction times on GOAT-THOUGHT, NURSE-THOUGHT and control trials in the identification task. As with the response time results of the discrimination task, this figure indicates that listeners were slowest at responding to GOAT-THOUGHT trials, as demonstrated by the peak of the data starting at a later timepoint than for the other vowel pairs, and by the spread of the data showing a greater number of slower responses when compared with other trials. This again points to listeners exhibiting some perceptual difficulty in identifying GOAT and THOUGHT stimuli. Unlike the results of the discrimination experiment, however, Figure 7.15 indicates that listeners reacted quicker to control trials compared to trials containing NURSE-THOUGHT stimuli in the identification task.

The plots in Figure 7.16 show the reaction times for GOAT-THOUGHT and NURSE-THOUGHT identification by listener age and sex. These plots compare the reaction times on GOAT-THOUGHT and NURSE-THOUGHT trials for each of the listener groups.

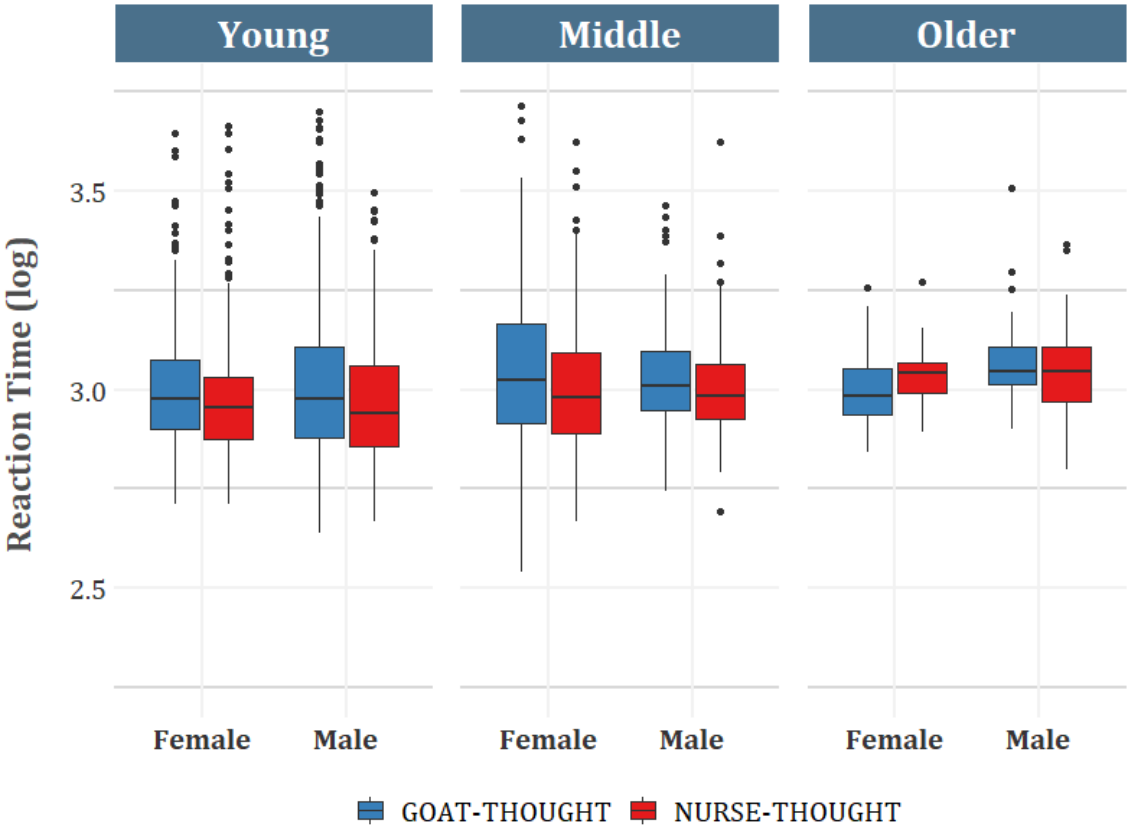


Figure 7.16: Reaction time of vowel identification by listener age and sex

The reaction time results for the young listeners in Figure 7.16 are similar for the males and females. As with the results of the discrimination task, the plot shows that young male and female listeners are comparable in terms of their response

times on the identification experiment, with both taking slightly longer to respond to GOAT-THOUGHT trials than NURSE-THOUGHT trials. Again, this indicates that the young Tyneside listeners find labelling GOAT-THOUGHT minimal pair stimuli more difficult than NURSE-THOUGHT stimuli.

Looking at the middle group in Figure 7.16, the male and female listeners show similarity in terms of the height of the boxes and the median reaction time values. Middle aged males and females also both show a distinction in their reaction times to GOAT-THOUGHT vs. NURSE-THOUGHT stimuli. As both exhibit slower response times in labelling GOAT and THOUGHT vowels, this indicates that the middle-aged listeners, as a group, found trials containing these vowels most difficult. Unlike the boxplots comparing the reaction times of the listener groups in the discrimination task (Figure 7.9), the above figure suggests a lack of an effect of sex on reaction time for the middle age group.

Figure 7.16 also shows that the older males and females behave similar to one another in terms of response time, again signalled by the similarity in the height of the boxes and the median reaction time values. However, the older listeners differ from the young and middle listener groups as they show little difference in response time between GOAT-THOUGHT vs. NURSE-THOUGHT trials, as in the case of the older males, while the older females appear to take longer to respond to NURSE-THOUGHT than GOAT-THOUGHT trials. This could indicate that the older listeners find it easier to identify GOAT and THOUGHT vowels than the younger age groups. While Figure 7.16 gives an indication of the effects of sex and age on reaction time, statistical analysis will determine whether these social factors have a significant impact on reaction times in the identification experiment.

7.2.4 Modelling Reaction Times

7.2.4.1 Overall Reaction Times

For a statistical comparison of reaction time across vowel pairs, a mixed-effects linear regression model was run on the data from the identification task. This model includes the log-transformed reaction times as the dependent variable, includes vowel pair (i.e. GOAT-THOUGHT, NURSE-THOUGHT, control pairs) as fixed effect, and listener and word pair as random effects. The output of the model is presented in Table 7.13 below. A t value of greater than +/-2 indicates statistical significance.

fixed effects	**Estimate**	**St. Error**	**t value**
(intercept)	2.956	0.014	206.267
vowel pair: GOAT-THOUGHT	0.0657	0.012	5.553
vowel pair: NURSE-THOUGHT	0.032	0.012	2.718
random effects	**Variance**	**St. Deviation**	
listener	0.006	0.076	
word pair	0.0007	0.026	

Table 7.13: Reaction Time ∼ Vowel Pair + (1| Listener) + (1| Word Pair) Model

The model output indicates that reaction time is significantly different depending on which vowels were presented in each trial of the identification task. The t values, shown in Table 7.13, suggest that listeners took significantly longer to respond to GOAT-THOUGHT stimuli when compared with the control vowel pairs used as stimuli in the task. This reflects what was found in the analysis of reaction times for the discrimination experiment, further indicating that these Tynesiders find it more challenging to label GOAT and THOUGHT vowels than the other audio stimuli used in the task. The results of the model also suggest that participants were significantly slower in the responses to trials containing NURSE-THOUGHT stimuli than those containing only control vowel pairs.

7.2.4.2 GOAT-THOUGHT Reaction Times

For a statistical analysis of reaction times in the identification task, linear regression models were compared using AIC values and ANOVA comparisons to find the optimal model. Table 7.14 shows the AIC for each model in the comparison. The models tested all included the log-transformed reaction times as the dependent variable, and all models contained predictors for the social factors of listener sex and age group. As before, both listener and word pair were incorporated into all models as random effects.

model	**AIC**
Reaction Time ∼ age group + sex + (1| listener) + (1| word pair)	-1471
Reaction Time ∼ age group * sex + (1| listener) + (1| word pair)	-1468

Table 7.14: Comparison of linear regression models for reaction time in discrimination of vowels

Comparing the models, Table 7.14 shows that the model which includes the fixed

effects of age group and sex has the lowest AIC value[1], and this best explains the data. ANOVA comparisons suggest that the inclusion of an interaction term between age group and sex does not improve the model ($p = 0.79$). The coefficients of the selected model are displayed below in Table 7.15.

fixed effects	**Estimate**	**St. Error**	**t value**
(intercept)	2.998	0.027	110.69
sex: male	0.001	0.027	0.05
age group: middle	0.036	0.029	1.23
age group: older	0.038	0.051	0.75
random effects	**Variance**	**St. Deviation**	
listener	0.006	0.081	
word pair	0.0007	0.026	

Table 7.15: Reaction Time ∼ + sex + age group + (1| listener) + (1| word pair) Model Output

As the model output indicates, neither sex nor age group significantly affects the time it takes for listeners to identify GOAT and THOUGHT vowels. While it seems that middle and older listeners take slightly longer to respond than younger listeners, this difference in reaction time is not significant. A similar lack of an effect of participant age group or sex was also found in the analysis of reaction times for the discrimination experiment.

7.2.5 Summary

Consistent with the results of the discrimination task, the results of the identification experiment indicate that listeners performed worse in terms of accuracy in identifying the GOAT and THOUGHT vowels, showing significantly fewer correct responses on trials containing GOAT-THOUGHT stimuli compared to other trials. Furthermore, listeners took significantly longer to identify whether a stimulus belonged to the GOAT or THOUGHT lexical set compared to other vowel pairs presented in the task. Again, as with the results of the discrimination experiment, no significant effect of listener age or sex was found. This indicates that the results of the perception experiment do not reflect the patterns found in the production data collected from the DECTE interviews, wherein young females

[1]In cases where the AIC value is negative, it is generally accepted that 'lowest' AIC value refers to the value which is the lowest number e.g. a more negative value indicates a better fit than a value closer to 0.

were found to show the greatest overlap of GOAT and THOUGHT in phonetic space. A comparison of the perception data with the results from the DECTE data will be explored in more detail in the following chapter, along with a discussion of whether these results can be taken as convincing evidence of a GOAT-THOUGHT merger in Tyneside English.

7.3 Further Observations

7.3.1 Misidentification of GOAT and THOUGHT

Additional analysis of the responses to the identification task shows a difference in listeners' ability to correctly identify GOAT vowels and THOUGHT vowels. This disparity is displayed in Figure 7.15 below.

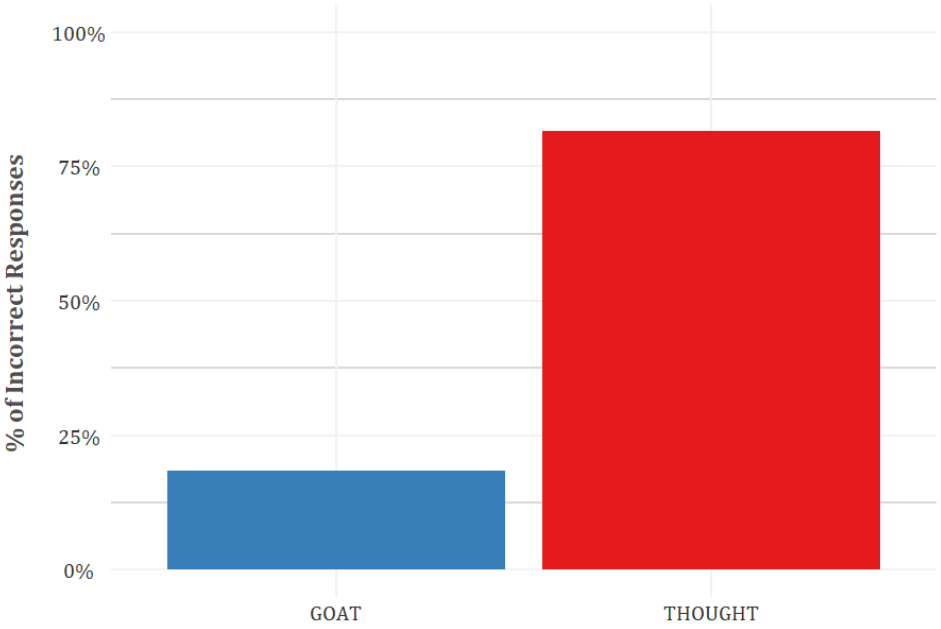


Figure 7.17: Percentage of incorrect GOAT and THOUGHT responses in the identification task

Figure 7.17 shows that, of the incorrect responses to trials containing GOAT-THOUGHT stimuli, listeners struggle to a greater extent to accurately identify THOUGHT vowels. Misidentified GOAT vowels make up only 18% of the incorrect responses to GOAT-THOUGHT trials, compared with the 82% for THOUGHT vowels mistakenly labelled as GOAT. This indicates that the large majority of errors come from miscategorisation of THOUGHT vowels as belonging to the GOAT vowel class.

As outlined above, the aim of identification experiments is typically to assess participants' knowledge of phonemic categories. The results shown in Figure 7.17 perhaps indicate that the phonemic boundaries of THOUGHT are changed or expanded for Tyneside listeners, and that listeners are more likely to accept THOUGHT stimuli as belonging to the GOAT class than vice versa. This result does not appear to be consistent with the results of the production analysis, presented in Chapter 5. As previously discussed, the findings from the DECTE data largely pointed to a lowering of GOAT in phonetic space to the position of THOUGHT. The evidence of asymmetric miscategorisation in the identification experiment would perhaps suggest, rather, that it is THOUGHT which has come to merge with the GOAT vowel. This, initially indicates that there is a mismatch between the results of the production and perception analyses in terms of evidence concerning the direction of the merger. Further discussion of the potential directionality of a Tyneside GOAT-THOUGHT merger – with reference to both the findings of the production and perception analyses – is presented in Chapter 8.

7.4 Perceptions of NURSE-THOUGHT

Stimuli contrasting the NURSE and THOUGHT vowels were included in both of the perception tasks as a point of comparison for GOAT-THOUGHT stimuli, and additionally to investigate whether there is any evidence of a perceptual merger between these vowels for the Tyneside listeners[2]. The findings presented within this chapter have indicated that listeners are more accurate and quicker at responding to NURSE-THOUGHT than GOAT-THOUGHT trials, suggesting that participants found it easier to discriminate/identify NURSE and THOUGHT vowels.

Results from the accuracy analyses, however, do indicate that listeners are slightly less accurate in responding to NURSE-THOUGHT trials than control trials. This could be taken as evidence of some lingering perceptual similarity between the NURSE and THOUGHT vowels due to a previous merging of these vowels in the speech community. If this were the case, it may be expected that the oldest listeners in the sample would show most difficulty in distinguishing these two vowels, and that the youngest listeners would display the largest perceptual distance between NURSE and THOUGHT. However, as shown in Figures 7.18 and 7.19 below, no strong effect of age on accuracy of NURSE-THOUGHT discrimination or identification is found.

[2]Minimal pair reading lists carried out with 24 of the listeners confirmed that no participant produced merged NURSE-THOUGHT vowels in their own speech

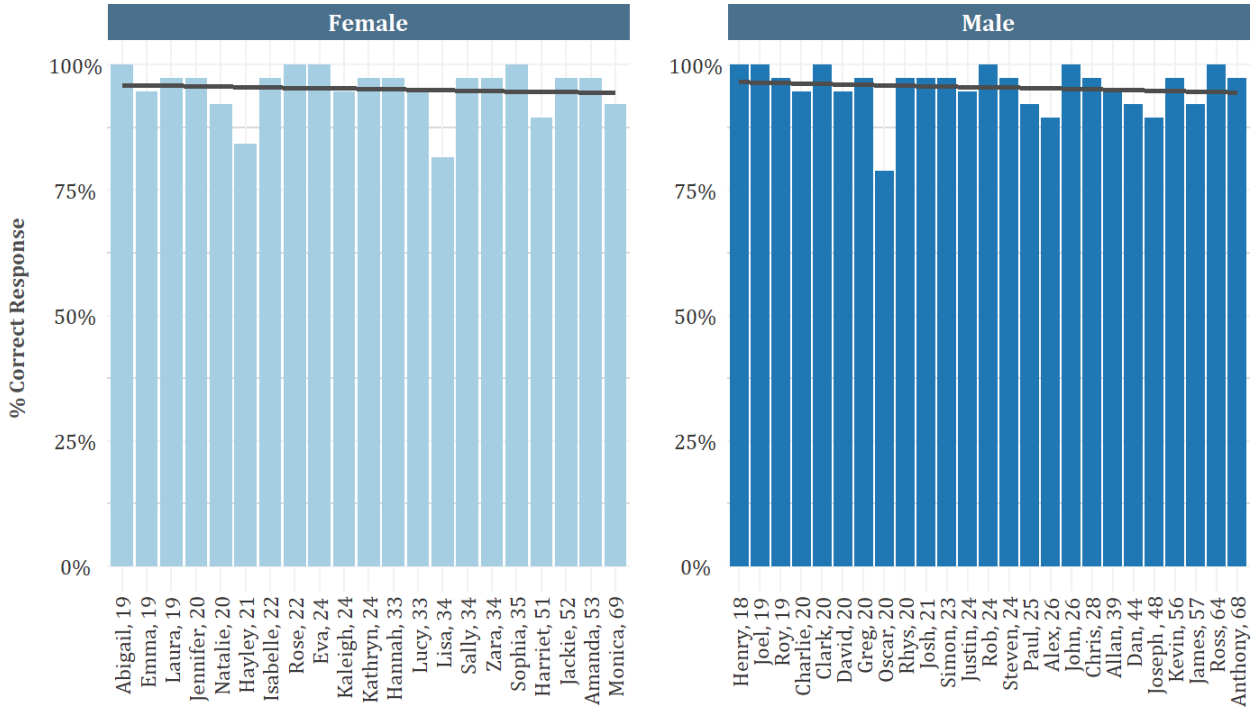


Figure 7.18: NURSE-THOUGHT discrimination by sex and age
(L-R youngest to oldest listener)

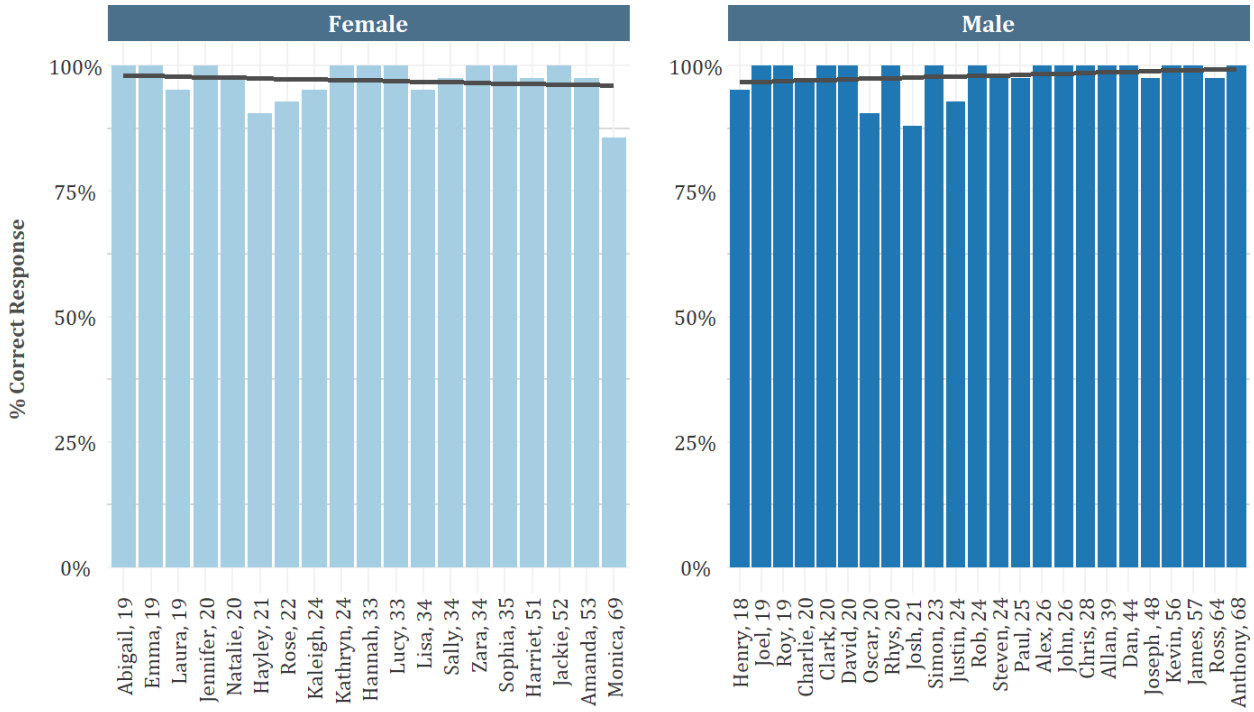


Figure 7.19: NURSE-THOUGHT identification by sex and age
(L-R youngest to oldest listener)

One possible explanation for listeners' marginally decreased accuracy in responding to NURSE-THOUGHT stimuli could be to do with the perception of NURSE-THOUGHT merging as a traditional feature of Tyneside English. While most recent research on Tyneside English reports an absence of NURSE-THOUGHT merging in the present-day speech community (Maguire 2007; Watt and Milroy 1999), many participants of the perception tasks showed an awareness of this feature. During the experiment session, several listeners commented that, although they themselves would not pronounce NURSE vowels as [ɔː]-like, e.g. *'shirt'* [ʃɔːt], they were aware that more 'broad Geordies' or speakers from 'more northern towns' such as Blyth and Ashington would pronounce these vowels identically. It appears that NURSE-THOUGHT merging has become a somewhat mythologised feature of Tyneside English, used by few contemporary Tyneside speakers but still closely associated with the Tyneside accent. As participants were recruited under the instruction that they were taking part in a perception experiment interested in the Tyneside accent, it may be possible that listeners were responding to some NURSE-THOUGHT trials based on how they thought a 'Geordie' would perceive the vowels, rather than how they themselves heard the stimuli.

7.5 Production Evidence

For a direct comparison of participants' perception and production of the GOAT and THOUGHT vowels, 24 of the 43 listeners who completed the perception tasks were also recorded reading a list of minimal pairs.

7.5.1 Comparing Accuracy with Production Evidence

Plotting the formant values of the GOAT and THOUGHT vowels from the minimal pair list readings, the majority of participants can be seen to show some degree of GOAT-THOUGHT overlap in the vowel space. This reflects the results obtained from plotting the overlap of GOAT and THOUGHT in the DECTE interview data, as shown in Chapter 5. For a comparison of participants' production and perception of the two vowels, Figures 7.20 and 7.21 show the F1-F2 values of GOAT and THOUGHT for the listeners who performed most and least accurately[3] across the two perception tasks. Pillai scores were also calculated for the participants who read the minimal pair lists. Used in the analysis of production data collected from DECTE, as presented in Chapters 4 and 5, Pillai scores indicate the overlap in the distribution of two

[3]This refers to the least and most accurate listeners in the perception tasks who also provided production evidence from minimal pair readings

vowels on a scale from 0-1 (scores closer to 0 indicate more overlap). Figures 7.20 and 7.21 also display the Pillai scores for the relevant speakers.

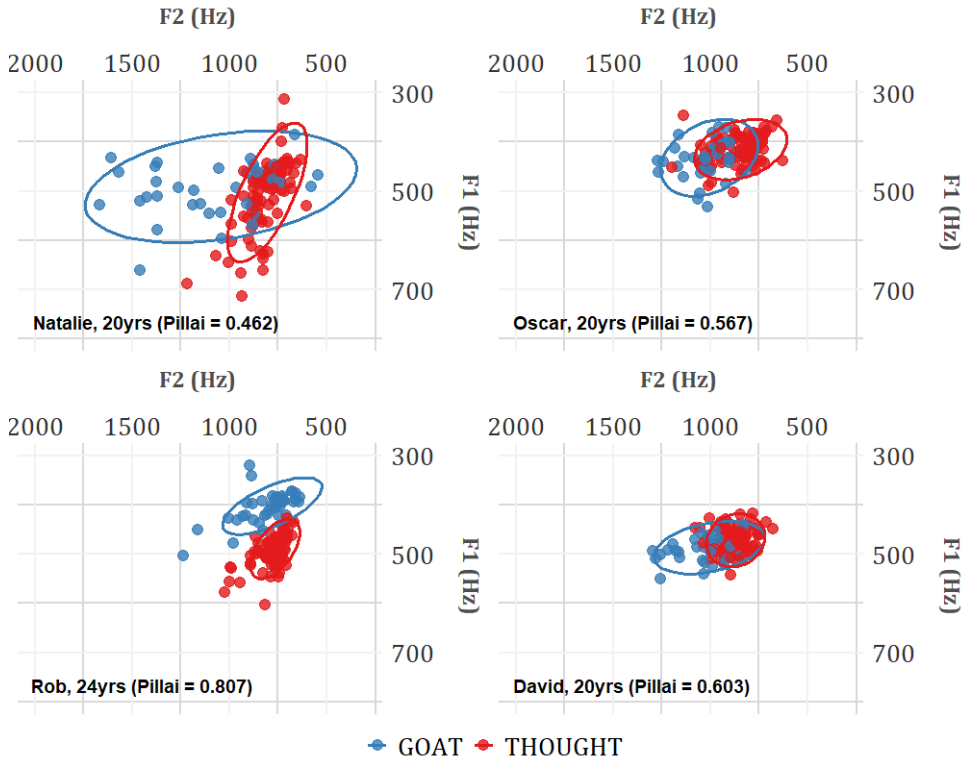


Figure 7.20: GOAT-THOUGHT Overlap of Least Accurate Listeners

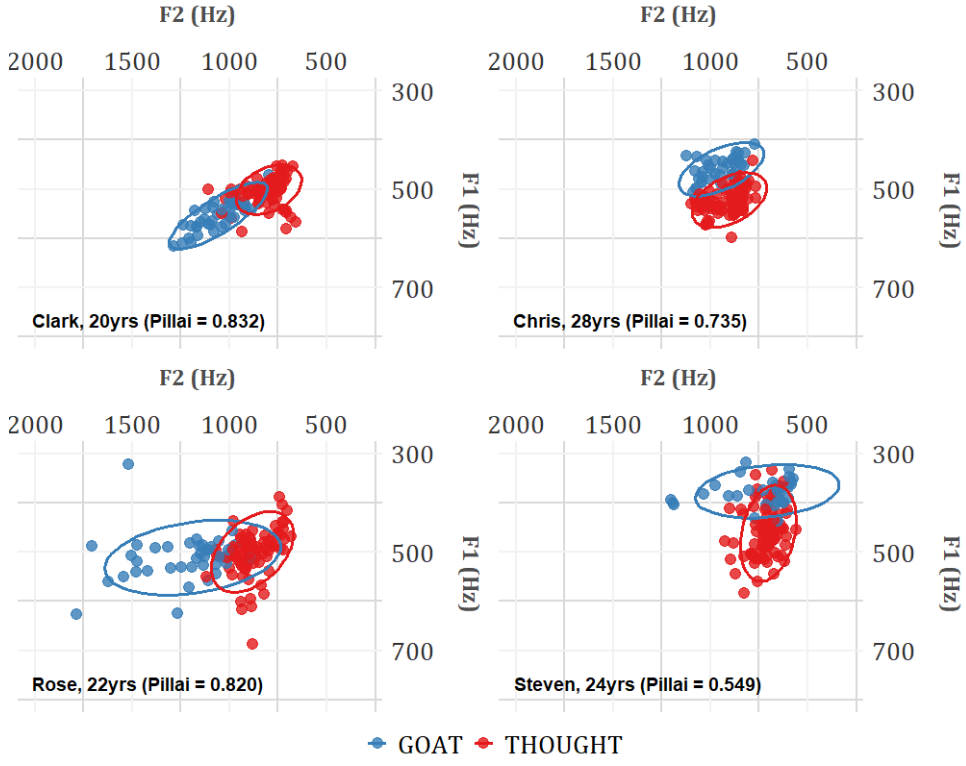


Figure 7.21: GOAT-THOUGHT Overlap of Most Accurate Listeners

Figure 7.20 shows the formant values of the GOAT and THOUGHT vowels along with the Pillai scores of the four participants who had the lowest accuracy rates across the discrimination and identification tasks. Three of the speakers show considerable overlap of the two vowels, and appear to show a lack of distinction between GOAT and THOUGHT in terms of vowel height. The Pillai scores, particularly for Natalie and Oscar, also indicate a good amount of overlap between the GOAT and THOUGHT vowels for these speakers. However, one of the least accurate listeners, Rob (24yrs), exhibits very little overlap of the vowels, with a GOAT vowel that is noticeably higher than THOUGHT. Additionally, Rob has a higher Pillai score than the other least accurate listeners, suggesting that he largely maintains a distinction between these two vowels in his own speech.

In Figure 7.21, which represents the most accurate listeners, all of the formant plots indicate some GOAT-THOUGHT overlap. Two of these participants, Clark (20yrs) and Chris (28yrs), show only slight overlap of the two vowels in the formant plots, with the Pillai scores indicating only a small amount of overlap between GOAT and THOUGHT for these speakers. Both Chris and Steven (22yrs) exhibit a higher GOAT vowel when compared with THOUGHT, although Steven's Pillai score suggests that he shows considerably more GOAT-THOUGHT overlap in the vowel space. The formant plot for Rose (22yrs) displays a lack of a height difference between GOAT and THOUGHT, although the Pillai score indicates that these vowels are largely distinct for this speaker. Many participants (i.e. Oscar, David, Natalie, Rose) are seen to show more variation in their GOAT vowel than in THOUGHT, particularly in terms of frontness. Similar results were also found in the acoustic analysis of the sociolinguistic DECTE interviews, as discussed in Chapter 5, where the distribution of GOAT tokens was shown to vary considerably in terms of vowel frontness.

Looking at Figures 7.20 and 7.21, there is an absence of an obvious difference in the realisations of GOAT and THOUGHT for the listeners who exhibited the highest and lowest accuracy rates in the perception tasks. It may be expected that the least accurate listeners struggle to distinguish GOAT from THOUGHT perceptually due to their lack of a distinction between these vowels in phonetic space. However, this does not appear to be the case. For these Tynesiders, simply looking at GOAT-THOUGHT overlap in the vowel space does not give a good indication of performance in the perception tasks, nor are accuracy rates a good predictor of GOAT-THOUGHT overlap in production.

7.5.2 Effect of Context on GOAT Production

A notable pattern emerged when plotting the formant values of participants' GOAT vowels in the minimal pair reading task. As shown in Figure 7.22, a number of participants were found to exhibit fronter GOAT realisations in open syllables, e.g. in *'low', 'flow', 'snow'* when compared to words in which the vowel is followed by a consonant. This pattern is found in both male and female speakers across the age range of the sample.

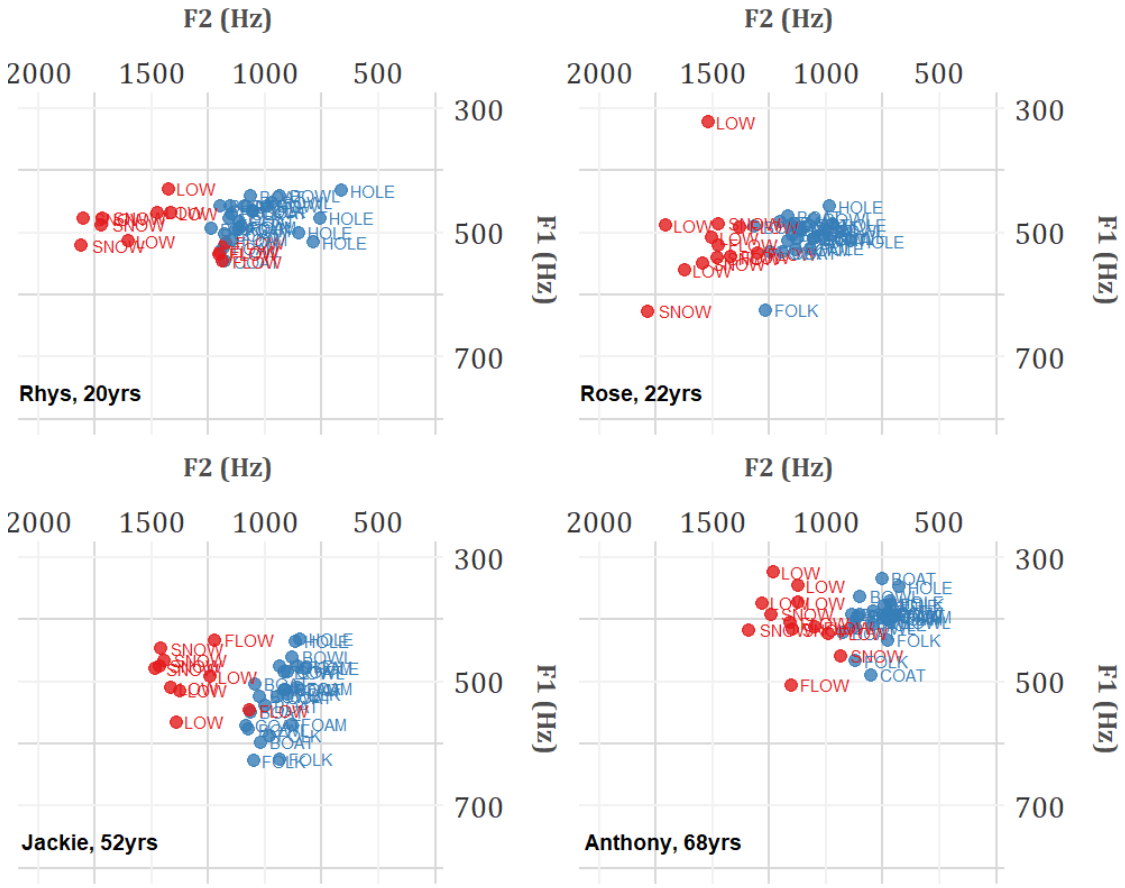


Figure 7.22: F1-F2 values of the GOAT vowels for four participants

The plots presented in Figure 7.22 indicate that, for the Tyneside participants, GOAT vowels in open syllables have higher F2 values, and thus fronter realisations than the same vowel in closed syllables. A similar pattern is reported by Labov *et al.* (2013: 45) in their study of GOAT-fronting in Philadelphia, in which speakers were found to show greater fronting of the vowel in open syllables. Watt and Tillotson (2001: 287) also find some evidence that GOAT is more prone to fronting in free syllables in their study of Bradford English speakers.

Figure 7.23 below shows the F2 values of each GOAT word read by participants in the minimal pair reading task. The plot shows that the words in which the GOAT

vowel appears in an open syllable have higher F2 values, and thus are more fronted compared to GOAT vowels in closed syllable environments. Additionally, the results of a t-test – performed with the `t.test` function in R – indicate that GOAT vowels are significantly fronter in free syllable environments ($p = <0.001$). Figure 7.23 also points to some general GOAT-fronting among these Tyneside participants. The plot indicates that GOAT is notably less fronted in following /l/ environments; a result often found in research on vowel fronting due to a retracting effect on the preceding vowel.

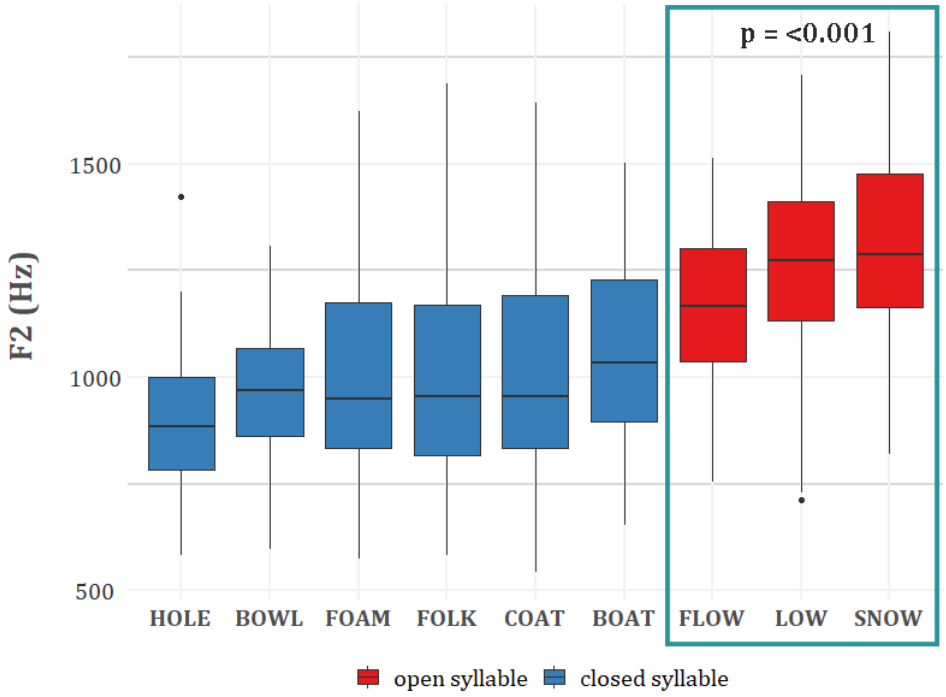


Figure 7.23: Vowel frontness by GOAT token

An effect of syllable type on GOAT realisations in Tyneside may help explain some of the results of the perception experiment. In both the discrimination and the identification tasks, listeners made fewer mistakes and had a higher chance of correctly responding to trials containing GOAT vowels in open syllable environments – as in the words *'low', 'flow', 'snow'* – indicating that listeners can more easily differentiate between GOAT and THOUGHT vowels when they are presented in open syllable environments. Therefore, the results of the perception tasks and the minimal pair readings suggest that these Tynesiders (a) show fronter pronunciations of GOAT in open syllables, (b) exhibit the largest amount of phonetic overlap between GOAT and THOUGHT when GOAT is in closed syllable position, and (c) are more accurate in their discrimination and identification of GOAT vs. THOUGHT vowels in trials containing open syllable GOAT stimuli.

One question arising from these results is whether it is possible that GOAT-THOUGHT merging in Tyneside may be the result of a conditioned merger in which only GOAT vowels in closed syllable position are merged with THOUGHT. However, given that the participants in the current study provided production data only for GOAT vowels contained in GOAT-THOUGHT minimal pair lists, just three distinct words in which GOAT appears in an open syllable – *'low', 'flow', 'snow'* – were available for analysis. To fully explore this potential effect of syllable type on GOAT-frontness, it would be necessary for participants to read a more comprehensive word list, which would include many examples of the GOAT vowel in a number of different phonetic and phonological environments. Although not possible with the data collected in the current study, the influence of free vs. checked syllables on Tyneside GOAT realisations may be a promising avenue of further research into Tyneside GOAT-THOUGHT merging.

7.6 Summary

This chapter has presented an analysis of the results of the discrimination and identification tasks which make up the perception component of the present study. Through an analysis of accuracy and reaction times, these findings show that the Tyneside listeners are consistently less accurate in responding to GOAT-THOUGHT trials, while also taking longer to respond to these trials. Such results indicate a degree of perceptual confusability between GOAT and THOUGHT, though it is not immediately apparent whether they can be taken as evidence of a Tyneside GOAT-THOUGHT merger. The following chapter will present a discussion of the findings of the main study, comparing the results from both the production and perception components of the main study to assess whether there is a merger-in-progress between GOAT and THOUGHT in the Tyneside speech community.

Chapter 8

Discussion

This chapter discusses the findings of the main study in line with the aims of the thesis. First, the primary goals and research questions of the thesis are restated in Section 8.1. Then, Section 8.2 recaps the results of the production and perception analyses, reviewing the evidence of a GOAT-THOUGHT merger in each individually before combining the production and perception findings to explore the overall indication of phonemic merger. Sections 8.3 and 8.4 then move on to present evidence to suggest the type and direction of the merger. The likelihood of a GOAT-THOUGHT overlap progressing to a full phonemic merger in Tyneside English is discussed in Section 8.5, along with evidence of any complications that may prevent this sound change from undergoing completion.

8.1 Recap of Goals and Research Questions

Before the results of the main study are discussed in further detail, the goals and research questions of the thesis, first presented in Chapter 1, are recapped here. The findings from both the production and perception analyses will be explored in response to the following aims and research questions:

8.1.1 Primary Thesis Goals

1. To assess the extent of overlap between the GOAT and THOUGHT vowels in the Tyneside speech community

2. To utilise multiple innovative methods of analysis and advanced statistical measures to investigate the merging of GOAT and THOUGHT in both speech production and speech perception

3. To determine whether the overlap between the GOAT and THOUGHT vowels is indicative of a current or ongoing phonemic GOAT-THOUGHT merger in Tyneside

8.1.2 Research Questions

1. To what extent does the evidence, from the production and perception analyses, support the idea of a GOAT-THOUGHT merger-in-progress in contemporary Tyneside English?

   i) Do the results obtained from the production data indicate convergence of the GOAT and THOUGHT vowels in phonetic space?

   ii) Do the results of the perception experiments point to a collapse of a phonemic GOAT-THOUGHT distinction?

   iii) Working within an apparent time framework, do the results show that GOAT and THOUGHT are becoming more merged over time?

2. If the findings of the main study are indicative of a merging of the GOAT and THOUGHT vowels, then:

   i) How do the results suggest that this merger developed (e.g. what type of merger is occurring)?

   ii) In which direction is the convergence of the merger (e.g. does GOAT move to coalesce with THOUGHT, or vice versa)?

   Else, if the results of the main study indicate that GOAT and THOUGHT are not merging, then:

   i) What may be preventing these two vowels from undergoing phonemic merger?

3. What is the nature of the relationship between the production and perception of the GOAT and THOUGHT vowels in Tyneside English?

   i) Is the amount of overlap between GOAT and THOUGHT in production reflected in perception?

   ii) Are the results consistent with previous accounts of the link between production and perception in situations of phonemic merger?

8.2 Evidence of Merger

In order to assess the evidence within the present study for a GOAT-THOUGHT vowel merger in contemporary Tyneside English, the results of the production and perception analyses must be discussed together and compared. This section first presents a summary of the findings of the production and perception investigations separately, before evaluating the overall evidence for the existence of a merger-in-progress in the speech community.

8.2.1 Summary of Production Results

The production component of this thesis utilised a number of different methods to examine the GOAT and THOUGHT vowels in present-day Tyneside speech, from auditory to acoustic analysis, making use of both static and dynamic data. Evidence indicative of a merging of the GOAT-THOUGHT vowels could be found in the results of several different measurement methods.

Firstly, formant plots showed considerable overlap of GOAT and THOUGHT for many of the Tynesiders. This was more evident for the females than the male speakers, which points to greater merging of these vowels among females. Analysis using Pillai scores, which calculate the overlap in distribution between two vowels, found that the majority of speakers scored less than <0.5, which indicates a lack of a clear distinction between these vowels in production. Dynamic analysis of vowel trajectories also suggested evidence of GOAT-THOUGHT merging among the speakers. In terms of vowel height, both young males and young females exhibited closer, more overlapping predicted trajectories of GOAT and THOUGHT than the older speaker groups. An examination of duration also found no significant difference between the two vowels. These results lend support to the argument that young Tynesiders, particularly females, are leading a sound change toward a merger of the GOAT and THOUGHT vowels.

The finding that the highest degree of overlap between GOAT and THOUGHT is manifest in the speech of females is consistent with Watt's (1998: 161) brief assessment of the merger. However, while the current study finds that young females exhibit the most overlap between these vowels, Watt (1998: 161) summarises that identical realisations of GOAT and THOUGHT were found predominantly in the speech of older and/or working-class females, as younger and middle-class females preferred to use the diphthongal [oʊ] variant of GOAT. Such a difference in results may be attributed to stylistic factors. The current study used sociolinguistic interview data while Watt focuses on minimal pair reading lists; the prestigious non-local [oʊ]

is likely to be found with increased frequency in more formal speech styles. The auditory analysis conducted on the DECTE data did show that young females had higher rates of [oʊ] usage than the middle and older female groups, but [oː] was still found to make up over 80% of GOAT tokens. Thus, the production results indicate that young females primarily use [oː] variants of GOAT, and also that they show greater overlap of GOAT and THOUGHT than any other speaker group.

Not all of the results found in the production analysis, however, neatly point toward a merger-in-progress. The Pillai scores in particular indicate that the young males in the sample typically have the largest distinction between GOAT and THOUGHT in phonetic space. Vowel trajectory comparisons also indicated that the young males have a greater distance between GOAT and THOUGHT, in terms of vowel frontness, than any other speaker group. This complicates the overall picture which suggests that a GOAT and THOUGHT are merging in apparent time. These results may be somewhat explained by the findings from the auditory analysis which show that the young males are the most variable speaker group in their use of different realisations of the GOAT vowel. Since this group produced the least amount of backed monophthongal [oː] vowels, there were fewer opportunities for overlap with [ɔː].

As the young males in Tyneside use a wider range of GOAT variants – and use them more frequently – compared with other speaker groups, then it may be the case that these speakers are not participating in the same GOAT-THOUGHT merging sound change as the young females in the speech community. However, the majority GOAT variant used by young male speakers is still the back monophthongal [oː]. There is some evidence within the results of the production analysis to suggest convergence of the GOAT and THOUGHT vowels among young Tyneside males. While GOAT and THOUGHT are both back vowels differentiated predominantly in terms of height, formant plots and analysis of vowel trajectories indicate considerable F1 overlap of the two vowels in the young male speakers. This suggests that, for these speakers, GOAT and THOUGHT now occupy a similar height position in the vowel space. So, while the young males appear to show a continued distinction in production between the two vowels in terms of frontness, it does seem that GOAT and THOUGHT have coalesced in the dimension of vowel height. It cannot be concluded conclusively, therefore, that young males on Tyneside are increasing the distinction between GOAT and THOUGHT, or that they are not at all participating in a converging of the GOAT-THOUGHT in the speech community.

8.2.2 Summary of Perception Results

As presented in this thesis, the investigation into the perception of the GOAT and THOUGHT vowels on Tyneside consists of both a discrimination and an identification task. The results of these tests were then analysed using different measures, including accuracy, reaction time, and sensitivity. From these analyses, some evidence to suggest a perceptual GOAT-THOUGHT overlap among the Tyneside listeners did emerge.

First, looking at accuracy, no speaker was able to discriminate GOAT from THOUGHT with 100% accuracy, and only one of the 43 listeners reached 100% accuracy in the identification task. This could be indicative of a community wide difficulty in distinguishing these vowels from one another, and in categorising these vowels correctly. However, accuracy measures are somewhat difficult to interpret when looking for conclusive evidence of a perceptual merger. As discussed in Chapter 2, discrimination tasks – often termed 'commutation tests' in sociolinguistic literature – have long been used to examine mergers in perception. Labov (1994: 356) proposes that an accuracy rate of 100% is suggestive of a phonemic distinction, while an accuracy rate of around 50% indicates phonemic merger. As the listeners in the current study range in accuracy from 45% to 95%, there is no clear cut-off point at which listeners may be judged to be perceptually 'merged' or 'non-merged'. Given that no speakers performed at 100% on the discrimination task, compounded with the presence of two listeners performing at or below chance level (i.e. $\leq$50%), the accuracy results do suggest a degree of confusability between GOAT from THOUGHT in Tyneside, and this may be taken as evidence to support the idea of a GOAT-THOUGHT merger-in-progress in the speech community.

The perception tasks were designed to allow for a comparison of GOAT-THOUGHT discrimination and identification with other vowel pairs. As such, the experiments included stimuli of filler vowel pairs, in addition to vowel pairs contrasting the NURSE and THOUGHT vowels, which were previously found to be merged – or at least near merged – in the Tyneside region. Mixed-effects logistic regression models show that, for both tasks, accuracy was significantly lower on GOAT-THOUGHT trials when compared with trials containing the other vowel pairs. Sensitivity measures which took into account the *Hit Rate* and the *False Alarm Rate* of each participant in each perception task showed that all listeners had lower *d'* values in discriminating GOAT from THOUGHT than in discriminating other vowels. This again indicates that listeners were least accurate when responding to GOAT-THOUGHT trials. Additionally, analyses of reaction time data, using

mixed-effects linear regression models, found that listeners were significantly slower in their response to GOAT-THOUGHT trials than when responding to trials containing NURSE-THOUGHT or filler vowel pairs. This increased response time indicates that listeners encountered more difficulty in distinguishing and identifying the GOAT and THOUGHT vowels than any other vowel pair presented in the tasks. Such results do suggest some degree of phonemic converging of GOAT and THOUGHT in Tyneside English, and, when combined with the accuracy results, strengthen the case for a community wide GOAT-THOUGHT merger.

While there is evidence to suggest a GOAT-THOUGHT merger in perception, not all of the results from the perception experiment point neatly to this conclusion. Firstly, the results do not find any effect of social factors on perception of GOAT and THOUGHT. If there is a sound change in progress toward a merger of these vowels on Tyneside, it might be expected that certain listener groups will show greater signs of merger than others, e.g. younger listeners might have decreased accuracy and increased reaction times due to difficulty in distinguishing GOAT from THOUGHT. This does not appear to be the case in the present study. Perhaps best exemplified by the performance of the middle female age group in the discrimination task, female listeners of a very similar age were found to occupy both ends of the accuracy spectrum. Thus, rather than patterns of GOAT-THOUGHT perception relating to social groups, the results show a considerable amount of variation between individuals.

Beyond social factors, the results of the perception experiments did point toward an effect of phonological context on accuracy. In both the discrimination and identification tasks, accuracy rates were increased on trials presenting GOAT and THOUGHT vowels in open syllable contexts, e.g. as in the words *'flow'* and *'flaw'*. This effect of environment on listener accuracy was further indicated by logistic regression models which incorporated word pair as a random effect, as calculations of probability from the model output suggested that participants had a higher chance of correctly responding to trials in which the GOAT and THOUGHT vowels were not followed by consonant. Such a finding indicates that the vowels in words like *'flow'* and *'flaw'* are more perceptually distinct, for these Tyneside speakers, than the same vowels in words such as *'coat'* and *'caught'* or *'hole'* and *'hall'*.

8.2.3 Comparison of Production and Perception Results

The current study has made use of data investigating the GOAT-THOUGHT merger in both speech production and speech perception. Individually, the results of the

the production and the perception analyses provide some evidence of merger. For a more comprehensive examination of this potential merger, the findings from both production and perception must be discussed in relation to one another. The production data obtained from DECTE may be compared to the perception data in order to look for sociolinguistic, community-wide patterns of GOAT-THOUGHT merging. However, as they made use of different sets of participants, a comparison of the production data from DECTE and the perception data from the experimental tasks would not allow for an analysis of the relationship between production and perception on an individual level. To resolve this, as outlined in Chapter 6, a subset of participants in the perception tasks were also recorded reading minimal pair lists. Therefore, with the data collected in the main study, it is also possible to investigate the production-perception relationship within individuals, as well as within the wider speech community.

In looking at the data showing individuals' production and perception of GOAT and THOUGHT, the relationship between production and perception does not appear to be neatly symmetrical. Shown in Figures 7.20 and 7.21 (Chapter 7), the formant values of GOAT and THOUGHT were plotted for the eight participants who performed most and least accurately in the perception tasks. Of the four least accurate listeners, three had more overlap in the distribution of GOAT and THOUGHT tokens, along with lower Pillai scores, than the most accurate speakers. This result would suggest that, typically, the participants who struggled the most on the perception tasks also had increased convergence of these vowels in production. However, one of the least accurate listeners was found to show minimal overlap of GOAT and THOUGHT, and, conversely, one of the most accurate listeners was found to have considerably more GOAT-THOUGHT overlap than the other participants who performed best in the perception task. This finding complicates the results, and indicates that, for a number of speakers, pronunciations of GOAT and THOUGHT are not a reliable predictor of accuracy in distinguishing and categorising the same vowels in perception tasks. Results therefore show variability in individuals' relationship between production and perception with regards to a GOAT-THOUGHT merger; for some production approximately reflects perception, for some the vowels appear more merged in production, and for others the vowels seem to show greater overlap in perception.

There are a couple of factors which may account for asymmetry between individuals' production and perception of GOAT and THOUGHT. Firstly, as noted above, listener accuracy in the perception tasks ranged on a gradient between 45% and 95%. This meant that, while there was variation in accuracy between

listeners, there was no clear cut-off point by which to determine whether participants exhibited signs of a perceptual GOAT-THOUGHT merger. Given that the perception results cannot neatly be interpreted as showing two distinct groups of 'merged' and 'unmerged' listeners, it can be difficult to directly compare the production with the perception results. This issue was curbed, to some extent, by comparing the production and perception results of only those participants who showed the lowest and highest accuracy rates on the perception tasks, i.e. those who displayed the greatest evidence of either a GOAT-THOUGHT split or a GOAT-THOUGHT merger. However, one complication in the present study is that not all of the listeners who took part in the perception task provided production data via minimal pair readings. As a result, some of the listeners at the most extreme ends of the accuracy spectrum could not be included in any analyses of a relationship between production and perception. The groups of 'least accurate' and 'most accurate' listeners, used in the comparison of individuals' production and perception, therefore included some participants who performed notably better/worse on the perception tasks than the others in the group. When compared with the production results, then, these participants may appear to show less, or greater, evidence of a GOAT-THOUGHT merger in their speech than their grouping of a 'least accurate' or 'most accurate' listener would suggest. This may provide a methodological explanation as to why some individuals' productions of GOAT and THOUGHT did not clearly match their perception of the same vowels.

A comparison of the perception results and the minimal pair readings provides evidence to suggest that syllable type has an effect on both the production and the perception of the GOAT and THOUGHT vowels. Just as listeners had higher rates of accuracy in responding to open syllable stimuli, formant analysis of GOAT tokens found that the participants produced significantly fronter GOAT vowels in free syllable environments. These findings indicate that speakers produce a greater distance between GOAT and THOUGHT in open syllables than in closed syllables due to a fronting of the GOAT vowel. This increased distinction between GOAT and THOUGHT in Tyneside speech may help to explain why participants performed better on trials containing open syllables; listeners are perhaps aware of the larger difference between the two vowels in free syllable contexts, and therefore have a larger perceptual distinction between GOAT and THOUGHT in this specific context.

In looking for sociolinguistic patterns of GOAT-THOUGHT merging in Tyneside, a comparison of the results of the DECTE production data with the perception task results finds little similarity between the two. As previously discussed, analyses of the DECTE interview data largely found that young female speakers exhibited the

greatest evidence of a merger between the GOAT and THOUGHT vowels. Results also indicated that young males had a larger distinction between the two vowels in phonetic space than any other speaker group due to the fronting of GOAT. However, the results of the perception tasks did not show any clear social patterns. From the results of both the discrimination task and the identification task, there is a lack of evidence to suggest that younger or female listeners show a more advanced perceptual merger of GOAT and THOUGHT than the other listener groups. Neither did the results indicate that young males had a greater perceptual GOAT-THOUGHT distinction. As such, the social patterns of Tyneside GOAT-THOUGHT merging found in the production data are not clearly reflected in the perception results.

It is perhaps not too surprising to find asymmetry between the production and perception results. Discussed in Chapter 2.7, there have been several examples of merger cases in which production and perception, in the individual and/or in the speech community, do not neatly reflect one another. This asymmetry may go either way, with some researchers finding that production is in advance of perception in situations of merger-in-progress (Thomas and Hay 2005; Arnold 2015), while others find that perception appears to lead production (Bowie 2000; Freeman 2019). Given that social patterns of GOAT-THOUGHT merging are visible in the production but not the perception data, it is possible to interpret the results of the present study as indicative of a GOAT-THOUGHT merger which is more advanced in production than perception. This could be consistent with Janson's (1983) theory that, in phonetically gradual sound changes, production is ahead of perception. He proposes that, in order for everyone in the speech community to continue to understand each other, perceptual categories cannot change radically from one generation to the next; while younger speakers may produce more innovative forms, their perception cannot stray too far from that of their parents' generation (Janson 1983: 31). However, as the perception results overall do suggest that listeners struggle somewhat in perceiving GOAT and THOUGHT stimuli, and that there is clear variation between listeners in terms of accuracy (e.g. a range from 45% to 95% in the discrimination task), the results of the present study perhaps cannot be so easily interpreted as showing a GOAT-THOUGHT merger led by production.

That the social patterns found in the DECTE production data are not reflected in the perception results may be attributed to methodological factors. More specifically, the lack of a comprehensive analysis of social class may have influenced these results. Some of the complications with analysing socioeconomic class in the present study were detailed in Chapter 4.3.3.2, and are discussed at further length in Section 9.2 below. As the convergence of the GOAT and THOUGHT vowels in

Tyneside appears to be a sound change that is occurring below the level of conscious awareness, it would be expected, as per Labov's (1966; 1990) curvilinear principle, for those speakers in the centre of the social class spectrum (e.g. the upper-working and lower-middle classes) to lead this change. While a brief analysis using Pillai scores suggested that overlap of GOAT and THOUGHT decreased as socioeconomic status increased, a full analysis including social class was not possible given issues related to sample size and stratification imbalances. It is therefore possible that social class may affect the overlap between GOAT and THOUGHT on Tyneside – either in production, perception, or both – but that the current study was not able to capture such patterns. Future analyses of Tyneside GOAT-THOUGHT merging which are able to fully incorporate socioeconomic status as an extralinguistic predictor may therefore find a clearer pattern in the effects of social class on merging, and may in turn find less of a mismatch between how these vowels are produced and the perceived across the speech community.

There is another possible explanation for the asymmetry in the production and perception of GOAT and THOUGHT in Tyneside, both at the individual and the community level. In Arnold's (2015) study of pre-/l/ vowel mergers in Ohio, she found that (a) participants' production of the mergers did not always reflect their perception, and (b) social patterns found within the community in the production data were not present in the perception data. This, therefore, is a similar case to the findings of the study presented in this thesis. Arnold (2015: 9) hypothesises that the apparent lack of a symmetrical production-perception relationship in her results may be due to listeners making use of perceptual cues other than vowel quality to distinguish between vowel classes. As the production component of the current study focuses mainly on vowel quality, with a brief look at vowel duration, it is possible that other factors (e.g. lip rounding) not included in the analysis are helping to maintain a distinction between GOAT and THOUGHT in production. If this were the case, then listeners could still utilise such cues to their advantage in the perception task, but, given the omission of an analysis of these cues, the production data would indicate that these vowels are pronounced more similarly than they actually are in the speech community. Therefore, it may also be the case that asymmetry in the production and perception of GOAT and THOUGHT in Tyneside English may be the result of unexplored factors, outside of vowel quality and duration, leading to an overstatement of the similarity in the pronunciation of these two vowels.

8.3 Type of Merger

The literature tends to agree, as detailed in Chapter 2, that phonemic mergers may arise via three distinct mechanisms; by-approximation, by-transfer, or by-expansion (Harris 1985; Herold 1990; Labov 1994). As the results of both the production and perception analyses indicate a merging of the GOAT and THOUGHT vowels in Tyneside English, it may also be possible to determine the mechanism behind this GOAT-THOUGHT convergence.

Mergers-by-transfer, which may occur either lexically (Trudgill and Foxcroft 1978; Labov 1994) or phonologically (Dinkin 2016), are typically described as a phonetically abrupt processes. Conversely, the transference of lexical items (or sounds in particular phonetic environments, in the case of phonological transfer) from one category to another happens gradually.

There is little to no evidence in the current study to suggest that a Tyneside GOAT-THOUGHT merger is taking place by-transfer. First, the evidence from the acoustic analysis of the production data suggests a more gradual coalescence of the vowels in phonetic space, thus would not be consistent with the 'phonetically abrupt' nature of a merger-by-transfer. Not only is there evidence which points to a phonetically gradual change, but there is also a lack of evidence indicating that word or phonological environment has an effect on the merging of the GOAT and THOUGHT vowels. Statistical models run on the production data found no significant effect of word frequency or phonological context on the realisation of the vowels under investigation. The one exception to this was that GOAT was found to be significantly more fronted word finally (e.g. *'no, snow'*), but this effect does not result in increased GOAT-THOUGHT overlap, rather, the fronting effect increases the distance between the two vowels. Finally, Labov (1994: 321) posits that mergers-by-transfer occur most often when a particular variable is subjected to social evaluation in the speech community. It is perhaps unlikely that Tynesiders have any awareness of an ongoing GOAT-THOUGHT merger in the local speech community, given that this sound change is not even discussed in most recent linguistic accounts of the Tyneside accent. Although not formally questioned about their awareness of a GOAT-THOUGHT merger in Tyneside, it is worth noting that many of the participants of the perception task expressed, post-experiment, that they were previously unaware that some Tynesiders may pronounce these two vowels identically. For these reasons, it does not appear that GOAT-THOUGHT merging on Tyneside is an example of a merger-by-transfer.

Cases of merger-by-expansion are noted for the speed at which the speech community can move from having two distinct sounds to showing a complete

merger of the two sounds. This is most notable in studies of different generations of family members, as in Herold (1990) and Johnson (2010), who discovered examples of parents producing distinct vowels where their children exhibited a merger. Another signal of mergers-by-expansion is that the fully merged vowel occupies a much larger, 'expanded' area, approximately spanning the entire space that the two previously distinct vowels took up in phonetic space.

The results of the current study are not indicative of a merger-by-expansion on Tyneside. As discussed, the findings from the production analysis indicate the gradual movement in phonetic space of one or more of the vowels toward the other. Moreover, analysis using formant plots and Pillai scores indicated that none of the speakers exhibited a full phonetic merger of the GOAT and THOUGHT vowels, and that certain speakers – and speaker groups – showed considerably more vowel overlap than others. This does not appear to fit the pattern of a merger-by-expansion, in which there is typically no intermediate stage between complete distinction and complete merger. Additionally, Herold (1990; 1997) proposes that mergers-by-expansion arise most frequently in high contact situations, where speakers with two distinct sounds come into contact with a language variety that exhibits a merger of these sounds. This would not accurately describe the situation on Tyneside, and, therefore, the findings of the present study indicate that the GOAT and THOUGHT vowels have not undergone a process of merger-by-expansion on Tyneside.

The final way in which mergers may develop is by the mechanism of merger-by-approximation. This type of merger is described as being phonetically gradual and lexically abrupt (Harris 1985; Labov 1994). As such, all words and phonological environments will be affected at the same time. In mergers-by-approximation, one phoneme may drift toward another, or they may both show movement toward each other to meet in the middle.

Given the findings from the present study, it appears most likely that GOAT and THOUGHT are coming to merge on Tyneside via a merger-by-approximation. Perhaps the biggest indication of a merger-by-approximation in the results is that they show evidence of movement of the vowels in phonetic space. This is only possible in by-approximation cases. Secondly, for much the same reason that the findings were not consistent with a merger-by-transfer, the lack of an effect of word or phonological environment on the vowels also points to a merger-by-approximation. In Harris' (1985: 310) diagram of a merger-by-approximation, presented as Figure 2.1 in Chapter 2, there is a stage in the process toward complete merger in which the two merging phonemes show variable overlap both phonetically and phonemically.

Looking at the results of both the production and perception analyses, there is an indication that many Tynesiders may be at this stage – which Harris labels Stage III – in the merger-by-approximation process. As discussed, the findings of the acoustic analysis illustrate that many speakers show a considerable GOAT-THOUGHT overlap in their speech production. The results of the perception tasks are also suggestive of some phonemic overlap of the GOAT and THOUGHT vowels, as listeners were found to be significantly less accurate – and significantly slower to respond – in their discrimination and identification of these two vowels compared with the other pairs of vowels included in the experiments. So, the production and perception components of this thesis show evidence of a GOAT-THOUGHT merger-by-approximation in Tyneside English, with the findings indicating, more specifically, that many speakers may be at Stage III in Harris' (1985: 310) schema of the by-approximation process of merger.

8.4 Direction of Merger

As the findings of the current study point to a merging of the GOAT and THOUGHT vowels, it may also be possible to determine the direction of this merger-in-progress. In other words, do the results suggest that GOAT is moving toward THOUGHT, that THOUGHT is merging to the position of GOAT, or are the two vowel coalescing in an intermediate position in the vowel space?

Discussed previously in Chapter 3, the literature reports that, in more contemporary Tyneside speech, the GOAT vowel is typically pronounced with the back monophthong [oː] (Watt 1998; Watt 2002), while THOUGHT is typically realised as the back open-mid [ɔː] vowel (Watt 1998; Beal *et al.* 2012). So, as distinct vowels, it would be expected that these two back vowels are distinguished predominantly by height, with GOAT occupying a higher position in the vowel space than THOUGHT. Therefore, in order for these two vowels to come together, there would need to be movement of one or more of the vowels in the direction of the other in the height (F1) plane.

The results from the production analysis primarily indicated that the GOAT vowel has lowered toward the position of THOUGHT in Tyneside speech. This was first signalled in the formant plots (Figures 5.1 and 5.2) which showed a GOAT-THOUGHT overlap in vowel height for many of the speakers. Looking at the position of each vowel, the THOUGHT vowel did not appear to be raised from its expected open-mid position, and the F1 values of the THOUGHT tokens were not raised in comparison with formant values reported in previous accounts of

Tyneside monophthongs (Ferragne and Pellegrino 2010). Rather, the formant plots suggested that the GOAT vowel had undergone a certain amount of lowering, leading to a phonetic overlap between GOAT and THOUGHT in terms of vowel height.

Generalised Additive Mixed Models (GAMMs), which output the predicted trajectories of the GOAT and THOUGHT vowels for different speaker groups, also found evidence of GOAT lowering among the Tyneside speakers. For both males and females, younger speakers were shown to exhibit considerably lower GOAT vowels than the middle and older speaker groups. In the case of the young males, GOAT had become lowered to such an extent that the GOAT and THOUGHT vowels were no longer significantly different heightwise. The figures illustrating the GAMM trajectories (Figures 5.11 and 5.12) also gave some suggestion of slight THOUGHT raising over time, which was most noticeable in comparing the young and middle speaker groups with the oldest speakers in the sample. This indicates that THOUGHT may have been subject to some slight raising in Tyneside English over the years, but that, most recently, the biggest change leading to a merging of the two vowels involves a lowering of the GOAT vowel.

Given the results illustrated in the formant and trajectory plots, the production results appear to show that younger speakers have a greater GOAT-THOUGHT overlap in height than the older speaker groups due to a lowering of the GOAT vowel in apparent time. However, the mixed-effects regression models run on the data did not clearly reflect these results. The model did not find that GOAT was significantly lowered for the younger speakers in comparison with the other speaker groups, but it did find that the older speaker group had significantly lower THOUGHT vowels than the younger speakers. As such, the model suggests that THOUGHT has undergone some slight raising in Tyneside, which is consistent with the results of the GAMM output.

Overall, the findings of the production analysis suggest that GOAT and THOUGHT overlap in terms of height for many of the Tyneside speakers, and that this is largely due to a lowering of the GOAT vowel to the position of THOUGHT. Again, there is some evidence to suggest that THOUGHT has undergone some slight raising over time, but it appears that GOAT-lowering is largely responsible for an increased phonetic overlap between these vowels in younger Tyneside speech.

Turning now to the perception analysis, one potential method of assessing the direction of the merger was to look at the the rates of misidentification of the GOAT and THOUGHT vowels. The results of the identification experiment found that, among the incorrect responses, listeners misidentified THOUGHT stimuli as belonging

to the GOAT lexical set over 75% of the time, while GOAT was misidentified as THOUGHT less than 25% of the time. Such a large degree of asymmetry in incorrect responses could indicate that the listeners perceive the THOUGHT vowel as sounding, traditionally, more GOAT-like. This may suggest, then, that the listeners have a raised phonemic boundary for the THOUGHT vowel, which in turn indicates that, at least in terms of perception, THOUGHT is raising toward GOAT. If THOUGHT is perceptually closer to GOAT in Tyneside, then this sound change would perhaps be more consistent with Labov's (1994: 123) Principle I of chain shifting, which states that long vowels – whether back or front – have a tendency to raise. However, as these results from the perception tasks do not accurately mirror the patterns found in the production data, it appears that the production and perception findings are not consistent in terms of directionality.

There are a number of factors to take into consideration with the examination of merger direction in the perception results. First, it is only the results of the identification task that show such skew in the misidentification of the vowels; there is no evidence to suggest that listeners were more prone to answer that THOUGHT stimuli sounded the same as the GOAT stimuli than vice versa in the 'same-different' discrimination task. Additionally, as mentioned in Chapter 7, the listeners generally performed quite well on the identification task, exhibiting higher accuracy rates than in the discrimination task. As only 10% of responses to GOAT-THOUGHT trials were incorrect responses, this is a relatively small sample size. Furthermore, while looking at the percentages of incorrect responses for each vowel may be able to provide some indication of the direction of a merger, it is clear that this is a rather rudimentary method of measurement. Outlined in Chapter 2, analysis methods which make use of resynthesized vowel continua, such as in Janson and Schulman (1983), are able to accurately pinpoint phonemic boundaries in perceptual space. Such methods are therefore advantageous in assessing changes to perceptual boundaries, as well as examining the direction of these changes. While this type of analysis was not possible in the current study given the experimental design, these factors do indicate that a more robust method is necessary in order to accurately assess the direction of a potential merger in perceptual space.

8.5 Is this a Merger-in-Progress?

It is clear from the results of the production and perception analyses that the GOAT and THOUGHT vowels have not yet fully merged in Tyneside English. While considerable overlap of these vowels was exhibited by many of the speakers, none

were shown to have a complete phonetic overlap of GOAT and THOUGHT in the speech production analysis. Similarly, while the results of the perception tasks indicated that a number of listeners struggled to discriminate and identify GOAT and THOUGHT, the majority performed above chance level on the two tasks. Such findings suggest that the situation of GOAT and THOUGHT resembles that of Stage III in Harris' (Harris 1985) diagram of the development of mergers-by-approximation, as shown in Figure 2.1. That is, there is some overlap in the phonetic realisation of these vowels, and in the perceptual categorisation of these vowels, but both GOAT and THOUGHT remain distinct.

As there is evidence to suggest a merging of the GOAT and THOUGHT vowels, and as this merger has not yet completed, it is tempting to conclude that this situation on Tyneside is an example of a merger-in-progress. It seems reasonable to believe, in such cases of overlap between two vowels in production and perception, that this coalescence will continue until, finally, the two vowels become one in the phonemic inventory of a speaker or speech community. Yet, this is not guaranteed; this is highlighted in the case of the historical MEAT-MATE near merger, discussed in Section 2.6.1, whereby, despite evidence of overlap, these vowels followed a path toward reversal rather than completion to full merger (Labov *et al.* 1972). Given such examples, Dinkin (2016: 165) argues that researchers must 'demonstrate that the change is likely to go to completion' in order to 'correctly' employ the term of 'merger-in-progress'. The likelihood of this GOAT and THOUGHT overlap progressing to a full phonemic merger in Tyneside English must therefore be evaluated.

8.5.1 GOAT Variation

One complicating factor in the completion of a GOAT-THOUGHT merger in Tyneside English is the presence of multiple variants of the GOAT vowel in the speech community. As Watt (1998: 161) states, homophony between theGOAT and THOUGHT vowels is dependent upon the usage of a back, peripheral monophthongal pronunciation of GOAT. The auditory analysis of GOAT presented in Section 5.2 of this thesis found that although monophthongal [oː] pronunciations were by far most frequent, other variants of the GOAT vowel were still present among the speaker sample. This raises the question of whether a GOAT-THOUGHT merger is able to progress toward completion, considering that one of the merging vowels has several realisations that are still used in the Tyneside speech community.

The results of the auditory analysis found examples of four different GOAT variants in the DECTE data. There was no statistical evidence to suggest that the

use of these variants is subject to any phonological or lexical conditioning. It may be the case that the coalescing GOAT and THOUGHT vowels will not be able to fully merge if these other variants of the GOAT vowel remain in the speech community. Two of the variants, [ʊə] and [ɵː], were found to be very infrequently used in the DECTE speaker sample. These variants are those which are the most localisable to the Tyneside area, and, as Watt (1998; 2002) found a growing preference on Tyneside for 'supralocal' forms such as [oː], it is possible that [ʊə] and [ɵː] may, in the future, fall out of usage altogether. Given their infrequency in the DECTE interviews examined in this study, and the overwhelming preference for the more generally northern [oː] variant, it is unlikely that local pronunciations such as [ʊə] and [ɵː] would pose a threat to the advancement of a GOAT-THOUGHT merger in future. However, the case of [ɵː] is somewhat complicated by evidence of GOAT-fronting in the region, as will be discussed in further detail below.

The second most favoured GOAT variant, after [oː], was the nationally used closing diphthong [oʊ]. Although the auditory analysis found that this was used most often by the upper-middle class speakers, tokens of [oʊ] were present in the interview recordings of speakers from across the socioeconomic class scale. This differs from Watt's (1998; 2002) findings that only middle-class speakers use [oʊ], suggesting a general increase in the use of this variant on Tyneside in recent years. Pronunciations of [oʊ] are unlikely to completely disappear from the Tyneside region, on account of its perceived prestige and its usage in many other British varieties of English. Nevertheless, given the dominance of [oː] in the present study, compounded by Watt's (1998: 288) assertion that it would be unexpected for Tyneside English to show levelling 'in the direction of prestige southern norms', the general northern [oː] variant is likely to remain as the primary variant of the GOAT vowel for the foreseeable future.

Results suggesting that [oʊ] is the second most frequent GOAT variant in Tyneside may have implications for the possibility of a phonemic GOAT-THOUGHT merger in the region. As the findings of the production analysis indicated that upper-middle class speakers were those most likely to use [oʊ], combined with the Pillai scores showing that upper-middle class speakers exhibited a larger distinction between the two vowels, it could be hypothesised that only working-class and lower-middle class Tynesiders are eligible for a GOAT-THOUGHT merger. However, as previously explained, the effects of social class on GOAT-THOUGHT overlap could not be fully explored in this thesis. The results of the auditory analysis also suggested that young speakers have greater usage of [oʊ] than the older speaker groups, and that, while the upper-middle class used this variant most frequently, it was still used to

some degree by speakers from the other socioeconomic groups. It may be the case, then, that a GOAT and THOUGHT overlap could never progress to a full phonemic merger on Tyneside, due to the existence of [oʊ] – in addition to the majority variant [oː] – in a speaker's own system, or within the wider speech community.

8.5.2 GOAT-Fronting

The completion of a GOAT-THOUGHT merger in Tyneside English may be thwarted by a competing sound change in the region. Fronting of the GOAT vowel is a relatively recent sound change which has been reported in many varieties of English across the globe, with reports in the UK from Milton Keynes and Reading (Williams and Kerswill 1999), York (Haddican *et al.* 2013), and Manchester (Baranowski 2017). The ubiquity of this sound change is often attributed to Labov's (1994: 116) third principle of chain shifting, which states that 'back vowels move to the front'. The majority of varieties affected by GOAT-fronting have a diphthongal realisation of the vowel. Docherty (2010: 67) asserts that fronting does not occur in accents with monophthongal variants of the GOAT vowel, such as in Scotland and parts of northern England. In support of this statement, Jansen (2019) finds no evidence of GOAT-fronting in her research into back vowel fronting in Carlisle English. Similarly, Haddican *et al.* (2013) find that, although results point to a change in progress toward a fronter GOAT vowel in York, this was only apparent in the speakers who used diphthongal GOAT pronunciations. These findings would suggest that, in general, varieties with monophthongal GOAT vowels are not participating in the recent fronting sound change which is taking place in many accents of English.

Tyneside English is a possible exception to this rule of no GOAT-fronting in monophthongal variants. As previously discussed, Watt (1998; 2002) noted the use of a centralised [ɵː] pronunciation among young middle-class male speakers in his dataset. Watt (1998: 264) goes on to describe this variant as both 'traditional and innovative', proposing that these younger males are using realisations that are at one time both signalling local identity and participation in a sound change affecting many other English varieties. This is explained by the fact that [ɵː] may be interpreted as a backing of the fronted [øː] associated with local Northumbrian speech, or alternatively, as an innovative fronting of [oː]. More recently, Wozniak *et al.* (2015) have also investigated GOAT-fronting in Tyneside English. They report findings of some fronted monophthongal [ɵː] tokens, though they give little indication as to how frequent these pronunciations are. Unlike in Watt's (1998; 2002) research, Wozniak *et al.* (2015) find no effect of speaker sex on [ɵː]; perhaps

this suggests that fronted GOAT vowels are more frequent, and less socially restricted, in Tyneside when compared to data from the 1990s. Research conducted by Watt (1998; 2002) and Wozniak *et al.* (2015) therefore does point to GOAT-fronting in Tyneside English. However, the Wozniak *et al.* (2015) study also included diphthongal pronunciations of GOAT, finding that monophthongs were subject to the least amount of fronting. From previous research, then, it is difficult to assess to what extent GOAT is being fronted in Tyneside.

The results of the auditory analysis conducted in the present study found tokens of fronted/centralised monophthongal GOAT variants in the speech of only two young males. Initially, the fact that this variant was used exclusively by two speakers in the sample would appear to suggest that fronting of monophthongal GOAT has not progressed on Tyneside. However, findings from the acoustic analysis do indicate that GOAT-fronting may be underway in the region. Firstly, the formant plots showing the overlap of the GOAT and THOUGHT vowels of the individual speakers, as illustrated in Figures 5.1 and 5.2 in Chapter 5, show that – for the majority of speakers – GOAT occupies a slightly fronter position than THOUGHT. Many of the speakers were still shown to exhibit overlap between GOAT and THOUGHT, only the position of the THOUGHT vowel typically overlapped with most backed tokens of GOAT. Furthermore, the results of the dynamic vowel analysis using Generalised Additive Mixed-Models (GAMMs) indicated that, younger speakers showed fronter GOAT vowels, across the vowel trajectory, than the middle and older speaker groups. This difference in the frontness of GOAT between the younger and the middle and older speakers proved to be significant in mixed-effects regression models run on the acoustic data; younger speakers had significantly higher F2 values for GOAT, thus the vowel is more fronted than in previous generations. The findings from the production data therefore suggest that, although young speakers appear to exhibit less of a height difference between GOAT and THOUGHT, the youngest members of the speech community also have fronter GOAT vowels when compared with older speakers.

In addition to the production analysis, this thesis' investigation into speech perception involved asking a subset of participants to read aloud a minimal pair list. Reflecting the results found in the DECTE data, many participants were found to have a slightly fronter GOAT vowel compared to THOUGHT, as displayed in Figure 7.22. These findings suggest that fronting of GOAT occurs in both conversational and citation speech styles. Additionally, the results indicated that GOAT is quite variable in terms of vowel frontness, as GOAT was found to be significantly more fronted in open syllable contexts compared to closed syllable

position. This suggests that GOAT is fronting in the speech community, and that it is particularly prone to fronting in free syllable environments.

The question remains, then, whether GOAT-fronting is likely to stop the progression of a GOAT-THOUGHT merger on Tyneside. While the results of the acoustic analysis showed young females to have the most overlap between GOAT and THOUGHT, formant plots, mixed-effects regression models and analysis of vowel trajectories showed that young male speakers are the most advanced GOAT-fronters in the Tyneside speech community. For the young male speakers, particularly in the formant plots and the figures showing the predicted trajectories of GOAT and THOUGHT output from the GAMM models (Figure 5.12), it is clear that this speaker group has the largest difference in F2 (frontness) between these two vowels. This indicates that while the young males are fronting GOAT, THOUGHT does not currently appear to be undergoing the same sound change. Results from the perception experiments also point to listeners being able to better discriminate and identify these vowels when the stimuli presented was a GOAT vowel in an open syllable context. This could suggest that listeners associate fronter pronunciations of GOAT with words in which the vowel is in a free syllable, making them more perceptually distinct from THOUGHT. These findings present some evidence that, although GOAT is fronting in the local speech community, THOUGHT is not following suit.

Future research into GOAT-THOUGHT merging on Tyneside must also take into account the advancement of GOAT-fronting in the region. At this stage it appears that young female speakers show most evidence of merger between GOAT-THOUGHT, and young males, although displaying some overlap in the position of the two vowels, are more likely to produce fronter tokens of GOAT. Based on the results of the current study, it is therefore difficult to speculate whether, in the future, Tyneside will see a merging of the GOAT and THOUGHT vowels, or whether increased fronting will lead to a greater distance in the positions of GOAT and THOUGHT in the vowel space. As GOAT-fronting is a sound change affecting many global varieties of English, it may be most probable that this fronting sound change will continue to progress on Tyneside. Additionally, a similar situation was reported in Bradford English, where Watt and Tillotson (2001) found that older speakers had an overlap between GOAT and THOUGHT in phonetic space, while the youngest speakers produced the GOAT vowel with a much more fronted realisation than the THOUGHT vowel. Such a finding provides evidence of fronting in a speech community which typically uses monophthongal realisations of GOAT, and, furthermore, these results also suggest that the

advancement of GOAT-fronting in apparent time may be responsible for the lack of a GOAT-THOUGHT overlap in the young Bradford English speakers. In Tyneside, then, it may be the case that the phonetically overlapping GOAT and THOUGHT vowels never advance toward a full phonemic merger due to future generations using increasingly fronted GOAT vowels. At present, given the evidence to suggest GOAT-fronting in the region, it is not clear whether a merger of the GOAT and THOUGHT vowels is likely to continue toward a complete merger in Tyneside, and thus use of the term 'merger-in-progress' to describe the current situation may not be entirely accurate here.

Chapter 9

Conclusion

This final chapter concludes the thesis with an evaluation of the strengths and limitations of the main study, in addition to a discussion of possible avenues of future research.

9.1 Strengths of the Thesis

This study has provided a thorough investigation into GOAT-THOUGHT merging in Tyneside English. Where there has previously been only brief speculation on an overlap between the GOAT and THOUGHT vowels (Watt 1998; Watt and Allen 2003), the present study has been the first to conduct a detailed examination of this potential merger using either speech production or speech perception data. As such, this study is also the first to present evidence of a merging of these two vowels in the Tyneside speech community, with the main findings suggesting a convergence of GOAT and THOUGHT in both production and perception.

With reference to the primary research questions of the thesis, the results of this investigation have provided information regarding the nature of the GOAT-THOUGHT merger and the quality of the merged vowel. Analysis of the speech production data unearthed compelling evidence to suggest that GOAT and THOUGHT are coalescing in Tyneside via a merger-by-approximation process. Additionally, the majority of the findings point to a lowering of the GOAT vowel to the position of THOUGHT as the direction of merging. Given the lack of previous research on this sound change, and the absence of phonetic detail in Watt's (1998: 161) short description of GOAT-THOUGHT homophony, very little had thus far been written on the development and quality of the potentially merged GOAT-THOUGHT vowel.

One of the methodological strengths of the main study is the use of multiple measures of merger, using both static and dynamic data. The results of the

auditory analysis confirmed that, while the monophthong [oː] was most frequently used by the Tyneside speakers, both monophthongal and diphthongal variants of the GOAT vowel were still used in the speech community (although more locally marked pronunciations were used infrequently among the speakers). Based on this finding, a dynamic analysis using vowel trajectory data was used to further assess the phonetic overlap of the GOAT and THOUGHT vowels. Explained in Chapter 4, analyses of monophthongs often use only one formant measurement per vowel, while studies of diphthongs typically take two or more measurements per vowel. While the main study did utilise methods of measuring merger that can accommodate only one-point vowel measurements (e.g. Pillai scores), the additional measurement of multiple formant values taken across the duration of the vowels therefore allowed for a measurement of mergedness which could comfortably incorporate the several different monophthongal and diphthongal variants of GOAT. The use of vowel trajectory data was advantageous as it reduced the need to exclude non-[oː] GOAT tokens from analysis; this resulted in a larger dataset, and a more realistic sample of the variation present within the wider speech community. As the results of the analysis of dynamic data were largely consistent with the results from other measures using only one-point formant values, the inclusion of an investigation into the vowel trajectories of GOAT and THOUGHT also gave credence to the findings from the static analyses methods.

Using a combination of methods and measures to assess mergedness in the present study also helped to reduce the risk of overconfident conclusions of merger. In a recent study by Turton and Baranowski (2020), statistical models suggested that Manchester English speakers showed a significant distinction between the FOOT and STRUT vowels in the F1 plane. However, the results of minimal pair readings indicated that speakers pronounced these vowels identically. Given the difficulties in accurately capturing any possible coarticulatory effects on the vowels, such as FOOT typically occurring in lower F1 environments than STRUT, the authors issue 'a warning to researchers attempting to diagnose the presence or absence of mergers by purely statistical means' (Turton and Baranowski 2020: 34). In an attempt to carefully consider many aspects of merger, the current study has utilised a number of different measures of merger; from formant values (both one-point and trajectory measures) and duration in the production data, to accuracy, sensitivity, and reaction time in the perception data. Again, this reduced the possible issue of over-reliance on certain results or statistical models in determining whether Tynesiders exhibited a phonemic GOAT-THOUGHT merger. For example, while the Pillai scores alone suggested that young male speakers had the largest distinction between the GOAT

and THOUGHT vowels, additional analysis of the vowel trajectories using GAMMs revealed that these speakers actually had a greater GOAT-THOUGHT overlap in terms of vowel height than many of the other speaker groups, and that they were in fact participating – by some measures – in the merging of the GOAT and THOUGHT vowels in the local speech community.

The current study also has some methodological implications for future research into mergers in phonetic space. Nycz and Hall-Lew (2013) compared the best measures of merger, but the methods discussed in the paper are only able to take one-point static formant measures as input. The production analysis, presented in Chapters 4 and 5, demonstrated that generalised additive mixed-models (GAMMs) offer an innovative approach to assessing mergedness using vowel trajectory data. GAMMs were used in the main study to predict the vowel trajectories of GOAT and THOUGHT for different speaker groups. These predicted trajectories may then be plotted for a visualisation of overlap across the duration of the vowel, and the model also provides information as to whether these trajectories are significantly different from one another. Analysis using trajectory data and GAMMs adds another dimension to measures of mergedness, as the shape of the trajectories of the two vowels may also be compared. Although the use of GAMMs in linguistic research has increased in recent years, they have rarely been employed in studies of merger (with the previously noted exceptions of Renwick 2017 and Renwick and Stanley 2020). This study proposes that GAMMs are an appropriate measure of merger, particularly if, as in the case of Tyneside, one of the potentially merging phonemes has a number of possible realisations.

A number of strengths in the analysis of the speech perception data may also be highlighted here. Firstly, methods more closely associated with laboratory phonology and speech perception research (i.e. the use of discrimination and identification tasks) were utilised to assess the perception of the GOAT and THOUGHT vowels. This approach to investigating mergers is in line with the increased importance placed on perception – and the more experimental methodologies employed to examine perception – exhibited in sociolinguistic studies over the past decade or so (Thomas and Hay 2005; Hay *et al.* 2006; Arnold 2015; Freeman 2019). As such, the present study adds to the growing body of research which demonstrates that perception is a necessary component in research of phonemic mergers, and that more experimental designs, such as those used in speech perception literature, can comfortably fit into sociolinguistic studies of merger.

Another advantage of the method of the current study is the use of natural

stimuli in the perception tasks. There are examples of sociolinguistic studies into the perception of mergers have used synthetic or resynthesised natural speech data as the basis of their experimental stimuli (Janson and Schulman 1983; Freeman 2019). However, as discussed in Chapter 6, the process of resynthesising speech may have certain consequences, for example, an unconscious removal of speech cues beyond those being investigated, which may be important for listeners to accurately perceive particular sounds (Labov *et al.* 1991: 50). Perhaps the main issue with synthesised stimuli is that it is extremely difficult to maintain the naturalness of the speech data after resynthesis, leading to stimuli which may have an 'unnatural' sound. As the current study utilised only natural speech in the perception experiments, the stimuli used were a more reliable, realistic sample of the sounds as found throughout the wider speech community. Thus, the perception investigation in the main study was able to make use of a more experimental type of methodology, while also remaining appropriate for a sociolinguistic study of GOAT-THOUGHT merging.

The use of multiple techniques of measuring mergedness in perception is another strength of the present study. Analyses using measures of accuracy, reaction time, and sensitivity all pointed to GOAT-THOUGHT stimuli being more difficult to distinguish and identify than the other vowel pairs included in the listening tasks. As a result, the consistent findings across these different measures strengthened the overall argument that there is some degree of phonemic overlap between GOAT and THOUGHT for the Tyneside listeners.

9.2 Limitations of the Thesis

One of the most notable limitations of the present study is perhaps the omission of an in-depth examination of social class. In the analyses of the speech production data obtained from DECTE, socioeconomic group was included as a predictor in the auditory analysis, which found that upper-middle class speakers exhibited increased usage of the prestigious [oʊ] variant of the GOAT vowel. This finding is consistent with previous research on Tyneside English (Watt 1998; Watt 2002; Beal *et al.* 2012). In terms of GOAT-THOUGHT merging, analyses using Pillai scores indicated that working class speakers demonstrated a greater phonetic overlap between the two vowels. While the production analysis does present some indication of the effects of social class on GOAT-THOUGHT convergence, it was not possible to include socioeconomic status in the speaker social groupings along with sex and age. Given the sample size, and the fact that some social classes are more well represented than

others in DECTE[1], it was not possible to socially stratify the data by sex, age and social class. As a result, this study could not investigate how social class interacts with speaker sex and age with respect to the merging of GOAT and THOUGHT.

An analysis of social class was not possible at all in the perception component of the main study. At the time of conducting the perception experiments, it had not yet been decided that participants who gave their occupation as 'student' would be assigned a socioeconomic class, in line with the National Statistics Socio-economic Classification guidelines (ONS 2010), based on the occupations of their parents. Participants were not required to complete information regarding parental occupations, and, as many students participated in the listening tasks, they could not be grouped into social classes using the same criteria as in the production analysis. Moreover, several participants did not supply information about their occupation (e.g. left a blank space on the form, or wrote N/A), causing further complications for any social class analysis. While socioeconomic status could not be included in the perception study, it is perhaps unlikely that such an analysis would have revealed anything about the perception of the GOAT and THOUGHT vowels, as the present study found little evidence to suggest that the other social factors of sex and age affected how these vowels were perceived.

One other disadvantage to the data collection methods utilised in the present study is the relatively low proportion of participants in the 'older' age groups. As the DECTE interviews are conducted by university students, the majority of participants fall into the youngest age group of between 18-30 years of age. Due to the comparatively low number of interviews with older speakers in the corpus, it was more difficult to find participants within this age group who fit the criteria of the study (e.g. male and female speakers over the age of 60 who were born in the Tyneside region, and who had subsequently lived in Tyneside for most of their lives). Similarly, recruitment for the perception tasks was carried out through university contacts and mailing lists. As a result, a much higher proportion of younger participants took part in the experiment, therefore, the younger and middle age groups in the production and perception analyses are notably more well represented in the data than the older age group. Another disadvantage is that the perception tasks were run in the speech lab at Newcastle University's campus on weekdays between the hours of 10am-5pm. While this was necessary due to the use of the university's campus-based recording equipment and experimental presentation software, the location may have had the unintended effect of

[1]As explained in Chapter 4, the interviews in DECTE are collected by university students, leading to a higher proportion of middle class interviewees in the corpus.

excluding participants who were unable or unwilling to travel to the university during the available times.

9.3 Suggestions for Further Research

The present study has utilised a variety of production and perception measures to investigate whether the GOAT and THOUGHT vowels are merging in Tyneside English. Beyond the goals and research questions of this thesis, the findings of the main study have highlighted a number of possible avenues for future research.

One potential direction of future research would be to conduct a similar analysis of both the production and perception of GOAT and THOUGHT for other English varieties in which overlap between these two vowels has been reported, most notably those found in some areas of Wales (Wells 1982), Bradford (Watt and Tillotson 2001), Lancashire (Ferragne and Pellegrino 2010), and Carlisle (Jansen 2015; Jansen 2018). While comments in the literature would suggest that these vowels frequently sound indistinguishable in a number of different accents – particularly those in the north of England which typically have monophthongal variants of the GOAT vowel – there has been a lack of any in-depth investigation into the potential merging of these vowels in these varieties. In these accounts, discussions of GOAT-THOUGHT overlap are largely restricted to simple formant analyses, while little consideration is given to the perception of these vowels. Thus, although previous research has suggested that speakers of several different varieties of English may not distinguish between the GOAT and THOUGHT vowels, there is an absence of strong evidence to suggest that this similarity is due to a phonemic GOAT-THOUGHT merger. The results of such research into other Englishes which may exhibit GOAT-THOUGHT merging could then be compared with the findings from this study of the Tyneside speech community to help contextualise just how 'merged' the Tynesiders are in comparison to speakers from other varieties which have been noted as having identical GOAT and THOUGHT vowels. Using a similar research design as that employed in the current study to examine GOAT-THOUGHT merging in other varieties would have the added benefit of testing the reliability of certain methods and measures in signalling 'mergedness', e.g. whether accuracy and reaction time analyses of the perception results are an accurate indicator of a listener's perceptual overlap between GOAT and THOUGHT.

The results of the current study do not demonstrate a neatly symmetrical relationship between the production and perception of the GOAT and THOUGHT vowels either at the level of the individual or at the level of the speaker group. Although these findings may indicate that, currently, GOAT and THOUGHT are

subject to greater overlap in production than perception for the majority of Tynesiders, further research into this potential merger should explore other factors which may explain why the social patterns found in the production data were not clearly reflected in the perception results. While it may not be feasible to investigate every aspect in which two sounds could differ, Gordon (2013: 206) emphasises the importance of widening research beyond the analysis of F1-F2 values in studies of phonemic merger. The current study attempted to conduct a broader analysis of the production data in its brief focus on vowel duration, in which no significant difference was found between the length of the GOAT and THOUGHT vowels for the Tyneside speakers. However, it is possible that there are other features of the production of these vowels, unexamined here, that serve to distinguish GOAT and THOUGHT. In a recent study by Maclagan *et al.* (2017), the NURSE and GOOSE vowels were found to overlap in phonetic space for many New Zealand English speakers, despite no prior suggestions of a merger between these two vowels in the community. To further examine factors which may serve to distinguish these vowels from one another, the authors also used measures of vowel duration, lip rounding, and vocal tract shape. Similar analyses of lip rounding or vocal tract shape in Tyneside may help to uncover whether there are any other factors which could be helping to maintain a distinction between the GOAT and THOUGHT vowels – in either production or perception – within the speech community.

Based on the results of the study presented in this thesis, perhaps the most obvious route of future work would be further research into GOAT-fronting in Tyneside English. An analysis of the DECTE interviews suggested evidence of fronted [oː] pronunciations of GOAT among the Tyneside speakers. While two young male speakers were clearly heard producing centralised [ɵː]-like variants during an auditory analysis of the data, analyses of formant values using GAMMs indicated that young speakers exhibited fronter GOAT vowels when compared with the older speaker groups. Additional research into GOAT-fronting on Tyneside is important as the advancement of GOAT-fronting in the region may prevent a complete phonemic GOAT-THOUGHT merger in Tyneside English. Thus, further studies of GOAT-THOUGHT merging must carefully consider GOAT-frontness in their analysis. Future studies of GOAT-fronting on Tyneside would also help to provide answers in the debate which asks whether the GOAT-fronting sound change, found in English varieties across the globe, can also take place in accents with monophthongal variants of GOAT (Docherty 2010; Haddican *et al.* 2013; Wozniak *et al.* 2015; Jansen 2019). The present study offers some strong evidence

of fronting of monophthongal GOAT, particularly in the finding that the vowel appears to be most fronted in word final/open syllable environments. This result that open syllables significantly promote fronting of the GOAT vowel indicates not that the majority of speakers are simply using a fronted variant of GOAT, such as a local [ɵː] pronunciation, but rather that the predominant variant of [oː] is undergoing a gradual phonetic fronting in open syllable contexts. In order to fully explore the effects of phonetic and phonological environment on the frontness of the GOAT vowel in Tyneside English, future studies must carefully examine GOAT in a wide range of different phonetic and phonological contexts.

9.4 Conclusions

This study investigated the potential merging of the GOAT and THOUGHT vowels in contemporary Tyneside English. Following Watt's (1998: 161) observation that these vowels are homophonous for some Tynesiders, this study examined the extent to which GOAT and THOUGHT are pronounced and perceived as identical in the Tyneside speech community. Using both static and dynamic measures, analysis of the speech production data indicated that many speakers, particularly young female speakers, exhibited a considerable – but not complete – GOAT-THOUGHT overlap in phonetic space. Findings from the perception study suggested some perceptual overlap between GOAT and THOUGHT, as participants were significantly slower and less accurate in responding to GOAT-THOUGHT stimuli when compared to other vowel pairs. Overall, the results point to a merging of these two vowels in Tyneside via a process of merger-by-approximation. Future research must pay attention to the advancement of GOAT-fronting, which may prevent progression toward a complete phonemic GOAT-THOUGHT merger in the region.

Bibliography

Arnold, L. (2015). 'Multiple mergers: production and perception of three pre-/l/ mergers in Youngstown, Ohio'. In *University of Pennsylvania Working Papers in Linguistics* 21: 1–10.

Ash, S. (2013). 'Social Class'. In J. Chambers and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*. 2nd Edition. Hoboken: Wiley Blackwell: 350–367.

Atkinson, J. (2011). 'Linguistic variation and change in a North-East border town: a sociolinguistic study of Darlington'. Unpublished PhD thesis, Newcastle University.

Baayen, R., D. Davidson, and D. Bates (2008). 'Mixed-effects modeling with crossed random effects for subjects and items'. In *Journal of Memory and Language* 59: 390–412.

Baayen, R., J. van Rij, C. de Cat, and S. Wood (2018). "Autocorrelated errors in experimental data in the language sciences: some solutions offered by generalized additive mixed models'. In D. Speelman, K. Heylen, and D. Geeraerts (Eds.), *Mixed-Effects Regresstion Models in Linguistics*. Cham, Switzerland: Springer: 49–69.

Baković, E. (2010). 'Fun with the low back vowel merger'. Last accessed 12th February 2020 from: *https://quote.ucsd.edu/phonoloblog/2010/08/23/fun-with-the-low-back-vowel-merger/*.

Baranowski, M. (2007). 'Phonological Variation and Change in the Dialect of Charleston, South Carolina'. In *American Speech* 82 (supplement 92): vii–268.

Baranowski, M. (2013). 'Sociophonetics'. In R. Bayley, R. Cameron, and C. Lucas (Eds.), *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press: 403–424.

Baranowski, M. (2017). 'Class matters: the sociolinguistics of GOOSE and GOAT in Manchester English'. In *Language Variation and Change* 29: 301–339.

Baranowski, M. and D. Turton (2018). *'Locating speakers in the socioeconomic hierarchy: towards the optimal indicators of social class'*. Paper presented at *New Ways of Analyzing Variation 46*. New York University.

Barnfield, K. and I. Buchstaller (2010). 'Intensifiers on Tyneside: longitudinal developments and new trends'. In *English World-Wide* 31: 252–287.

Barras, W., P. Honeybone, and G. Trousdale (2007). *'Is word-frequency a factor in phonological demergers? The sociophonology of* SQUARE *and* NURSE *in Lancashire English'*. Paper presented at *2nd International Conference on the Linguistics of Contemporary English*. University of Toulouse.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). 'Fitting Linear Mixed-Effects Models Using lme4'. In *Journal of Statistical Software* 67: 1–48.

Bauer, L. and P. Warren (2004). 'New Zealand English: phonology'. In E. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton (Eds.), *A Handbook of Varieties of English: Phonology*. Berlin: Mouton de Gruyter: 580–602.

Baxter, L. (2010). 'Lexical diffusion in the early stages of the merry-marry merger'. In *University of Pennsylvania Working Papers in Linguistics* 16: 11–19.

Beal, J. (1985). 'Lengthening of a in Tyneside English'. In R. Eaton, O. Fischer, W. Koopman, and F. van der Leek (Eds.), *Papers from the 4th International Conference on English Historical Linguistics*. Amsterdam: John Benjamins: 31–44.

Beal, J. (1999a). *English Pronunciation in the Eighteenth Century*. Oxford: Clarendon Press.

Beal, J. (1999b). 'Geordie nation: language and regional identity in the Northeast of England'. In *Lore and Language* 17: 33–48.

Beal, J. (2000). 'From Geordie Ridley to Viz: popular literature in Tyneside English'. In *Language and Literature* 9: 343–359.

Beal, J. (2008). 'English dialects in the north of England: phonology'. In B. Kortmann and C. Upton (Eds.), *Varieties of English: The British Isles*. Berlin: Mouten de Gruyter: 122–144.

Beal, J. (2009). 'Enregisterment, commodification, and historical context: "Geordie" versus "Sheffieldish"'. In *American Speech* 84: 138–156.

Beal, J. and L. Burbano-Elizondo (2012). '"All the lads and lasses": lexical variation in Tyne and Wear: a discussion of how the traditional dialect terms *lad* and *lass* are still used in the modern urban dialects of Newcastle upon Tyne and Sunderland'. In *English Today* 28: 10–22.

Beal, J., L. Burbano-Elizondo, and C. Llamas (2012). *Urban North-Eastern English: Tyneside to Teesside*. Edinburgh: Edinburgh University Press.

Bell-Berti, F., L. Raphael, D. Pisoni, and J. Sawusch (1978). 'Some relationships between speech production and perception'. In *Phonetica* 36: 373–383.

Bigham, D. (2004). *'The PIN-PEN vowel merger in Southern Illinois English'*. Paper presented at *New Ways of Analyzing Variation 33*. University of Michigan.

Bigham, D. (2010). 'Correlation of the low-back vowel merger and TRAP-Retraction'. In *University of Pennsylvania Working Papers in Linguistic* 15: 21–31.

Blake, R., C. Shousterman, and L. Newlin-Lukowicz (2015). 'African American Language in New York City'. In S. Laneheart (Ed.), *African American Language*. Oxford: Oxford University Press: 280–298.

Boberg, C. and S. Strassel (1995). 'Phonological Change in Cincinatti'. In *University of Pennsylvania Working Papers in Linguistics* 2: 25–35.

Boersma, P. and D. Weenink (2019). Praat: doing phonetics by computer. Version 6.1.05 [Software]. http://www.praat.org/.

Bowie, D. (2000). 'Perception and production in processes of merger'. In *Proceedings of the Deseret Language and Linguistic Society Symposium* 26: 137–143.

Bradlow, A., L. Nygaard, and D. Pisoni (1999). 'Effects of talker, rate, and amplitude variation on recognition memory for spoken words'. In *Perception and Psycholinguistics* 2: 206–219.

Brunellière, A., S. Dufour, N. Nguyen, and U. Frauenfelder (2009). 'Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception'. In *Cognition* 111: 390–396.

Buchstaller, I. (2011). 'Quotations across the generations: a multivariate analysis of speech and thought introducers across 4 generations of Tyneside speakers'. In *Corpus Linguistics and Linguistic Theory* 7: 1–18.

Buchstaller, I. and A. Mearns (2018). 'The Effect of Economic Trajectory and Speaker Profile on Lifespan Change: Evidence from Stative Possessives on Tyneside'. In N. Braber and S. Jansen (Eds.), *Sociolinguistics in England*. London: Palgrave Macmillan: 215–241.

Burbano-Elizondo, L. (2006). 'Regional variation and identity in Sunderland'. In T. Omoniyi and G. White (Eds.), *The Sociolinguistics of Identity*. London: Continuum: 113–128.

Burbano-Elizondo, L. (2015). 'Sunderland'. In R. Hickey (Ed.), *Researching Northern English'*. Amsterdam: John Benjamins: 183–204.

Chambers, J. (2013). 'Patterns of variation including change'. In J. Chambers and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*. 2nd Edition. Hoboken: Wiley Blackwell: 297–324.

Chen, J. (2007). MARCS Auditory Laboratories. Version 1.0. [Software]. University of Western Sydney.

Chirrey, D. (1999). 'Edinburgh: descriptive material'. In P. Foulkes and G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*. London: Routledge: 223–229.

Colbourne, B. (1982). 'A Sociolinguistic Study of Long Island, Notre Dame Bay, Newfoundland'. Unpublished Masters thesis, Memorial University of Newfoundland, St. John's.

Cole, A. and P. Strycharczuk (2019). 'The PRICE-MOUTH crossover in the "Cockney diaspora". In *Proceedings of the 19th International Congress of Phonetic Sciences*: 602–606.

Coretta, S. (2017). *'Vowel duration and tongue root advancement in Italian and Polish'*. Paper presented at *Ultrafest VIII*. University of Potsdam.

Corrigan, K. (2012). 'GOAT vowel variants in the Diachronic Electronic Corpus of Tyneside English (DECTE)'. In T. Nevalainen and E. Traugott (Eds.), *The Oxford Handbook of the History of English*. Oxford: Oxford University Press: 1–5.

Corrigan, K., I. Buchstaller, A. Mearns, and H. Moisl (2012). *The Diachronic Electronic Corpus of Tyneside English*. https://research.ncl.ac.uk/decte/. Newcastle University.

Corrigan, K., A. Mearns, and H. Moisl (2014). 'Feature-based versus aggregate analyses of the DECTE corpus: phonological and morphological variability in Tyneside English'. In B. Szmrecsanyi and B. Wälchi (Eds.), *Aggregating Dialectology, Typology, and Register Analysis:linguistic variation in text and speech*. Berlin: Walter de Gruyter: 113–150.

Crinson, J. and J. Williamson (2004). 'Non-standard dialect in the formal speech of 15-year-olds on Tyneside'. In *Language and Education* 18: 207–219.

Cruttenden, A. (2014). *Gimson's Pronunciation of English*. 8th Edition. London: Routledge.

Cukor-Avila, P. and G. Bailey (2013). "Real time and apparent time'. In J. Chambers and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*. 2nd Edition. Hoboken: Wiley Blackwell: 239–267.

De Jong, K. and Y. Hao (2017). 'Featural generalization in second language identification performance: comparing learners with different L1s'. In *Journal of Phonetics* 66: 15–27.

Di Paolo, M. (1992). 'Hypercorrection in response to the apparent merger of (ɔ) and (ɑ) in Utah English'. In *Language & Communication* 12: 267–292.

Di Paolo, M. and A. Faber (1990). 'Phonation differences and the phonetic content of the tense-lax contrast in Utah English'. In *Language Variation and Change* 2: 155–204.

Di Paolo, M., M. Yaeger-Dror, and A. Wassink (2011). 'Analyzing vowels'. In M. Di Paolo and M. Yaeger-Dror (Eds.), *Sociophonetics: a student's guide*. London: Routledge: 87–106.

Dinkin, A. (2011). 'Weakening resistance: progress toward the low back merger in New York State'. In *Language Variation and Change* 23: 315–345.

Dinkin, A. (2016). 'Phonological transfer as a forerunner of merger in Upstate New York'. In *Journal of English Linguistics* 44: 162–188.

Dixon, P. (2008). 'Models of accuracy in repeated-measures designs'. In *Journal of Memory and Language* 59: 447–456.

Docherty, G. (2010). 'Phonological innovation in contemporary spoken British English'. In A. Kirkpatrick (Ed.), *The Routledge Handbook of World Englishes*. Oxford: Routledge: 59–75.

Docherty, G. and P. Foulkes (1999a). 'Derby and Newcastle: instrumental phonetics and variationist studies'. In P. Foulkes and G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*. New York: Oxford Univrsity Press: 47–71.

Docherty, G. and P. Foulkes (1999b). 'Sociophonetic variation in 'glottals' in Newcastle English'. In, *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco: 1037–1040.

Doernberger, J. and J. Cerny (2008). 'The low back merger in Miami. In *University of Pennsylvania Working Papers in Linguistics* 14: 11–15.

D'Onofrio, A. (2019). *'Age-based dynamics of the perception-production link'*. Paper presented at *UK Language Variation and Change 12*. Queen Mary University of London.

Drager, K. (2010). 'Speaker age and vowel perception'. In *Language and Speech* 54: 99–121.

Drager, K. (2014). *'Perception and language attitudes'*. Paper presented at *New Ways of Analyzing Variation 43*. Chicago.

Dufour, S., A. Brunellière, and N. Nguyen (2013). 'To what extent do we hear phonemic contrasts in a non-native regional variety? Tracking the dynamics of perceptual processing with EEG'. In.

Ebabel, M., M. Emcauliffe, and G. Ehaber (2013). 'Can mergers-in-progress be unmerged in speech accommodation?' In *Frontiers in Psychology* 4: 1–14.

Eberhardt, M. (2009). 'African American and White vowel systems in Pittsburgh'. In *The Publication of the American Dialect Society* 94: 129–157.

Eckert, P. (2013). 'The whole woman: sex and gender differences in variation'. In *Language Variation and Change* 1: 245–267.

ELAN. (2018). Max Planck Institute for Psycholinguistics, The Language Archive. Version 5.2. [Software]. Nijmegen, The Netherlands.

Fehringer, C. and K. Corrigan (2015a). 'The rise of the "going to" future in Tyneside English: evidence for further grammaticalisation'. In *English World-Wide* 36: 198–227.

Fehringer, C. and K. Corrigan (2015b). '"You've got to sort of eh hoy the Geordie out": modals of obligation and necessity in fifty years of Tyneside English'. In *English Language and Linguistics* 19: 355–381.

Ferragne, E. and F. Pellegrino (2010). 'Formant frequencies of vowels in 13 accents of the British Isles'. In *Journal of the International Phonetic Association* 40: 1–34.

Foulkes, P. (1997). 'English [r]-sandhi: a sociolinguistic perspective'. In *Historie Epistemologie Language* 19: 73–96.

Foulkes, P. (1999). 'Rule inversion in a British English dialect: a sociolinguistic investigation of [r]-sandhi in Newcastle upon Tyne'. In *University of Pennsylvania Working Papers in Linguistics* 4: 259–270.

Freeman, V. (2019). 'Prevelar merger in production vs. perception'. In *Proceedings of the 19th International Congress on Phonetic Sciences*: 457–461.

Fridland, V., K. Bartlett, and R. Kreuz (2004). 'Do you hear what I hear? experimental measurement of the perceptual salience of acoustically manipulated vowel variants by Southern speakers in Memphis, TN'. In *Language Variation and Change* 16: 1–16.

Fridland, V. and T. Kendall (2012). 'Exploring the relationship between production and perception in the mid front vowels of U.S. English'. In *Lingua* 122: 779–793.

Fridland, V., T. Kendall, and C. Farrington (2014). 'Durational and spectral differences in American English vowels: dialect variation within and across regions'. In *The Journal of the Acoustical Society of America* 136: 341–349.

Frieda, E., A. Walley, J. Flege, and M. Sloane (2000). 'Adult's perception and production of the English vowel /i/'. In *Journal of Speech, Language, and Hearing Research* 43: 129–143.

Fruehwald, J. (2011). *Praat Handcoder*. `https://github.com/JoFrhwld/FAAV/blob/master/-praat/handCoder.praat`.

Fruehwald, J. (2016). 'The early influence of phonology on a phonetic change'. In *Language* 92: 376–410.

Garde, P. (1961). ‘Réflexions sur les différences phonétiques entre les langues slaves’. In *Word* 17: 34–62.

Gordon, E. and M. Maclagan (2001). ‘"Capturing a sound change": a real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English’. In *Australian Journal of Linguistics* 21: 215–238.

Gordon, M. (2008). ‘The West and Midwest: phonology’. In E. Schneider (Ed.), *Varieties of English: The Americas and the Caribbean*. Berlin: Mouton de Gruyter: 129–143.

Gordon, M. (2013). ‘Investigating chain shifts and mergers’. In J. Chambers and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*. 2nd Edition. Hoboken: Wiley Blackwell: 203–219.

Gordon, M. (2014). ‘Exploring chain shifts, mergers and near-mergers’. In P. Honeybone (Ed.), *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press: 173–190.

GOV.uk (2018). ‘Regional ethnic diversity’. `https://www.ethnicity-facts-figures.service.gov.-uk/uk-population-by-ethnicity/national-and-regional-populations/regional-ethnic-diversity/latest`. Last accessed: 5th February 2019.

Haddican, B., P. Foulkes, V. Hughes, and H. Richards (2013). ‘Interaction of social and linguistic constraints on two vowel changes in northern England’. In *Language Variation and Change* 25: 371–403.

Hall-Lew, L. (2010). ‘Improved representation of variance in measures of vowel merger’. In *Meeting of the Acoustical Society of America, Proceedings of Meetings on Acoustics*. Vol. 9. Baltimore: pp. 1–10.

Hall-Lew, L. (2013). ‘"Flip-flop" and mergers-in-progress’. In *English Language and Linguistics* 17: 359–390.

Hall-Lew, L., A. Cardoso, Y. Kemenchedjeva, K. Wilson, R. Purse, and J. Saigusa (2015). ‘San Francisco English and the California vowel shift’. In *Proceedings of The 18th International Conference of the Phonetic Sciences*. Glasgow.

Harris, J. (1985). *Phonological Variation and Change: studies in Hiberno-English*. Cambridge: Cambridge University Press.

Hay, J., P. Warren, and K. Drager (2006). ‘Factors influencing speech perception in the context of a merger-in-progress’. In *Journal of Phonetics* 34: 458–484.

Herold, R. (1990). ‘Mechanisms of merger: the implementation and distribution of the low back merger in Eastern Pennsylvania’. Unpublished PhD thesis, University of Pennsylvania.

Herold, R. (1997). 'Solving the actuation problem: merger and immigration in eastern Pennsylvania'. In *Language Variation and Change* 9: 165–189.

Hickey, R. (1999). 'Dublin English: current changes and their motivation'. In P. Foulkes and G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*. New York: Oxford Univrsity Press: 265–281.

Hickey, R. (2004). 'Mergers, near-mergers and phonological interpretation'. In S. Pfenninger, O. Timofeeva, A. Gardner, A. Honkapohja, M. Hundt, and D. Schrier (Eds.), *Contact, Variation, and Change in the History of English*. Amsterdam: John Benjamins: 125–137.

Hickey, R. (2015). 'The North of England and Northern English'. In R. Hickey (Ed.), *Researching Northern English*. Amsterdam: John Benjamins: 1–24.

Hoenigswald, H. (1960). *Language Change and Linguistic Reconstruction*. Chicago: University of Chicago Press.

House, A. (1961). 'On vowel duration in English'. In *Language Variation and Change* 33: 1174–1178.

Hughes, A., P. Trudgill, and D. Watt (2012). *English Accents and Dialects*. London: Hodder Education.

Irons, T. (2007). 'On the status of low back vowels in Kentucky English: More evidence of merger'. In *Language Variation and Change* 19: 137–180.

Iverson, P. (2012). 'Measuring phonetic perception in adults'. In A. Cohn, C. Fougeron, and M. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Jansen, S. (2015). 'Carlisle and Cumbria'. In R. Hickey (Ed.), *Researching Northern English*. Amsterdam: John Benjamins: 227–250.

Jansen, S. (2018). *'Linguistic repercussions of political and socioeconomic changes: sound changes in a West Cumbrian peripheral community'*. Paper presented at *The 8th Northern Englishes Workshop*. Newcastle University.

Jansen, S. (2019). 'Change and stability in GOOSE, GOAT and FOOT: back vowel dynamics in Carlisle English'. In *English Language & Linguistics* 23: 1–29.

Janson, T. (1983). 'Sound change in perception and production'. In *Language* 59: 18–34.

Janson, T. and R. Schulman (1983). 'Non-distinctive features and their use'. In *Journal of Linguistics* 19: 321–336.

Jenkins, J. and W. Strange (1999). 'Perception of dynamic information for vowels in syllable onsets and offsets'. In *Perception and Psychophysics* 61: 1200–1210.

Johnson, D. (2007). 'Stability and change along a dialect boundary: the low vowels of Southeastern New England'. Unpublished PhD thesis, University of Pennsylvania.

Johnson, D. (2010). 'Stability and change along a dialect boundary: the low vowels of Southeastern New England'. In *American Dialect Society* 85 (supplement 95): v–235.

Johnson, D. and J. Nycz (2015). 'Partial mergers and near-distinctions: stylistic layering in dialect acquisition'. In *University of Pennsylvania Working Papers in Linguistics* 21: 109–117.

Jones-Sargent, V. (1983). *Tyne Bytes. A computerised sociolinguistic study of Tyneside*. Frankfurt am Main: Peter Lang.

Jones-Sargent, V. (1985). 'Tyneside Syntax: a presentation of some data from the Tyneside Linguistic Survey'. In W. Viereck (Ed.), *Focus on England and Wales*. Amsterdam: John Benjamins: 163–177.

Kendall, T. and V. Fridland (2017). 'Regional relationships among the low vowels of U.S. English: evidence from production and perception'. In *Language Variation and Change* 29: 245–271.

Kerswill, P. (1984). 'Social and linguistic aspects of Durham (e:)' in *Journal of the International Phonetic Association* 14: 13–34.

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*.

Kerswill, P. and A. Williams (2002). 'Dialect recognition and speech community focusing in New and Old Towns in England: the effects of dialect levelling, demography and social networks'. In D. Preston and D. Long (Eds.), *Handbook of Perceptual Dialectology. Vol. 2*. Amsterdam: John Benjamins: 173–204.

Kirkham, S., C. Nance, B. Littlewood, K. Lightfoot, and E. Groarke (2019). 'Dialect variation in formant dynamics: the acoustics of lateral and vowel sequences in Manchester and Liverpool English'. In *The Journal of the Acoustical Society of America* 145: 784–794.

Klatt, D. (1976). 'Linguistic uses of segmental duration in English: acoustic and perceptual evidence'. In *The Journal of the Acoustical Society of America* 59: 1208–1221.

Klatt, D. (1980). 'Software for a cascade/parallel formant synthesizer'. In *The Journal of the Acoustical Society of America* 67: 971–995.

Kleber, F., J. Harrington, and U. Reubold (2011). 'The relationship between the perception and production of coarticulation during a sound change in progress'. In *Language and Speech* 55: 383–405.

Kökeritz (1953). *Shakespeare's Pronunciation*. New Haven: Yale University Press.

Kontra, M. (1993). 'The messy phonology of Hungarians in South Bend: a contribution to the study of near mergers'. In *Language Variation and Change* 5: 225–231.

Koops, C., E. Gentry, and A. Pantos (2008). The effect of perceived speaker age on the perception of pin and pen vowels in Houston, Texas. In *University of Pennsylvania Working Papers* 14: 93–101.

Kurath, H. and R. McDavid (1961). *The Pronunciation of English in the Atlantic States: based upon the collections of the linguistic atlas of the Eastern United States*. Ann Arbour: University of Michigan Press.

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'. In *Language Variation and Change* 2: 205–254.

Labov, W. (1994). *Principles of Linguistic Change, Volume 1: Internal Factors*. Oxford: Blackwell.

Labov, W. (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Oxford: Blackwell.

Labov, W. (2010). *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors*. Ann Arbour: University of Michigan Press.

Labov, W., S. Ash, and C. Boberg (2006). *The Atlas of North American English*. Berlin: Walter de Gruyter.

Labov, W. and M. Baranowski (2006). '50msec'. In *Language Variation and Change* 18: 223–240.

Labov, W., M. Karen, and C. Miller (1991). 'Near-mergers and the suspension of phonemic contrast'. In *Language Variation and Change* 3: 33–74.

Labov, W., I. Rosenfelder, and J. Fruehwald (2013). 'One hundred years of sound change in Philadelphia: linear incrementation, reversal, and reanalysis'. In *Language* 89: 30–65.

Labov, W., M. Yaeger, and R. Steiner (1972). *A Quantitative Study of Sound Change in Progress. Vol. 1*. US Regional Survey.

Lass, R. and M. Laing (2016). 'Q is for WHAT, WHEN, WHERE?: the 'q' spellings for OE hw-'. In *Folia Linguistica Historica* 37: 61–100.

Lindsey, G. (2019). *English After RP*. Cham, Switzerland: Palgrave Macmillan.

Llamas, C. (2007). '"A place between places": language identities in a border town'. In *Language in Society* 36: 579–604.

Macaulay, R. (1978). 'Variation and consistency in Glaswegian English'. In P. Trudgill (Ed.), *Sociolinguistic Patterns in British English*. London: Edward Arnold: 133–143.

MacKenzie, L. and D. Turton (2020). 'Assessing the accuracy of existing forced alignment software on varieties of British English'. In *Innovative Methods in Sociophonetics, special issue of Linguistics Vanguard* 6.

Maclagan, M., C. Watson, R. Harlow, J. King, and P. Keegan (2017). 'Investigating the Sound Change in the New Zealand English Nurse Vowel /ɜ:/'. In *Australian Journal of Linguistics* 37: 465–485.

Macmillan, N and C. Creelman (2005). *Detection Theory: A User's Guide*. 2nd Edition. New Jersey: Lawrence Erlbaum Associates Inc.

Maguire, W. (2007). 'What is a Merger, and can it be reversed? The origin, status and reversal of the "NURSE-NORTH merger" in Tyneside English. Unpublished PhD thesis, Newcastle University.

Maguire, W., L. Clark, and K. Watson (2013). 'What are mergers and can they be reversed?' In *English Language and Linguistics* 17: 229–239.

Majors, T. (2005). 'Low back vowel merger in Missouri speech: acoustic description and explanation'. In *American Speech* 80: 165–179.

Makowski, D. (2018). 'The psycho package: an efficient and publishing-oriented workflow for psychological science'. In *Journal of Open Source Software* 3: 470.

Mathôt, S., D. Schreij, and J. Theeuwes (2012). 'OpenSesame: an open-source, graphical experiment builder for the social sciences'. In *Behavior Research Methods* 44: 314–324.

McGuire, G. (2010). A brief primer on experimental designs for speech perception research. In: 77.

Mearns, A. (2015). 'Tyneside'. In R. Hickey (Ed.), *Researching Northern English'*. Amsterdam: John Benjamins: 161–182.

Mellesmoen, G. (2018). 'A remedial path to merger: merger by phonological transfer in British Columbia English'. In *Toronto Working Papers in Linguistics* 40: 1–15.

Milroy, J. and J. Harris (1980). 'When is a merger not a merger? the MEAT/MATE problem in present-day English vernacular'. In *English World Wide* 1: 199–210.

Milroy, L. and M. Gordon (2003). *Sociolinguistics: Method and Interpretation*. Oxford: Blackwell.

Milroy, L., J. Milroy, G. Docherty, P. Foulkes, and D. Walshaw (1999). 'Phonological variation and change in contemporary English: evidence from Newcastle upon Tyne and Derby'. In *Cuadernos de Filologia Inglesa* 8: 35–46.

Minkova, D. (2014). *A Historical Phonology of English*. Edinburgh: Edinburgh University Press.

Moisl, H. and W. Maguire (2008). 'Identifying the main determinants of phonetic variation in the *Newcastle Electronic Corpus of Tyneside English*'. In *Journal of Quantitative Linguistics* 15: 46–69.

Mok, P., D. Zuo, and P. Wong (2013). 'Production and perception of a sound change in progress: tone merging in Hong Kong Cantonese'. In *Language Variation and Change* 25: 341–370.

Montgomery, C. (2007). 'Northern English Dialects: A Perceptual Approach'. Unpublished PhD thesis, University of Sheffield.

Moody, T. (2007). *The Mid-Northumbrian Dialect*. Morepeth: Morpeth Northumbrian Gathering Committee and the Northumbrian Language Society.

Niedzielski, N. (1999). 'The effect of social information on the perception of sociolinguistic variables'. In *Journal of Language and Social Psychology* 18: 62–85.

Nunberg, G. (1980). 'A falsely reported merger in eighteenth-century English: a study in diachronic variation'. In W. Labov (Ed.), *Locating Language in Time and Space*. New York: Academic Press.

Nycz, J. (2011). 'Second Dialect Acquisition: Implications for Theories of Phonological Representation'. Unpublished PhD thesis, New York University.

Nycz, J. and L. Hall-Lew (2013). 'Best practices in measuring vowel merger'. In *Meeting of the Acoustical Society of America, Proceedings of Meetings on Acoustics*. Vol. 20. San Francisco: pp. 2–19.

Office for National Statistics (2010). `https://www.ons.gov.uk/methodology-/classificationsandstandards/otherclassifications/-thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010`. Last accessed: 5th February 2019.

Orton, H. and E. Dieth (1962-1971). *Survey of English Dialects*. Leeds: Arnold.

Påhlsson, C. (1972). *The Northumbrian Burr*. Lund: C.W.K. Gleerup.

Pearce, M. (2009). 'A Perceptual Dialect Map of North East England'. In *Journal of English Linguistics* 37: 162–192.

Pearson, H. (1994). *The Far Corner*. London: Waverly.

Pellowe, J., G. Nixon, B. Strang, and V. McNeany (1972). 'A dynamic modelling of linguistic variation: the urban (Tyneside) linguistic survey'. In *Lingua* 30: 1–30.

Pollack, I. and D. Pisoni (1971). 'On the comparison between identification and discrimination tests in speech perception'. In *Psychonomic Science* 24: 299–300.

Renwick, M (2017). 'Dynamic trajectories of tense vs. lax vowels in the American South'. In *The Journal of the Acoustical Society of America* 142: 2680–2680.

Renwick, M. and J. Stanley (2020). 'Modeling dynamic trajectories of front vowels in the American South'. In *The Journal of the Acoustical Society of America* 147: 579–595.

Reubold, U. and J. Harrington (2017). 'The influence of age on estimating sound change acoustically from longitudinal data'. In S. Wagner and I. Buchstaller (Eds.), *Panel Studies of Variation and Change*. New York: Routledge: 129–151.

Rickford, J. and S. King (2019). 'Language and Linguistics on Trial'. In J. Rickford (Ed.), *Variation, Versatility and Change in Sociolinguistics and Creole Studies*. Cambridge: Cambridge University Press: 245–300.

Rij, J. van, M. Wieling, R. Baayen, and H. van Rijn (2017). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. R package version 2.3.

Rosenfelder, I., J. Fruehwald, K. Evanini, and J. Yuan (2011). Forced Alignment and Vowel Extraction (FAVE). Version 1.2.2. [Software].

Sankoff, G. (2004). 'Adolescents, young adults and the critical period: two case studies from "Seven Up"'. In C. Fought (Ed.), *Sociolinguistic variation: Critical reflections*. Oxford: Oxford University Press Oxford: 121–39.

Schleef, E. and D. Turton (2018). 'Sociophonetic variation of *like* in British dialects: effects of function, context and predictability'. In *English Language and Linguistics* 22: 35–75.

Sloos, M. (2013). 'The reversal of the BÄREN-BEEREN merger in Austrian Standard German'. In *The Mental Lexicon* 8: 353–371.

Sneller, B. (2015). 'Using rates of change as a diagnostic of vowel phonologization'. In A. Tuzzi, M. Benesová, and J. Maçutek (Eds.), *Recent Contributions to Quantitative Linguistics*. Berlin: Walter de Gruyter: 169–184.

Sóskuthy, M. (2017). *Generalised Additive Mixed Model for Dynamic Analysis in Linguistics: A Practical Introduction*. `http://eprints.whiterose.ac.uk/113858/`. Working paper. Last accessed: 5th February 2019.

Strang, B. (1968). 'The Tyneside linguistic survey'. In *Zeitschrift für Mundartforschung, NF* 4: 788–794.

Strang, B. (1970). *A History of English*. London: Methuen & Co LTD.

Strelluf, C. (2016). 'Overlap among back vowels before /l/ in Kansas City'. In *Language Variation and Change* 28: 379–407.

Stuart-Smith, J. (1999). 'Glasgow: accent and voice quality'. In P. Foulkes and G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*. London: Routledge: 203–222.

Stuart-Smith, J., R. Lennon, R. Macdonald, D. Robertson, M. Sóskuthy, B. José, and L. Evers (2015). 'A dynamic acoustic view of real-time change in word-final liquids in spontaneous Glasweigan'. In *Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow.

Sussman, J. and V. Lauckner-Morano (1995). 'Further tests of the "perceptual magnet effect" in the perception of [i]: identification and change/no-change discrimination'. In *The Journal of the Acoustical Society of America* 97: 539–552.

Tamminga, M. (2018). 'Modulation of the following segment effect on English coronal stop deletion by syntactic boundaries'. In *Glossa: A Journal of General Linguistics* 3: 86.

Thiel, A. and A. Dinkin. (2017). *'Escaping the* TRAP*: losing the Northern Cities Shift in real time'*. Paper presented at *New Ways of Analyzing Variation 46*. The University of Wisconsin-Madison.

Thomas, B. and J. Hay (2005). 'A pleasant malady: the ELLEN/ALLAN merger in New Zealand English'. In *Te Reo* 48: 69–93.

Thomas, E. (2001). *An acoustic analysis of vowel variation in New World Englishes*. Publication of the American Dialect Society 85. Durham, NC: Duke University Press.

Thomas, E (2002). 'Sociophonetic approaches of speech perception experiments'. In *American Speech* 77: 115–147.

Thomas, E. (2011a). 'Collecting data on phonology'. In W. Maguire and A. McMahon (Eds.), *Analysing Variation in English*. Cambridge: Cambridge University Press: 7–29.

Thomas, E. (2011b). *Sociophonetics: an introduction*. New York: Palgrave Macmillan.

Trudgill, P. (1972). 'Sex, covert prestige and linguistic change in the urban British English in Norwich'. In *Language in Society* 1: 179–195.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Trudgill, P. (1982). *On Dialect: social and geographical perspectives*. New York: New York University Press.

Trudgill, P. (1999). *The Dialects of England*. 2nd Edition. Cambridge, Massachusetts: Blackwell.

Trudgill, P. and T. Foxcroft (1978). *'On the sociolinguistics of vocalic mergers: transfer and approximation in East Anglia'*. P. Trudgill (Ed.). London: Edward Arnold.

Trudgill, P., D. Schreier, D. Long, and J. Williams (2004). 'On the reversibility of mergers: /w/, /v/ and evidence from lesser-known Englishes'. In *Folia Linguistica Historica* 24: 23–45.

Turton, D. (2017). 'Categorical or gradient? an ultrasound investigation of /l/-darkening and vocalization in varieties of English'. In *Laboratory Phonology* 8: 1–31.

Turton, D. and M. Baranowski (2020). 'Not quite the same: the social stratification and phonetic conditioning of the FOOT-STRUT vowels in Manchester'. In *Journal of Linguistics*: 1–39.

Van Heuven, W., P. Mandera, E. Keuleers, and M. Brysbaert (2014). 'SUBTLEX-UK: a new and improved word frequency database for British English'. In *Quarterly Journal of Experimental Psychology* 67: 1176–1190.

Venables, W and D. Ripley (2002). *Modern Applied Statistics with S*. 4th Edition.

Viereck, W. (1966). *Phonematische Analyse des Dialekts von Gateshead-upon-Tyne/Co. Durham*. Hamburg: Cram, de Gruyter & Co.

Wade, L (2017). 'The role of duration in the perception of vowel merger'. In *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8: 1–34.

Warburton, J. and D. Turton (2017). *'Reanalysis of Tyneside linking /r/: are younger speakers reversing the trend toward word- level deletion?'* Poster presented at *The 4th International Workshop on Sound Change*. University of Edinburgh.

Warren, R. and S. Fulop (2014). 'The merged vowel of PIN and PEN as realized in Bakersfield, California'. In *Journal of the Acoustical Society of America* 135: 2292.

Wassink, A. (2006). 'A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties'. In *Journal of the Acoustical Society of America* 119: 2334–2350.

Watson, K. and L. Clark (2012). 'How salient is the NURSE SQUARE merger?' In *English Language and Linguistics* 17: 297–323.

Watt, D. (1998). 'Variation and change in the vowel system of Tyneside English'. Unpublished PhD thesis, Newcastle University.

Watt, D. (2000). 'Phonetic parallels between the close-mid vowels of Tyneside English: are they internally or externally motivated?' In *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'. In *Journal of Sociolinguistics* 6: 44–63.

Watt, D. and W. Allen (2003). 'Tyneside English'. In *Journal of the International Phonetic Association* 33: 267–271.

Watt, D. and P. Foulkes (2017). 'Tyneside English'. In R. Hickey (Ed.), *Listening to the Past: audio records of accents of English*. Cambridge: Cambridge University Press: 142–170.

Watt, D., C. Llamas, and D. Johnson (2013). In.

Watt, D. and L. Milroy (1999). 'Patterns of variation in Newcastle vowels'. In P. Foulkes and G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*. London: Arnold: 25–46.

Watt, D. and J. Smith (2005). 'Language Change'. In M. Ball (Ed.), *Clinical Sociolinguistics*. Malden, MA: Blackwell: 101–116.

Watt, D. and J. Tillotson (2001). 'A spectrographic analysis of vowel fronting in Bradford English'. In *English World-Wide* 22: 269–302.

Wells, J. (1982). *Accents of English*. Cambridge: Cambridge University Press.

West, H. (2009). '"I'm not Geordie! I'm not actually anything!" convergent and divergent trends: dialect levelling and the struggle for identity in a South Durham new-town'. Unpublished Masters thesis, University of Edinburgh.

Wieling, M. (2018). 'Analyzing dynamic phonetic data using generalized additive mixed modelling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English'. In *Journal of Phonetics* 70: 86–116.

Williams, A. and P. Kerswill (1999). 'Dialect Levelling: Milton Keynes, Reading, Hull'. In P. Foulkes and G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles*. London: Routledge: 141–162.

Wong, A. (2012). 'The lowering of raised-THOUGHT and the low-back distinction in New York City: evidence from Chinese Americans'. In *University of Pennsylvania Working Papers in Linguistics* 18: 157–166.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*. 2nd Edition. Chapman and Hall/CRC.

Wozniak, K., B. Haddican, P. Foulkes, V. Hughes, and M. Sóskuthy (2015). *'The exceptionality of Tyneside back vowel fronting'*. Poster presented at *New Ways of Analyzing Variation 44*. Toronto.

Wyld, H. (1936). *A History of Modern Colloquial English*. Oxford: Blackwell.

Yeni-Komshian, G. (1998). 'Speech perception'. In J. Gleason and N. Bernstein Ratner (Eds.), *Psycholinguistics*. 2nd Edition. Fort Worth: Harcourt Brace College: 107–156.

Zhang, J. (2019). 'Feature-specific advantages in L3 phonological acquisition'. In. Melbourne: 3740–3744.

Appendix A

Information Sheet & Consent Form

School of English Literature, Language and Linguistics

Dear _____,

I am a PhD student at Newcastle University. I am conducting a study on language innovation and change in the Tyneside dialect.

My project requires North-Eastern participants like you to use your knowledge and intuitions of local Tyneside pronunciations. I am trying to find out how dialects in the North East may be changing, which dialect features are dying out, and whether there are any new features being used by the community. As a local speaker, you can help me to track the history of the Tyneside dialect, and gain a better understanding of what it means to sound 'Geordie' in the 21st century.

Participation in my project involves a short experiment, lasting for a total of approximately 15 minutes. The experiment is a listening task. You will be asked to listen to speech examples from local speakers, and to communicate what you think you heard the speakers say.

Please be reassured that the answers you give will be kept entirely confidential, and you will remain entirely anonymous. Access to data from these sessions will be restricted to academic researchers who follow the same procedures for keeping the data confidential and anonymous.

It is important to note that your participation in the project is entirely voluntary. You may withdraw at any time if you choose.

By participating in this study you are contributing to important, exciting new knowledge and information regarding the history and the changing nature of the Tyneside dialect. If interested, at the end of my project I would be happy to send you a summary of the results of my study. For any queries/further discussion of my project or your participation in it, please contact me at j.warburton@newcastle.ac.uk.

Thank you,
Jasmine Warburton

Newcastle University

Participant Consent Form

I understand that to take part in the study I will participate in an audio-taped reading task and a listening experiment.

I understand that my participation will take approximately 30 minutes, or however long I wish. I understand that the experiment can occur at a time and that is convenient for me.

I understand that I am under no obligation to participate and that I may refuse to answer any questions, to stop the experiment at any time, or withdraw from the study at any time.

I understand that my answers and comments will be kept entirely confidential. I understand that neither my name nor my address will be identified in any report or presentation that may arise from the study.

I understand that only academic researchers will have access to the information collected during the study and they too will be required to keep all the personal information confidential.

I understand that I may obtain information of the results of this study by contacting the principal investigator.

I understand what this study involves and agree to participate. I have been given a copy of the participant information sheet.

Signed _____ Date_____

Appendix B

Participant Questionnaire

Participant Questionnaire

Name:

Age: **Gender:**

Occupation:

Birthplace:

Current residence
(and how long you have lived there):

Length of time living in North East
(if you moved, please give details of when/where/how long):

Both parents born and raised in North East? ☐ yes ☐ no
If no, please give details below:

Appendix C

Perception Task Experimental Stimuli

C.1 Discrimination Task

GOAT-THOUGHT **Pairs**	NURSE-THOUGHT **Pairs**	Control Pairs
boat - bought	burn - born	bear - beer
bone - born	curse - course	chair - cheer
foam - form	heard - hoard	full - fool
hole - hall	shirt - short	sun - soon
low - law	work - walk	–

C.2 Identification Task

GOAT-THOUGHT **Pairs**	NURSE-THOUGHT **Pairs**	Control Pairs
bowl - ball	bird - board	pit - pet
coat - caught	fur - for	pin - pen
folk - fork	nurse - Norse	bad - bed
flow - flaw	stir - store	had - head
snow - snore	worm - warm	–