# Children's developing awareness of regional accents:

## A socioperceptual investigation of pre-school and primary school children in York

Ella Rose Webb Jeffries

PhD

University of York

Language and Linguistic Science

April 2016

# Abstract

This thesis explores children's developing awareness of regional accents, a relatively under-researched area of socioperceptual work. A series of four experiments are run with children in York between the ages of 2 and 9. These experiments are designed in order to investigate the process by which children progress from the ability to recognise familiar speakers to the ability to group speakers according to their regional accent. The Identification experiment establishes pre-school children's ability to recognise familiar speakers, while the Recognition experiment finds that features of a familiar speaker's accent forms part of the recognition process. The Grouping experiment goes on to investigate pre-school children's ability to group speaker guises according to phonological regional variables based on a Yorkshire/Standard Southern British English (SSBE) accent distinction. Finally, the Second Grouping experiment explores older, primary school children's ability to group different speakers according to phonological regional variables based on different accent distinctions (Yorkshire/SSBE, Yorkshire/Scottish and Yorkshire/North East).

Throughout these experiments, independent variables relating to the children's backgrounds are found to play a role in their abilities in the tasks. Generally, the girls perform better than the boys and there is an improvement throughout the ages. Furthermore, the children's exposure to regional variation is found to significantly affect their performance in the Grouping and the Second Grouping experiments. Children with regular exposure to non-local speakers are found to perform better in these tasks overall. It is proposed that the findings from all four experiments are best explained by interpreting them through an exemplar theoretic account. In such an account, speaker categories develop from the abstraction across social-indexical properties of phonetic variation which accumulate through an individual's experience with variation in their linguistic input.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

I have learned much more from these past few years of doing my PhD than I ever could have imagined. I most definitely could not have got to the point that I am at now, submitting the final version of my thesis, without the network of support that I have had from people in different aspects of my life.

First and foremost I would like to express my gratitude to my two supervisors, Dr Carmen Llamas and Professor Paul Foulkes, for their guidance over the past three and a half years. Thank you for helping me to hone my skills in many different areas (writing, experiment design, conference presentations – among many others). I have learned a lot and feel very different to the person who started out on the PhD road a few years ago.

I would also like to thank other members of the department for their help and support in carrying out the work for this thesis. In particular, Huw Llewelyn-Jones for technical expertise and assistance, Márton Sóskuthy for help with statistical models and getting to grips with R and Tamar Keren-Portnoy, the other member of my thesis advisory panel, for advice on experimental design and all things children related. I would also like to thank my fellow PhD colleagues, many of whom are now good friends, who have helped me in particular times of need. Special thanks to Jessica Wormald, Rachelle Albright, Rebecca Woods, Ania Kubisz and Daniel Lawrence. I hope that I have somehow managed to repay the favour!

Thank you to the AHRC for funding my PhD research and for providing additional financial support in the form of the research training support grant which allowed me to participate in my first international conference. Thanks also to the Philological society for providing me with a valuable travel bursary which gave me the opportunity to communicate my research on the other side of the Atlantic.

One of the best outcomes of doing my PhD has been meeting so many interesting people and making so many great friends, from around the world as well as from different academic disciplines. I won't name you all here but thanks to all of those with whom I have spent lunch times, coffee times and drink times and who have made the whole process much more enjoyable and sociable and certainly a lot less lonely than it could have been.

The contributions of this thesis only exist because of the participants. I would like to thank each and every one of the children who took part in the four experiments carried out here. I have learned a lot from working with you all and I am grateful for your appreciation of sticker rewards! Of course thanks are also due to the schools, nurseries, teachers and parents who agreed for their children to participate, gave me space to carry out the experiments and were very cooperative in allocating me some of their time. I really appreciate your help and hope that my academic and public contributions will go some way to paying it back. The experiments would also not have existed without the speakers who are in them. A big thank you to all of the many speakers who I recorded – I hope I have managed to make your voices heard.

Last but definitely not least, I would like to thank my close friends and family. Thank you to those nearest and dearest to me for your understanding, kind words, light-hearted humour and encouragement which have been invaluable throughout this process. In particular, thank you to my parents for your unwavering patience and support - emotionally, intellectually and financially!

Most importantly, I cannot thank my mum enough. Thank you for expressing your unequivocal pride, your excitement and your belief in me, even when I didn't believe in myself. You're more of a role model to me than you'll ever know (despite the fact that I'm yet to finish reading one of your books…!).

# Author's declaration

I declare the content of this thesis to be my own work and that all other content from outside sources has been appropriately referenced. This work is original and has not been previously submitted for an award at any institution. Some parts of this thesis have been presented elsewhere at conferences and/or through publication in an online journal.

The findings from chapter 3 have been published in an online journal:

Jeffries, E. (2015). Pre-school children's identification of familiar speakers and the role of accent features. *York Working Papers in Linguistics 14.*

The findings from chapters 3 and 4 have been presented at the following conferences:

(2015). Jeffries, E., Foulkes, P. & Llamas, C. *Pre-schoolers' categorisation of speakers by regional accent* (oral presentation). UK Language Variation and Change 10 (UKLVC 10), University of York, UK.

(2015). Jeffries, E., Foulkes, P. & Llamas, C. *Pre-schoolers' categorisation of speakers by phonological variables* (oral presentation). International Child Phonology Conference (ICPC 2015), Memorial University, Newfoundland, Canada.

(2014). Jeffries, E., Foulkes, P. & Llamas, C. *Pre-schoolers' categorisation of speakers by regional accent* (oral presentation). Variation in Language Acquisition 2 (ViLA 2), Université Grenoble 3, France.

# Chapter 1: Introduction

This thesis reports on research into the development of children's awareness of regional accents of their language. Experimental methods were used to investigate the process by which children between the ages of 2 and 9 progress from recognising familiar speakers to grouping speakers according to their regional accent. The results are interpreted in the framework of exemplar theory, which proposes an explanation for the development of speaker groups based on social-indexical properties of phonetic variation.

The research took place in York, North Yorkshire, in the North of England. Apart from the children in one of the pilot studies (chapter 5), all of the children who took part in the research were living in York at the time of their participation. The research focused on investigating children's awareness of the differences between regional accent variants typically used in York, compared to those typically used in other parts of the country.

## 1.1 Research motivations

The linguistic variants that make up an individual speaker's accent can convey many different aspects of their social background. These range from broader attributes, such as whether the person is a native speaker of the language they are speaking, their regional origin, their social class background or their educational background, to more specific attributes, such as their friendship groups and even the kinds of television programmes they watch (cf. Stuart-Smith et al., 2013). All of these factors contribute to the variation inherent in the speech that we hear every day. The research reported in this thesis is motivated by the question of how we, as listeners, decipher aspects of a speaker's social background, based on these linguistic variations.

In particular, this thesis addresses the development of regional accent awareness and how we develop the ability to categorise speakers into groups based on where they are from. A common experience for many British people is to encounter a speaker and try to work out where they are from on the basis of their accent. Often, a listener will comment on the fact that they are able to name a speaker's place of origin because they sound like someone else that the listener knows from the same region. These connections, between familiar and unfamiliar speakers and between individual and group-level properties, lie at the foundation of the research reported in this thesis.

Adults vary in their ability to differentiate between regional accents of their native language. Some individuals are very good at deciphering where a speaker is from and can even pinpoint the features of their accent that are diagnostic of their regional origin. Other individuals are less good at this skill and, although they may be able to tell that a speaker sounds 'different', cannot identify the features particular to a speaker's accent or where the speaker might be from. This variable perceptual awareness amongst adult individuals is likely to be due to their overall experience with different regional accents (as found by Clopper and Pisoni, 2004b), as well as perhaps some aspect of individual skill at detecting phonetic detail. Generally, studies show that adults are better at differentiating accents based on features that are diagnostic of broad regional distinctions, rather than features confined to a smaller geographical area (see section 2.3.3). For the purposes of the present study, it is important to note that, in general, British adults are aware of the phonetic variables indicative of a north/south distinction (Lawrence, 2014 – see section 2.3.3). As Lawrence found, British adult listeners performed well in an accent awareness task based on hearing single-word stimuli featuring stereotypical north/south accent differences (such as the vowels in the BATH, STRUT, FACE and GOAT lexical sets – see section 2.1.3). Listeners were asked to place tokens on a map to indicate where they thought the speaker was from and, overall, they were accurate at placing northern-sounding tokens in the north and southern-sounding tokens in the south. While studies have been carried out in order to investigate the extent of adults' regional accent awareness, little is known about how this awareness develops and what mediating factors may play a role in its development.

In order to understand how we become able to group speakers based on their accent, it is imperative to investigate the development of sociolinguistic skills amongst children. These skills involve children learning the social rules of variation – figuring out who uses particular linguistic variants, in which contexts and what they represent. This knowledge forms an integral part of language acquisition as it helps us to learn important social cues, such as whether we are speaking to someone in authority or whether a speaker comes from a different social class background and how that might affect our social interaction with them. Despite their importance, however, little is known about how these sociolinguistic skills develop or how this knowledge is cognitively stored, as the subject is rarely addressed in sociolinguistic or child language development research. Therefore, a better understanding

of how social-indexical knowledge develops alongside linguistic knowledge is needed. As Nardy and colleagues (2013:257) remark, the acquisition of sociolinguistic variation amongst children "remains a nascent field"; there is a need for further research into how children learn language alongside learning about how language is socially variable (according to the speaker's background or the social situation, for example). This will give us a better insight into the basis of the links between language and social-indexical information, which, in turn, leads to the ability of a listener to group speakers according to this information.

## 1.2 Research context

This research is situated in the expanding area of research investigating sociophonetic variation in speech perception. In addition to work on sociolinguistic production, speech perception studies provide essential evidence for understanding the association between linguistic and social information, which in turn has broader implications, such as understanding how linguistic variation is cognitively stored and accessed and how this might lead to the formation of linguistic stereotypes (see Drager, 2010:473). Understanding more about how the links between linguistic variables and their social indications develop amongst children has wider implications, helping to inform us about how social judgments based on language arise. The current research is not in itself an investigation into the social attitudes of speakers and listeners but, rather, a test of children's ability to recognise accent distinctions. However, this ability is seen as an essential first step in identifying the processes which lead to the formation of such attitudes.

Research into the topic of this thesis requires the combining of evidence accumulated from previous work in different fields of research. As remarked upon by Foulkes (2010:26), language development research has tended to focus on the acquisition of abstract linguistic forms rather than on the acquisition of knowledge about variation. In the meantime, studies in speech processing have reported extensive evidence about the ways in which indexical information (for example, pertaining to an individual speaker's identity) is processed alongside speech. However, these studies are often not concerned with how this information might relate to a speaker's social background or how the listener's social background might play a role in their processing of this indexical information in speech. Furthermore, as Nardy and colleagues (2013:256-7) observe, while studies in sociolinguistics have generally

focused on the social variation inherent in the speech of adults, studies in child language acquisition have often failed to be concerned with the variation inherent in the speech that children encounter in their linguistic input. Therefore, it is important to consider evidence from the fields of sociolinguistics, speech perception and studies of child psychology in order to get a broader perspective on the different aspects involved in children's sociolinguistic development.

There has been little previous research based on children's perception in the complementary fields of speech perception and sociolinguistics. While work in sociolinguistic production has found that children acquire patterns of structured variation early in their language development, this research has not been mirrored in relation to perception. Therefore very little is known about children's developing awareness of linguistic variation and their ability to create meaningful categories based on this variation.

One increasingly popular approach to explaining the parallel development of linguistic and social knowledge is exemplar theory (ExT), a theory of memory and categorisation originally developed in psychology. ExT proposes that we store detailed episodic traces in memory and that these memory traces affect how we process and interpret our future experiences. There has been a rise in the advocacy of exemplar models of memory and conceptualisation in fields of linguistic research relevant to the current study (e.g. sociolinguistics, phonological development, speech perception). These models differ in the extent to which they determine memories as being purely episodic-based, or whether they also involve a process of abstraction. For example, in the speech perception literature, Goldinger (1998) employs an extreme episodic model, named MINERVA 2 (proposed by Hintzman, 1986; 1988), to explain the results of a 'word-shadowing' task in which participants echoed the speaker-specific acoustic details of words they were asked to listen to and repeat. Goldinger interprets these results as indicating that our memory retrieval process fundamentally relies on accessing episodic traces, stored with details such as speaker-specific information.

On the other hand, studies in speech perception have highlighted the problems with advocating a purely episodic-based memory system. For example, studies by Norris and colleagues (2003) and McQueen and colleagues (2006) found that listeners rely on abstracted lexical knowledge in the learning and categorising of speech sounds. In their

experiments, the authors found that listeners interpreted an ambiguous fricative [f-s] sound differently, depending on their prior exposure to this sound. Listeners who were primed with prior exposure in which the ambiguous sound represented [f] in f-final words, were more likely to interpret the ambiguous sound as [f] in testing. The equivalent result was found when the ambiguous sound represented [s]. The authors interpret their results as an indication that feedback from lexical abstractions influences the perception of incoming speech sounds. Therefore, they dispute individuals' ability to rely upon purely episodic traces in memory and, instead, advocate models which depend on both episodic memories and abstractions to explain speech perception on the phonological level.

ExT models used to explain the parallel storing of social-indexical information alongside linguistic information rely on both episodic memories and abstractions across these episodes. In such models, abstractions lead to the formation of categories which we form by comparing new stimuli that we encounter by comparing the stimuli to individual instances (exemplars) we have stored in memory. Therefore, cognitive categories are thought to be composed of the accumulation of individual exemplars with similar properties. These similarities can be due to a range of factors, such as the auditory similarity of spoken stimuli (e.g. their phonetic similarity), or the similarity of the context in which the stimuli were encountered (e.g. in a situation where the listener was being reprimanded). Smith and Zárate (1992) use ExT as a model of our development of social judgments in social psychology. They present evidence suggesting that the social judgments we make about a person can be influenced even by a single previous encounter. For example, if we come across someone who has a very similar hairstyle to someone else with whom we have previously had a negative encounter, we may initially judge the new individual in a similarly negative light. In other words, the hairstyle and the social judgment are linked in our stored exemplar of this encounter. This can also account for the social judgments that a listener makes based on the way that someone speaks. Of particular relevance to the current study, ExT has been proposed as a way to explain how we develop categories pertaining to social-indexical distinctions, such as those based on speakers' regional accents. Foulkes (2010) hypothesises that we cognitively categorise speakers based on the accumulation of individual speakers' exemplars and therefore that differences between individual speakers form the basis for the development of these speaker categories. In order to explore this

theoretical model of the development of cognitive categories, it is crucial to explore the early stages of this proposed link between individual speakers' exemplars and the creation of speaker groups. Therefore, the current research starts the investigation into children's developing awareness of regional accents by investigating children's ability to recognise speakers and the role of a speaker's accent in this ability, before moving on to investigate their ability to use speakers' accents as categorising criteria.

## 1.3 Research questions

This thesis aims to address the question of children's developing awareness of regional accents by investigating pre-school and primary school children's abilities to differentiate and group speakers based on regionally distinctive phonetic accent features. Through a series of four experiments, the overall aim is track the children's anticipated progression from the recognition of familiar speakers to the ability to group unfamiliar speakers according to their regional accent. The research also considers the effect of the children's age (and therefore the potential impact of maturational factors) on this ability. Furthermore, in accordance with an exemplar theoretic account of the building up of social categories from the storing of individual instances in memory, other independent variables anticipated to play a role in this process are considered throughout the analyses of the experiments. The four main research questions addressed in this thesis are:

(1) Is there development in children's awareness of regional accent in the pre-school and primary school years?

(2) Do independent factors, such as the children's age, sex and exposure to regional variation, contribute to their developing awareness of regional accent?

(3) What do the answers to (1) and (2) tell us about how social/speaker-specific information is processed alongside linguistic information?

(4) Can we account for the cognitive storage of this social/speaker-specific information in a theoretical model which relies upon the development of speaker categories building up from individual exemplars?

Altogether, in answering these questions, this research intends to go some way towards developing an understanding of this hitherto under-researched area concerning young children's perceptual awareness of regional accent distinctions.

## 1.4 Aims of the research and implications of the findings

Very few studies have investigated young children's perceptual awareness of regional accents. There is a gap between research investigating perceptual awareness of regional accents amongst young infants (e.g. Schmale et al., 2010; Butler et al., 2011; Best and Kitamura, 2012) and research investigating adults' abilities to categorise speakers on the basis of their accent (e.g. Williams et al., 1999; Clopper and Pisoni, 2004a/b); an account of the development of this awareness in the extensive range of ages between infancy and adulthood is largely missing. Therefore, the primary aim of this research is to investigate this previously under-investigated stage of development.

Furthermore, the few studies in this area of research that have been undertaken (cf. Floccia et al., 2009; Wagner et al., 2014; Beck, 2014) have presented inconsistent results and explanations of their data (see section 2.3.4). This is likely to be due to their contrasting methodologies, and is indicative of the difficultly in designing experiments suitable for young children. Therefore, the research reported in this thesis also aims to contribute to the development of experimental methods for working with young children, including the kind of task design appropriate for their cognitive capabilities (e.g. working memory limitations).

The analysis of the effects of independent variables forms another major focus of this investigation. It is anticipated that findings related to the children's maturation (taking account of their age and their sex) as well as their exposure to variation will serve to inform a usage-based account of regional accent awareness, in which ExT is advocated as the best way of explaining this development.

In broader terms, this research intends to form the basis of further research into the development of social judgments based on speakers' accents. As Drager (2010) suggests, studies in sociolinguistic perception, in tandem with studies investigating sociolinguistic production, can help us to better understand the link between social and linguistic information and how it is cognitively processed and stored. In turn this knowledge is important for investigating the link between speaker-specific and group-level attributes and the formation of linguistic stereotypes based on this link. As described above, ExT can account for the development of these cognitive connections. An ExT approach can explain the development of speaker categories based on the cognitive storing of social-indexical information embedded in individual linguistic exemplars. Furthermore, ExT explains how

social judgments can be formed based on our experience with individual instances, as the judgments associated with our stored exemplars affect our processing of incoming exemplars (see section 2.4.2.).

## 1.5 Thesis structure

The following chapter (**Chapter 2**) presents a review of the literature, which, as indicated above, covers a wide range of relevant subject areas. This is in order to bring together evidence from a variety of different research fields which, overall, support the rationale for undertaking the current study. The literature review starts by giving an overview of accent diversity in Britain, focusing in particular on the accents and features investigated throughout the experiments in this study. Then the review goes on to consider how children acquire an accent in production, principally focusing on studies looking at children's sociolinguistic variation in their production as well as studies which have investigated children's acquisition of a second accent/dialect. . Following on from considering the kinds of variation children use in their linguistic productions, the chapter goes on to situate the little that is already known about children's accent perception in the more general field of their social preferences and in comparison to studies on adults' awareness of regional accents. Next, the theoretical side of the investigation is put under review and exemplar theory is underlined as the most promising way of explaining children's acquisition of social-indexical knowledge alongside their linguistic knowledge. In such an exemplar account, it is proposed that the building of speaker categories develops initially from the storing of familiar speakers' exemplars. Therefore, the chapter finishes by summarising studies in speaker recognition and speech perception, in particular exploring the differences in the processes of familiar speaker recognition and unfamiliar speaker discrimination as well as describing how a speaker's accent is relevant in the process of recognition.

**Chapters 3-5** present the experiments of this thesis which aimed to track the chronological development of accent awareness, from the ability to recognise familiar speakers and the role of the speaker's accent in this process to the ability to categorise unfamiliar speakers based on their regional accents.

**Chapter 3** presents the first two experiments, which were run with pre-school (2-4-year-old) children. The 'Identification experiment' investigated children's ability to identify

familiar nursery teachers' voices from short audio stimuli. The 'Recognition experiment' then went on to investigate children's ability to recognise one particular nursery teacher's voice from amongst stimuli of other unfamiliar speakers with different accents.

Following on from these two experiments, **Chapter 4** presents the 'Grouping experiment', also run with pre-school children (although this time with 3-4-year-olds). The Grouping experiment investigated children's ability to group speaker guises, from an unfamiliar speaker, into categories based on a regional accent distinction. The final experiment, the 'Second grouping experiment', is presented in **Chapter 5**. This experiment builds on the Grouping experiment by investigating the ability of older, primary school children (5-9-year-olds) to categorise different unfamiliar speakers on the basis of different regional accent distinctions.

In **Chapter 6**, the General Discussion consolidates the results and discussions of the experiments described in chapters 3-5. The chapter starts by presenting an overview of the investigation. It then goes on to discuss in more detail how the findings from the experiments contribute to knowledge about development in children's accent awareness in the pre-school and primary school years and the role of independent variables pertaining to the children's maturation and their exposure to variation in their linguistic input. Following on from this, the chapter details a theoretical account of the experimental findings, based on the tenets of exemplar theory and incorporating research from studies in voice perception. Finally, this chapter discusses the limitations of the research.

**Chapter 7** forms the overall Conclusion. This chapter presents a brief overview of the thesis and the overall research findings along with a summary of the contributions this research has made to the relevant research fields. The chapter ends with considering future directions for the research; in particular how this work may underpin further research investigating how social evaluations about speakers with different accents develop amongst children. This is considered to be an important step in investigating the formation of linguistic stereotypes.

# Chapter 2: Literature Review

This chapter will review the literature in five broad areas of research, each relevant to the usage of the current study. Section 2.1 introduces the topic of accent diversity and gives an overview of English accent variation in the UK before describing the accents and accent features relevant for the thesis. Section 2.2 gives an overview of work on accent production, in particular investigating how children acquire their accent and how they acquire sociolinguistic variation in their production. Section 2.3 moves on to evaluate work on accent perception, tracking both children's emerging social preferences and their ability to group speakers by their regional accent. Section 2.4 approaches the theoretical side of the investigation, highlighting exemplar theory as the most promising way to account for the acquisition of social-indexical information. In an exemplar account it is proposed that the building of speaker categories develops initially from the storing of familiar speakers' exemplars. Therefore, Section 2.5 explores speaker recognition, focusing on what makes a speaker's voice familiar to a listener and how a speaker's accent is relevant in the process of recognition.

## 2.1 Accent diversity

### 2.1.1 Defining the term 'accent'

At the most general level, the term 'accent' refers to "variations in pronunciation" (Hughes et al., 2012:3) and is therefore based on phonetic and/or phonological distinctions. These patterns of variation are typical of particular communities of speakers who are in some way socially or regionally distinct from other communities of speakers. While geographical region is the most common distinction used to define an accent community, other distinctions such as social class or ethnicity may also underlie accent differences.

As noted by Hughes and colleagues (2012), in studies of English varieties, the term 'accent' is often used interchangeably with the term 'dialect'. Although the definition of a dialect is not consistent across different languages (see Britain, 2004), the term usually refers to a language variety differentiated from other varieties through its distinct use of grammar and vocabulary, as well as features of pronunciation. Therefore, dialect varieties often go together with accent varieties, particularly when both are regionally defined. For example,

someone who speaks the Yorkshire dialect will also have a Yorkshire accent (although the reverse would not necessarily be true).

In this thesis, I am investigating the awareness of regional accent differences based on phonetic/ phonological criteria. However, many of the studies that I reference throughout the thesis are based on broader studies of regional dialect distinctions. Therefore I use the double term 'accent/dialect' in order to indicate the relevance of these studies to both accent and dialect varieties.

### 2.1.2 Studying English accent/dialect variation

The study of accents and dialects has a long tradition in the UK, with dialectology fieldwork starting in the 19th century (e.g. Wright 1898-1905, Orton 1962). This fieldwork saw the start of the creation of dialect maps which tracked the differences in regional dialects by documenting them as isoglosses on a map. Since this time, however, many other ways of investigating the influence of different social factors on language use have developed. Whereas the earlier dialectology studies tended to focus on the speech of older men, in rural locations, Labov (1966) led the way in the study of urban speakers who were sampled to represent a mixture of people in society. This has resulted in the modern focus of sociolinguistics, which has been to investigate speakers' language use in line with different aspects of their social background, such as their gender, age, social class and ethnicity as well as their regional location (e.g. Trudgill 1974).

Furthermore, different discoveries about group patterns of linguistic behaviour have led to different definitions of how particular groups can be defined as 'speech communities' – for example as social networks (Milroy, 1992) and as communities of practice (Eckert, 2000). An individual's social network refers to all of the social connections that they have with other individuals and Milroy (1992) argues that linguistic change is more likely to occur when these connections are loose and weak (i.e. when not all individuals know each other/ they don't know each other very well). On a smaller scale, a community of practice refers to a self-defined social group that arises from mutual engagement in a common interest (such as members of a sports team). For example, Eckert (2000) found that individuals in a U.S. high school used linguistic variants in order to mark themselves as belonging to particular communities of practice - the school-oriented 'Jocks' vs. the anti-establishment 'Burnouts'.

Overall, the various different ways that speech communities have been defined in sociolinguistics indicate the far from straightforward relationship between a speaker's way of talking and their regional location. The different kinds of linguistic variation that result from different social criteria demonstrate that speakers' accents cannot be precisely categorised in terms of just one of these social factors. Therefore, any distinctions between groups of speakers that are made along regional divisions are based on gradient criteria relating to the statistical use of certain linguistic forms by one particular group more than another (Foulkes and Docherty, 2006).

The distinction between standard and non-standard regional accents/dialects in British English is a further demonstration of the complicated relationship between different varieties throughout the country. Kerswill (2006) defines 'Standard English' as a social dialect, closely related to the standardised written form of English and regarded as the most prestigious variety of the language by the majority of the population. Therefore this form of Standard English relates mainly to grammatical and lexical language use. As a dialect, Standard English can be spoken with different standard (or non-standard) accents.

'Received Pronunciation' (RP) is an accent which arose in Victorian England as a marker of high social standing and because of its prestige became publicly regarded as the 'standard' accent of English in England. Although RP is primarily a social accent and therefore can be spoken by people from different regions, because of its origins in London high society, it is phonologically a South Eastern accent (Kerswill, 2006). Features of RP have changed over the years and its use has spread to speakers other than the upper classes of society. Therefore, the newer term 'Standard Southern British English '(SSBE) is now commonly used by linguists to refer to the modern instantiation of this prestige accent (and this will be the preferred term used throughout this thesis). This term also highlights the regional origin of the accent and its overlap with southern English accents more generally.

In Scotland, there is a separate Scottish standard accent, 'Scottish Standard English' (SSE). SSE is the accent of many Scottish people who speak the Standard English dialect (Carr 1999: 156). In comparison to standard accents such as SSBE and SSE, all other accents (including all regional varieties) can be defined as 'non-standard'.

In England, speakers from the South generally have an accent more similar to SSBE than speakers from the North (and are therefore often regarded as 'posh' by Northerners). However, because SSBE is primarily a social accent, the use of SSBE is usually indicative of a speaker's higher social class (where social class can be defined by measures such as a person's education, income and/or occupation). The relationship between accent and social class is again a gradient one; people of a lower class tend to have much more regional variation in their pronunciations than people from a higher social class, who tend to have a more 'standard' (SSBE-like) accent. In other words, speakers from a lower social class are more likely to be heard using non-standard accent features in their speech. This is also true of the relationship between SSE and non-standard Scottish regional varieties. As Stuart-Smith (2003) describes, many Scottish speakers move along a continuum of using more or less SSE features vs. Scots vernacular accent features and this can be an indication of their social class background.

Alongside the ubiquity of standard forms used in standard accents such as SSBE and SSE, there is also evidence of the spreading of non-standard accent features throughout England and Scotland. For example, several studies have reported an increase in TH-fronting, a non-standard feature from London English, in different British regional varieties. These include: Milton Keynes, Reading and Hull (Williams and Kerswill, 1999), Glasgow (Stuart-Smith, 1999) and Derby and Newcastle (Docherty and Foulkes, 1999). The spread of TH-fronting is a demonstration of the influence of the covert prestige associated with non-standard varieties and their features. As well as the overt prestige of standard forms, therefore, language change can be driven by the covert prestige of non-standard forms and these different ideological motivations provide the population at large with "a multiplicity of linguistic norms" (Kerswill 2006:3).

### 2.1.3 Accents and features relevant to the current study

Despite the unstable nature of accents, their varying distinctions and their changing features, there are accent contrasts in Britain which are representative of a large enough proportion of speakers and which have been stable for a long enough time to consider them as typical for speakers from a certain region.

Hughes and colleagues (2012) describe Britain as having an 'accent continuum' in which different regional accents can be measured on a scale of similarity, as neighbouring regions sharing certain accent features but not others. While careful not to overstate the comprehensiveness of regional accent isoglosses, they divide the country into broadly defined groups which mark the use of certain features (see Figure 2.1). I will use this map to differentiate the main regional accent distinctions and features focused on in this thesis; these are the Yorkshire, SSBE, North Eastern and Scottish accents. In Figure 2.1 Yorkshire takes up the majority of the Central North (CN) region marked on the map and the North East region is marked as such (NE). Scotland is also differentiated from England along with Northern Ireland. As mentioned above, SSBE is phonologically a South Eastern accent (marked SE on the map).

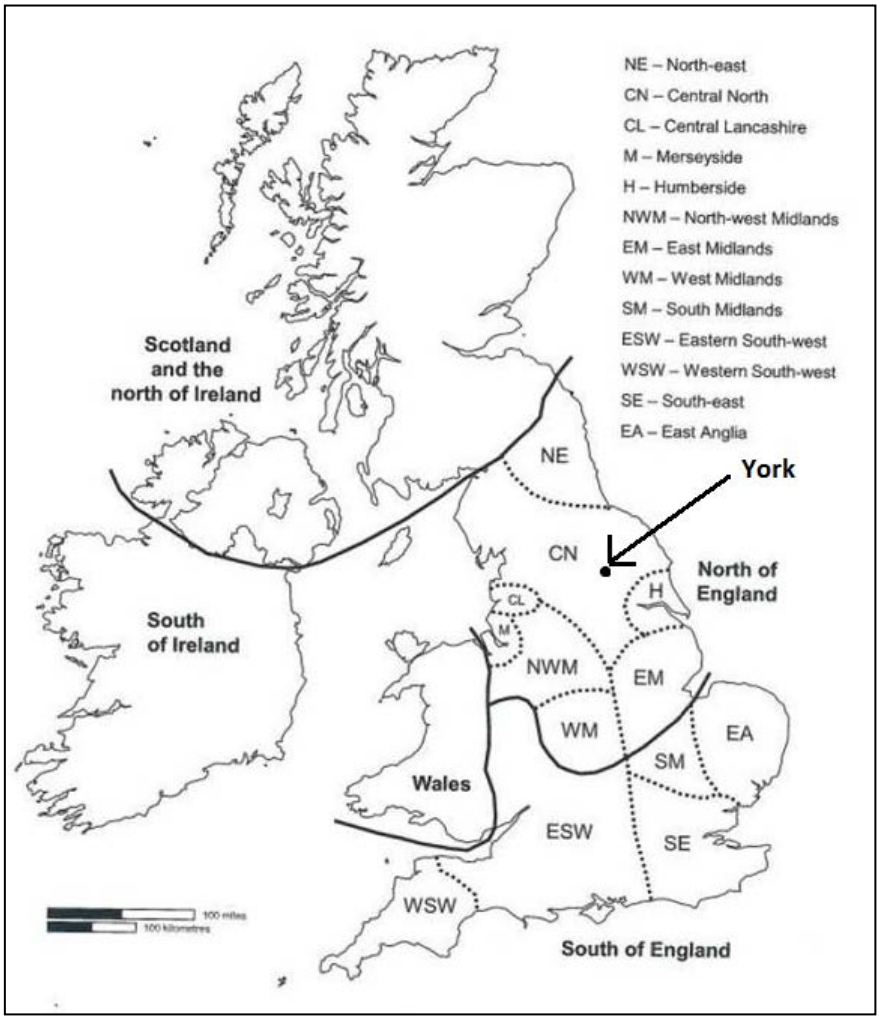



**Figure 2.1: Regional accent groups (adapted from Hughes et al. 2012:71) with York pinpointed**

The focus of this thesis is on children from York, North Yorkshire, which is situated in the county of Yorkshire, in the Central North of Britain, as indicated in Figure 2.1. For the purposes of this thesis, I will use the 'Yorkshire accent' as the point of reference for the regional accent of speakers from this area. The Yorkshire accent also extends to cover the accent of speakers from the other ridings (i.e. county subdivisions) of Yorkshire (West and East) which the city of York itself borders (see Figure 2.2). Although there are differences in the accents between speakers from the different ridings, there are many more commonalities. Therefore, studies based on the York accent (e.g. Haddican et al., 2013; Tagliamonte and Roeder, 2009) often use results from studies across the rest of Yorkshire as a comparison with their findings. Furthermore, the accent features referred to in this thesis are not reported as differing across the regions of Yorkshire and therefore the reference to Yorkshire speakers and the Yorkshire accent is thought to be justified.


**Figure 2.2: Map showing the relationship between current administrative areas and the Yorkshire boundary, with York circled (adapted from The Yorkshire Ridings Society, 2013)**

In Yorkshire, the vowel in the BATH lexical set (as devised by Wells, 1982a) patterns with the vowel in TRAP and is pronounced [a]. This is also true for speakers from other regions in the north of England, such as the North East and most speakers of Scottish Standard English. In contrast, for speakers from the south of England and speakers of SSBE, this vowel patterns with the vowel in START and is pronounced [ɑ:]. The vowel in STRUT patterns with FOOT in the north of England and is pronounced [ʊ], whereas in the South the vowel in STRUT splits from the FOOT vowel and is instead pronounced [ʌ]. In Scotland, STRUT is pronounced [ʌ] like SSBE, while FOOT patterns with GOOSE and is pronounced [u].

The differences in pronunciation of the STRUT and BATH vowels are often seen as linguistic stereotypes (in Labov's 1972 terms) of the north/south of England as they are often overtly commented on by lay listeners and are known to characterise speakers from these two broad geographic regions. Their pronunciations also appear to largely cross social class variation and the non-standard Northern variants are used by both middle class and working class speakers from the North (although with some gradient variation and stylistic shifting, as described by Wells, 1982b:351-353).

Other features of the Yorkshire accent are more socially differentiated. The monophthongal vowels [e:] and [o:] found in the lexical sets FACE and GOAT, are the 'mainstream' Northern variants (Watt, 2002) and are used by speakers in the Central North and North East of England as well as in Scotland. However, they are often not used by middle-class speakers in these areas, who use the more typically Southern/SSBE diphthongal pronunciations [eɪ] and [əʊ]. Despite their general use across the north of Britain, it has been found that lay listeners are more likely to associate the monophthongal pronunciations with speakers from the North East rather than the Central North of England (Lawrence, 2014 - see section 2.3.3). This is perhaps a reflection of the rise of the use of the diphthongal variants in Central North regions such as York, a change in progress, as reported by Haddican and colleagues (2013).

A stereotypical feature differentiating Scottish (and Northern Irish) accented speakers from English speakers (apart from a minority of speakers in the South West and North West) is the use of postvocalic /ɹ/ after vowels in the NURSE and lettER lexical sets, amongst others. Although recent studies have found some decline in the use of rhoticity in large cities such as Glasgow (e.g. Stuart-Smith, 2003), there is still a pervasive link between rhoticity and an English/Scottish distinction. This is evidenced by the frequent reporting of (r) as a

diagnostic feature of Scottish speech by speakers from the English/Scottish border (as described by Llamas, 2010).

This section has explored accent diversity in Britain as a basis for the investigation of how an awareness of these accent differences develops in childhood. Because children's perception of variation has been found to be intimately linked to their own production (see Barbu et al., 2013), the next section investigates the process of children's accent acquisition.

## 2.2 Acquiring an accent: children's production

This section explores children's acquisition of sociolinguistic variation in relation to their accent. This reveals an important link between the linguistic input that children receive and their acquisition of sociolinguistic variation. The section goes on to review studies in second dialect acquisition which have shown the limitations of accent acquisition.

### 2.2.1 The emergence of sociolinguistic variation in production

When children acquire a language, they are also acquiring an accent. They learn to speak their native language(s) with particular pronunciations. In order to understand how an accent is acquired it is necessary to question where these pronunciations come from (children's parents, their peers, or perhaps both?) and when these pronunciations become fixed, if at all (as addressed in section 2.2.2).

Even within a region and a community there is much linguistic variation for a child to learn. This includes local vs. non-local forms and forms which indicate a speaker's age, gender and social class, all of which can be reflected to varying degrees. In his original developmental model of the acquisition of standard spoken English, Labov (1964) posits that production of local vernacular forms occurs around the age of 5 and before children pick up on the social significance of different forms. However, this has since been disputed (cf. Nardy et al., 2013) as studies have found that children from a pre-school age acquire complex sociolinguistic variation of different kinds from their community. Foulkes and colleagues (1999) found that 2-4-year-olds in Newcastle acquired accent-specific phonological variation. They followed adult-like patterns of production in their realisations of the variants of /t/, which differ according to whether they are in initial, intersonorant or pre-pausal position. Similarly, Roberts (1997) and Roberts and Labov (1995) found that 3-4-year-olds in

Philadelphia had successfully learned the variable rules of their community regarding the deletion of final /t/ and /d/ in consonant clusters and most had acquired a sound change in progress (the raising and tensing of short *a* in Philadelphia). Pre-school children have even shown stylistic shifting; for example, Smith and colleagues (2007) found that some children used more of the local Scottish lexical variant *hoose* [hu:s]when in more informal contexts (e.g. when playing with their parents), compared to *house* [haus] when they were situated in a more formal context (e.g. when being taught something by their parents). The acquisition of stylistic shifting was found to be related to the amount of stylistic shifting the children were exposed to by their caregivers, emphasising the importance of children's linguistic input.

The importance of the input for indexing social meaning in relation to gender has been uncovered by Foulkes and colleagues (2005). They analysed the Child Directed Speech (CDS) of 39 mothers in Tyneside speaking to their 2-4-year-old children and found that when speaking to girls, the mothers used more standard forms in their speech. Caregiver's speech was analysed in relation to community norms by Smith and colleagues (2013). Their study in Buckie, Scotland investigated the differences between the use of local forms in caregiver speech compared to the community norms but also in comparison with their (2-4-year-old) children's speech. They found that fewer local phonological and lexical variants were used by the caregivers compared to the community, whereas morphosyntactic features were used at similar rates. These differences were reflected in the children's speech, as they tended to use fewer local phonological and lexical variants but the same proportion of local morphosyntactic variants. These results show the importance of the caregiver's input but also the over-riding importance of community patterns, with regard to some variables. Evidence from other cultures in which speech directed to children is not the norm (Shneidman and Goldin-Meadow, 2012) reveals that even in this situation, the speech directed from adults to children is a good predictor of children's word learning. In their study of a Mayan village community, Shneidman and Goldin-Meadow found that children's vocabulary at 35 months was related to the number of different word types that they had heard in the speech specifically directed to them at 24-months, rather than the number of word types that they overheard in the speech around them. This suggests that child-directed

speech has a cross-cultural importance in children's linguistic development and therefore that each child's individual exposure to variation is an important line of enquiry.

The importance of linguistic input will be one of the main focuses of the experiments in this thesis. As the experiments in this thesis are primarily investigating children's accent awareness, section 2.3 explores what is already known about accent perception amongst both children and adults. Firstly, the next section considers how fixed a speaker's accent from childhood becomes, and the extent to which it can change over time.

### 2.2.2 A 'critical period' in accent/dialect acquisition?

The previous section established that from a young age, children develop sociolinguistic variation in their production and that their exposure to linguistic input plays an important role in this acquisition. A leading question from these findings relates to the extent to which these variations in children's productions are influenced by age and maturation factors, and how likely children's productions are to change if they are exposed to new variation from different linguistic input (i.e. a different linguistic community).

Support for the idea that pronunciations are fixed from an early age comes from the well-known 'critical period hypothesis' (CPH), made popular by Lenneberg, 1967. This hypothesis stipulates that humans are biologically programmed to learn language at a young age and after adolescence the ability to learn a language to a native-like standard rapidly declines. The linguistic system we learn as a child is therefore thought to be more stable than anything linguistic we might learn later in life. The existence of a critical period to such a degree of innateness has been heavily criticised and disputed ever since. As DeKeyser and Larson-Hall (2005) summarise, although there is general agreement amongst researchers that a learner's age of acquisition (AoA) is negatively correlated with their overall language proficiency, how this age effect is influenced by other factors remains controversial. In presenting a fairly balanced overview of research in this field, DeKeyser and Larson-Hall cite a number of studies dating from the 1970s which have presented evidence for the critical period, mostly in the form of grammaticality judgement tests. The authors also present research with counterarguments to the CPH, such as the fact that the linguistic input that children receive is more likely to be more extensive and/or varied than adults,' and that children's attitude towards language learning is less likely to be affected by psychological

effects such as being self-conscious. However, the authors' general conclusion is that these other factors do not do enough to explain the AoA effects and that the CPH cannot, therefore, be dismissed altogether.

Other researchers are more critical of the CPH, particularly in terms of the critical period not extending beyond adolescence. For example, Birdsong (2005) presents evidence that age effects on language acquisition persist even after adolescence and argues that a decline in linguistic proficiency is due to the aging process more generally, rather than maturational processes leading up to adolescence. Furthermore, Birdsong and Vanhove (2016) present concerns over how we can objectively measure the 'native-likeness' of language acquisition and how analysing the results of acquisition data in different ways can lead to contrasting conclusions.

Despite Birdsong and Vanhove's overall conclusion that there is, thus far, no convincing evidence for a critical period in language acquisition that ends after adolescence, this argument is mainly based on evidence that the difficulties in language acquisition persist after adolescence and are therefore still age-related. The effect of an age-related difficulty on a speaker's acquisition of accent features is relevant when considering whether it is possible to change our accent or acquire a new one throughout childhood and beyond. Work on second dialect acquisition has tried to find answers to these kinds of questions by investigating children's learning of pronunciations as well as grammatical forms and lexical items particular to their dialect. Studies have shown that children are able to acquire a second dialect to a certain extent but that this depends on the children's age as well as the particular features they are acquiring. Payne (1980) looked at a community in King of Prussia, Philadelphia, and studied the phonetic variables of out-of-state children who had moved from elsewhere (such as New York) at various ages. Payne found that phonetic variables that could be acquired by the addition of a simple rule such as: [-low] → [-back] / [-cons, -back] (which fronts the vowels in the diphthongs <uw> and <ow>), were acquired successfully by more than 50% of the out-of-state children. There was an effect of age and length of time in the community though, as children who had moved to King of Prussia at an earlier age and children who had lived there longer tended to be more successful in acquiring these phonetic features. Of particular significance was the lack of acquisition of

these variables by the 10-14-year-old age group, who moved to the area around the time of adolescence.

However, Payne found that children of an even younger age were largely unable to produce another Philadelphian phonetic variable: the split of short-ɑ into a tense and lax distinction. Rather than a simple rule addition, this is a rule that is phonetically, grammatically and lexically conditioned. Only children whose parents were also born and raised in the area were able to successfully acquire these particular variants. Similarly, in his Norwich study, Trudgill (1981) found that speakers who were born and raised in Norwich were unable to acquire the phonetic [u:] / [ʌu] distinction (between words such as *moan* and *mown* for example) unless their parents were also from Norwich.

Building upon these studies, attempts have been made to measure the difficulty of acquiring particular dialect features at particular times. For example Chambers (1992) sets out eight 'principles' of dialect acquisition. These principles hypothesise the process by which a dialect is acquired, and account for observations such as those found in Payne and Trudgill's studies. For example, Principle 3 states that "Simple phonological rules progress faster than complex ones" (Chambers, 1992: 682), which explains why more fine-grained distinctions might only be picked up when a child has been in the community for a long time. Kerswill (1996) focuses on the relative difficulty of acquiring different kinds of second dialect features, which he puts into a hierarchy. In his ranking, phonological rules are posited as the most difficult features to acquire, whereas new lexical items are much easier to acquire and therefore continue to be acquired throughout the lifetime.

While these measures indicate the hierarchical difficulty of learning certain features, there is much individual variation in the ways and means by which children develop their accent/dialect. A child's family, their peers, milestones such as starting school and the influence of the standard variety all have an effect on the linguistic features they encounter, as well as which ones they pay more attention to. Trudgill's (1981, 1986) study of 7-year-old twins who moved to Australia from Reading found that the children acquired Australian accent features both in different orders and with different levels of regularity. Tagliamonte and Molfenter (2007) tracked the acquisition of British linguistic variables by three Canadians who moved to York under the age of 5. They found that although the phonological rules the children acquired did not reach categoricity, their acquisition accelerated at

particular points, such as starting school. This demonstrates the strong influence that children's peers have on their accent development at this age, which is due to both the children's increased exposure to local forms, as well as their social motivation to sound like their peers.

Although children respond to community and peer influences, particularly around the start of school, Chambers (2002) and Hazen (2002) claim that children are not overtly aware of any differences between their parents' speech and the speech of their friends. Hazen's (2002) study found that a boy born and raised in the Northern United States acquired the accent features of his community rather than his Southern parents but thought that he sounded the same as the rest of his family. In a similar vein, Chambers' (2002) research centered on a study of 'Ethan', a boy raised in Toronto with Eastern European immigrant parents. Ethan's accent did not show any sign of influence from his parents' non-native accents and, until early adolescence, he did not show any awareness that his parents' accent was different to those in their community. Conversely, bilingual children's pronunciations have been found to show influences from their parents' non-native accents, particularly during code-switching. Khattab (2007) found that English-Arabic bilingual children produced some vowels and consonants more similar to their parents' non-native accent, but only when they were speaking in bilingual mode. For example, when only speaking in English, the children pronounced the GOAT vowel in the same way as monolinguals in their local community, as [ou]. However, when code-switching between English and Arabic, they were found to use a close monophthong realisation of the GOAT vowel [o:], heard in their parents' non-native accent. Khattab (2013) describes how these switches between different variants have been found to depend on the child's use of language (whether the child is monolingual or bilingual mode) but also the communicative situation itself (whether the child is trying to establish closeness or distance from their parental interlocutor). These influences on the children's phonetic realisations indicate that although children might not be able to express their overt awareness of native/non-native accent differences, they are still able to use this extra element of variation in their linguistic input as a communicative tool.

The relationship between children's acquisition of a second dialect and their age does not appear to be as straightforward as the critical period hypothesis suggests. In fact, some

aspects of a speaker's accent can change even after adolescence; again this very much depends on the individual's experience and is often motivated by expressions of their linguistic identity/attitudes. For example, as described by Labov (1972), Swiss German women are known to change their dialect if they get married and move to another village because they risk being mocked for using their native dialect. Other examples of accent changes in adulthood present instances of less conscious adjustments on a smaller scale. Sankoff (2004) found that one speaker (originally from the north of England) increased his pronunciation of the Southern [ɑ:] in BATH rather than the Northern [a] between the age of 28 and 35 after spending time in other parts of the country, including London. The changing of individual features of an adult's accent/dialect in this way is usually referred to as 'accommodation' (from Giles, 1973), whereby a speaker converges to sound more like their interlocutor/speech community in order to decrease social distance. While the processes of accommodation and acquisition are usually considered separately, Trudgill (1986) puts them on the same scale, with 'full accommodation' amounting to 'acquisition' of a second dialect. Nevertheless, he still posits an adolescence-related age restriction on what he calls 'full accommodation' (i.e. the acquisition of a second dialect). He hypothesises that under the age of 8, full accommodation is possible, while between 8-14 it is a lot more difficult and over 14 it is very unlikely.

Rather than a 'critical period' in dialect acquisition based on age, however, it could be that perceptual interference from a speaker's first dialect affects their ability to acquire a second dialect. Evans and Iverson (2007) found a perception-production link in their study of adolescents from the north of England. The participants were studied before and after attending university in different parts of the country, where they were immersed in a multidialectal environment. The study found that those participants who produced more Southern-sounding vowels in their own speech were more likely to choose Southern-sounding 'best' exemplars in perception. Therefore, the authors conclude that having a specific dialectal perceptual phonetic space may inhibit the perception and production of phonetic variables outside of this space. While this space is still being moulded during childhood (see perceptual studies of accent/dialect variation in infants - section 2.3.2), it is more rigid in adulthood and this is why we tend to see a difference between children and adults' acquisition of accent/dialect variation. However, this perceptual space can still be re-

moulded and adjusted during adulthood. This is more likely to occur when an individual is exposed to more variation in their linguistic input (e.g. Logan et al., 1991 - see section 2.5.4).

Overall, studies in second dialect acquisition have found that the pre-adolescent years (before the age of about 11/12) are important in shaping a child's developing accent/dialect. Throughout these years, community patterns are prevalent and often children acquire linguistic variants from the wider speech community rather than solely from their parents.

## 2.3 Accent perception

While children show sociolinguistic variation in their production from a young age, the question of whether they can perceive this variation has not been comprehensibly investigated. While experiments with infants have investigated their ability to discriminate between speakers with different accents (see section 2.3.2), studies with adults have focused on the harder task of categorising speakers according to their accent (section 2.3.3). The few studies which have investigated the development of accent awareness between the ages of infancy and adulthood are described and critiqued in section 2.3.4. The main criticism of these studies (as well as of the infant studies) is that most of them show a lack of linguistic rigour in both their experimental design and in their analyses. They do not narrow down the particular accent features under investigation, making it very difficult to evaluate what the children are actually responding to during the experiment. First of all, in order to situate children's perception of accents in a more general cognitive framework, the next section will explore children's developing social preferences.

### 2.3.1 Children's developing social preferences

Studies in developmental psychology have found that children from 5-6-years old show an in-group bias towards those of their own race (Baron and Banaji, 2006; McGlothlin and Killen, 2010), nationality (Rutland et al. 2005) and native accent (Kinzler et al. 2007). This has been found to be linked to exposure, with children from more homogenous school environments more likely to display this bias (McGlothlin and Killen 2010). Furthermore, Howard and colleagues (2014) have found a link between neighbourhood exposure and social preference. In their study, infants who were exposed to more non-native speakers in their neighbourhood showed more inclination to attend to instructions from a non-native speaker. Phillips and Shonkoff (2000) expand on this claim by emphasising the wide-

reaching effects of neighbourhood diversity in a child's pre-school years. The authors make a general observation that, to some extent, children's early social perception is influenced by interactions with those other than their parents.

Children's attitudinal development towards different regional accents is another manifestation of their early social bias. While children have been found to start out with a general in-group preference, this attitude changes by around the age of 10 when they become aware of the prestige of the standard variety of their language and of negative stereotypes towards non-standard accents/dialects. Giles and colleagues (1983) found a change in attitudes in a group of 7-10-year-old children from Bristol. While the younger children showed a more positive attitude towards the non-standard Welsh accent, rating the accent higher for traits such as *clever* and *successful,* the older children rated the RP accent higher for these traits. Similarly, Kinzler and DeJesus (2013) found a change between 5-6 and 9-10-year-olds from Northern and Southern America. The 5-6-year-olds from both regions made more positive judgements about their local accent, whereas the 9-10-year-olds from Southern America made judgements akin to widely held adult stereotypes (e.g. Northerners sound intelligent, Southerners sound friendly). While these studies find an age effect, they do not investigate the children's linguistic background. Exposure to different kinds of linguistic variation would be worthy of investigation here, both in terms of parental input and the potentially wide-reaching effects of the media. This would test assertions such as the one made by Lippi-Green (2012), who claims that Disney's use of non-native accents to represent 'bad' characters teaches children to discriminate. This claim has been widely disseminated as a warning of the accountability of film and television characters for children's misinformed prejudices. However, this assertion does not appear to have been validated through evidence from perceptual experiments with children.

Even younger children have shown a preference for standard over non-standard forms in France. Barbu and colleagues (2013) found that French pre-school children as young as 3-years-old were more likely to positively evaluate a standard variant (realisation of the variable liaison e.g. [pətituʁs] / *'small bear'*) rather than the non-standard variant (zero realisation [pətiuʁs]). They found a link between the children's social class and this preference: children from a higher class background, with more exposure to the standard form themselves, showed this preference at an earlier age than those from a lower class

background. However this study highlights the fact that cross-cultural differences may play a role in the perception and strength of the standard. Whereas in France, Standard French is still the expected form of spoken language in educational, work and formal settings, in Britain this is increasingly not the case. Although SSBE is still the most highly regarded accent, it does not hold the same level of prestige it once did – when it was more usually referred to as RP (Received Pronunciation). The wider acceptance of non-standard regional British accents is reflected in their increasing representation in the media. As described by Stuart-Smith and colleagues (2013), the increased use of speakers with non-standard regional accents in the broadcast media means that listeners have become much more aware of regional diversity. This has caused some studies to propose possible effects of the media on language change, usually with respect to non-standard forms which carry covert prestige and are subsequently adopted by young people (Trudgill, 1986; Williams and Kerswill, 1999; Stuart-Smith et al., 2013). What does not appear to have been investigated, however, is how listeners' exposure to accent variation in the media might affect their ability to perceive accent differences more generally. This is a hard question to investigate empirically as it would require keeping very detailed accounts of listeners' exposure to different speakers on the television and on the radio etc. However as generations of listeners are spending increasing time listening to broadcast media, it is becoming more and more important to consider the media as a source of linguistic input in its own right.

Overall, the studies in this section show that young children between the ages of 3 and 10 are developing preferences and social attitudes for certain ways of speaking, pertaining to regional/non-standard vs. standard language varieties. However, these attitudes and preferences do not indicate their level of awareness of the different linguistic features of these varieties. This awareness is important in order to better understand the foundation of these developing evaluations. Such evaluative judgements about accent are particularly important in the UK situation, where there is a complex relationship between accents, regions, social class and what is considered 'standard' (as described in section 2.1.2).

### 2.3.2 Accent perception of infants

Studies with infants show the effects of their familiarity with their local accent as compared with other, non-local accents. Butler and colleagues (2011) found that in a Head-Turn Preference Procedure (HPP), infants at 5 months old showed a preference towards their

local West Country accent compared to unfamiliar Welsh and Scottish accents. The HPP measures infants' attention to an audio stimulus by calculating the amount of time they turn their head and look towards it compared to the amount of time they turn their head towards other audio stimuli. A longer head-turn time therefore indicates a preference for a particular stimulus (see Kemler Nelson et al., 1995). In the authors' interpretation of these results, the children show the ability to discriminate between accents by a familiar/different distinction whereas they are unable to discriminate between two different accents. However, it is unclear whether the children's attention was drawn to the speakers' accents in particular; they heard different passages from different speakers and there was no systematic analysis of the accent features used in the passages. Furthermore, the speakers were told to read the passages in an infant-directed style and therefore their intonation patterns were likely to vary considerably. Therefore, the prosodic variation of the individual speakers and their accents could have played a role in the results of this experiment.

Studies of infants' word learning have found that accent differences prevent the recognition of familiar words when they are heard in an unfamiliar accent (cf. Schmale et al., 2010; Best and Kitamura, 2012). Although the exact age is disputed, these studies have found that generally by 18 months, infants are able to abstract across different and unfamiliar accents in order to understand familiar words. Van Heughten and Johnson (2015) highlight the importance of task design when interpreting infants' responses. They found that 28-month-old Canadian toddlers were better able to understand task instructions heard in an unfamiliar Scottish accent when the stimuli were presented in full sentences rather than isolated words. Therefore perhaps the ability to abstract over accent differences is more easily achieved when children hear longer stimuli both because they have more content to analyse and because the surrounding context can help to disambiguate words heard in isolation. However, this may be true of word learning in general and have little to do with the presence of an 'unfamiliar' accent; in this study of Canadian toddlers, the children scored above chance in both conditions of the experiment and therefore seemed to be relatively comfortable interpreting the Scottish accent generally. Furthermore, the authors' transcriptions show that there were only small differences between the Scottish and Canadian pronunciations of the words used in the experiment.

There are conflicting results in the infant perception literature regarding the role of children's previous exposure to accent varieties. Van Heughten and Johnson found that prior exposure to the Scottish accent (through listening to a story passage in the laboratory) did not result in children more accurately understanding the accent. However, Schmale and colleagues (2012) found that two minutes of prior exposure to an unfamiliar (Spanish) accent helped 24-month-old Americans to learn words in this accent. These inconsistent results could be interpreted as creating difficulty for an ExT account, in which prior linguistic exposure plays a pivotal role in the processing of speech. However, these studies do not show that the children understood what the speakers were saying in the prior exposure part of the task. Therefore, it is problematic to generalise the results from this part of the task on to the second part, as the prior exposure can only be interpreted as having an effect on children's word learning if they understood the words they heard during the exposure phase. Although the results of these studies rely on atypical lab-based exposure, similar inconsistent results have been found in studying children's immersive experience of an accent.

Floccia and colleagues (2012) found that 20-month-old children brought up in a rhotic community were quicker at recognising words pronounced in a rhotic form, regardless of whether they were 'mono-accentual' (with two rhotic parents) or 'bi-accentual' (with at least one non-rhotic parent). They interpret this as suggesting that a child's phonological representations are conditioned by their community rather than by their parents. However, van der Feest and Johnson's (2015) study of 24-month-old Dutch children found a different result. This study investigated two groups of children. One group (the 'Uniform Input' group) had grown up with exposure to only their local variety of Dutch - in this variety all fricatives are devoiced. The other group (the 'Mixed Input' group) had grown up with parents who spoke a different variety of Dutch - in which a fricative voicing contrast is maintained. Therefore, the Mixed Input group were exposed to both a devoicing variety of Dutch and a voicing contrast variety of Dutch. The study found that both groups of children were able to recognise words (featuring voiced/devoiced fricatives) by speakers of both varieties. Unlike Floccia and colleagues' (2012) study, this suggests that the children's input from the community did not outweigh the input from their parents. Furthermore, the Mixed Input group were able to adapt their processing of the speech from the speakers with the

different accents. This group of children were able to detect mispronunciations by the speaker with the voicing contrast accent (their parents' accent) whereas the Uniform Input group of children ignored these mispronunciations because they weren't familiar with the voicing contrast in the first place. Therefore, the Mixed Input group appear to have an advantage over the Uniform Input group; rather than ignoring the input from their parents (as Floccia et al.'s study suggests), they are able to utilise their mixed evidence of a phonological contrast to decide when this contrast is relevant (i.e. according to the speaker's accent).

Van der Feest and Johnson describe a careful procedure for categorising the children as belonging to either the Uniform input or the Mixed input group. Children in the Mixed Input group had both parents from outside the region who both had an accent with a fricative voicing contrast. It was ensured that these children also received regular exposure to their local (fricative devoicing) variety through contact with speakers from their local community and watching programmes on television featuring this variety. Furthermore, children in the Uniform Input group were only chosen if they didn't have parents or secondary caregivers who spoke a variety with fricative voicing contrasts and therefore there was very little chance of them being exposed to a variety with this contrast. The authors' grouping of these children is in line with second dialect acquisition studies (see section 2.2.2) in which children are usually classified as bidialectal if they move to a new town from elsewhere and have two non-local parents. Therefore, the linguistic input from their parents is different from the linguistic input from their community.

The results from Floccia and colleagues' study should be treated with more caution as their methods for differentiating between mono-accentual and bi-accentual children are questionable. Floccia and colleagues label children as 'bi-accentual' even if they only have one parent from outside the local area. As a result, children with two non-rhotic parents and only one non-rhotic parent are grouped together. This makes it more difficult to generalise about the exposure of these children to rhotic forms as some of the children will have more exposure to the community norms at home than others. Moreover, all of the children have some exposure to rhotic forms in their environment and therefore the extent to which they are exposed to rhotic forms is a matter of degree. Furthermore, this mono/bi-accentual classification ignores complications such as the dominant effects of the primary caregiver's

linguistic input (which studies such as Foulkes et al., 1999 have shown to influence young children's use of linguistic variants).

Overall, these studies with infants highlight the difficulty in measuring children's exposure to variation in a consistent way. Although these studies seem to equate a speaker's accent with one particular pronunciation pattern, variation in pronunciation can occur because of many other factors (as mentioned in section 2.2.1) and therefore caution should be taken against categorising accents in a definitive way. These studies with infants are based on speaker discrimination and/or word learning which reflects the infants' familiarity with the individual or accent that they are hearing. While the results of these studies are variable, their focus on speaker discrimination and/or word learning as a result of accent familiarity is a preliminary step to understanding more about how speakers might then become grouped together conceptually as children get older.

### 2.3.3 Categorising accents: adolescents and adults

Adults vary considerably in their ability to (1) identify a speaker's accent and (2) categorise speakers into groups based on their accent. These tasks are dependent on differences in the perceptual attentiveness of listeners, as well as the strength of the speaker's accent and the listeners' familiarity with the accent in question. Atagi and Bent (2013) found that in a free classification task, adult listeners were more likely to group together non-native speakers based on the strength of their foreign accent rather than distinctions between the accents themselves. The strength of the speakers' accents were measured in a previous study (Atagi and Bent, 2011), which asked listeners to rate speakers on a scale of 1-9 (where 1 = 'no foreign accent' and 9 = 'very strong foreign accent'). Adults were also found to be better at identifying the accent (e.g. Spanish, Japanese, Korean) of non-native speakers whose first languages were more similar to their own (Atagi and Bent, 2014). While the results of Atagi and Bent (2014) show a familiarity effect, the overall finding indicates the listeners' lack of knowledge about how geographical areas relate to accent distinctions, rather than revealing anything about their linguistic ability to differentiate between varieties.

A similar familiarity effect has been found for listeners' grouping of speakers according to their regional accent/dialect. In a forced categorisation task in which listeners were asked to group speakers into six dialect regions in Wales, listeners were found to be best at

categorising speakers from their own dialect region (Williams et al., 1999). However, the listeners still only managed to correctly group speakers from their own region 45% of the time. Clopper and Pisoni (2004a) found that U.S. listeners' categorisation of speakers into six dialect regions was similarly not very accurate (average 30% correct). Both of these studies again predominantly highlight the listeners' naivety about geographical boundaries rather than determining their perceptual linguistic skills.

However, on further investigation, Clopper and Pisoni found that broader categorisations arose out of the listeners' groupings. A clustering analysis revealed that listeners' responses centered around three major U.S. regions: New England, South and West. Similarly, in a free classification task in which listeners created their own groups of similar sounding speakers, Clopper and Pisoni (2007) found that a corresponding three-way regional division was used. This study also found that mobile listeners who had moved around the country created more groups than non-mobile listeners, on average. This implies that exposure to more linguistic variation helps in perception, and is further evidenced by Clopper and Pisoni's (2004b) forced categorisation task. In this study, 'Army brats' (adolescents who had lived in at least three different states) performed better at categorising accents than 'homebodies' (those who had lived in the same state their whole lives).

Clopper and Pisoni (2004a) conclude that naïve listeners are aware of distinctive differences between speakers of different dialects of American English and can use this knowledge to group speakers on the basis of broad regional distinctions. Montgomery (2007) investigated broad regional differences in England in his study of the perception of the linguistic North/South divide by naïve listeners. Montgomery's study asked participants to divide the country into a linguistic North/South and also place speakers on a map corresponding to where they thought the speakers were from. On the map, six city locations were provided in order to counteract the participants' lack of geographical knowledge. The results found that the informants relied upon on sociocultural factors, rather than just linguistic facts, when making their decisions. These factors included: cultural salience (the salience of particular cities such as Manchester and their associated accent), claiming/denial (whether particular speakers were desirable enough to be associated with the informant's home town) and proximity (for example, Crewe-based informants shifting the North/South divide to ensure that they were included in the North). Although the informants were presented with city

locations on the maps given to them, this task was still primarily driven by listeners' knowledge of the link between geographical areas and accent distinctions; the choice of the city locations provided likely influenced their choice of North/South boundary.

Lawrence (2014) investigated the effect of particular phonetic variants on listeners' differentiation of Northern and Southern speakers in Britain. He found that in general, when asked to place tokens of sound files on a map, listeners correctly placed Northern tokens of the BATH, STRUT, FACE and GOAT vowels in the North and Southern tokens in the South. This supports the general view amongst linguists that these vowels are diagnostic of a general North/South divide (see section 2.1.3). Furthermore, he found that the Northern BATH and STRUT tokens were more associated with the Central North (such as Manchester), whereas Northern GOAT and FACE tokens were more likely to be placed further north, particularly in the North East. This could indicate lay listeners' awareness of the spreading of the diphthongal GOAT and FACE forms throughout regions of the North (see section 2.1.3) and therefore their association of the monophthongal forms as being even further removed from the South. However, as the regional background of the listeners in this experiment is not known it is impossible to decipher how familiar they were with these accent varieties. Therefore, it cannot be concluded whether their answers were based on their actual experience with the accent varieties or on a familiar/unfamiliar distinction.

Overall, the studies summarised in this section show that adults, although varying in their ability, are able to broadly categorise speakers according to their regional accent and that this ability is affected by several extra-linguistic, as well as linguistic factors. However, how this ability develops and whether there is any evidence of this ability emerging amongst children, is a question which has been under-investigated.

### 2.3.4 Younger children's perception and categorisation of accents

Only a few studies have investigated children's awareness of regional accent between infancy and adolescence and the findings do not appear to be consistent or conclusive.

Nathan and colleagues (1998) found that younger children showed a greater phonetic sensitivity to regional accent differences compared to older children. The children were asked to repeat and define words heard in their own (London) accent and a Glaswegian accent. Whereas the 4-year-olds tended to repeat the Glaswegian words phonetically rather

than in their own accent, the 7-year-olds were more likely to repeat the words in their own accent. The 7-year-olds were also more likely to correctly define the words than the 4-year-olds, indicating a better comprehension of the Glaswegian accent. The authors explain this as due to the younger children having less phonologically abstract representations, on account of their relative lack of experience with other accents.

Floccia and colleagues' (2009) study provides further evidence for a lack of phonological awareness regarding regional accent distinctions before the age of 7. Their study investigated whether British children (from Plymouth) could categorise speakers based on phonological differences differentiating regional accents. They found that 7-year-olds but not 5-year-olds were able to group speakers according to home vs. regional accent criteria (and also that the 7-year-olds were significantly better at grouping speakers according to home vs. foreign accent criteria). Furthermore, they found that 'bidialectal' 7-year-olds (with at least one parent who spoke a different variety of British English to their home variety) performed better than monodialectals.

The authors explain that the younger children's inability to differentiate a home accent from a regional accent is due to their 'perceptual threshold' being higher; their threshold for variation is higher and therefore they don't notice systematic differences based on this variation. This explanation somewhat contradicts the findings in Nathan and colleagues' (1998) study, where it was discovered that the younger children were more likely to pick up on phonetic differences between accents. Rather than their perceptual threshold being higher, therefore, it appears to be the case that younger children (under 7) have not made a phonological connection between the different realisations of phonemes in different accents across different speakers. This could be due to the fact that older children have had more experience and exposure to different accents and that this has made them more aware of the differences which can be generalised across groups of speakers.

However, there are problems with Floccia and colleagues' (2009) study which limit the reliability of the results. Predominantly, the task design itself may have contributed to children's lack of success. The task was difficult for young children as they only heard two example sentences in each accent before they were then asked to categorise 'aliens from elsewhere' into two groups based on these examples. Therefore, the task relied upon

children's working memory as well as their understanding of what aliens are and their understanding that people from different places speak differently.

Furthermore, there is a lack of rigorous linguistic analysis of the results. The different children heard different stimuli and therefore it isn't clear which particular linguistic features the children may have been attending to when categorising the speakers. Additionally, although measurements were taken of two vowels (/æ/ and /ɪ/) to show differences between the accents, it isn't explained why these particular vowels were chosen and how discernible the pronunciation differences were. Furthermore, it isn't known how many examples of these vowels the children heard from each of the accented speakers as they each heard different sentences in a random order.

Beck (2014) deals with some of the experimental design issues raised here in her experiment with 5-7-year-olds in the U.S. She used an easier experimental design, an ABX discrimination task, to test children's regional accent awareness. In an ABX design, a listener hears two 'known' speakers (A and B) and then has to decide whether an 'unknown' speaker (X) is more similar to A or B. The children heard single words which were chosen according to the vowel quality differences between two different accents (Philadelphia vs. General Southern). They heard three speakers pronounce each word and were asked which two speakers sounded most alike. Beck found that the children in her study were able to discriminate between a familiar, local (Philadelphian) and a non-familiar, non-local (General Southern) regional accent, scoring 64% correct on average.

Beck also investigated exposure-related correlations but didn't find that the children's experience and exposure to other accents aided them in the ABX task. She did, however, find a link between children's performance in a separate 'awareness' task and their ability in the ABX task. The awareness task asked children questions such as: *Does this person sound like he lives here? Can you guess where this person is from?* Beck found that children who correctly identified a local speaker as local were more likely to perform better in the ABX task. This showed that these children seemed to have a heightened awareness of a 'local' vs. 'non-local' distinction. A complementary result was not found for the awareness question, which asked children to identify where a non-local speaker was from, however. Beck (2014) interprets this discrepancy as being due to the children having a 'familiar' or 'local' exemplar category but not necessarily an 'unfamiliar' or 'non-local' exemplar category. She

hypothesises that an established category built up from experienced exemplars is needed in order to interpret what the variation represents. This relies upon the theory that the development of exemplar categories develops from interactions with specific individuals (see section 2.4.2).

Another task carried out by the children in Beck's study supports this explanation. In a 'Similarity Judgement Task' the children listened to pairs of local/non-local speakers saying the same word. From these pairs, the children were asked to decide which speaker they thought sounded most like themselves. Beck then analysed these results in comparison with the children's answers to the awareness task questions. 'Insiders' (children with at least one parent from the local region) were more likely to identify the local accented speaker as sounding like themselves if they had performed better in the awareness questions as well. However, for the Outsiders (children with both parents from outside the local area), the opposite effect was found and those with more awareness were less likely to identify the local speakers as sounding like themselves. Beck interprets this as the Outsider children having a confused 'local' speaker category because although they and their parents are local, they don't sound the same as the other locals.

Wagner and colleagues (2014) come to a similar conclusion to Beck. They suggest that rather than creating an in/out distinction, children group speakers in terms of familiarity. Rather than a familiarity with individuals, however, they refer to a familiarity with a home dialect more generally. From the results of their study, they propose that children compare regional and second language dialects to their home dialect on a gradient scale. Their study compared 5-6-year-olds' categorisation of regional dialects to their ability to associate regional dialects with cultural items. In a similar experimental design to Floccia and colleagues (2009), they found that children could successfully categorise speakers by a home (Midland American) vs. second language dialect distinction (Indian English) but not a home vs. regional dialect (British English) or a regional dialect vs. second language dialect distinction. However, they found that when asked to group speakers with either a familiar or an unfamiliar cultural item (such as familiar vs. unfamiliar clothing), children made a distinction between regional dialect speakers and second language dialect speakers; they grouped regional dialect speakers with the familiar cultural items and second language dialect speakers with the unfamiliar cultural items. Therefore, for these children, a second

language dialect was unfamiliar enough to be associated with unfamiliar cultural items and to be distinguished from a home dialect. However, a regional dialect was familiar enough to be associated with familiar cultural items but was not familiar enough to be distinguished from a second language dialect in purely linguistic terms. The difference in children's performance in these two tasks essentially relies on the different designs of the tasks. While the authors propose that the link between dialects and unfamiliar/familiar objects forms a starting point from which categorisation of the dialects might develop, they also admit that children might achieve better results with simply an easier accent categorisation task.

## 2.4 Theoretical approaches to the acquisition and storing of social-indexical information

Sections 2.2 and 2.3 have explored how sociolinguistic variation is acquired by children. Studies described in these sections have found evidence of sociolinguistic variation in pre-school children's production and perceptual preferences. Therefore, alongside their acquisition of language, children seem to be acquiring the social indexicality of linguistic variation in their early years. In order to support these findings, it is necessary to account for how social-indexical information may be acquired alongside the linguistic in a theoretical model.

This is a question for theoretical accounts of language acquisition and phonological development but also involves speech processing, speech perception and more general cognitive processes of categorisation and conceptual development. This section will review formalist vs. functionalist models of language acquisition and whether or not they account for the learning of social-indexical information. The role of exemplar theory in usage-based accounts of language acquisition will be proposed as the most promising way to account for the storing and accessing of social-indexical information.

### 2.4.1 Formalist vs. functionalist models of language acquisition

Formalist approaches to phonological development in the process of first language acquisition focus on the abstractions that children make with regards to phonemic categories and their relationship to distinctive features, syllabic and prosodic structures (Selkirk, 1980 ) or rule-based constraints (e.g. in Optimality theory, Prince and Smolensky, 1993). In focusing on the end result – an adult's knowledge of the rules governing the use of

abstract phonological contrasts - these accounts are largely uninterested in the detail and variability inherent in language, and indifferent to the particular linguistic input that a child receives. Instead they rely on the concept of Universal Grammar (UG) (Chomsky, 1967), which is grounded in the hypothesis that all languages are governed by a set of principles which children have the innate capacity to learn. Vihman (2014) summarises the different formalist approaches and their shortcomings. She highlights the fact that while there is experimental evidence to show the effects of the prelinguistic period on children's later linguistic development (cf. Velleman and Vihman, 2003), this period is often disregarded in formal accounts. Furthermore, an underlying assumption of formalist accounts is that child and adult phonologies are essentially the same, but this does not fit with data showing that there is a high degree of variability amongst children and their developing phonologies (cf. Sosa and Stoel-Gammon, 2006; Vihman and Croft, 2007). In order to address this issue in particular, McAllister Byun and colleagues (2016) propose a more integrated account of children's phonological acquisition; 'The Articulatory – map model". The A-map model combines a formalist account of motor-control constraints based on optimality theory with a functionalist account of an exemplar-based grammar. This combination of approaches is used in order to explain how phonological development is shaped by an individual's experience and therefore why children have different phonologies from adults.

In order to understand the means by which linguistic abstractions are arrived at, functionalist or emergentist approaches to language acquisition focus on what children are able to learn from the language that they experience, aided by their developing social skills and cognitive abilities (Behrens, 2009). Generally these accounts emphasise the role of the linguistic input that children receive as well as children's speech and motor limitations/developments. Again, Vihman (2014) gives an overview of these accounts, subcategorising them into 'Self-Organising Models' and 'Usage-Based Models'. Self-organising models, such as dynamic systems theory (Thelen and Smith, 1994) and the frame/content model (MacNeilage and Davies, 1990), propose that it is the combined development of cognitive, neural and motor skills of an individual, along with changes in their environment, which together enable the individual's learning of language. Dynamic systems theory is a general theory of development which aims to describe how a system (such as a human being) develops through a combination of perception and action, reacting

to the environment and gaining knowledge from practice. This general framework has been used to explain the nonlinear, variable and individualistic aspects of child phonological development (cf. Szreder-Ptasinska, 2012). The frame/content model was specifically developed to account for children's speech development. This model proposes that the mechanical opening and closing of the jaw, used in children's first babbling vocalisations, forms the foundation for further language development in the form of syllable frames. Content is then added to these frames as new sounds are learned, according to mechanically determined combinations of vowels and consonants (Vihman, 2014:281). Like dynamic systems theory then, the frame/content model emphasises a child's individual consolidation of their phonological system.

While usage-based models (such as Pierrehumbert, 2003) similarly advocate the importance of children's individual development, more specifically they focus on the idea that a child's language develops as it is being used, in relation to their linguistic exposure. Therefore, with their emphasis on evaluating linguistic input, usage-based models seem to be the best equipped for considering children's exposure to linguistic variability and therefore the link between social-indexical properties of speech and language acquisition. This is because usage-based models are able to provide an account of how children use implicit, probabilistic methods of learning from their linguistic input, as well as more explicit methods of forming higher-level abstractions (see section 2.5.1).

Furthermore, usage-based models emphasise the domain-general cognitive skills which children can use in language acquisition. This means that the children's learning process when acquiring language is not specific to language learning but applies more generally to their cognitive development as a whole. As described by Tomasello (2003), evidence for domain-general skills such as 'intention reading' and 'pattern finding' have been found by studies in psychology and cognitive science as well as linguistics. Intention reading refers to the ability to share attention with someone else as well as the ability to direct their attention to something. This is important when learning language, in order to understand the symbolic relationship between linguistic forms and their meanings. The relationship between form and meaning needs to be mutually understood in order for communication to be successful. Pattern finding refers to the processes of categorisation and conceptual organisation, which are important in order for children to generalise across their linguistic

input and establish abstractions at different levels (e.g. grammatical, phonological). Rather than an innate set of language-specific principles aiding children to learn the complexities of language (as a formalist account would argue), usage-based accounts focus on what children can learn from their linguistic surroundings, using the cognitive skills they have and use for other purposes. Therefore, in usage-based accounts there is a focus on language as a social development and dependent on the available linguistic input. As described in the next section, many usage-based accounts of the cognitive processes involved in language acquisition now advocate exemplar theory (ExT) as the best way to explain the storing of both linguistic and social information.

### 2.4.2 Integrating exemplar theory to account for the acquiring, storing and processing of social-indexical knowledge

Originally developed in psychology, ExT is a theory of categorisation based on the storing of specific memories (exemplars). The claim is that our mental categories contain representations of specific instances we have encountered. What makes these representations specific depends on the details that we encode as part of the encounter. These details can include such information as our reactions to the encounter at the time and/or the social context in which the encounter was made. Which particular details we encode alongside each representation is mediated by multiple factors; these factors depend upon our evaluations of the encounters at the time, which in turn are affected by our past experiences, shaping the way that we perceive of the encounter. New instances we encounter are categorised according to the other exemplars that they most resemble and these categories can be formed at different linguistic and social levels. In describing how our social judgements can be accounted for in an exemplar-based model of memory, Smith and Zárate (1992) emphasise the important role of the attention we pay to the experiences we encounter and how this attention depends on different social, motivational and contextual factors. These factors are of the kind often investigated in social psychology; in particular the authors point towards Turner and colleagues' (1987) work on self-categorisation theory, which describes the links between a person's individual identity and their social identity at a higher-order group level. Our desire to maintain both an individual identity and a group-level identity influences the attention we pay to particular events in particular situations. Smith and Zárate (1992:13-16) present some examples of the potential social

motivations for paying attention to particular stimuli; they suggest influences such as the 'perceiver's self relevant attributes', 'in-group/out-group dynamics' (from Tajfel and Turner, 1985) and 'cultural "default" values'. The 'perceiver's self relevant attributes' refers to the likelihood that we will evaluate something along a dimension that we identify with ourselves (for example if we identify ourselves along a dimension of intelligence, we are more likely to remember another person along this same dimension). In terms of 'in-group/out-group dynamics', it is posited that we are more likely to pay attention to individual aspects of people who form part of our social in-group but attention to more general aspects of members of a social group to which we do not belong. Finally, we are likely to remember a stimulus by its differences from an expected cultural/social norm. These suggestions from Smith and Zárate (in tandem with social psychological discourse from those such as Turner et al., 1987) assume a link between stored exemplars and social judgments, and are therefore also relevant to a sociolinguistic account of the building up of indexical knowledge from stored exemplars. As described in section 2.3.1, children's social preferences change and develop in their early pre-school and primary years; these developments can be explained in terms of such an exemplar account.

As Foulkes (2010:6) describes, because of its broad applications, proponents of ExT come from an extensive range of linguistic research areas, including phonetics/phonology, syntax, pragmatics, sociolinguistics and speech perception. Pierrehumbert (2003) proposes an ExT account of the learning of phonology. This account explains the building of linguistic categories at the level of positional variants of phonemes (allophones) rather than phonemes themselves, which are not seen to play much of a role in adult speech perception (Vihman and Croft, 2007:710). Furthermore, Pierrehumbert's account focuses on the lexicon as the target of phonological acquisition and therefore emphasises the importance of word forms in the process of statistical learning. Pierrehumbert describes each speech stimulus as a token, (in the form of a word), which can be defined by its perceptual phonetic detail. Category labels are formed through their association with these tokens which are stored as exemplars in memory; the strength of each category depends on how many exemplars it has encoded and how recently they have been activated. Therefore the probability of the occurrence of phonetic patterns forms the basis from which categories of positional variants develop. These are language specific developments because the phonetic

patterns are language specific. Pierrehumbert supports her analysis with evidence from studies showing the level of detailed phonetic information which native speakers necessarily pick up in their own language in order to create category boundaries. For example, the phonetic distinction between whether a sound represents a consonant or a vowel can differ between languages. This is true of the interpretation of glottalisation (as measured by a reduced amplitude and F0 disturbance) in English compared to Coatzospan Mixtec, a language in which glottalisation can represent vowel contrasts; some instances of glottalisation interpreted as consonants in English would be interpreted as vowels in Mixtec (Gerfen and Baker 2005, cited in Pierrehumbert 2003:121).

Although Pierrehumbert's model includes an account of how categories of exemplars develop (at the level of positional variants of phonemes), the model does not propose how these categories might influence incoming exemplars. The focus of the model is on the influence of individual episodes and does not advocate a role for abstracted categories until after the exemplars have been stored (postlexically). Another extremely episodic-based model is proposed by Goldinger (1998) who tests the predictions of Hintzman's (1986; 1988) model, MINERVA 2, against the results from a series of speech perception/production experiments. Hintzman's model "takes episodic storage to a logical extreme, assuming that all experiences create independent memory traces" (Goldinger 1998:254). This is in opposition to the mainstream view that listeners create an abstract mental lexicon by normalising over the details of speech and forming average prototypes of word representations in memory (see section 2.5.4 for further explanation of prototypes vs. exemplars in speech perception). Goldinger found that participants in a word-shadowing task were more likely to accurately imitate the acoustic properties of words that they were asked to repeat, when the words were relatively low frequency words. Goldinger interprets this finding as an indication that indexical information is processed alongside the phonetic properties of a word, and that, over time, the indexical information stored alongside words that have been experienced many times by different speakers, is abstracted over so that the particular details of a speaker's voice may become obscured to the listener. Therefore, in the word-shadowing task, high frequency 'old' words are less likely to be repeated with acoustic similarity to the speaker being shadowed because the listener has encountered many other instances of these words over time. Therefore, the listener has many stored exemplars of

these words to influence their perception and production of these words. Others working in speech perception have disputed the extent to which episodic traces can account for how listeners process and interpret variation in the speech signal. Cutler and colleagues (2010) suggest that only a hybrid model, which includes a role for both abstractions and episodes in the processing of speech, can account for evidence showing that listeners adjust their interpretation of phonemes after limited exposure to deviant realisations. This evidence comes from studies by Norris and colleagues (2003) and McQueen and colleagues (2006) who carried out lexical decision tasks with Dutch listeners. These studies found that listeners who were exposed to an ambiguous fricative sound (between [f] and [s]) replacing [f] in f-final words were both more likely to then interpret ambiguous [f]/[s] sounds as [f] in testing (Norris et al. 2003), and were also quicker at deciding whether a stimulus was a word or a non-word when it featured [f] as opposed to [s] (McQueen et al. 2006). The results from these experiments suggest that listeners retune their phonological categories based on realisations that demonstrate lexical distinctions and that these lexically-driven changes then influence their pre-lexical processing of incoming speech. The authors argue that these results challenge extreme episodic models, such as Goldinger (1998), which do not advocate a role for abstractions in the process of interpreting speech pre-lexically, but only for abstractions to occur post-lexically(after the words have been stored as episodes). While these authors acknowledge that episodic traces in memory do play a role in speech processing, they argue that these episodes cannot be the only way in which we interpret incoming speech and therefore that a level of abstraction is also needed to explain the process as a whole.

Goldinger (2007) concedes that purely episodic-based models of memory are not entirely adequate and applies a 'complementary-systems' approach, (based on McClelland et al. 1995), to account for both abstract and episodic processes in speech perception. This approach proposes that these different processes can be explained by the evolution of different systems in the brain, which are specialised in dealing with these different tasks. As Norman and O'Reilly (2003) explain, while the hippocampal system of the brain supports the rapid learning and memorisation of new stimuli, the neocortex is a slower system, computing regularities and assigning similarities between stimuli. These systems have a reciprocal, inter-dependent relationship. Therefore, while traces stored in the hippocampus

are also influenced by input from the neocortex, the representations that develop in the neocortex are created by the accumulation of episodic traces in the hippocampus. Goldinger uses the complementary-systems approach to explain how listeners can quickly adjust to deviant pronunciations of a phoneme, such as in the experiments described above (Norris et al. 2003 and McQueen et al. 2006); the abstract information in the listener's neocortex 'corrects' the input (and therefore the listener understands the word) while the episodic traces containing speaker-specific information in the hippocampus means that the listener is able to attribute this divergent pronunciation as specific to the speaker.

According to Docherty and Foulkes (2014:46), such hybrid models, incorporating the role of episodes and abstractions are now the consensus view amongst proponents of ExT. In their focus on sociophonetic variability, Docherty and Foulkes merit an exemplar approach in being able to account for the connection between language and its social context, and, in particular, how phonetic properties can be aligned with social referents – such as particular speakers and particular accents. From this sociolinguistic perspective, ExT proposes that when exemplars are encoded in memory, social information about the speaker is simultaneously embedded with the phonetic information. Therefore, alongside linguistic generalisations such as positional variants of phonemes (as Pierrehumbert describes), social generalisations which enable us to categorise speakers into social groups can develop from the storing of these exemplars. As part of a usage-based theory, this process involves both statistical learning and pattern finding, as it is the frequency with which exemplars are encountered and the familiarisation of exemplars to other exemplars which activates their memory. Similarly to Pierrehumbert (2003), these exemplars are usually assumed to be encoded as words (see Pierrehumbert 2002 for a justification of the storing of exemplars at the word level by describing word-specific phonetic effects). Studies in sociophonetics have uncovered the correlation between the use of particular phonetic variants and social information pertaining to a speaker, such as their age, sex, social class or regional background. The use of these sociophonetic variants has been found to be acquired at a young age in production (as described in section 2.2.1) but the developing awareness of the social meaning of this variation has been less thoroughly explored (see section 2.3.4). Therefore, an account of how children's knowledge of social groups arises from their experience with variants that these social groups use is a necessary link. Exemplar theoretic

accounts such as Foulkes and Docherty (2006), Foulkes (2010) and Foulkes and Hay (2015) offer an explanation of how this social-indexical information is learned. These approaches conjecture that the initial awareness of individual talker differences may provide the basis from which knowledge about broader groups develops. As more experience with more speakers accumulates, more exemplars are stored and an increased number of abstract speaker categories emerge, based on grouping these exemplars according to their linguistic similarities. How likely a particular exemplar is to be accessed in memory is dependent not only on the frequency with which that particular exemplar has been encountered but also how recently the exemplar has been encountered(cf. Drager, 2010; 2011) and the degree of attention the exemplar was given when being encoded. As Foulkes and Hay (2015) clarify, rather than the storing of every single experience, linguistic or otherwise, exemplar theory predicts that it is only the important ones which are actually stored as episodic memories. The importance of particular exemplars can be measured by how informative they are (e.g. content words are more likely to be stored because they are more informative than function words). As Pierrehumbert (2006) describes, the importance of exemplars can also be measured by the attention they are paid at the time of encoding. Attention can be paid due to many different factors, such as the novelty of the stimulus in a particular context (e.g. hearing a speaker with an unusual voice quality), the significance of the social situation (e.g. a romantic proposal) or the personal connection a perceiver has to the context (e.g. their personal relationship with an interlocutor).

Foulkes (2010) hypothesises that the recognition of familiar individuals marks the start of the process of social category formation. Children show a preference for their mother's voice from infancy (see section 2.5.2), as exemplars from their mother's speech are likely to be the most regularly encountered and stored in memory. Over time, other individual speakers are recognised as exemplars from their speech are stored in memory as well. Categories of speaker then develop through the grouping together of exemplars which are similar in some way. Phonetically transparent categories, which are more likely to be influenced by biological factors, such as speaker sex, develop early on. These categories are more transparent because the cues that they rely on are more distinct, with little overlap. For example, there is little overlap between the fundamental frequency (f0) of male vs. female voices; Foulkes (2010:9) cites Klatt and Klatt's (1990) study, which found that the

average f0 of female voices is 1.7 times that of male voices. On the other hand, less transparent categories, which are more dependent on social distinctions such as speaker accent, are less clear-cut. As described in section 2.1.2, accents cannot be categorised in a definite way but rely on gradient distinctions to differentiate between them. These kinds of social categories, which are dependent on speaker tendencies, develop as children get older and are exposed to more social variation. Studies discussed in section 2.3.4 indicate that between the pre-school years and around the age of 7, children are developing cognitively stored speaker groups indexing social characteristics such as a speaker's accent. Children seem to be aware of these accent differences before they are able to understand the reasons for these differences.

This distinction between awareness and understanding is supported by research in children's conceptual development, which has generally found that children's judgment about category membership is more sophisticated than their explanation of it (summarised by Murphy, 2004). For example, Blewitt (1994) found that children from 2-years-old can accept objects into more than one hierarchical category; they were willing to categorise a monkey as both a monkey and an animal. Johnson and colleagues (1997) found that 3-year-olds showed an understanding of the relationship between subordinate categories that are part of the same basic category; they were able to attribute properties of one subordinate category (e.g. a whelk as a kind of seashell) to other subordinate categories (e.g. other kinds of seashell). These examples show that although pre-school children are not able to explain why these kind of categorical rules apply, they show quite a developed awareness of these rules.

Murphy concludes that children's conceptual abilities are similar to those of adults, but differ in terms of: experience with category members, domain knowledge and processing capacity/fluency (Murphy, 2004:376). Experience with category members refers to the fact the children have simply not experienced as much of the world as adults because they have not been alive as long. Therefore they have had a more limited opportunity to be able to form categories based on the things that they have experienced. Domain knowledge refers to the impact that an increased general knowledge has on our conceptual ability. Knowledge of the world that we gain through education and parental input as well as life experience helps us to interpret things in the world around us and ultimately to create categories based

on similarities between these things. Finally, processing capacity/fluency refers to the fact that adults generally have a better working memory than children and are able to remember more, at a quicker rate and for a longer period time. These three components (experience, knowledge and processing ability) are comparable to the tenets of an ExT account of the building of social categories, as described above. In explaining how exemplars are encoded, an ExT account emphasises the importance of variability and frequency in the input as well as the role of attention and the recency of exemplar activation. These factors rely on experience, knowledge and processing ability to varying extents.

This ExT approach advocates a link between the storing of individual speakers' voices and the storing of abstract speaker categories based on these individuals. This proposition relies upon understanding what makes a voice distinctive and/or recognisable in the first place. Therefore, the different but connected processes of familiar speaker recognition and unfamiliar speaker discrimination will be explored in the next section. The section then summarises studies in speech perception which link the process of speaker recognition to an exemplar theoretic account of speech processing. Finally, the section finishes with a summary of some of the ongoing issues with exemplar theory.

## 2.5 Speaker Recognition

A distinction can be made between speakers who are familiar to an individual in the long-term and those who are made familiar in the short-term. Long-term familiar speakers, such as close friends and family, are well-known to an individual and have become familiar over a long period of time. On the other hand, a previously unfamiliar speaker who has been heard recently and/or multiple times (e.g. as part of an experiment), might become familiar in the short-term; a listener remembers their voice a short time after hearing it but will be less likely to remember it after a longer period of time. This section looks at the connection between these distinctions by comparing familiar speaker recognition to unfamiliar speaker discrimination as there are different accounts of the extent to which these processes are separate or linked. One way of testing the connection between these processes is by investigating the role that a speaker's accent/dialect plays in how easily they are recognised and whether this role is affected by a listener's level of familiarity with the speaker. While the contribution of a speaker's accent/dialect in the recognition process has been studied amongst adult listeners, research investigating child listeners appears to be absent.

### 2.5.1 Speaker discrimination vs. speaker recognition

The ways in which the voices of different speakers vary and can be distinguished from one another are numerous. In principle, these variations are either due to biological factors (such as a speaker's vocal tract size) or social factors (such as a speaker's regional location). This distinction implies that there is a clear divide between variations that we are born with and can thus not easily be changed, and variations that are prone to change with our social circumstances. While in practice there is no such clear division between biological and social aspects of a speaker's voice, some features can be said to more biologically rooted and others more socially rooted. For example, acoustic cues such as a speaker's fundamental frequency can indicate biological categories such as their age or sex (Foulkes and Docherty, 2006). Also, a speaker's voice quality largely depends on their physical make up and therefore is a fairly individual aspect of their voice (Laver, 1980), although links have also been found between voice quality and social background (e.g. Esling, 1978). On the other hand, a speaker's pronunciation of segmental accent features indicates their social background to a larger extent; their pronunciations reflect the acquisition of patterns learned from their community. Overall, despite the examples given, the multitude of (un/)controllable features of a voice can be difficult to disentangle when defining what makes a speaker's voice distinguishable or recognisable. Furthermore, this leads to the question of the connection between processes of speaker discrimination and speaker recognition.

The processes of (1) discriminating between speakers and (2) recognising speakers are linked in the work of forensic speech analysts. One of their most common tasks is to compare voice samples in order to determine the likelihood that they come from the same speaker (Rose, 2002:9). Therefore their job is to both discriminate between different speaker samples and to recognise same speaker samples by studying distinguishable features of the voice (such as their voice quality, fundamental frequency and realisation of phonetic segments). The two processes of recognition and discrimination are similarly combined in the testing of lay listeners' ability to identify the voices of speakers they have heard before. This ability has been found to develop with age. For example, in Mann and colleagues' (1979) study, listeners of different ages heard a three/four word utterance spoken by a target speaker. They were then asked to identify the target speaker out of two

speakers whom they heard saying either the same or a different three/four word utterance. The study found that the ability to identify the target speaker improved throughout childhood from age 6-years-old and then reached an adult level by 14-years-old. [1]

A similar developmental effect has been found when listeners are asked to identify a speaker after a long delay. In order to investigate the reliability of earwitness testimonies from child witnesses to crime, Öhman (2013) tested the ability of children to recognise a voice after a period of two weeks. She found that although the overall recognition rate was low, 11-13-year-old children performed similarly to, or better than, adults. This was compared to 7-9year-olds, who performed worse. For this study, the listeners were asked to listen to a voice-line up in which seven different voices were heard. They were then asked to decide which (if any) of the voices was the speaker that they had heard two weeks earlier. The listeners were thus being asked both to discriminate between and recognise aspects of the speakers' voices.

The processes of discrimination and recognition are unavoidably linked when investigating the recognition of speakers in the short-term through a speaker comparison task. However, understanding how we become able to recognise familiar speakers in the long-term is a different question and has been found to rely on different neural processes. Van Lancker and Kreiman (1985; 1987) report that different regions of the brain are activated during familiar speaker recognition compared to unfamiliar speaker discrimination (see Öhman, 2013 for an overview of such studies). They found that listeners with damage to the right hemisphere of the brain performed significantly less well in a recognition task (of familiar famous speakers) than both normal listeners and those with damage to the left hemisphere. Conversely, the ability to discriminate between unfamiliar speakers was impaired in listeners with either/both damage to the right and left hemispheres of the brain. The authors explain this finding by describing the different brain mechanisms involved in the tasks. Whereas recognising a long-term familiar speaker relies heavily on overall pattern recognition (primarily using the right hemisphere of the brain), discriminating between

[1] A decline was found between age 10 and 13, however. This dip in the developmental process corresponds with a dip found in the developmental stages of the encoding of familiar faces by Carey et al. (1980), indicating a possible link between these two processes.

voices involves more detailed analysis of the features of the voice (using the left hemisphere of the brain), as well as pattern recognition.

Kreiman and Sidtis (2011) build this finding into a general framework of voice perception, which they describe as an interaction between 'featural processing' and 'Gestalt pattern recognition'. When perceiving and recognising a familiar voice, the unique pattern of features of the voice undergoes Gestalt pattern recognition; it is processed holistically. The principle of Gestalt psychology says that each time our mind creates a mental percept (impression of something), this percept is held in the mind as a whole, independent of its parts. When it comes to discriminating between unfamiliar voices, however, as well as perceiving the voice as a whole, it is advantageous to compare specific features of the voice. This is because the unique pattern of features of an unfamiliar voice is not associated in memory with a particular individual. Therefore, a set of unfamiliar voices are more homogenous to a listener because their patterns of features are not known to the listener. On the other hand, familiar voices are more heterogeneous because their patterns are familiar and this makes their voices more easily recognisable. Kreiman and Sidtis connect these parameters (homogeneity-heterogeneity and unfamiliarity-familiarity) to the preference for using pattern recognition or featural analysis in order to process a voice. The more familiar and heterogeneous a set of voices, the more likely pattern recognition will be used (using the right hemisphere of the brain) and the more unfamiliar and homogenous a set of voices, the more likely featural analysis will be used (using the left hemisphere of the brain). Kreiman and Sidtis emphasise that this is a continuum (see Figure 2.3) and therefore that the processing of a voice will not exclusively involve one side of the brain or the other, but both sides to different extents.

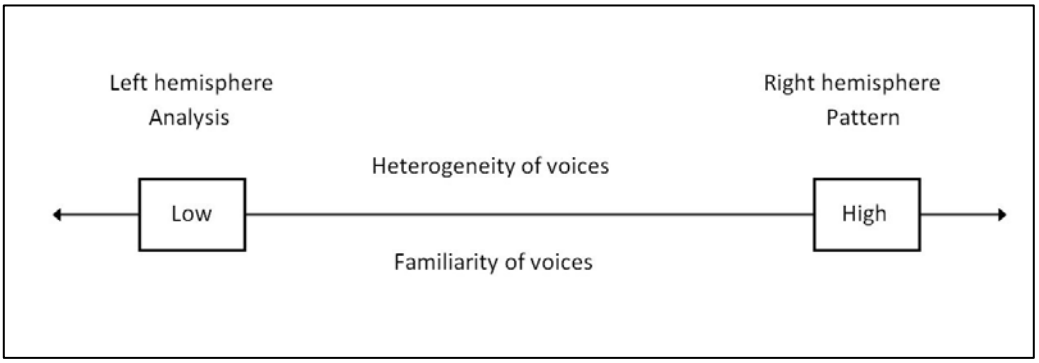



**Figure 2.3: Continuum of the parameters of homogeneity-heterogeneity and unfamiliarity-familiarity in voice perception, correlated with left-right brain hemisphere preferences (from Kreiman and Sidtis, 2011:234)**

The preference for voice recognition of familiar speakers to occur in the right hemisphere of the brain is compatible with other processes thought to occur here, such as "social context, emotional meanings, complex patterns, and personal relevance" (Kreiman and Sidtis, 2011:230). All of this information can help contribute to the listener's recognition of a speaker. Belin and colleagues (2004, 2011) expand on the claim that different parts of the brain and different neural pathways are used in speaker recognition. They present a model of speaker recognition based on a model of facial recognition (originally proposed by Bruce and Young, 1986). They differentiate between three cerebral pathways which deal with separate issues in the process of recognition: the speech information itself (see section 2.5.4 for further explanation of speech perception studies), affective aspects of the voice and the identity of the speaker. Although using different parts of the brain, these processes are posited as "likely to interact to build increasingly abstract representations" (Belin et al., 2004:131). However, the details of how these abstract representations ultimately lead to speaker discrimination/recognition are left somewhat unexplained.

An account of the abstraction process is posited by Öhman (2013), who describes a 'memory model' of voices. This model suggests that unfamiliar voices which are encountered are compared to a stored average voice 'prototype'. The evidence for this view comes from a study in which participants were asked to listen to a speaker and then try and identify that speaker from a ten-voice lineup after increasing periods of time (cf. Papcun et al., 1989; Kreiman and Papcun, 1991) . The study found that the more time had passed (they were

tested after one, two and four weeks), the more likely the listeners were to choose an 'average sounding' voice as the target voice. The measurement of an average sounding voice was calculated from the results of a previous listening experiment in which participants were asked to rate voices on a seven-point scale, ranging from 'very easy to remember' to 'very hard to remember'. The authors interpret the listeners' choice of an average sounding voice as an indication that they remembered the unfamiliar voices as deviations on an average voice prototype. Over time, these deviations were forgotten and therefore the listeners chose the most average sounding voice at testing. The authors extend the protoype account to the recognition of familiar voices, which they propose are recognised from their deviations alone. Therefore the unique pattern of features which make up a familiar voice can be seen as deviations from a general voice prototype.

The prototype approach is compatible with the finding that unfamiliar speaker discrimination and familiar speaker recognition rely to different extents on different neural processes. However, this account poses the question: where does the average voice prototype come from for each individual listener? This question does not seem to be addressed by studies in voice perception. Presumably, a voice prototype must be built up from our experience of speakers' voices. Therefore, in the first instance, this protoype is based upon the voices of speakers familiar to us. This creates a somewhat circular process; our storing of familiar speakers' voices is based on how they deviate from a prototype, which is itself based on familiar speakers' voices. This process therefore also relies upon a constant updating of the protoype as we encounter more speakers. An explanation of the updating process is best explained by returning to an exemplar theory account, which can explain speaker discrimination and recognition processes in combination with the findings from voice perception studies.

As described in section 2.4.2, an exemplar theoretic account of the building of cognitive speaker categories suggests that these cateogories start from the storing of individual familiar speakers' exemplars. These exemplars come from linguistic encounters, primarily in the form of words, which are stored in memory alongside other non-linguistic information about the encounters. This may include social information about the speaker, for example, their voice quality or the particular segmental features used (Foulkes and Docherty, 2006). These exemplar stores build up to form categories pertaining to individual

speakers and it is a familiar speaker's category of exemplars which is then accessed in speaker recognition. The process of accessing a speaker's category of exemplars therefore corresponds to the 'pattern recognition' process described in voice perception studies.

The particular speaker categories that are created and the particular exemplars that are stored are based on the importance that a listener attaches to them. This importance can be defined in terms of 'attention' and 'salience' (Pierrehumbert, 2006). The more attention that a listener is paying at the time, the more likely an exemplar is to be encoded and the more novel (salient) an exemplar (without being so novel that it cannot be interpreted), the more likely it is to be encoded. While these contributory factors are difficult to measure and define (Foulkes and Hay, 2015:307), they help to explain the intrinsic link between the social information and the linguistic information; social information (i.e. *why* was attention being paid and *what* was salient about the exemplar/the encounter?) available at the time of the linguistic encounter is built in to the stored exemplar, and therefore the social and the linguistic information are accessed together.

An ExT account proposes that groups of exemplars can be abstracted at further levels to create categories of speakers. Both the groups of exemplars corresponding to an individual speaker and the categories of speaker created by abstracting across these groups of exemplarscorrespond to the 'prototypes' posited by the memory model of voices described above. Therefore, when an unfamiliar speaker is encountered, their exemplars are compared to the exemplars of an individal speaker and/or a group of speakers that a listener has stored in memory (this corresponds to the 'featural analysis' as described by the voice percption studies above). Depending on the task at hand, this comparison can be with regards to what is different to the stored exemplars or what is similar. For example, when discriminating between unfamiliar speakers, a listener may be more likely to attend to features of the unfamiliar speaker's voice that stand out (in comparison to their stored speaker exemplars). On the other hand, in a categorisation task, they are more likely to focus on the similarities of the speakers' voices (also in comparison to stored speakers exemplars). In both cases, the larger the store of exemplars, the easier the task, as the listener has more points of comparison to make.

Whether the memory retrieval process used in the recognition of voices is based on accessing particular exemplars stored in memory or something more abstract, such as an

abstracted group of exemplars or an average voice prototype, is very hard to test empirically. It is likely that both these processes are used, suggesting that recognition can involve remembering both something specific and something more abstract. This is advocated by 'Dual process' theories of recognition, which investigate recognition more generally (not just the recognition of voices). As Diana and colleagues (2006) explain, dual process theories claim that two types of memory process are involved in recognition: 'recollection' and 'familiarity'. The familiarity process is primarily based on the activation of conceptual information which has been stored alongside previous encounters with the familiar item. The recollection process goes further and accesses stored episodic information as well as conceptual information. This results in the retrieval of specific instances of the remembered item.

As described by Jacoby (1991), these two different processes of recognition are usually evidenced by direct vs. indirect tests of memory. Recollection is tested through direct methods which rely upon attention-based, intentional retrieval of remembered items, for example asking a listener to recollect a detail from a specific linguistic encounter. On the other hand, familiarity is tested through indirect methods which aim to capture non-attention based, automatic processes of recognition. This is harder to evidence and relies upon methods such as asking subjects to perform an attention-based task while also having their memory tested in a separate task, to try and ensure that their responses to the memory test are as automatic and unconscious as possible. In his review of these indirect tests of memory, Jacoby highlights the problem of assuming that at an unconscious process is fully unconscious. From the results of his own series of three experiments, Jacoby (1991) advocates a 'process dissociation framework', which aims to demonstrate the separate processes involved in recollection and familiarity. His experiments exposed participants to a list of words they had to read, some of which were normal words (referred to as 'read wrds') and some of which were anagrams they had to solve before reading the word. Participants were then tested by looking at a new list of words and deciding whether they had seen them before in the previous list. The participants were divided into two conditions, those who could pay full attention to the task, and those who had their attention divided by participating in a distractor task at the same time. Jacoby found that the recognition of the anagrams, but not the read words, was affected by these different test conditions. Whereas

the read words were recognised with roughly the same amount of accuracy in both conditions, the anagrams were less likely to be recognised in the divided attention condition. Jacoby explains this result by differentiating recollection from familiarity-based judgments. Whereas recollection, the type of memory involved in remembering the anagrams, was impeded by dividing the attention at test, familiarity, the more automatic type of memory, was not impeded and therefore demonstrates its unconscious nature. Therefore, Jacboby concludes that "Recollection, as a basis for recognition memory judgments, requires prior processing of a sort that will support later recollection", whereas "Familiarity-based judgments require relatively little processing capacity" (Jacoby 1991:530). These two distinct memory processes support the two different types of learning often associated with language acquisition. As Vihman (2014:44) describes, one learning mechanism, employed by both children and adults, involves probabilistic, statistical processing of the input which appears to be implicit and non-attention based. This learning mechanism relies more upon familiarity-based memory processes. The other learning mechanism is explicit and involves attending to the input in order to abstract regularities and, ultimately, to create categories from the input (e.g. in infant word learning). This learning mechanism relies more on recollection-based memory processes.

These two different memory processes, linked to two different learning mechanisms, can be used to describe how familiar speaker recognition and unfamiliar speaker discriminationrely on pattern recognition and featural analysis, and therefore both more abstract and more detailed analysis, to different extents. Despite familiar speaker recognition relying on a more holistic approach, some details about the speaker's voice will be more or less salient in the recognition process. The particular features that make a speaker's voice idiosyncratic will depend both on the individual listener and the individual speaker. Therefore the next section looks at listeners' variability in recognising familiar speakers and more specifically the role of accent/dialect in speaker recognition.

### 2.5.2 Variability in familiar speaker recognition

Adults' ability to recognise familiar speakers is very variable and depends considerably on both the individual listener and the individual speaker. Overall success rates as low as 60-70% have been found for familiar speaker identification (Ladefoged and Ladefoged, 1980; Foulkes and Barron, 2000). In these studies, the speakers' voice quality, pitch and accent

were found to contribute to how easily they were identified (section 2.5.3 further explores the role of accent/dialect in speaker recognition). The length of the stimulus can also have an effect on a listener's ability to identify a speaker. Schweinberger and colleagues (1997) found that the recognition of famous speakers increased when the stimulus length was extended by small increments in the first second of speech. Ladefoged and Ladefoged found an improvement over longer speech samples, with better familiar speaker identification from a 30 second stretch of speech compared to one sentence or one word.

Young infants have been to show a preference for their mother's voice from as early as 2 days old (DeCasper and Fifer, 1980), suggesting the start of a process of long-term familiarity. By measuring infants' sucking activity on a non-nutritive nipple, DeCasper and Fifer found that the infants would suck more in order to hear their mother's voice rather than a stranger's voice. However, beyond displaying a preference, only a few studies have looked at young children's abilities to identify familiar speakers. Bartholomeus (1973) tested 4-year-old nursery children in a voice-face matching task, involving other children and staff at the nursery. During the testing procedure, the children heard a speech sample and were asked to choose the corresponding picture of the speaker from a choice of 21. Each child took part in 17-19 trials altogether and overall the children scored just under 60% correct. Spence and colleagues (2002) found that 3-5-year-old children scored above chance in matching a cartoon voice with its picture. The children listened to 20 voice samples and chose from a closed set of 6 pictures. They found an improvement in this ability between 3-year-olds (61% correct overall) and 4-year-olds (81% correct overall). Also, they found that there was a significant improvement when the characters were more familiar to the children.

While these studies show a certain level of familiar speaker identification by young children, as we will see in the next section, they do not further investigate which aspects of a speaker's voice the children are using in the identification/recognition process (cf. Creel and Jimenez, 2012).

### 2.5.3 The role of accent/dialect in speaker recognition

Voice recognition of unfamiliar speakers has been found to be impaired when the speaker has an unfamiliar accent/dialect. Stevenage and colleagues (2012) label this the 'other-

accent' effect; listeners are better at identifying voices with their own accent over voices with a different ('other') accent. They tested listeners from Southampton and Glasgow who were exposed to a target voice for 30 seconds before hearing a six-person line-up of 8 second speech samples from which they had to identify the target speaker. A significant interaction between the accent of the target speaker and the listener was found, with much higher correct identification rates when the listener and the speaker had the same accent. This suggests that listeners pay attention to a familiar aspect of an unknown voice when storing it and later retrieving it from memory.

Sjöström and colleagues (2009) present further evidence for the important role of accent/dialect in speaker recognition. In their study, a bidialectal speaker was not recognised across his two Swedish dialects. Participants were familiarised with the speaker in one of his two dialects by listening to a reading passage. They were then asked to identify the speaker from a voice line-up, with four foil voices alongside the target. All speakers were heard saying phrase-length utterances. When the target speaker shifted dialects, listeners found it much more difficult to identify him than when he was heard speaking in the same dialect as the reading passage. While accent can be enough to disguise a speaker, it can also be enough to recognise a speaker. Foulkes and Barron (2000) found that within a group of 10 young male friends of diverse backgrounds, the two most consistently identified were those with the 'strongest' accents, i.e. the accents least typical of the group as a whole (Tyneside and London). A similar finding was made in a study by Ladefoged and Ladefoged (1980) which tested one of the author's ability to identify familiar speakers from a sample of 29 friends, family and acquaintances. Ladefoged specifically refers to hearing 'North Country vowels' in one sample and then being able to identify the speaker because he knew that they came from the North. These findings suggest that a speaker's accent can play a dominant role in the recognition process. However, the questions of which specific features and to what extent they contribute to identification have yet to be fully addressed.

Nolan (1983) suggests that an accent cannot be properly defined outside of the speaker who uses it. He proposes that 'personal' and 'accentual' information in speech are intertwined with one another. Therefore an individual's use of accent features depends on the style of speech they are employing and the more familiar we are with a speaker, the more likely we are to know their stylistic range. Another finding from the Ladefoged and Ladefoged study

was that Ladefoged failed to correctly identify his own mother from both one-word and one-sentence samples. On reflection, Ladefoged highlights the role of accent here, as he struggled to differentiate his mother's voice from those of the other similarly-accented RP speaking women in the sample. In this case, then, accent had a negative impact on his ability to identify a speaker relative to the other possible candidates in the task. This suggests that although we use features of a speaker's accent in order to recognise them, the more familiar we are with someone, the less we tend to notice their accent.

Esling (1998) describes this in terms of using 'us' as a reference point from which we then judge other speakers. Therefore if we know someone well, they do not differ as much from 'us' and we don't notice the differences between 'us' and 'them' as saliently. What we might notice about a speaker's accent is relative to the experience we have, both with them and their accent, as well as other accents more generally. If, based on an ExT account, we categorise accents based on ones we have heard before (see section 2.4.2), to some extent we must be taking account of the accent features of familiar speakers, even if this is fairly unconscious. Although there are many different identifying features of a speaker's voice, based on both biological and social factors (as described in section 2.5.1), "if they didn't have a distinguishable ensemble of accent features, we couldn't tell their voice apart from other people's" (Esling, 1998:173). This supports Nolan's description of the 'personal' and 'accentual information in a speaker's voice being intertwined: the more familiar a speaker is to us, the more difficult it is to tell what it is about their voice that we are distinguishing from others and recognising as theirs. This view supports the findings in the differences between speaker discrimination and speaker recognition from the voice perception literature described above. While an overall pattern recognition is more likely to be employed to recognise a familiar speaker, more analytical featural analysis takes place for the voices of more unfamiliar speakers.

Whether a speaker's accent is an important identifying criterion for children is still a question for further investigation. From an early age, children can identify an individual speaker through highly conspicuous aspects of their voice, pertaining to their age or sex. For example, Creel and Jimenez (2012) found that after a short familiarisation phase, 3-6-year-olds were significantly better at identifying two cartoon characters' voices when the characters were of different sexes or ages. In an ExT approach, these results suggest that

children of this age have established cognitive speaker categories based on age and sex. They are therefore able to use this categorical distinction to differentiate and identify speakers. At what point categories based upon accent are established and how this may play a role in children's identification of speakers is a question that the current research aims to address.

Exemplar theory can also be used to connect studies in speech perception with studies of voice/speaker recognition. As mentioned above, Belin and colleagues (2004; 2011) propose that there is an interaction between speech processing and the processing of details about a speaker's voice at an abstract level. Studies in speech perception have gone on to find an even more integral link between these processes, focusing on how the speech signal indexes talker-specific information and how this can affect our processing of spoken language

### 2.5.4 Speech perception studies

Nygaard and Pisoni (1998) summarise the previous division between studies in speech perception and studies in speaker identification. Studies in speech perception have tended to take an abstractionist view. This view advocates processes of normalisation in which abstractions are made over the variability in speakers' voices in order to form linguistic representations (e.g. phoneme categories). Studies in speaker identification, on the other hand (such as those discussed in section 2.5.1), treat the speech signal as talker-specific but do not integrate this fact into a model of how speech is interpreted linguistically. Instead these studies endorse the notion that talker information is held in a different part of the speech signal to the linguistic information. Despite the division between these two approaches, there are now studies bridging this gap and integrating the ideas of talker-specific learning in a model of speech perception, linking the personal properties of speech to the linguistic.

Nygaard and Pisoni's (1998) experiments found that the learning of particular voices helped listeners to recognise words. In a set of three separate experiments, they trained listeners by familiarising them with 10 individual voices over nine days of exposure in one hour training sessions. They then tested the listeners on their ability to name the speakers they heard and also their ability to identify words/sentences heard mixed in continuous white noise. They found that when listeners were trained to recognise the speakers through isolated words,

they were able to recognise novel words heard in noise, whereas when they were trained to recognise speakers through sentences, they could only recognise novel sentences and not novel words. From these results, the authors conclude that the perception of personal properties in the speech signal is intrinsically linked to the linguistic, but that these properties are remembered at the level which will be of most relevance. So in the case of word-learning, the listeners remembered word-specific speaker attributes but in the case of sentence-learning, the distinctions were at a sentence level. These results can be interpreted through an exemplar theory account because the storing of speaker-specific information encoded in memory appears to be helping in the processing of new exemplars. The role of attention is highlighted here, as through the familiarisation training, the listeners were being asked explicitly to pay attention to the speech, whether at the word or sentence level.

Although these results show convincing evidence that linguistic exemplars are encoded with speaker-specific information, from a sociolinguistic point of view this does not tell the full story. As the authors acknowledge, the speakers selected to create the stimuli were not controlled for age or accent and there is no account of how these factors could have had an effect. As explained in the previous section, ExT emphasises the importance of an individual's developing experience with different speakers. Therefore, it is likely that the listeners in Nygaard and Pisoni's experiments had different amounts of exposure to speakers of different ages and speakers with different accents and that this may have affected how easily particular exemplars were activated and accessed by the individual listeners. Furthermore, as Foulkes (2010:31-32) states, such laboratory experiments, involving unnatural exposure to voices heard in artificial contexts (in white noise in this case), can only give us a limited understanding of how speaker recognition works in real life and we should be cautious not to overstate our interpretations of the results from these experiments.

Nevertheless, it is the step away from abstraction and the consideration of variation in the input which links these newer approaches in speech perception to the exemplar theoretic accounts of language acquisition. Pisoni (1997) and Johnson (1997) specifically reject the normalisation process typically presumed to occur in speech perception, in favour of an exemplar-based model. Normalisation refers to a generalisation process which is thought to occur in order for listeners to be able to interpret between-speaker variation; listeners are

able to abstract across different pronunciations of the same word or the same phoneme in order to understand that the same word or phoneme is being referred to. A consequence of normalisation is that specific detail about individual instances (such as speaker-specific information) is lost through this abstraction process. In an exemplar model, however, details about the individual instances are stored in memory as part of the exemplars representing these instances.

Johnson (1997) details how an exemplar model of speech perception can be executed through a set of formulae. These formulae calculate the auditory similarity between exemplars; the degree to which a particular exemplar triggers the memory of another exemplar; and finally the likelihood of an exemplar being associated with a particular category, depending on its activation of similar exemplars in the same category.

Based on these formulae, Johnson ran vowel identification simulations and created a model that was able to accurately identify vowels (from 39 different speakers) 80% of the time. He also found that a model based on these formulae could interpret variation in the frequency of the input. He therefore hypothesises that other sources of variability, such as accent/dialect, could be handled in the same way, meaning that accent/dialect variability would be stored in remembered exemplars.

Having used exemplar theory to explore vowel categories, in a later study Johnson (2006) explored categorisation of speakers in an exemplar model. This study looked at the categorisation of speakers as male or female, proposing that these categories are generated based on individual voices rather than a normalized vocal tract measurement. Johnson uses an 'exemplar resonance' model to explain how input speech activates stored exemplars with an auditory similarity. In this model (see Figure 2.4), incoming exemplars (at the bottom of the diagram) are encoded in memory and activate similar stored exemplars which in turn activate categories built on the basis of these exemplars. The 'resonance' part of the model indicates that activation spreads the other way too; from the category to other exemplars within the category. Therefore social categories such as gender which are devised from groups of exemplars storing this social information, can feed back and affect the categorisation of incoming exemplars. This process creates more categorical responses overall and can explain why listeners respond more quickly when asked to repeat words heard by stereotypical, rather than non-stereotypical sounding males and females (as found

by Strand, 2000). This finding suggests that listeners have a pre-conceived idea of what speakers of different genders sound like and that reinforcement of these gender norms results in a quicker processing speed.

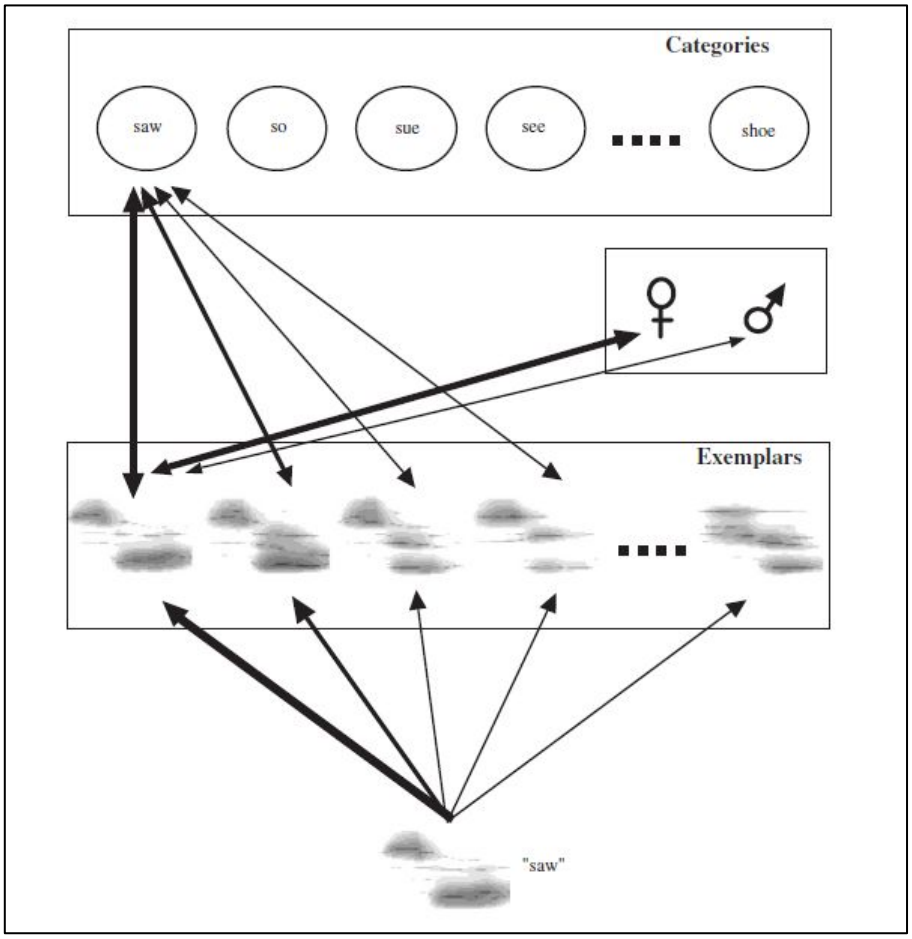



**Figure 2.4: Schematic illustration of an exemplar-resonance model of speech perception (from Johnson 2006:493)**

As with other linguistic proponents of exemplar theory, Johnson's account defines exemplars as being stored at the word level. Therefore, it is posited that incoming exemplars at the level of the word activate similar exemplars, which in turn then activate the social information stored as part of these exemplars. . However, others working in speech perception take a different, 'dual route' approach. Sumner and colleagues (2014) link the acoustic properties of words in incoming speech directly to social information as well as to other linguistic information. Therefore they advocate a direct link between word forms and 'social salience', suggesting that strongly 'idealised forms' are held more substantially in

long-term memory because they are more heavily 'socially weighted'. These terms are not defined by the authors and therefore their meanings are hard to discern; they can only really be interpreted from the examples given. 'Social salience' and 'social weighting' are used interchangeably to mean something akin to the amount of attention that a particular pronunciation of a word is given due to it being spoken in a standard accent and/or in careful speech. The 'idealised forms' then refer to these particular pronunciations, which are held with high regard because they are pronounced in a careful manner/by someone with a standard accent.

To support this analysis, the authors use evidence from studies such as Sumner and Samuel (2009), which found that non-rhotic New York listeners performed similarly to rhotic New York listeners (and other rhotic Americans) in recognising rhotic variants more quickly and accurately than non-rhotic variants. Sumner and colleagues explain this by arguing that although the rhotic tokens may be atypical to the non-rhotic speakers, they are idealized as the standard and therefore hold a special place in memory because they are paid more attention during encoding. Consequently, it is predicted that prestigious accents are more strongly encoded and therefore speakers with a prestigious accent are more easily accessed in memory than non-standard speakers. The authors also use a similar explanation for why native English listeners rate non-native English speakers lower than native English speakers on scales of comprehension and reliability (as found in Lev-Ari and Keysar, 2010). They suggest that linguistic comprehension is intrinsically tied up with both phonetic and social information, so that when social information is triggered (i.e. that this is a non-native speaker), phonetic information is accessed in a particular way. A non-native accent is not prestigious enough to hold a strong social weighting and a special place in memory and comprehension. Therefore, listeners rate a non-native accent low for comprehension and reliability due to its lack of social weighting rather than their lack of experience with the accent per se. Again, there is a lack of explanation for the terms used by the authors here. How the prestigiousness of a speaker and an accent can be measured and how it may differ from person to person is not considered. Instead, a few examples presumed to apply across the board are given (e.g. a non-rhotic variant produced by a British English speaker being more prestigious than a rhotic variant).

One of the main purposes of Sumner and colleagues' approach is to dispute raw frequency-driven approaches to word recognition and speech processing. They propose that idealised forms are more likely to be remembered even if they are not used very often. However, in an ExT account (as described in section 2.4.2), the encoding of exemplars is not purely frequency based; it depends on the attention paid at the time of the encounter and the importance of the encounter as well as how frequently and recently it has been accessed. Therefore, an ExT approach can account for the results of experiments such as Sumner and Samuel (2009) on an individual level; if an individual regards a particular exemplar as important, they will therefore pay more attention to it and be more likely to encode it.

This still leaves the questions of when and how certain exemplars take on a higher level of importance and what this importance relies upon. One instantiation of the importance associated with certain forms is the distinction between standard and non-standard forms. As mentioned in section 2.3.1, some children from the age of 3 have started to show perceptual awareness of standard forms and are more likely to positively evaluate these forms. However, the importance attached to standard forms varies cross-linguistically and cross-culturally and very much depends on an individual's experience and exposure, both throughout childhood and adulthood.

The influence of variation in an individual's input on their speech perception is evidenced by studies in second language (L2) acquisition. These studies have found that experience with different speakers makes a listener more flexible in processing speech. For example, Logan and colleagues (1991) found that Japanese L2 learners of English were able to create much more robust /r/ vs. /l/ categories after a training phase in which they were trained to identify minimal pairs contrasting the /r/ and /l/ sounds by listening to five different speakers produce these minimal pairs. The study also found that listeners' performances were dependent on the talker heard in training and at test; some speakers' words were significantly better at being identified than others. From these results the authors conclude that talker-specific information is stored in long-term memory alongside the speech information. Therefore, in order to best represent the listeners' real experience with a variety of speakers in real life, variability in their input through linguistic training will help L2 learners to better acquire this distinction. Further evidence for talker-specific details being encoded along with words perceived in speech comes from Goldinger's (1996; 1998)

speech perception experiments. Goldinger (1996) found that listeners were better at remembering words heard previously when they heard them spoken again by the same speaker, rather than a new speaker, and Goldinger (1998) found that listeners echoed talk-specific acoustic details of words they were asked to listen to and repeat in a word-shadowing task (see section 2.4.2). Both of these studies are interpreted as evidencing episodic memory traces which encode speaker details in their representations and which are accessed in speech perception.2.5.5 Ongoing issues with an exemplar theory account

Although we have seen support for an exemplar-based account of cognitive processing in studies of speech perception, language acquisition and sociolinguistics more generally, there are still general issues with an exemplar account remaining to be answered. One major issue, as pointed out by Foulkes and Hay (2015), is that of the link between stored episodic memories and the abstractions based upon them. While most exemplar theoretic accounts now suppose the storing of abstract categories as well as phonetically detailed instances (cf. Docherty and Foulkes, 2014), it isn't entirely clear what exactly is being stored and how the exemplars are linked together. Different perspectives in sociolinguistics (see section 2.4.2) and speech perception (see section 2.5.4) point to different interpretations.

What both studies in speech perception (Sebastián-Gallés, 2005) and sociolinguistics (Foulkes 2010) have emphasised, however, is the need to look at children's development to further understand these processes. For example, Sebastián-Gallés says: "speech perception studies, from a developmental point of view, almost exclusively finish somewhere between 12 and 24 months of age…a particularly interesting field of research should be how speech perception is modified during childhood" (Sebastián-Gallés, 2005: 559).

## 2.6 Summary of the literature

This chapter has reviewed literature from a range of research areas in order to contextualise the scope of the current study. This previous research has found that important linguistic developments are occurring for children in the pre-school and early primary school years. Studies have found sociolinguistic variation in children's productions from a young age, shadowing the accent patterns of their local community.

In terms of perception, studies with infants have found that by around the age of 18-months, they are able to recognise familiar words in unfamiliar accents. However, the limited studies

investigating children's ability to categorise speakers according to their regional accents present inconsistent results regarding the abilities of children between the age of 5 and 7. Furthermore, there appear to be no categorisation tasks carried out with younger, pre-school children in order to investigate their abilities.

Accounts advocating exemplar theoretic models of the development of linguistic and social categories in the literature appear to best explain the process of children's developing accent awareness as their acquisition of social-indexical knowledge builds up from individuals to speaker groups. These accounts integrate the role of episodic memories, which contain social details of stored linguistic encounters, with the role of abstracted social categories, which develop as a result of the shared properties of these exemplars. As such, these accounts illustrate the importance of both implicit and explicit modes of learning in children's interpretation of their linguistic input. Non-attention-based, implicit, probabilistic processes result in familiarity-based judgements, while attention-based, explicit recall encourages the formation of categorical abstractions such as those relating to regional accent distinctions. ExT accounts are also supported by evidence from studies in speech/voice perception and speaker recognition which go some way to explaining the kinds of indexical information tied up in speech as well as the way in which we learn to recognise individual familiar speakers. In the current work, it is proposed that an exemplar theoretic model of speaker categorisation can be strengthened by incorporating the results of such studies.

Each of the experiments in the following chapters aims to investigate a different stage of an ExT account of the building of social-indexical knowledge based on regional accent distinctions. Therefore, the experiments aim to track the sequential progression of accent awareness. As this process is hypothesised to stem from the exemplars of familiar speakers (Foulkes 2010), chapter 3 starts by investigating children's recognition of familiar speakers and the role of accent features in this process. Chapters 4 and 5 then go on to investigate children's ability to categorise speakers according to their accent and the effect of independent factors, such as the children's exposure to linguistic variation, on this ability.

# Chapter 3: Pre-school children's identification of familiar speakers and the role of accent features

## 3.1 Introduction

The two experiments in this chapter investigate the first stage in the proposed progression of social-indexical knowledge, from familiar speaker recognition to the ability to group speakers according to regional accent differences. Thus, the extent to which children are able to recognise familiar speakers is explored in the first experiment and then the second experiment goes on to investigate the role of accent features in children's recognition of familiar speakers. This is in order to establish whether a speaker's accent forms part of the recognition process for the child listeners. In turn, the results from the experiments in this chapter will support the design of the following experiments, which lead on to investigate children's perception of regional accent features amongst unfamiliar speakers.

As described in section 2.3.2, studies have shown that 4–5 month old infants demonstrate a familiarity and preference not just for the voice of their mothers (whose speech they are able to process faster than the speech of unfamiliar speakers; cf. Purhonen et al., 2005), but also more generally for their own accent over other regional accents (Butler et al., 2011). Furthermore, when it comes to word learning, infants have been found to be better at recognising familiar words in their local accent, before being able to extrapolate this ability across other regional accents as they grow older (cf. Schmale et al., 2010). Although the design of these studies means that the reasons for their results can be disputed (see section 2.3.2), they do indicate that speakers' accents play a role in speech perception and speaker discrimination from an early stage in a child's linguistic and cognitive development.

However, beyond the age of infancy (if we take this roughly to mean post-18 months), very little has been done to investigate the further development of both speaker and accent recognition and the link between these perceptual skills in early childhood. A certain level of familiar-speaker recognition amongst pre-school children has been established (Bartholomeus, 1973; Spence et al., 2002). As described in section 2.5.1, familiar speaker recognition has been found to rely upon the recognising of a familiar pattern of idiosyncratic features of their voice. However, neither Bartholomeus or Spence and colleagues investigated which idiosyncratic features of the voice the children used to recognise and identify each familiar speaker. Both these studies use a relatively large range of speakers

(around 20 in each case) and utterances of 12 words (Bartholomeus) or 4 seconds (Spence et al.). Therefore, each child is likely to have heard a varied range of voice qualities, pitches and accent differences (amongst other speaker-specific distinctions) during these experiments and the potential effects of such differences on speaker recognition are not explored in either study. Furthermore, studies have found that adult listeners are varied in their ability to recognise familiar speakers; this ability has been found to be affected by factors such as the level of familiarity between the listener and the speaker, the speaker's accent and the length of the audio stimulus that the listener is exposed to (see section 2.5.2). Thus, it is pertinent to investigate the effect of these factors on children's recognition of familiar speakers as well and to compare children's and adults' abilities.

The two experiments in this chapter focus on analysing the perception of speaker differences. They address the question of pre-school children's ability to identify familiar speakers, and the role of regionally-based phonetic accent features in this process. Two related experiments were undertaken in order to investigate the different aspects involved in familiar speaker recognition. The first experiment was carried out in order to establish the level at which the pre-schoolers were able to identify familiar speakers. In this experiment, named the 'Identification experiment', the children were asked to identify familiar speakers in a voice-face match procedure, and the results address the following questions:

(1) Are pre-school children from the age of 2.5 years able to identify familiar nursery teachers from short audio stimuli?

(2) Does this ability improve with age throughout the pre-school years?

(3) Is this ability affected by the amount of exposure the child has had to the speakers?

(4) Does this ability improve when the children hear a longer audio stimulus?

(5) Are particular speakers, with more idiosyncratic features of voice, such as their pitch or voice quality, more recognisable than others?

While the results from the Identification experiment establish children's ability to identify familiar speakers, the second experiment goes on to investigate the role of particular accent features in speaker recognition. In this experiment, named the 'Recognition experiment', children were asked to listen to different stimuli and give yes/no answers as to whether

they recognised the speaker as a particular nursery teacher or not. In order to test the role of accent in the recognition process, the different stimuli featured the nursery teacher in two different accent guises (her normal accent and a disguised accent) as well as unfamiliar speakers with different accents. Following on from the Identification experiment, the Recognition experiment addresses the questions:

(6) Are the same children able to recognise one particular familiar nursery teacher (with a strong regional accent) from single-word stimuli?

(7) Is this ability influenced by external social factors, such as the child's age/sex or their exposure to different languages and/or accents at home?

(8) Are the children able to distinguish other, unfamiliar speakers from this familiar speaker based on the same single-word stimuli?

(9) Are some unfamiliar speakers with different regional accents more easily distinguished from the familiar speaker than others?

(10) Does a disguised accent result in a reduced ability to recognise the familiar speaker?

## 3.2 Identification experiment

Taken together, all of the experiments reported in this thesis investigate the claim that the awareness of speaker groups based on regional accent stems from the recognising of individuals and the storing of exemplars of their speech in memory (see section 2.4.2). Therefore the Identification experiment aims to establish pre-school children's ability to identify familiar speakers in order to lay the foundations for the subsequent experiments (chapters 4 and 5), which focus on the over-arching question of children's developing awareness of regional accents.

The children's performance in the Identification experiment will be further analysed by investigating independent variables which are anticipated to play a role in their successful identification of the speakers. These variables are both participant-dependent (the children's age, their amount of exposure to the speakers) and stimuli-dependent (the length of the stimuli, the idiosyncrasies of the individual speakers).

### 3.2.1 Methodology

The overall aim of the Identification and Recognition experiments is to investigate pre-school children's recognition of familiar speakers and whether regional accent cues play a role in this process. The experiments were designed as two 'listening games' for the children to play. This was in order to encourage the children to take part and keep them entertained. The games consisted of several audio stimuli for the children to listen to, featuring both familiar and unfamiliar speakers' voices. Attendees of a nursery group in York were recruited to participate in the experiments and a personal connection with a parent of a child at the nursery group helped to initiate communication with the nursery staff.

#### *3.2.1.1 Participants*

Twenty two children aged between 2 and 4 years (2.4 to 4.10) were tested. Three children were excluded from the analysis as they gave little or no response to the stimuli even after several efforts to help them understand and engage with the task. Therefore 19 children's results were analysed altogether, 11 female and 8 male. All children were born and brought up in York, apart from one child who was born in Germany but had lived in York from the age of 5 months.

#### *3.2.1.2 Background questionnaire*

The children's parents were asked to answer some background questions about their child's age, the languages and accents their child was exposed to at home, the length of time the parents had lived in York, and how many days/hours a week their child attended the nursery (see Appendix 1). This data was collected in order to be used in the analysis of the Identification and the Recognition experiments (to address research questions (2), (3) and (7).

#### *3.2.1.3 Speakers and stimuli*

Stimuli for the experiment were taken from all of the six teachers who worked at the nursery. Recording of a seventh nursery teacher was also taken but her sample was discarded after pilot testing the experiment. She left the nursery part-way through the experimental design process and some of the children in the pilot study failed to recognise her, even from her picture. The teachers all worked regularly at the nursery and so were in

contact with all of the children. They were all women, aged 21-48; five had lived in York all their lives or came from elsewhere in Yorkshire (Knaresborough, Leeds). The exception was one teacher (Jane[2]), who was born in Leeds but grew up in Nottingham and Kent before moving back to Yorkshire when she was 12.

All the teachers were recorded in a quiet room at the nursery using a Zoom H4n recorder which was set to record at a 32bit 96kHz sampling rate. All speakers were recorded producing the same stimuli, which consisted of a story passage and 15 short phrases (see Appendix 2), which they were asked to read as naturally as possible in a relaxed, informal style. The story passage consisted of 177 words from the start of the children's book *The Gruffalo* (Donaldson and Scheffler, 1999), a book regularly read to the children at the nursery and written in rhyming couplets such as (1).

(1) A mouse took a stroll through the deep dark wood.
A fox saw the mouse and the mouse looked good.

The 15 phrases that were also recorded were chosen as examples of the kind of phrases the nursery teachers use on a daily basis at the nursery. Piloting the experiment revealed that the children found it difficult to concentrate for the full length of time and therefore only seven of the nursery phrases were used in the final experiment (see Table 3.1). The phrases chosen ranged from 1 to 7 syllables in length. This was in order to be able to investigate the potential effect that the length of the stimulus might have on the children's ability to identify the speaker (research question (4)).

#### *3.2.1.4 Experimental design*

Audacity software (Audacity Team, 2012) was used to edit the sound files into the stimuli. The Gruffalo passage was divided into seven separate and continuous stimuli, with one stimulus taken from each nursery teacher (apart from Jane, who provided two stimuli). Also, from the phrases recorded, one phrase was taken from each nursery teacher (apart from Leanne who provided two), (see Table 3.1).

[2] Pseudonyms are used throughout.

| Stimulus | Phrase | Nursery teacher |
|---|---|---|
| 1 | *Gruffalo 1* | Jane |
| 2 | *Gruffalo 2* | Wynne |
| 3 | *Gruffalo 3* | Leanne |
| 4 | *Gruffalo 4* | Kristina |
| 5 | *Gruffalo 5* | Alice |
| 6 | *Gruffalo 6* | Claire |
| 7 | *Gruffalo 7* | Jane |
| 8 | *Hello, Nursery?* | Leanne |
| 9 | *Line up to wash your hands* | Alice |
| 10 | *Daddy's here to collect you* | Claire |
| 11 | *There's a good girl* | Jane |
| 12 | *Where's teddy?* | Wynne |
| 13 | *No!* | Alice |
| 14 | *Bye!* | Leanne |

**Table 3.1: Stimuli and speakers for the Identification experiment**

The experiment was constructed using the software package Psychopy (Peirce, 2007) and was designed to be run on a laptop computer. Each audio stimulus was presented separately while a picture relating to its content was displayed on screen. This was in order to keep the task entertaining for the children, therefore retaining their attention. Also, a picture of all the nursery workers was displayed at the bottom of the screen throughout the experiment (see Figure 3.1). The experiment was designed to be manually controlled (by me, the experimenter) and each new visual stimulus was displayed when a key was pressed on the keyboard , with the audio starting after one second each time. The responses (including any null responses) were logged by the pressing of a corresponding answer key. The stimuli were presented in the same sequence for each child as they followed a chronological order.

**Figure 3.1: Screen shot from the Identification experiment[3]**

### *3.2.1.5 Experimental procedure*

I ran the experiment with the children individually in a quiet corner of the nursery, using my laptop computer. The children wore headphones designed for child use(JVC HA-KD5-Y –E) for the duration of the experiment. I also wore headphones (SONY MDRXB400W) in order to monitor the experiment. Additionally, I had a microphone to talk through which was connected to the computer via an amplifier mixer. This allowed me to be heard by the child in order to give instructions and prompt them for a response if necessary. The use of this audio equipment also helped to minimize distraction from the rest of the room.

I asked each child to sit at the computer with me and I gave them some initial instructions: *'You are going to hear your teachers reading a story and I would like you to point at the picture of who you think is talking.'* As we were both wearing headphones, this allowed the child to indicate a response without having to speak. Also, this ensured that s/he was matching voices and faces, which has previously been found to be an easier task than having to name the speaker outright (Bartholomeus, 1973).

Before the experiment started, I checked that the child recognised all of the nursery teachers by pointing at their pictures and asking the child to name them. This ensured that all of the children reached a baseline level of familiarity with the speakers. The experiment was

[3] Faces pixelized at http://www.facepixelizer.com/

started and the first audio clip played. If the child did not initiate pointing at one of the nursery teachers, I prompted them: *'Who was that speaking? Can you point to their picture?'*

I logged the child's response by pressing the corresponding answer key on the keyboard. This answer was then automatically recorded onto a data sheet by the experimental software. The child was asked to do the same again with the rest of the stimuli. The Identification experiment took around five minutes for the children to complete and at the end of this task, the headphones were removed in order to give further instructions for the Recognition experiment (see section 3.3).

### 3.2.2 Identification experiment results

#### *3.2.2.1 Results overall*

Table 3.2 displays each child's overall number of correct answers for the Identification experiment. Background information pertaining to the amount of time the children have spent at the nursery is also shown in the table. Overall, the children's mean score was 8.53 out of 14 (SD = 3.64), equating to 60.9% accuracy (chance = 16.7%).

| Child | Age (years) | Correct answers (/14) | Correct answers (%) | Years at nursery | Hours a week at nursery |
|---|---|---|---|---|---|
| F9 | 3.8 | 14 | 100.0 | 3 | 45 |
| M7 | 4 | 13 | 92.9 | 2 | 12 |
| F10 | 4.9 | 12 | 85.7 | 2 | 15 |
| F12 | 3.8 | 12 | 85.7 | 3 | 36 |
| F8 | 4.8 | 12 | 85.7 | 4 | 17 |
| F11 | 2.4 | 10 | 71.4 | 1.5 | 30 |
| F13 | 2.8 | 10 | 71.4 | 2 | 16 |
| F2 | 3.8 | 10 | 71.4 | 4 | 24 |
| F6 | 4.6 | 9 | 64.3 | 2 | 10 |
| M1 | 3.6 | 9 | 64.3 | 3 | 32 |
| M3 | 3.6 | 9 | 64.3 | 2 | 20 |
| M4 | 4.3 | 9 | 64.3 | 2 | 9 |
| F3 | 3 | 8 | 57.1 | 2 | 22 |
| M5 | 4.3 | 7 | 50.0 | 3 | 15 |
| F4 | 4.5 | 6 | 42.9 | 2.5 | 5 |
| M8 | 4 | 6 | 42.9 | 2 | 24 |
| M6 | 2.7 | 3 | 21.4 | 2 | 17 |
| M2 | 2.5 | 2 | 14.3 | 1.5 | 24 |
| F7 | 3.3 | 1 | 7.1 | 1.5 | 27 |
| **Average** | 3.7 | 8.5 | 60.9 | 2.4 | 19 |

**Table 3.2: Identification experiment results (scores in rank order) and background information for each child: F=Female, M=Male**

#### *3.2.2.2 Statistical analysis of Identification experiment results*

A stepwise backward regression method was used in a binary, mixed effects logistic model, run in R using the lme4 library (R Core Team, 2013). A logistic model is used when there are two possible outcomes, in this case a correct or incorrect answer, in order to show the likelihood that the predictor variables entered into the model predict a particular outcome (Baayen, 2008). A mixed effects model includes random variables as well as fixed predictor variables in order to account for individual variation. In this model, each individual's responses are represented by their own coefficient, ensuring that the potential for one particular individual's results to warp the overall results is greatly reduced (Drager, 2011).

The children's responses, correct or incorrect, were turned into a factor so that each response from each individual was used in the model analysis. The dependent variable was therefore a binomial factor distinguishing between correct and incorrect answers, with the

default set to correct answers. This means that the model presents the log odds of a correct answer. Independent, fixed predictor variables of stimulus length, years at nursery and hours at nursery were included in the model. The children's sex was also originally included in the model, however it was not found to be a significantly predicting variable and was therefore removed from the final model. All of the predictor variables in the final model were measured and entered as continuous variables. Additionally, the variable 'child' was treated as a random factor in the model. Two-way interactions between all the fixed predictors were also included. Age was not included as a fixed variable as it was found to correlate highly with the children's years spent at nursery. Therefore the effects of these predictors cancelled each other out when entered into the model together. The children's years spent at nursery were included in the model at the expense of age. This was because their years spent at nursery were a more explanatory predictor in pre-testing as judged by both a lower Akaike information criterion (AIC) and Schwarz's Bayesian criterion (BIC). Both the AIC and BIC measure the goodness-of-fit of a model (i.e. how well the model prediction fits the actual data), taking into account the model's complexity (its number of predictions). The BIC is a more conservative measure than the AIC (Field et al. 2012:868); the number of predictions in a model affects the outcome of the BIC more severely and therefore it tends to be used for larger data sets.

Only the interaction between stimulus length and hours spent at the nursery failed to significantly affect the predictive power of the model. Therefore this was the only interaction removed from the final model (see Table 3.3).

| | **Estimate** | **Std. Error** | **z value** | **Pr(>\|z\|)** | **Sig.** |
|---|---|---|---|---|---|
| (Intercept) | 5.810 | 3.294 | 1.764 | 0.078 . | . |
| STIMULUS LENGTH | -0.494 | 0.265 | -1.865 | 0.062 . | . |
| YEARS SPENT AT NURSERY | -2.709 | 1.341 | -2.020 | 0.043 * | * |
| HOURS A WEEK AT NURSERY | -0.269 | 0.144 | -1.862 | 0.063 . | . |
| STIMULUS LENGTH*YEARS | 0.334 | 0.124 | 2.706 | 0.007 ** | ** |
| YEARS*HOURS | 0.116 | 0.057 | 2.049 | 0.041 * | * |

**Table 3.3: Mixed effects logistic regression model fit to the data for the Identification experiment (Significance level: '.' = 0.1, '*'= 0.05, '**' = 0.01)**

Table 3.3 shows the effect of the dependent variables on the log odds of the children giving a correct answer. The variables of stimulus length and years at nursery enter into a significant interaction, as do years at nursery and hours spent at nursery every week. Figure 3.2 plots the significant interaction between the length of the stimulus and the number of years that the children have been attending the nursery. As the majority of the children had attended the nursery for 2, 3 or 4 years, these are the years plotted. This shows that while both a longer stimulus and more years at nursery result in a higher probability of a correct answer from the children, these are not independent of one another. As the number of years the children have been attending the nursery increases, the length of the stimulus has a stronger effect on their predicted ability. In other words, for the lowest performing younger children who have been attending the nursery for fewer years, the length of the stimulus has some effect. This effect is strengthened for the older children, who have been attending the nursery for more years and who perform better overall.

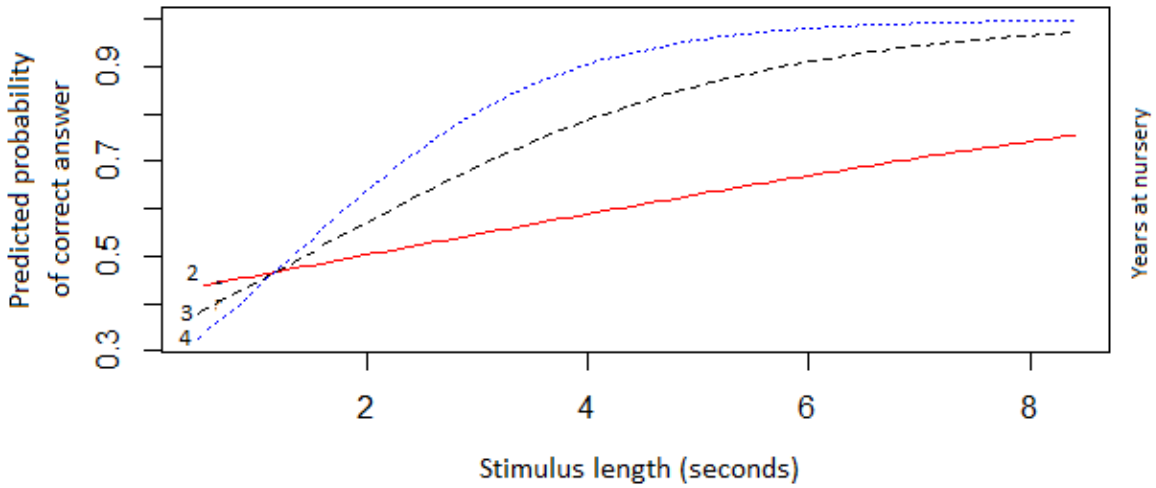



**Figure 3.2: The interaction between stimulus length and the years the child has been attending the nursery. (The line plots the predicted probability of a correct answer (i.e. 0 = an incorrect answer is predicted and 1 = a correct answer is predicted).**

Figure 3.3 plots the significant interaction between the number of years that the children have been attending the nursery and the number of hours a week that they attend the nursery. Those who have attended the nursery for longer are more likely to give a correct answer overall. Of the children who have been attending the nursery for 3 or 4 years, those who spend more hours per week at the nursery have a higher probability of achieving a correct answer in this task. The children who have been attending the nursery for 2 years, however, are not showing the same pattern of improvement, possibly due to a couple of low-scoring individuals driving this effect (cf. M2 and M6 in Table 3.2).

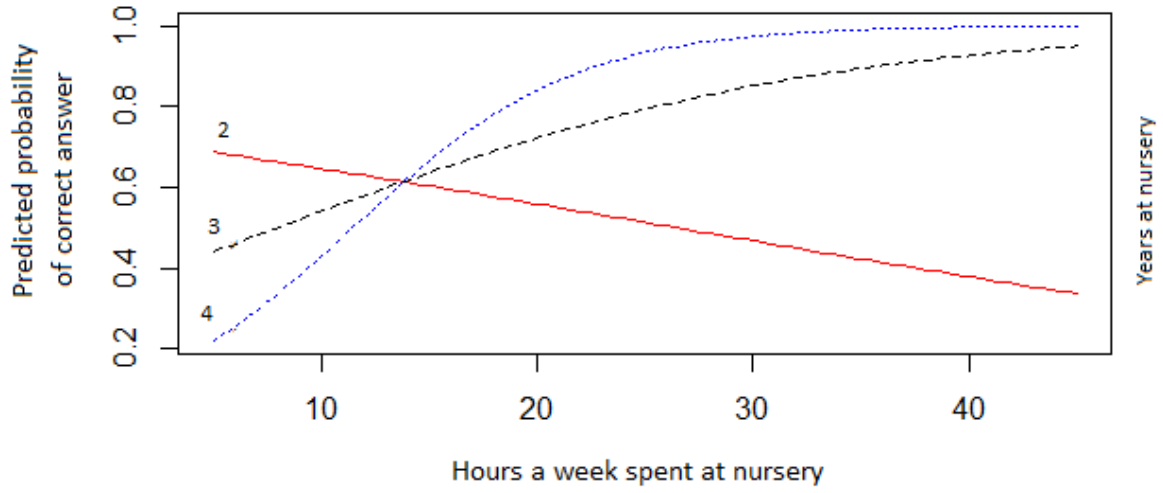



**Figure 3.3: The interaction between the number of hours the children attend the nursery a week and the number of years they have been attending the nursery altogether.**

### *3.2.2.3 Speaker effects*

The rate of identification of each individual speaker was further analysed in order to address research question (5).

Certain speakers were more readily recognised by the children than others. Table 3.4 shows the number of children who correctly identified each speaker for each of their stimuli.

| Number of children correct (/19) | Speaker | Stimulus |
|---|---|---|
| 16 | Jane | *Gruffalo 7* |
| 16 | Jane | *There's a good girl* |
| 15 | Kristina | *Gruffalo 4* |
| 15 | Alice | *Gruffalo 5* |
| 14 | Leanne | *Gruffalo 3* |
| 14 | Claire | *Gruffalo 6* |
| 13 | Claire | *Daddy's here to collect you* |
| 12 | Jane | *Gruffalo 1* |
| 12 | Leanne | *Hello, Nursery?* |
| 11 | Alice | *Line up to wash your hands* |
| 10 | Leanne | *Bye!* |
| 8 | Wynne | *Where's teddy?* |
| 5 | Wynne | *Gruffalo 2* |
| 2 | Alice | *No!* |

**Table 3.4: Number of children with correct answers for each speaker, shown in rank order**

Excluding Kristina, who only one stimulus token as opposed to two or three from all the others, Jane was the most often recognised, with an average of 14.67 correct responses. Wynne was the least well recognised, with an average of 6.5 correct responses. The very low number of correct responses for Alice's one word stimulus (*No!*) is likely to be due to the shortness of the stimulus (0.46 seconds), even compared to Leanne's one word stimulus (0.9 seconds).

As Jane's voice stood out as being much lower in pitch than the others, this was taken as the basis for further investigation. Jane was also the only nursery teacher not from Yorkshire (as mentioned in section 3.2.1.3) and therefore her accent could be anticipated as playing a role in her voice being recognised. However, as Alice was the only teacher with a broad Yorkshire accent (see section 3.3.1.1) and Jane had such a noticeably low-pitched voice, it was most likely Jane's pitch that distinguished her from the other speakers. The teachers' fundamental frequencies were measured to explore the effect of this feature in more detail across the whole group of speakers (cf. Compton, 1963; Foulkes and Barron, 2000 and for previous work showing f0 to be one of the most readily distinguishable characteristics of a voice). F0 was measured by calculating each speaker's mean fundamental frequency across all of their stimuli used in the experiment.

The box plot in Figure 3.4 shows each speaker's average f0 (in Hz). From this figure it is clear that Jane's average f0 is conspicuous. Whereas Jane's f0 falls near the typical male average of 120Hz, the other speakers are around the female average of 225Hz (cf. Fry 1979:68). Although Leanne's pitch is also noticeably different from the others, as it is above rather than below the average pitch for a woman as well as more variable in pitch range overall, it is arguably less conspicuous as her voice is obviously female-sounding.

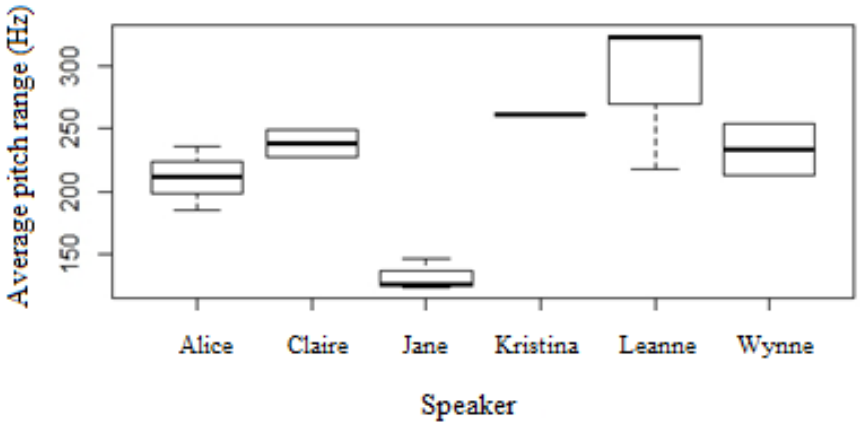



**Figure 3.4: Each nursery teacher's average f0 (pitch) range over the stimuli used in the Identification experiment**

Jane also has a very distinctive voice quality, which under Laver's VPA protocol might be classified as a deep, harsh, whispery, creaky voice (Laver, 1968:48). Therefore altogether it appears that Jane has a relatively idiosyncratic f0 and voice quality, which can both be predicted to mark her out as easy to identify relative to the other speakers. In comparison, Wynne, who was the least well recognised speaker, has an average f0 right in the middle of the average female pitch range. It is therefore probable that Wynne's pitch, coupled with inconspicuous voice quality, made her voice harder to identify.

#### *3.2.2.4 Discussion of Identification experiment results*

In addressing the research questions (1) – (5), results from the Identification experiment show that pre-school children are able to identify six individual familiar speakers with a high rate of accuracy. Overall, they performed well above chance level (with only two children below chance) and at a very similar level to the children in Spence and colleagues' (2002) study. Also, similarly to Spence and colleagues, there appears to be a developmental

effect, as the older children outperformed the younger children. However, this improvement is not purely age-related but is also affected by the amount of exposure the children have had to the speakers (measured by the amount of time that they have spent at the nursery). This exposure is both in terms of the years that they have been attending the nursery, combined with their regular attendance at the nursery on a weekly basis. Not surprisingly, both of these exposure criteria have an effect. This supports previous findings that a higher level of familiarity results in better recognition rates. Spence and colleagues showed this to be true for children in terms of how familiar they appeared to be with cartoon characters (e.g. children showed a stronger familiarity if they could name the characters without prompting). The present experiment, however, shows this a little more rigorously via a basic quantification of exposure.

The children's ability to identify the correct familiar speaker improves with a longer audio stimulus. This confirms earlier findings from adults (Ladefoged and Ladefoged, 1980; Schweinberger et al., 1997). In turn, this finding suggests that better identification relies on a number of different cues, as a longer stimulus is also likely to provide more acoustic indications pertaining to the identity of the speaker. In terms of these cues, f0 and voice quality were found to be likely to play a role in the success of the best identified teacher. Jane's overall f0 and voice quality were very distinct from the other five nursery teachers, demonstrating the importance of idiosyncrasies in a speaker's voice in the identification process.

The next question this research addresses is whether accent is also a factor in recognising a familiar speaker from shorter stimuli.

## 3.3 Recognition experiment

The Identification experiment has established that pre-school children are able to identify familiar speakers. The Recognition experiment builds on this by investigating the link between familiar speaker recognition and regional accent distinctions. This is in order to find out whether children attend to features of a familiar speaker's accent in the recognition process, something which has not been previously investigated with this age group. If it is found that children use features of a familiar speaker's accent in order to identify them, this would indicate that these features are remembered as part of stored examples of speech

from the speaker. These features then have the potential for being accessed in speech processing when new speakers are encountered and consequently used as criteria to group speakers. Therefore findings from this experiment are expected to support the claim that speaker groups based on accent may develop from the storing of speech of familiar individuals.

In further support of this claim, independent variables relating to the children's individual experience and exposure will be analysed in relation to their performance in this task. These variables include the children's age, their sex and their exposure to regional variation.

### 3.3.1 Methodology

#### *3.3.1.1 Stimuli and speakers*

Stimuli used for the experiment were taken from recordings of one of the nursery teachers (Alice) and eight other speakers, unknown to the child participants. These speakers included seven female and one male, all aged 20-29 and with different regional accents. The male speaker was included in order to test the children's understanding of the task. As described in section 2.5.3, it has previously been found that pre-school children can use a speaker's sex to categorically distinguish between speakers' voices. Therefore the children in this experiment were predicted to find it relatively easy to distinguish the male speaker's voice from Alice's, and their ability to do this will establish that they have understood the instructions given to them.

Alice was chosen as she was the teacher judged impressionistically to have the strongest Yorkshire accent. For example, she had a consistently monophthongal vowel quality in the GOAT and FACE lexical sets. Therefore, being the most distinctly 'Yorkshire', her accent provided the clearest benchmark for comparison with the other broad accents of the unfamiliar speakers. Also, she was of a comparable age (21 years) to the unfamiliar speakers recorded. The speakers were all recorded reading the same story passage and phrases as those in the Identification experiment.

#### *3.3.1.2 Experiment design*

Eight words were chosen from the recordings to use as stimuli. These words captured differences between the Yorkshire accent and the accents of the other speakers recorded.

Seven of these were based on vowel quality differences and one based on the rhotic/non-rhotic accent distinction. These words were chosen based on the main differences between the regional accents of the unfamiliar speakers compared to Alice. The decision of which words to use was also largely based on the quality and clarity of recordings, once the individual words were extracted and listened to in isolation. Using single-word stimuli helped to ensure that the listeners focused on the speakers' pronunciation of the segmental accent features rather than suprasegmental information about the speaker's voice (such as their pitch or voice quality) or the semantic content of the utterance.

For each word, three tokens were extracted as stimuli: one token from Alice and one each from two unfamiliar speakers with different accents. Six of the other speakers had distinctly different regional accents compared with Alice (two Standard Southern British English, one North American (Californian), one North East of England, one Northern Irish, one Scottish) while two of the speakers had similar accents to Alice's. These two were another Yorkshire speaker and a speaker from Lancashire, which, as another area of the Central North, shares many accent features with Yorkshire (see map in section 2.1.3). Stimuli taken from these two speakers were therefore used in order to investigate whether the similarity in their pronunciation to Alice's would cause the children to misidentify these speakers as Alice. Additionally, Alice was recorded saying two of the stimuli for a second time (*gruffalo* and *coat*). She was asked to disguise her accent by pronouncing these words with an accent from the south of England, using the diphthong [əʊ] as opposed to the monophthong [oː]. These recordings were taken in order to investigate whether Alice would be miscategorised as an unfamiliar speaker when using the phonetic realisation of a different accent from her own. Table 3.5 shows the full list of stimuli and each speaker's phonetic realisation of the vowel quality/consonant distinction under consideration. Alice's disguised tokens are indicated with *.

| Word | Associated lexical set/ feature of accent | Alice: Phonetic realisation | Unfamiliar speaker 1: Accent and phonetic realisation | Unfamiliar speaker 2: Accent and phonetic realisation |
|---|---|---|---|---|
| *gruffalo* | GOAT | [oː] | SSBE M [əʊ] | Lancashire F [oː] |
| *never mind* | Rhoticity | No /r/ | Northern Irish F /r/ realisation | American F /r/ realisation |
| *know* | GOAT | [oː] | SSBE F [əʊ] | North East F [ɔː] |
| *fox* | LOT | [ɒ ] | Scottish F [ɔː] | American F [ɑː] |
| *mouse* | MOUTH | [aʊ] | Northern Irish F [aʉ] | Scottish F [au] |
| *food* | GOOSE | [uː] | Northern Irish F [ʉ] | Scottish F [ʉ] |
| *coat* | GOAT | [oː] | SSBE F [əʊ] | Yorkshire F [oː] |
| *good* | FOOT | [ʊ] | Northern Irish F [ʉ] | Scottish F [ʉ] |

**Table 3.5: Stimuli used in the pilot Recognition experiment, along with each speaker's phonetic realisation of the vowel/consonant distinction. SSBE= Standard Southern British English, M = male, F= female**

In addition, two longer stimuli were taken from the recordings and used at the beginning of the Recognition experiment (see Table 3.6). These phrases were included at the beginning of the experiment as another means to establish children's understanding of the task - the longer stimuli were anticipated to make it easier for the children to recognise Alice.

| Phrase | Familiar speaker | Unfamiliar speaker 1 | Unfamiliar speaker 2 |
|---|---|---|---|
| *Come on, it's story time soon* | Alice | Northern Irish male | Scottish female |
| *Well done* | Alice | SSBE female | Northern Irish female |

**Table 3.6: Phrases used as stimuli at the beginning of the pilot Recognition experiment, along with the speakers' accents**

#### *3.3.1.3 Participants*

The same children from the Identification experiment were tested. The youngest two children were excluded due to not responding to this second experiment. This left 17 children, aged 2.7-4.9 years.

#### *3.3.1.4 Experiment procedure*

A pilot version of the experiment was carried out with four nursery children, aged 3-4 years. In this version of the experiment, the children were asked to listen to three 'aliens' (represented by different speakers) and decide which of the three was Alice. However, following the pilot runs of the experiment with these four children, and in response to difficulties that the children encountered, the experiment design was changed to better suit the children's capabilities.

Three of the four children show insufficient understanding of the pilot Recognition experiment task. They needed to be prompted to choose an alien each time and therefore it was apparent that their choice was not based on remembering the voices that they had just heard. The cognitive processes involved, including the abilities to hold voices in working memory and to compare the similarity between them, appeared to be too difficult for children this age. Furthermore, as suggested by one of the nursery teachers, these children might not yet be aware of the concept of an 'alien' and therefore the difficulty might also have been due to struggling to understand what the pictures represented.

Due to these issues, the main version of the experiment was run in a different way and the children who took part in the pilot did not take part in the main version of the experiment. Therefore their results are not included in the main analysis of either the Identification or the Recognition experiments.

After finishing the Identification experiment and in preparation for the Recognition experiment, I told each child *'Now you are going to hear some more people speaking. It might be Alice that you hear, or it might be someone that you don't know. Each time you hear a voice, I would like you to tell me whether you think it is Alice talking or not. If you think it is Alice, say 'Alice' but if you think it is someone else that you don't know, say 'No', okay? Let's have a go.'* The headphones were then again placed over the child's head.

A picture of Alice (with the headline 'Is it Alice?') appeared on screen to remind the child who they were listening out for. Then the screen was changed to a picture of a teddy bear and the child heard the first stimulus. After they had heard the stimulus, I asked the child *'Was that Alice speaking?'* and I logged their response. Next, a different teddy bear appeared on screen and a different speaker was heard saying the same stimulus. Again, the child was asked if this was Alice speaking. The question was intentionally worded in this way; the children were being asked to think of Alice specifically and to compare their representation of Alice held with the voice of the speaker they had just heard. This has important consequences for the results as analysed in an ExT account, advocating the storing of individual speakers' exemplars which are accessed in speech processing (see section 3.4). The experimental process described above was repeated a third time with a third teddy bear and a third speaker (see Figure 3.5). This whole process was then repeated for each set of three stimuli.

No trial runs were included in the experiment as piloting this version of the experiment found that the children were able to understand the task but struggled to concentrate if the task was any longer than the final version used for the experiment. Two versions of the experiment were created, with different orders of the speakers in each set of three; eight children took version 1 of the experiment and the other nine children took version 2. Overall, the Recognition experiment took three to four minutes for each child to play and they were rewarded with a sticker for playing the game.

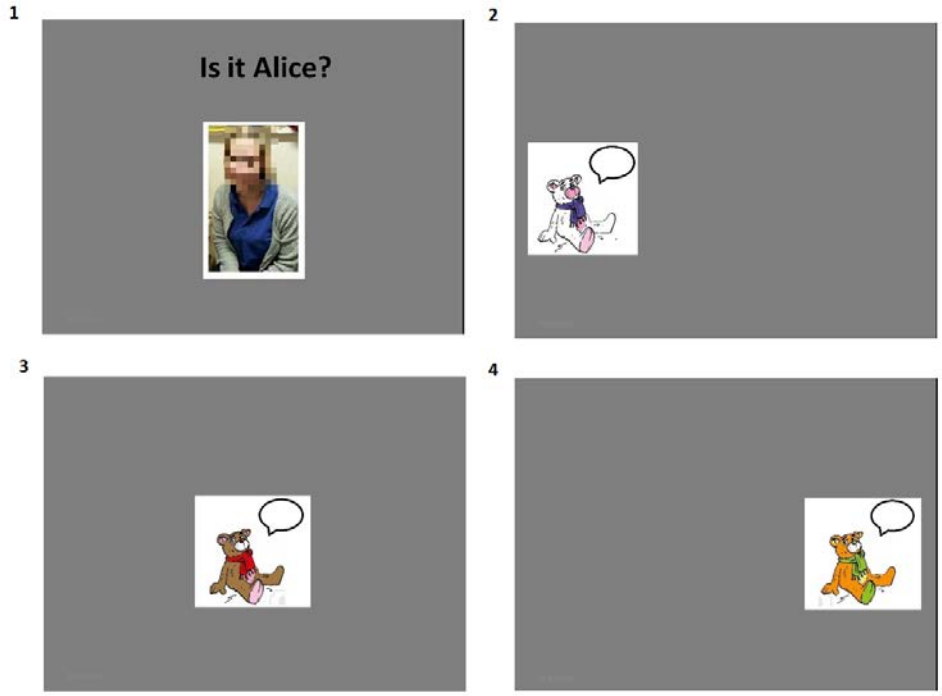



**Figure 3.5: Screens showing the procedure for the Recognition experiment**

### 3.3.2 Recognition experiment results

#### *3.3.2.1. Results overall*

Table 3.7 presents each child's results along with relevant background information. Overall, the children's mean score was 21.53 out of 33 (SD = 6.06), equating to 65.2% accuracy (chance = 50%).

| Child | Age (yrs) | Correct answers (/33) | Correct answers (%) | Langs spoken at home | Yorkshire parents | Parents from |
|---|---|---|---|---|---|---|
| F9 | 3.8 | 30 | 90.9 | 2 | 0 | Derby, Lebanon |
| F2 | 3.8 | 29 | 87.9 | 1 | 0 | Liverpool, London |
| F13 | 2.8 | 28 | 84.8 | 1 | 1 | Sheffield, Chesterfield |
| F10 | 4.9 | 27 | 81.8 | 2 | 0 | Ireland |
| F3 | 3.0 | 27 | 81.8 | 1 | 1 | Burnley, Doncaster |
| M7 | 4.0 | 27 | 81.8 | 1 | 1 | York |
| F8 | 4.8 | 26 | 78.8 | 1 | 0 | Birmingham, Durham |
| F12 | 3.8 | 24 | 72.7 | 2 | 0 | Northampton, Germany |
| F6 | 4.6 | 21 | 63.6 | 1 | 0 | York |
| M5 | 4.3 | 20 | 60.6 | 1 | 1 | Middlesbrough, Richmond |
| M1 | 3.6 | 17 | 51.5 | 2 | 1 | York, Japan |
| M8 | 4.0 | 17 | 51.5 | 1 | 0 | Liverpool, Devon |
| F4 | 4.5 | 16 | 48.5 | 1 | 2 | Doncaster, Selby |
| M3 | 3.6 | 16 | 48.5 | 1 | 1 | York, London |
| M4 | 4.3 | 16 | 48.5 | 1 | 1 | Sheffield, Mansfield |
| F7 | 3.3 | 14 | 42.4 | 2 | 1 | East Yorkshire, Surrey |
| M6 | 2.7 | 11 | 33.3 | 1 | 0 | Birmingham, Durham |
| **Av.** | 3.9 | 21.5 | 65.2 | | | |

**Table 3.7: Recognition experiment results (scores in rank order) and background information for each child: F=Female, M=Male, Av.=Average**

A post hoc comparison showed no difference between the two versions of the experiment, therefore the results from both versions were combined for the analysis. A significant positive correlation was found between the children's scores for the Identification experiment and those for the Recognition experiment ($r = 0.78$, $p<0.01$), indicating that those who performed well in the first experiment were more likely to score well in the second.

In addressing research question (6), *(Are pre-school children able to recognise one particular familiar nursery teacher (with a strong regional accent) from single-word stimuli?),* the overall results show that the children's performance was above chance and therefore to some degree they were able to recognise Alice and distinguish her from the unfamiliar speakers. Questions (8)-(10) will be further investigated in section 3.3.2.4 which will take a closer look at the cases in which Alice was missed and unfamiliar speakers were misidentified as Alice. First, question (7) (repeated below) is addressed in the next section.

> (7) Is the children's ability to recognise one familiar nursery teacher influenced by external social factors, such as their age/sex or their exposure to different languages/accents at home?

#### *3.3.2.2 Statistical analysis*

As with the Identification experiment, statistical analyses were carried out in R. A stepwise backward regression method was used in a binary, mixed effects model. Again, the dependent variable was a binomial factor distinguishing between correct and incorrect answers, with the default set to correct answers. Main effects predictors of age, sex and Yorkshire parentage were included in the model. The predictor pertaining to languages spoken at home was left out of the model as this did not show a strong or significant correlation with the children's correct answers. Additionally the children were all dominant in English and varied in terms of their exposure and fluency with the other languages.

Age was measured and entered into the model as a continuous variable, while sex and number of Yorkshire parents were included as binary independent variables. The children were split into a binary distinction according to whether they had at least one Yorkshire parent (1 or 2), or no Yorkshire parents (0). The region of Yorkshire was defined as set out in section 2.1.3. This distinction was made in line with second dialect studies which compare children who have two non-local parents to those with both parents from the local area (see section 2.2.1). Furthermore, as there was only one child with two Yorkshire parents (F4), it was deemed most appropriate to group this child together with those who also had at least one Yorkshire parent. This made for a more equal point of comparison, with eight children in the (0) Yorkshire parent group and nine children in the (1 or 2) Yorkshire parent group.

For the Yorkshire parent predictor, the default was set to no Yorkshire parents; for the sex predictor, the default was set to female. Therefore the coefficient in the 'estimate' column shows the effect of each of the default dependent variables on the log odds of the children scoring a correct answer. Two-way interactions between all the main effect predictors were also included. Only the interaction between age and Yorkshire parents was found to be non-significant. Therefore this was the only interaction removed from the final model (see Table 3.8 below). As the interactions between the predictor variables were found to be significant, the main effects are interpreted as part of these interactions rather than individually.

| | **Estimate** | **Std. Error** | **z value** | **Pr(>\|z\|)** | **Sig.** |
|---|---|---|---|---|---|
| (Intercept) | 4.41 | 1.31 | 3.37 | 0.0007 *** | *** |
| AGE(Years) | -0.70 | 0.30 | -2.37 | 0.02 * | * |
| SEX(Male) | -6.65 | 2.02 | -3.3 | 0.00097*** | *** |
| YORKSHIRE PARENT(S) (True) | -1.39 | 0.43 | -3.22 | 0.0013** | ** |
| AGE(Years):SEX(Male) | 1.27 | 0.53 | 2.4 | 0.02 * | * |
| SEX(Male):YORKSHIRE PARENT(S) (True) | 1.71 | 0.66 | 2.61 | 0.009 ** | ** |

**Table 3.8: Mixed effects logistic regression model fit to the data for the Recognition experiment (Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

Table 3.8 shows that the final model finds two significant interactions. Figure 3.6 plots the significant interaction between the children's sex and whether they have a parent from Yorkshire. The model shows that girls generally outperform boys but this is also dependent on whether they have parents from Yorkshire. The model predicts that the likelihood of getting a correct answer decreases for girls if they have at least one parent from Yorkshire. For boys the opposite is true but this is over a much smaller range and their probability of getting a correct answer stays below 0.6.

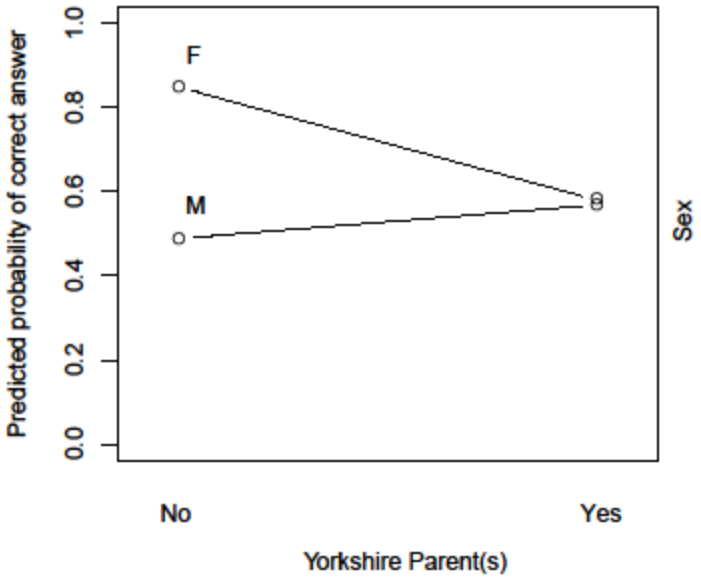



**Figure 3.6: Interaction of sex and Yorkshire parents (f=female, m=male)**

Figure 3.7 shows the interaction of age and sex. Older boys are predicted by the model to perform better than younger boys. A higher performance is maintained for the girls, although there is a slight drop through the age groups. This is probably caused by a few individuals as there is a particularly high scoring 2-year-old and 3-year-old (F13 and F3) and a low scoring 4-year-old (F4) (see Table 3.7). Interpreting Figure 3.6 and Figure 3.7 together it is clear that while sex is an important predictor it is not something that can be analysed independently of other factors. With this relatively small sample size it is impossible to be conclusive but it appears that girls are better in the recognition task generally, and while age affects the boys' results, exposure to other accents has more of an effect on the girls' results.

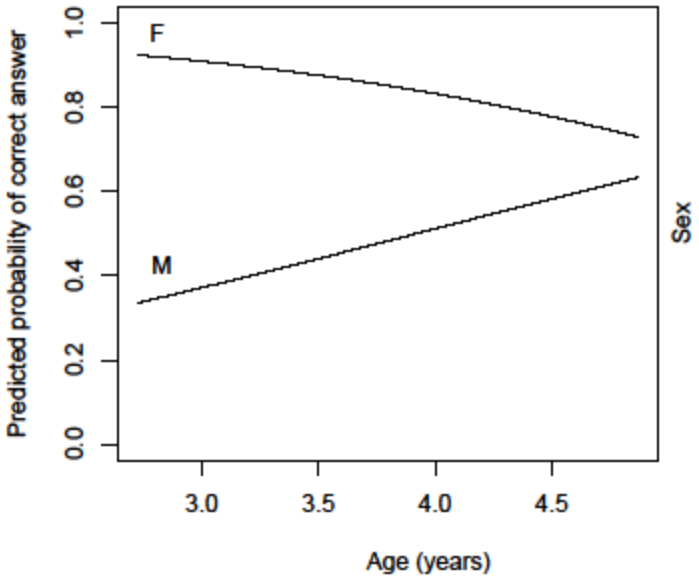



**Figure 3.7: Interaction of age and sex (f=female, m=male)**

#### *3.3.2.3 Individual performances*

The seven children who scored chance level (50%) or below in the Recognition experiment were excluded from further analysis (from M1 down to M6 in Table 3.7). As the next part of the analysis is based on individual responses and due to the relatively low number of children who took part in the Recognition experiment, this decision was made after careful consideration of each child's performance and whether their contribution could justifiably be seen as showing an understanding of the task. Understanding of the task was also determined by taking into account the children's responses to the first two, longer sets of stimuli which were included for this purpose. On average, the seven lowest scoring children achieved 52.4% correct for these stimuli compared to an average of 90% correct by the ten highest scoring children.

There are commonalities amongst the backgrounds of the seven lowest scoring children, compared to the ten highest scoring children (see background data in Table 3.7). Whereas most of the highest scoring children are girls (8/10), most of the lowest scoring children are boys (5/7). Furthermore, while most of the highest scoring children have two non-Yorkshire parents (7/10), most of the lowest scoring children have at least one Yorkshire parent (5/7).

Therefore, the majority of these individuals' performances are predicted by the statistical analysis in the previous section; their results support the interpretation that it is generally the girls and those with non-Yorkshire parents who perform better in the Recognition experiment.

The non-Yorkshire parents of the highest scoring children came from a range of different UK regions (see Table 3.7). A couple of these parents' regional varieties broadly correspond to the accents of the unfamiliar speakers (see Table 3.7) - for example Burnley/Lancashire and Middlesbrough/North East. However, as this was only true for a few individuals and there were relatively few stimuli featuring each of the unfamiliar speakers, there was not enough data to investigate the potential role of individual accents.

#### *3.3.2.4 Misses vs. false alarms*

As Table 3.9 shows, the top 10 scoring children were more likely to miss the familiar speaker (miss) than wrongly identify an unfamiliar speaker (false alarm).

| Error type | Number of tokens | Percentage of tokens |
|---|---|---|
| Miss | 44/110 | 40% |
| False alarm | 28/220 | 12.7% |

**Table 3.9: Number and percentage of error types for the Recognition experiment**

Table 3.10 shows the number of missed tokens and false alarms for the two longer sets of stimuli at the beginning of the experiment. The children's understanding of the task is indicated by the very high performance for the first phrase, for which there is only one missed token and no false alarms. There are a few more mistakes made by the children in the second set of stimuli. This is unsurprising as the two unfamiliar speakers for this set were both female, compared to the first set in which one of the unfamiliar speakers was male.

| Phrase | Number of missed tokens | Number of false alarms |
|---|---|---|
| *Come on, it's story time soon* | 1 | 0 |
| *Well done* | 4 | 1 (N Irish female) |

**Table 3.10: Missed tokens and false alarms for the first two sets of stimuli**

*False alarms*

Overall, the false alarm rate is low. In answering research question (8) this indicates that, in general, the children are able to distinguish the unfamiliar speakers from Alice. However, in addressing question (9), it is noticeable that the speakers with the highest number of false alarms are the Yorkshire and Lancashire female speakers (see Table 3.11).

| Lexical set | Word | Unfamiliar speaker 1: phonetic realisation | No. of false alarms | Unfamiliar speaker 2: phonetic realisation | No. of false alarms |
|---|---|---|---|---|---|
| GOAT | *gruffalo* | SSBE M [əʊ] | 1 | Lancs F [əʊ] | 3 |
| rhoticity | *never mind* | N Irish F [ɹ] | 2 | American F [ɹ] | 1 |
| GOAT | *know* | SSBE F [əʊ] | 1 | North East F [oː] | 0 |
| LOT | *fox* | Scottish F [ɔː] | 1 | American F [ɑː] | 1 |
| MOUTH | *mouse* | N Irish F [aʉ] | 1 | Scottish F [au] | 1 |
| GOOSE | *food* | N Irish F [ʉ] | 2 | Scottish F [ʉ] | 0 |
| GOAT | *coat* | SSBE F [əʊ] | 2 | Yorkshire F [ɵː] | 3 |
| FOOT | *good* | N Irish F [ʉ] | 2 | Scottish F [ʉ] | 1 |
| GOAT | *gruffalo* | Yorks F [ɵː] | 3 | SSBE F [əʊ] | 2 |

**Table 3.11: False alarms for each unfamiliar speaker**

In relation to question (7), this suggests that the children were more likely to misidentify an unfamiliar speaker as Alice if they heard a female speaker with a similar regional accent and vowel pronunciation. Although the Lancashire female pronounced the GOAT vowel with a diphthong rather than a monophthong, this diphthong has less formant movement in the

offglide than the diphthongs pronounced by the other unfamiliar speakers. This is evident from a comparison of the Euclidean distance between the F1 and F2 formants of all the stimuli featuring a GOAT vowel. The Euclidean distance between the nucleus value and the offglide value was calculated using the equation:

$$\sqrt{(F1nucleus - F1offglide)^2 + (F2nucleus - F2offglide)^2}$$

The vowels were hand measured at the 20% and 80% vowel duration points in Praat (see Figure 3.8). Figure 3.8 shows that the SSBE speakers had a high level of offglide movement, suggestive of highly diphthongal [əʊ] vowels. The familiar speaker had a much lower level of movement, indicating a monophthongal vowel. As well as monophthongal, this vowel is a fronted GOAT vowel, sounding more like [ɵ:] than its backed equivalent [o:] (see Watt and Tillotson, 2001 and Haddican et al., 2013 for accounts of GOAT fronting in Yorkshire English) . The unfamiliar Yorkshire speaker's GOAT vowels were similarly monophthongal and fronted, and while the Lancashire speaker's GOAT vowel showed more movement, this was still less than that produced by the SSBE speakers. The North East speaker also produced a monophthong with little movement, but this was a more back vowel: [o:] rather than the fronted variant. Therefore, it is probable that this difference meant that she was not mistaken for the familiar speaker. The high degree of diphthongal movement of the familiar speaker's disguised GOAT tokens most likely accounts for why she is not correctly identified from these tokens by most of the children, as further explored below.

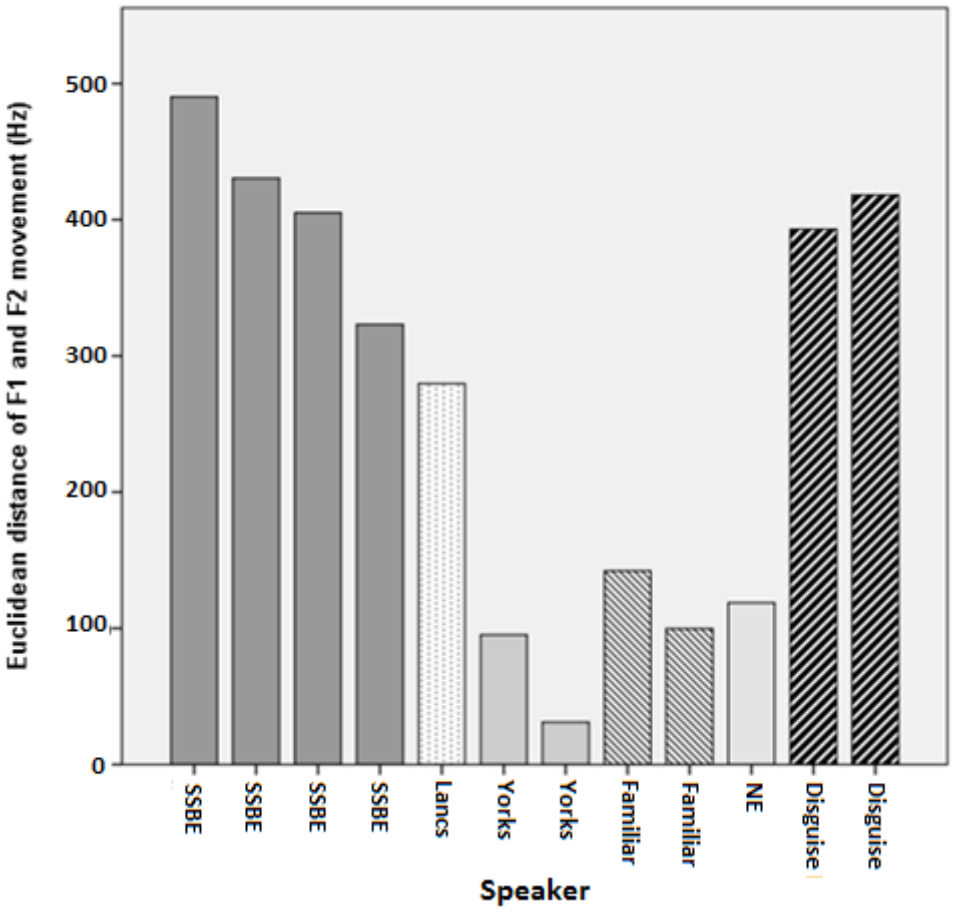



**Figure 3.8: Euclidean distance between the F1 and F2 nucleus and offglide values of GOAT vowels**

*Misses*

Table 3.12 shows the number of misses for each word spoken by the familiar speaker. The words for which a disguised accent was used are indicated with *. Apart from *good,* these words result in the most misses. Whereas none of the children fail to identify Alice when she used a monophthongal GOAT vowel (in *gruffalo*), the diphthongal quality of her GOAT vowel in both the disguised words resulted in many misses. Therefore, an accent disguise appears, as predicted, to inhibit recognition for most of these children.

| Word | Feature | Pronunciation | Misses (/10) |
|---|---|---|---|
| *good* | FOOT | [ʊ] | 9 |
| **coat* | GOAT | [əʊ] | 8 |
| **gruffalo* | GOAT | [əʊ] | 7 |
| *food* | GOOSE | [uː] | 6 |
| *know* | GOAT | [ɵː] | 4 |
| *mouse* | MOUTH | [aʊ] | 3 |
| *fox* | LOT | [ɒ ] | 1 |
| *never mind* | RHOTIC | No /r/ | 1 |
| *gruffalo* | GOAT | [oː] | 0 |

**Table 3.12: Misses for each word from the familiar speaker**

### 3.3.3 Discussion of Recognition experiment results

In addressing research questions (6)-(10), results from the Recognition experiment show that pre-school children vary considerably in their abilities in what was expected to be a difficult and unusual task for them. Some children were able to recognise the familiar speaker with a strong Yorkshire accent from single word stimuli, while others were not. This ability is somewhat affected by external social factors. The girls generally perform better than the boys but the sexes are affected differently by other external factors. The boys improve with age, indicating an important developmental effect. This development could be in terms of their improvement in understanding the task, their level of concentration and/or their amount of exposure to the familiar speaker. Girls who have neither parent from the region performed better than those with at least one Yorkshire parent. This suggests that those who have had more exposure to different accents are better able to distinguish different pronunciations. These findings, indicating a different development in abilities between the sexes, would need to be tested with a larger dataset consisting of a greater range of ages.

The children who perform above chance level overall are good at distinguishing unfamiliar speakers from the familiar speaker, as shown by the low false alarm rate. The false alarms that do occur are mainly from the Yorkshire and Lancashire speakers, who use a more monophthongal pronunciation of the GOAT vowel, similar to that of the familiar speaker. An accent disguise confuses the children to the extent that many are unable to recognise the

familiar speaker. Although this is similar to findings from Sjöström and colleagues (2009), the present study has a narrower focus - one-word accent-based phonetic realisations as opposed to whole phrases. Therefore this limits the range of features in the voice that the children could be using to recognise the speaker.

Wider implications of the findings from this experiment relate to how speakers' voices are stored cognitively and the kind of social information that may be stored alongside the linguistic information. This is discussed in the next section.

## 3.4 General discussion

This chapter has presented two experiments, investigating pre-school children's recognition of familiar speakers and the role of accent features. In the Identification experiment it was found that pre-school children are able to identify six familiar speakers and that their ability to do so improves with age and exposure to the speakers. Children's ability in this task also improved with a longer stimulus. There was variation in the level of recognition for the individual speakers and the speaker with the most distinctive pitch and voice quality was the most easily identified. The Recognition experiment asked children to recognise one particular familiar speaker from short one-word utterances. Thus this task focused on the pronunciation of the segmental features in the word rather than suprasegmental cues pertaining to the speakers' pitch or voice quality. This was a harder task, reflected in the children's much more varied results. Overall, an improvement with age was found for the boys while girls with parents from elsewhere performed better than those with both parents from Yorkshire. Accent was found to play a role in children's performance as the familiar speaker was most likely to be mistaken for a different speaker with a similar accent. Furthermore, a disguise in the familiar speaker's accent resulted in the children failing to recognise her.

Results from these two experiments show that, overall, age and exposure play a key role in children's ability to recognise and identify familiar speakers. Although this corresponds to intuitions about the nature of speaker recognition, the findings can be explained if we interpret them through cognitive models of recognition. As described in section 2.1.1, Diana and colleagues' (2006) framework suggests that the recognition process can involve both 'recollection' of stored instances and 'familiarity' of more abstract conceptual information.

This framework can be applied to the present study, to explain the processes involved in identifying a familiar speaker. In being able to name a particular speaker, the children are recalling other stored instances of the speaker from their long-term memory and therefore accessing both conceptual and episodic information. Children's conceptual abilities are qualitatively not all that different from those of adults (see section 2.4.2). Therefore improvement of conceptual ability with age is due to factors such as the general developmental advances that come with increased experience and exposure to members of categories, a better understanding and knowledge of the world, and increased processing speed that develops with maturation of the brain (Murphy, 2004). This helps to explain the results of the Identification experiment in which the older children, who have been attending the nursery for longer and have therefore had more exposure to the nursery teachers, perform better than the children who have had less exposure to the nursery teachers.

The improvement in the Identification experiment as a function of the pre-schoolers' years attending the nursery can also be explained by the dramatic change in episodic memory during this time (cf. Newcombe et al., 2000). The remembering of particular episodes uses explicit memory which, unlike implicit memory, is strongly age dependent. For example, Drummey and Newcombe (1995) found that 5-year-olds outperformed 3-year-olds in an explicit memory task (remembering having seen particular animal pictures three months previously). However they found no age-related improvement in an implicit memory task (which involved identifying animals from blurry pictures which became increasingly focused). Therefore, in the Identification experiment, the older children are more easily able to access explicitly remembered episodes from the familiar speakers, which leads to their correct identification.

In the Recognition experiment, age and exposure were found to play a different role for the girls and the boys. Overall, the girls performed better in this task than the boys, while the boys improved with age. This supports previous research which has found that pre-school girls are generally better at tasks requiring the accessing of phonological information in long-term memory and quick processing speed in perception (Halpern, 1997; Sternberg, 2004). However, more studies are needed to support this reasoning behind the sex differences found here, as other studies have failed to find significant differences in

performance between girls and boys in similar tasks (cf. Bartholomeus, 1973; Creel and Jimenez, 2012; Beck, 2014). The girls with parents from outside of Yorkshire performed better than those with at least one parent from Yorkshire. Therefore it appears that their increased exposure to accent categories outside of Yorkshire increases their conceptual ability to differentiate between individuals producing regionally-based phonetic accent features.

These two experiments also found that different features of the voice played a role in identification/recognition of the speakers. The results from the Identification experiment revealed that the speaker with the most distinctive pitch and voice quality was the most easily identified. This supports previous research, for example, from Foulkes and Barron (2000), who found that speakers with relatively high or low pitch were more easily identified within a close-knit group of ten friends. In the Recognition experiment, the speakers' accents were the focus of investigation. On the basis of short, one-word stimuli, it was found that a characterising feature of the Yorkshire accent (a monophthongal GOAT [o:] vowel) was used to identify the familiar speaker.

These results could be interpreted through a 'prototype' model of speaker recognition (from Papcun et al., 1989 - see section 2.1.1), which suggests that each voice heard is compared to an 'average' speaker that the listener has built up through their experience of different speakers. It is the idiosyncrasies of a familiar speaker's voice, when compared to this average speaker, which are then used to recognise them. Therefore, in the Identification experiment, the speaker with the most characteristic voice and the one which deviates most from the average prototype in terms of pitch and voice quality (Jane), is the most readily recognised. In the Recognition experiment, Alice's characteristic monophthongal GOAT vowel would be the means for comparison with this average prototype. Something to investigate further would be the effect of including another deep female voice in the speaker line-up for the Identification experiment. Another speaker who similiarly deviates from an average prototype would potentially be confused as Jane. Perhaps in this situation the listener would rely on other, less salient cues in order to identify the speaker as Jane over another, similar sounding speaker. Due to the forced choice design of the experiments, however, any decisions made by the listener are relative to a finite set of alternatives and therefore are not necessarily able to be generalised beyond this context.

Alternatively, the results from these two experiments can be explained through an exemplar-based account, which instead of relying upon the concept of one 'average' voice prototype, describes how a listener compares a speaker to their stored categories of exemplars from previous encounters with other speakers. Foulkes (2010) hypothesises that speaker groups based on social criteria develop initially from storing exemplars of individual familiar speakers in memory. Exemplars are stored with detail, such as their phonetic pronunciation as well as social information about the speaker. The exemplars of a familiar individual cluster together, forming a category of exemplars particular to that individual. This helps the individual to be recognised in future encounters: the incoming exemplars of their speech activate the stored exemplars with similar phonetic detail. In turn this activates the social information pertaining to the speaker's identity. This explains the results of the Identification experiment in which individual familiar speakers were recognised. This account also explains the inaccuracies of the children in the Recognition experiment. When Alice's voice is disguised, the segmental phonetic information does not closely match that of the children's stored exemplars of Alice. Therefore there is impairment in the children's ability to use the phonetic information to access the social information regarding who the speaker is. Furthermore, an unfamiliar speaker may be misidentified as Alice if phonetic information is similar enough to activate stored exemplars of Alice. This mistaking of an unfamiliar speaker for someone familiar has been found previously in studies with adults and can be explained in the same way (cf. Ladefoged and Ladefoged, 1980).

The advantage of an exemplar-based account is that it can help to explain the connection between the performances in the two experiments (as a significant correlation was found between these performances – see section 3.3.2.1). A better overall familiarity with the speakers (indicated by a high performance in the Identification experiment) translates into a better ability to use phonetic information in order to recognise a speaker. In exemplar terms, better familiarity with a speaker means that a listener has more exemplars of their speech stored in memory and therefore more instantiated information pertaining to the phonetics of their speech as well as social information about the speaker themselves. These numerous and frequently activated exemplars are then more easily activated by incoming exemplars in perception, helping the listener to decipher whether the phonetic information

in the exemplar they are processing relates to stored exemplars of the potential familiar speaker who they might be listening to.

An exemplar theoretic account suggests that both phonetic and social information pertaining to these individuals can then be abstracted to form speaker social groups, such as ones based on regional accent. Therefore the next step for the current work is to investigate (as suggested by Foulkes, 2010) how children advance from the storing of accent information as part of e familiar speakers' exemplars, to the forming of speaker groups based on social differentiations such as regional accent.

## 3.5 Conclusion

These experiments have investigated pre-school children's ability to identify familiar speakers and in particular, which aspects of a speaker's voice seem to play an important role in the recognition process. The Identification experiment revealed that the familiar teacher with the most distinctive voice pitch and voice quality was most readily recognised by the children. This indicates the importance of these voice attributes in identifying an individual who has a particularly idiosyncratic voice pitch/quality. The Recognition experiment found that regionally-based phonetic accent features also appear to contribute to speaker recognition. The children were more likely to misidentify an unfamiliar speaker as their familiar teacher if they used the same phonetic accent features. Additionally, the children generally failed to recognise their teacher when she disguised her accent and used a phonetic realisation different to her own.

This study also looked at the effect of external influencing factors on the abilities of the children in both tasks. Generally, there was an improvement in both tasks throughout the pre-school years as the oldest children outperformed the youngest children. However, this was not found to be a straightforward relationship and must be considered along with other factors. In the Identification experiment, the amount of exposure the children had to the nursery teachers interacted with their years spent at nursery. This meant that overall, older children who spent more time at the nursery on a weekly basis were better able to identify their teachers. In the Recognition experiment, the children's sex seemed to have more of an impact on their ability. For the boys, there was a significant improvement in the task throughout the pre-school years, showing an age-related development. The girls, who

generally outperformed the boys, did not show an improvement with age but with an exposure to different regional accents at home. This suggests a more advanced conceptual ability to differentiate speakers with different accents amongst those who have increased exposure to members of accent categories outside of their home region. Further work to investigate these findings and their implications is needed; as a start, chapter 4 considers pre-school children's ability to group speakers by regional accent and the impact of their linguistic background on this ability.

# Chapter 4: Pre-school children's categorisation of speakers by regional accent

## 4.1 Introduction

Chapter 3 described two experiments run with pre-school children in York. These experiments found that children's ability to identify familiar nursery teachers relied on aspects of the speakers' voices such as their pitch, voice quality and regional accent. The next step in investigating children's awareness of regional accent, therefore, is to test their ability to perceive regional accent distinctions amongst unfamiliar speakers and utilise these distinctions in grouping the speakers together.

The ability to categorise speakers according to regional accent is something of a life-long skill which develops as we encounter more people from different places and our experience of the world helps us to interpret the linguistic nuances that we hear. How this categorisation process develops is the question addressed in this chapter. This is addressed by investigating whether the ability to group speakers according to their accent can be seen to emerge in pre-school age children.

Perceptual categorisation studies have found that the accuracy with which adults can group speakers into accent/dialect categories varies according to how broadly defined the groupings are as well as the age of the listeners themselves and their experience of variation relative to their residence in different regions of the country (cf. Williams et al., 1999; Clopper and Pisoni, 2004a; 2004b). While differences between adolescents and teachers have been found (Williams et al., 1999), the age at which children can use regional accent features in order to group speakers, and how this ability develops, is not clearly understood. Although a few studies have investigated young children's awareness of regional accents, their differing methodologies and assumptions deliver conflicting conclusions. While Floccia and colleagues (2009) found that children under the age of 7 were not able to group speakers by their regional accent, Beck (2014) found that children from the age of 5 were able to match regionally distributed pronunciations (see section 2.3.4). However, no studies appear to have looked at the ability of children under the age of 5 in this task. Therefore, the question for the current research is whether children of an even younger, pre-school, age show an emerging awareness of accent and how this might be manifested in their ability to group speakers according to regional accent features.

Although accent perception has not been studied in pre-school age children, the development of other sociolinguistic skills has been uncovered by previous studies (see section 2.2.3). These studies have shown that variation in the input is something that is experienced from the very beginning of language acquisition and that this relates to young children's own competencies in sociolinguistic variation, both in production and perception. Therefore, further investigation into the input that children receive is important for understanding their growing sociolinguistic competence.

The importance of linguistic input is at the heart of an exemplar model of indexical learning (see section 2.4.2). Such a model describes how children may advance from the storing of social information in the exemplars of familiar speakers' voices, to the forming of speaker groups based on social differentiations such as regional accent. An exemplar model emphasises the storing of experienced instances and therefore the importance of exposure. There is still a lot of scope for investigation into how this process might occur and the experiment in this chapter (named the Grouping experiment), along with the findings from chapter 3, aims to look into some of these issues. The overall aim of the Grouping experiment is to investigate whether pre-school children are able to use regional linguistic variation to group unfamiliar individuals and how their exposure to linguistic variation might have an effect on this ability. The exposure that the children have to linguistic variation is measured by whether their parents come from the local area, as well as more generally by investigating any age-related developments. In light of the findings from chapter 3, the Grouping experiment also explores whether the children's sex plays a significant role at this stage in their perceptual development.

The Grouping experiment investigates pre-school children's awareness of accent features indicative of the distinction between speakers from the north and south of England; the BATH, STRUT, FACE and GOAT vowels are the focus as they are diagnostic of these broad geographical areas (see section 2.1.3). As Lawrence (2014) found (see section 2.3.3), adults perform well at correctly labelling northern and southern tokens of these vowels based on single-word stimuli. This indicates the salience of these vowel distinctions for adults, and consequently these distinctions are expected to be the most salient for children as well; one of the aims of the Grouping experiment is to test the salience of these vowel distinctions for pre-school children. The Grouping experiment is designed with three levels of difficulty in

order to establish the different degrees of children's accent awareness. The first difficulty level tests children's ability to group speakers based on different pronunciations of the same word (e.g. 'b[a]sket' and b[ɑ:]sket'). This is similar to Beck's (2014) study, although the Grouping Experiment uses sentence-length stimuli rather than single words, and a grouping task rather than an ABX discrimination task (see section 2.3.4 for an explanation of the ABX task). These adaptations make the task slightly harder, but the results can be said to indicate more reliably that the children are using the differences to group speakers rather than simply matching sounds. Despite this, however, the design of the task for difficulty level 1 means that the results are still limited as they show children's ability to group speakers according to phonetic differences but not necessarily that these differences are linked in a more abstract way (i.e. that they belong to the same phonemic category). Therefore, the second difficulty level asks the children to group speakers based on different pronunciations of the same phoneme but in different words (e.g. 'p[a]th'/'p[ɑ:]th' and 'gr[a]ss'/'gr[ɑ:]ss'). This task is therefore testing something more advanced: the children's awareness that the same phoneme can be present in different words and can be pronounced differently. The task for the third difficulty level further tests the children's awareness of regional accent differences by asking them to group speakers across different phonemes (e.g. 'gl[a]ss'/'gl[ɑ:]ss' and 'br[e:]k'/'br[eɪ]k'). This task therefore investigates whether the children can group together speakers in an even more abstract way: based on the speakers' pronunciation of different categories of sound which together correspond to a regional accent distinction.

The specific research questions that the Grouping experiment addresses are:

(1) Can 3-4 year-olds group speakers by **phonological variables** indexing regional accents (difficulty level 1)?

(2) Can they do this when the phoneme is embedded in **different words** (difficulty level 2)?

(3) Can they group speakers using **different phonemes** (difficulty level 3)?

(4) To what extent do these abilities vary with **age, sex** and **input** from **different regional accents**?

## 4.2 Methodology

### 4.2.1 Participants

Twenty pre-school children took part in the experiment. These children were all attending one of two different nurseries in York; nine children aged 3.1 years to 4.5 years from one nursery and eleven children aged 3.2 to 4.6 years from the other nursery. An additional four children attempted the experiment but their results were excluded due to their lack of understanding of the task and not responding to the stimuli. Eighteen of the twenty children who took part were born in York; one child moved from Germany to York aged 5 months and another child moved from London and had been living in York for 17 months. The regional backgrounds of these children are considered in the analysis of the results (section 4.3).

### 4.2.2 Experiment stimuli

Stimuli for the experiment were recorded from one speaker who used two different accent guises. Using the same speaker helped to ensure that the children would focus on the phonological accent variation of the speaker guises during the experiment, rather than making decisions based on other characteristics of individual speakers' voices.

The speaker used to record the stimuli was a 25-year-old female from Chester. She was chosen as she has a fairly non-specific accent due to moving around the country and living in both the south and the north of the UK (as well as abroad) for various lengths of time. Due to her mobility and familiarity with both Northern and Southern accents, this speaker felt comfortable in switching between using vowel pronunciations typical of both Yorkshire and the South East of England. The speaker was recorded reading a list of sentences, using a Zoom H4n recorder which was set to record at a 16 bit 44.1kHz sampling rate.

The stimuli sentences were specifically constructed for use in the experiment, in order to include words featuring four regional accent differences between Yorkshire and Standard Southern British English (SSBE). These accent differences are the realisations of the vowels in the BATH, STRUT and FACE lexical sets (Wells, 1982a). Stimuli featuring the GOAT lexical set were also originally recorded but not used in the final version of the experiment. This was due to having to keep the experiment short in order to hold the children's attention. The BATH vowel (and its realisation as [a] or [ɑ:]) and the STRUT vowel (and its realisation as

[ʊ] or [ʌ]) are described as amongst the most conspicuous accent features in differentiating a Northern accent from a Southern one (cf. Wells, 1982a/b; Hughes et al., 2012). These distinctions differentiate most speakers from Southern England, up to the East and West Midlands, from speakers in the north of England. The FACE vowel (and its realisation as [e:] or [eɪ]) differentiates between speakers from further north in England (referred to as the 'Central North' in Hughes et al., 2012) and those from both the Midlands and the south of England. Together, therefore, these three vowels form a distinction between the vowels used in a local Yorkshire accent (situated in the Central North) and those used in a SSBE accent (pronunciations typical of the South East of England).

The sentences used for the experiment were designed with the target word (i.e. the word featuring the accent difference) at the end of each sentence in order to emphasise the target words and draw the children's attention to their phonetic realization during the experiment. The rest of the words in each sentence were chosen carefully in order to avoid including any words with any salient accent differences.

The sentences were constructed in groups of four. These groups consisted of two reference sentences and two grouping sentences. The two reference sentences were the example sentences presented to the children at the beginning of the experiment. The children were then asked to categorise the grouping sentences with either one of these example sentences (see experimental procedure, next section). The reference sentences in each set were therefore worded exactly the same as each other, so that the only difference between them was the target vowel pronunciation. For example, the BATH vowel ([a]/[ɑ:] distinction) in the sentences: 'This is my b[a]sket' and 'This is my b[ɑ:]sket'.

The grouping sentences, however, were designed with three different levels of difficulty. For the first difficulty level (the same word condition), the grouping sentences ended in the same word as the reference sentences and focused on the same target vowel pronunciation. For example, the reference sentences featured the BATH vowel in the word 'basket': 'This is my b[a]sket/ b[ɑ:]sket' and the grouping sentences also featured the BATH vowel in the word 'basket': 'Put me in a b[a]sket/ b[ɑ:]sket'. For the second difficulty level (the same phoneme condition), the grouping sentences ended in a different word to the reference sentences but the word featured the same phoneme (from the same lexical set) as the reference sentences. For example, the reference sentences featured the BATH vowel in the

word ‘path’: ‘We need to walk on the p[a]th/p[ɑ:]th’ and the grouping sentences also featured the BATH vowel in the word ‘grass’: ‘I want to walk on the gr[a]ss/gr[ɑ:]ss’. For the third difficulty level (the different phoneme condition), the grouping sentences ended in a different word from the reference sentences and, in addition, the final word featured a different phoneme. For example, the reference sentences featured the FACE vowel in the word ‘break’: ‘What did you br[e:]k/br[eɪ]k’, but the grouping sentences featured the BATH vowel in the word ‘glass’: ‘It was a gl[a]ss/gl[ɑ:]ss’.

The speaker was asked to read all of the sentences, first with a SSBE pronunciation and then with a Yorkshire pronunciation of the target word. In order to keep the pronunciation of the rest of the stimulus consistent, she was instructed to read the rest of the sentence naturally, in her normal accent (which was quite standard and not regionally distinctive). Additionally, a set of English/French sentences was recorded, with a reference and a grouping sentence spoken in English and then in French. This set of stimuli was recorded in order to be used at the beginning of the experiment, to test the children’s understanding of the task. Floccia and colleagues (2009) and Wagner and colleagues (2014) found that children performed well at grouping speakers by a foreign/native accent distinction (see section 2.3.4) and therefore it was anticipated that the children in the current experiment would find it relatively easy to group together English vs. French stimuli.

Table 4.1 shows each group of sentences used in the experiment, including the difficulty level (DL) and the phonetic realisations of the vowel in the last word. For DL1, four sets of stimuli were used, featuring differences in the pronunciations of three vowels: BATH, STRUT, and FACE. For DL2 and DL3, only pronunciations of the BATH and FACE vowels were featured. This was due to constraints on time as the task was designed to take no longer than 15-20 minutes and therefore it was not possible to include combinations of all vowels.

| Group | Reference sentences | Grouping sentences | DL |
|---|---|---|---|
| 1 | This is my TOY BOX | Put me in a TOY BOX | N/A |
| 1 | Ceci est mon COFFRE À JOUETS | Mettez-moi dans une COFFRE À JOUETS | N.A |
| 2 | This is my BASKET [a] | Put me in a BASKET [a] | 1 |
| 2 | This is my BASKET [ɑ:] | Put me in a BASKET [ɑ:] | 1 |
| 3 | This is my BUS [ʊ] | Put me on a BUS [ʊ] | 1 |
| 3 | This is my BUS [ʌ] | Put me on a BUS [ʌ] | 1 |
| 4 | This is my GATE [e:] | Put me by a GATE [e:] | 1 |
| 4 | This is my GATE [eɪ] | Put me by a GATE [eɪ] | 1 |
| 5 | We need to walk on the PATH [a] | I want to walk on the GRASS [a] | 2 |
| 5 | We need to walk on the PATH [ɑ:] | I want to walk on the GRASS [ɑ:] | 2 |
| 6 | Which fruit do you want on your PLATE? [e:] | I would like some GRAPES [e:] | 2 |
| 6 | Which fruit do you want on your PLATE? [eɪ] | I would like some GRAPES [eɪ] | 2 |
| 7 | What did you draw this AFTERNOON? [a] | I drew a CASTLE [a] | 2 |
| 7 | What did you draw this AFTERNOON? [ɑ:] | I drew a GIRAFFE [ɑ:] | 2 |
| 8 | Where is the SNAKE? [e:] | It is by the TABLE [e:] | 2 |
| 8 | Where is the SNAKE? [eɪ] | It is by the FIREPLACE [eɪ] | 2 |
| 9 | What did you BREAK? [e:] | It was a GLASS [a] | 3 |
| 9 | What did you BREAK? [eɪ] | It was a GLASS [ɑ:] | 3 |
| 10 | What did you eat this AFTERNOON? [a] | I had some CAKE [e:] | 3 |
| 10 | What did you eat this AFTERNOON? [ɑ:] | I had some CAKE [eɪ] | 3 |
| 11 | What do you want to do LATER? [e:] | I want to PLAY [e:] | 2 |
| 11 | What do you want to do LATER? [eɪ] | I want to DANCE [ɑ:] | 3 |

**Table 4.1: Stimuli recorded for use in the experiment**

### 4.2.3 Experimental design

The experiment was designed to be run in Microsoft PowerPoint as a slideshow, with pictures and sound clips. Each slide consisted of a different grouping task, with twelve tasks altogether. The first five grouping tasks (for the French/English sentences and the DL1 sentences) involved cartoon bears. In each of these tasks, the two mummy bears were displayed at the bottom of the screen. They were made distinguishable by having different

colour patches on their fur (see Figure 4.1). Each mummy bear was linked to a sound clip of one of the reference sentences and next to each mummy bear was a picture of the subject of the sentence, for example a toy box in the case of the first sentence. At the top of the screen, pictures of five identical baby bears were displayed, each linked to an audio clip of one of the grouping sentences. For the sake of consistency, in each task, three of the baby bears were linked to the Yorkshire sound clip and two of the bears were linked to the SSBE sound clip. The baby bears were arranged in a random order each time, to prevent children forming a pattern based decision.

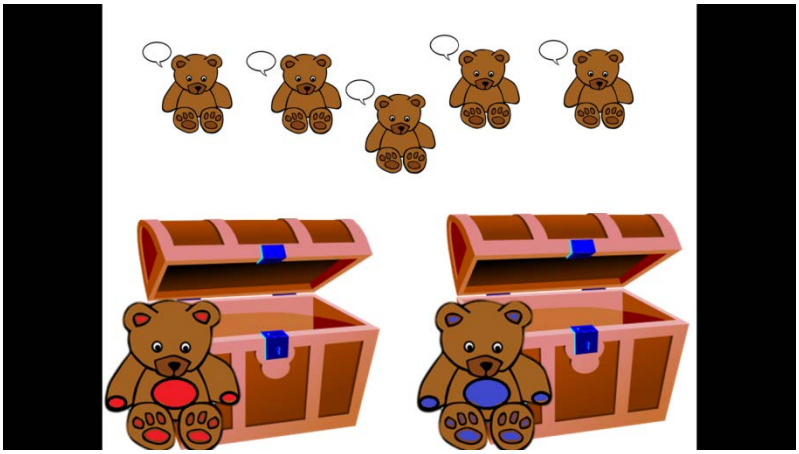


**Figure 4.1: Screen shot of first 'teddy bear' grouping task**

To keep the task varied and interesting, the grouping tasks for DL2 and DL3 used pictures of cartoon mothers and their daughters, instead of teddy bears. The grouping tasks were primarily the same, with two mothers (distinguished by different colour dresses) displayed at the bottom of the screen and five daughters arranged randomly at the top of the screen (see Figure 4.2). Similarly, a picture relating to the content of the sentence was displayed at the bottom of the screen, in between the mothers. Three different sets of mothers and daughters were used throughout these seven grouping tasks.

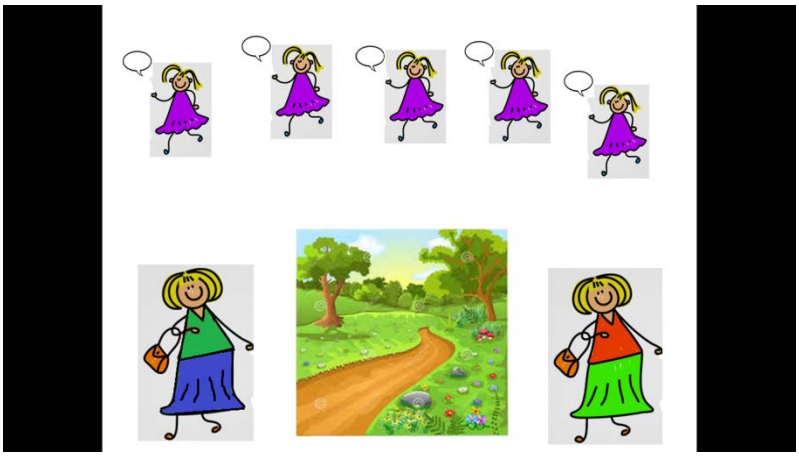


**Figure 4.2: Screen shot of first 'mother and daughter' grouping task**

### 4.2.4 Experimental procedure

The children took part in the experiment individually and I ran the experiment with each child, either in a quiet corner of the nursery or at the child's home with their parent(s) present. The experiment was presented and run on my laptop computer while both the child and I wore headphones in order to listen to the audio stimuli. Throughout the running of the experiment I was able to communicate with the child by speaking through a microphone. I asked the child if they would like to play a game on the computer in which they would be asked to group lost baby bears with their mummy bears and lost daughters with their mums. They sat at the computer with me while I controlled the running of the experiment. The PowerPoint show was started and the child was presented with their first 'sorting screen'. I told the child, '*We are going on a teddy bear's picnic but the baby bears have lost their mummy bear and need to be put with the right one. So what I would like you to do is to listen carefully to the mummy bears talk and then listen to each baby bear and tell me which mummy you think they belong to.*' [4]

[4] This wording was used as a way to avoid asking the children to group the bears based on the way that they sounded, as this could have been confusing to the children because the bear voices were spoken by the same person and therefore sounded very similar. This wording does have a potential problematic issue for children whose parents are not from the local area and who therefore sound different to their parents. However, as described in section 2.2.2, previous work has found that children are not overtly aware of differences between their parents and their own pronunciations. Furthermore, the results (section 4.3) showed that the children with non-local parents performed better overall, and therefore their parents' non-local status seemed to benefit, rather than hinder, their performance in this task.

I clicked on the picture of each mummy bear and the corresponding sound clip of one of the reference sentences played while the picture of the bear moved on screen to indicate that it was speaking. After the mummy bears had been heard, the first baby bear was played and I asked the child '*Which mummy bear does the baby belong to? Can you point to her?*' I then waited for a response from the child and dragged the picture of the baby bear over to the mummy bear the child indicated. This process was then repeated with each of the other four baby bears. If the child hesitated in responding and seemed stuck, I asked if the child would like to hear the baby bear again or whether they would like to hear the mummy bears speak again. This usually resulted in a 'yes' response and therefore I played the relevant sound clip(s) again. Each child responded differently to the stimuli, some giving an answer almost immediately and others needing more prompting and reminding of the task in hand. Therefore, it was necessary for me to respond on the spot to the requirements of each child. Those who took longer to provide an answer were more likely to need to hear a repetition of the mummy bear audio clips. If the child changed their mind, I registered the child's final answer.

I logged the responses by saving the slide with the sorted baby bears where the child had placed them (see Figure 4.3). If the child did not respond at all, even after prompting and a repetition of the audio clips, the baby bear was left where it was, at the top of the screen. The same procedure was followed for the rest of the grouping tasks. The five tasks involving the teddy bears followed a 'teddy bear's picnic' story, to keep the child interested. This culminated in an image of the teddy bears displayed altogether having a picnic (see Figure 4.4). At the end of the teddy bear task, I praised the child on their performance and asked if they would like to carry on and do the same thing again by grouping together daughters who had lost their mothers. Children who did not want to continue and those who failed to understand the task did not take part in the second part of the experiment (involving the stimuli of DLs 2 and 3). Out of the 20 child participants from the first part of the experiment, 15 went on to do the second part of the experiment.

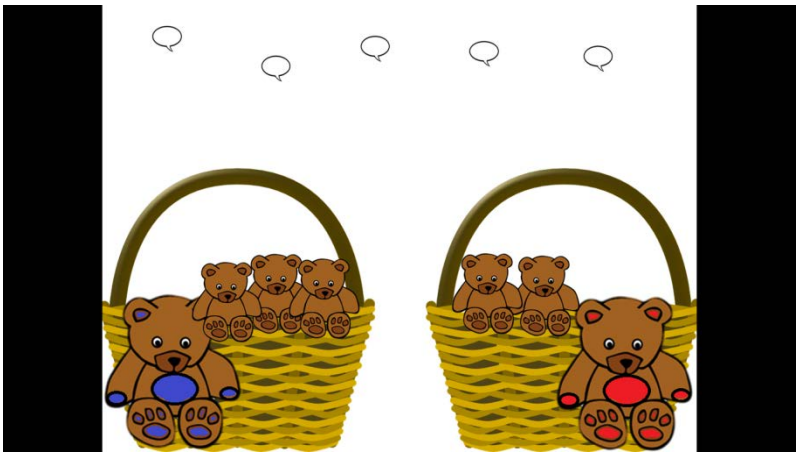


**Figure 4.3: Example of a 'sorted' screen**

**Figure 4.4: End screen for DL1: 'Teddy bear's picnic'**

## 4.3 Results and analysis

The overall results are presented and then the results for each DL are further analysed. The children's performance in the first set of stimuli (the English/French stimuli) showed a lot of variation. The inclusion of bears speaking different languages appeared to confuse some of the children and did not relate to their performance in the rest of the task. Therefore the children's responses to this first set of stimuli were disregarded for the rest of the analysis.

### 4.3.1 Overall results for each difficulty level

Table 4.2 shows the children's overall results for each DL along with the background data to be analysed statistically. The results are presented separately for each DL as not all of the children completed all of the DLs in the experiment. Furthermore, as each DL is testing a slightly different ability, the results for each DL are statistically analysed separately. This

allows for the fact that is likely that different factors play a role in the children's success in each DL.

| Child | Nursery | Age | Yorkshire Parents | (1) Same Word (% correct) | (2) Same Phoneme (% correct) | (3) Different Phoneme (% correct ) |
|---|---|---|---|---|---|---|
| F1 | 1 | 4.38 | 0 | 100 | 91.30 | 75 |
| F2 | 1 | 4.41 | 0 | 80 | 69.57 | 91.67 |
| F3 | 1 | 3.7 | 1 | 60 | 34.78 | 25 |
| M1 | 1 | 3.11 | 1 | 40 | 52.17 | 25 |
| F4 | 1 | 3.07 | 2 | 66.67 | 39.13 | 33.33 |
| F5 | 1 | 4.52 | 0 | 86.67 | 69.57 | 91.67 |
| M2 | 1 | 4.27 | 1 | 60 | 91.30 | 25 |
| F6 | 1 | 3.61 | 1 | 53.33 | 56.52 | 91.67 |
| M3 | 1 | 3.54 | 0 | 80 | | |
| F7 | 2 | 3.38 | 1 | 53.33 | | |
| F8 | 2 | 3.37 | 0 | 66.67 | | |
| F9 | 2 | 3.59 | 2 | 73.33 | 47.83 | 66.67 |
| F10 | 2 | 3.56 | 0 | 60 | 39.13 | 58.33 |
| M4 | 2 | 3.2 | 0 | 53.33 | | |
| M5 | 2 | 4.5 | 0 | 53.33 | | |
| F11 | 2 | 4.44 | | 86.67 | 56.52 | 100 |
| M6 | 2 | 4.41 | 2 | 60 | 60.87 | 58.33 |
| F12 | 2 | 4.64 | 2 | 46.67 | 86.96 | 50 |
| M7 | 2 | 4.2 | 0 | 60 | 56.52 | 91.67 |
| M8 | 2 | 4.59 | 1 | 60 | 47.83 | 58.33 |

**Table 4.2: The children's background information and percent correct answers for each DL**

Table 4.3 and Figure 4.5 show the mean percentage correct results for each difficulty level (DL). For each DL, the mean is above chance at 50%. However the t-value for each DL decreases, indicating a higher performance in DL 1 which is lower in DL 2 and then even lower in DL 3. DL 3, however, shows a much wider spread of scores indicating much more individual variability.

| DL | Mean % correct (SD) | t-value |
|---|---|---|
| 1 (Same word) | 65% (15) | 4.44 ($p<0.001$) |
| 2 (Same phoneme) | 60% (18) | 2.01 ($p=0.055$) |
| 3 (Different phoneme) | 63% (27) | 1.83 ($p=0.089$) |

**Table 4.3: Mean percentage correct for each DL in comparison to chance level (50%)**

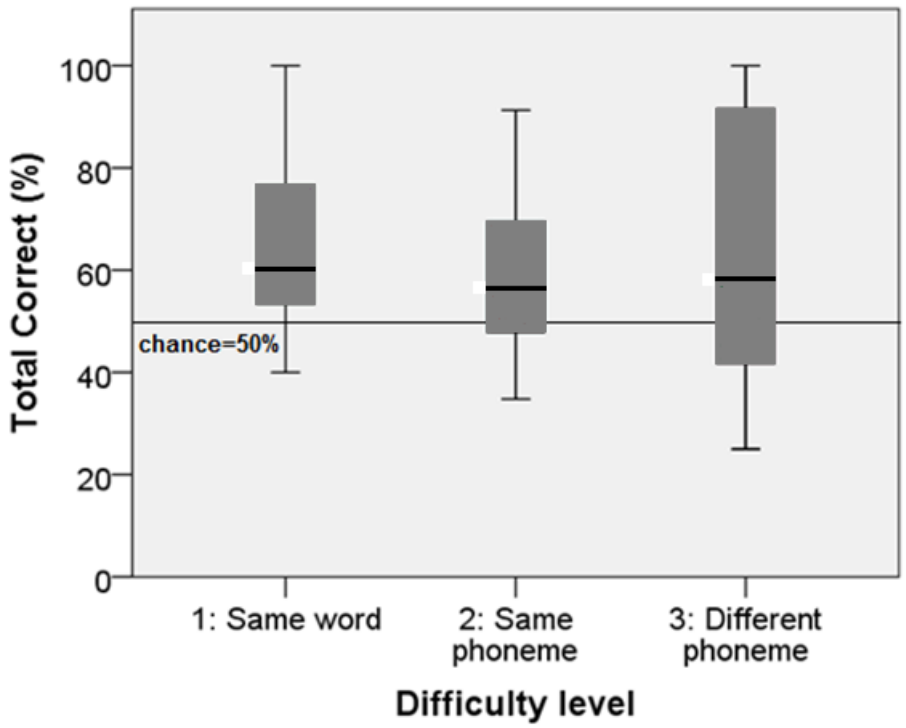



**Figure 4.5: Mean percentage correct for each DL**

Although the children's overall performance in DL1 is the highest, there is a lot of variation in how each individual performs across the difficulty levels (see Table 4.2). Only six of the children perform best in DL1 and out of these six, only three perform progressively worse across the DLs (i.e. second best in DL2 and worst in DL3). Of the remaining nine children, four perform best in DL2 and five perform best in DL3. Because of this variation between the performances across the DLs, henceforth the analysis of the results will consider the DLs separately.

### 4.3.2 Effect of independent variables on the scores for each difficulty level

Before analysing the results through statistical modelling, the raw results for each DL were plotted against each of the independent variables outlined in the research questions. This was in order to capture any correlations between the independent variables and the children's performance, and therefore provide a rationale for investigating the predictive power of these variables statistically (see section 4.3.3).

#### *4.3.2.1 Age*

A significant positive correlation was found between the children's age and their percentage of correct answers overall (Spearman's rho=.61, one-tailed p=0.002). As the correlation graph in Figure 4.6 shows, there is a clear division between the performance of the 3-year-olds and the performance of the 4-year-olds. Only one 4-year-old child scores below chance level, compared to seven 3-year-olds. Due to this age division, the children were split into two age groups for further analysis.

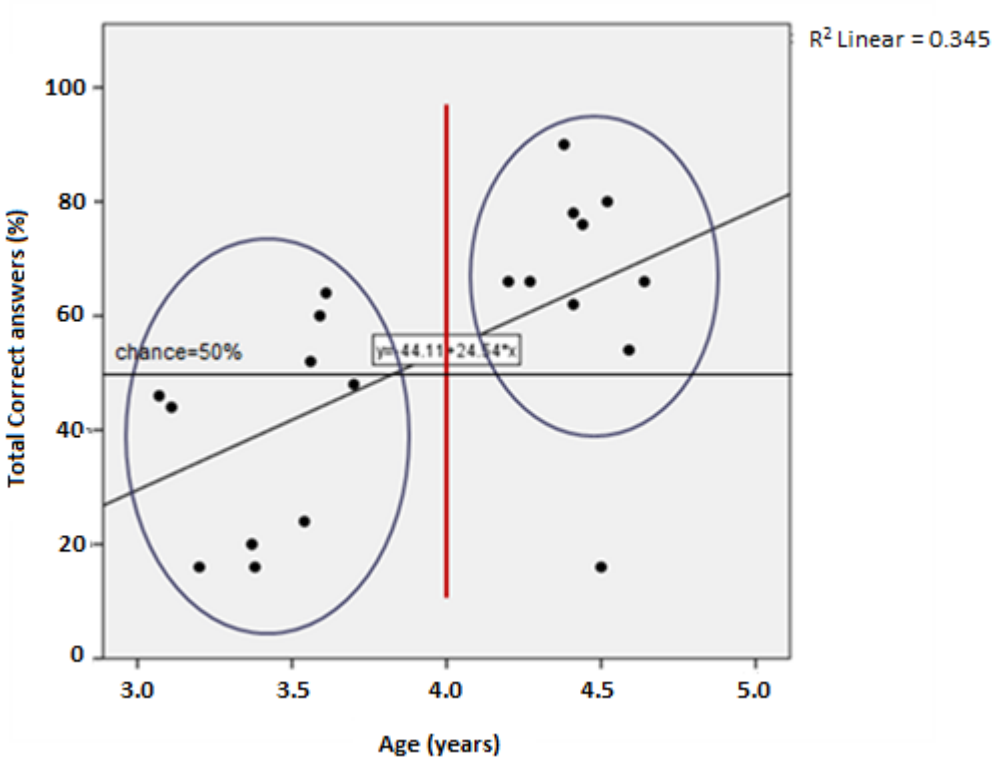



**Figure 4.6: Correlation between age and correct answers overalls overall**

Figure 4.7 shows the spread of scores across the DLs for the 3-year-olds compared to the 4-year-olds. Whereas the 4-year-olds score consistently above chance and at around the same level for each of the DLs, the 3-year-olds' scores are much more variable. They score above chance in DL1, below chance in DL2 and then around chance but with a very large range in DL3.

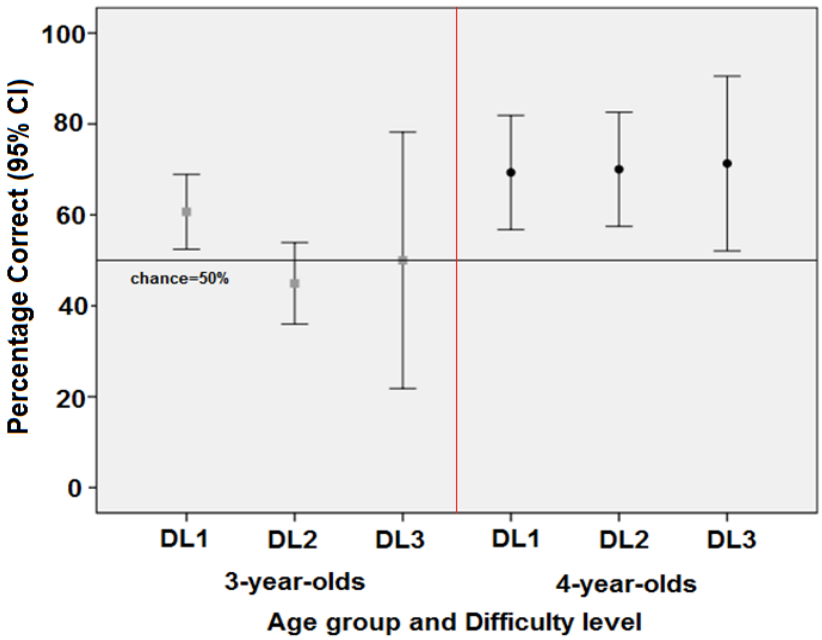



**Figure 4.7: Correct answers for each DL, divided by age group**

#### *4.3.2.2 Sex*

Figure 4.8 shows the spread of scores between the sexes for each DL. The girls score higher than the boys in general but the boys have a much wider spread of scores, particularly in DL3.

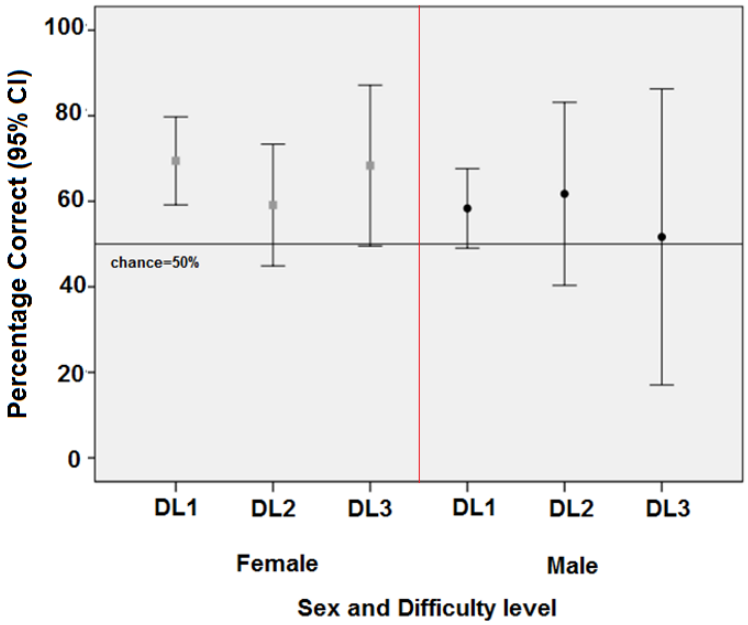



**Figure 4.8: Correct answers for each DL, divided by sex**

#### *4.3.2.3 Yorkshire parents*

Figure 4.9 shows the spread of scores relating to linguistic input. To be consistent with the Recognition experiment in chapter 3, this is measured by whether the children have any parents from Yorkshire (1+), or no parents from Yorkshire (0). Children who have no Yorkshire parents scored higher in each DL, compared to those who have at least one parent from Yorkshire. There is a particularly stark contrast for the DL3 results, in which those with no Yorkshire parents score on average around 80% correct, whereas those with at least one Yorkshire parent score less than 50% correct.

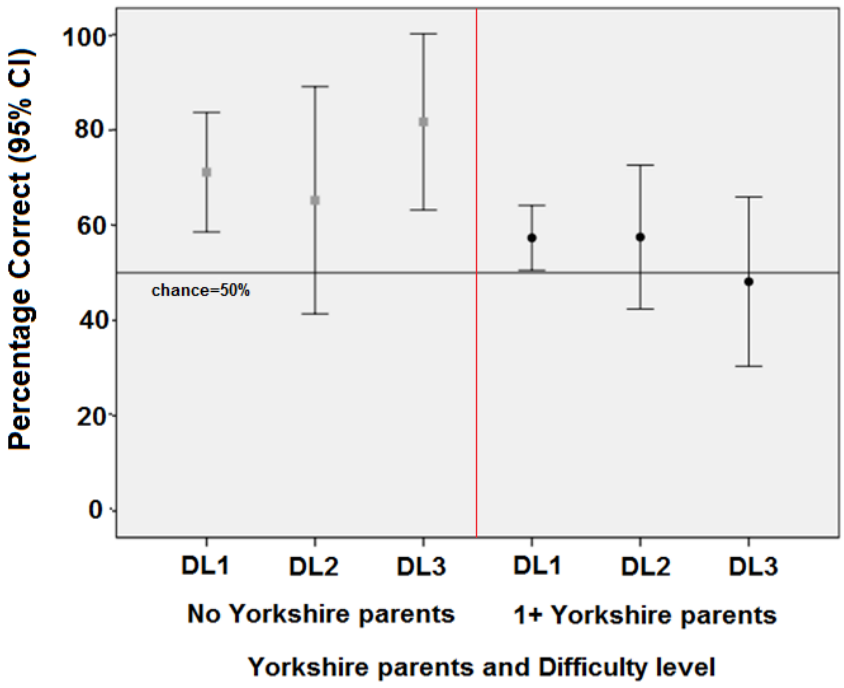



**Figure 4.9: Correct answers for each DL, divided by children with vs. children without parents from Yorkshire**

### 4.3.3 Mixed effects statistical modelling

Binary mixed effects logistic models were run in R (R Core Team, 2013) using a stepwise backward regression method. Separate models were run for each DL and due to the relatively small number of participants only three independent variables were included in each model. The three independent variables under investigation were included as binary independent variables: Age group (3-year-old/4-year-old), Yorkshire parent (Yes/No) and Sex (Female/Male) with the default values in the model amounting to '3-year-old girl with no Yorkshire parent(s)'. These default values acted as a baseline against which the model could measure the rest of the results (i.e. 4-year-olds' results were measured in comparison to 3-year-olds' results and boys' results were measured in comparison to girls' results). The difference in performance between the two different nurseries was also tested but not found to be statistically significant. As age was found to be the most significant predicting factor in the preliminary analysis, it was deemed important to consider how other independent variables may influence children's ability across the age groups. Therefore two-way interactions between age and sex and age and input were also included in the model, in

order to test for any potential relationships in the effects of these variables. Individual child was included as a random effect, in order to account for individual variation.

### 4.3.4 DL 1: Same word results

Table 4.4 shows the results of the mixed effects model fit over the results from DL 1 of the experiment. Although interactions were tested in the original model, these were not found to be significant and therefore the model was run with just three fixed, independent variables along with child as a random variable. Both age group and sex were found to significantly predict the probability of the children scoring a correct answer.

| | **Estimate** | **Std. Error** | **z value** | **Pr(>\|z\|)** | **Significance** |
|---|---|---|---|---|---|
| (Intercept) | 0.78 | 0.25 | 3.16 | 0.002 | ** |
| AGE (4-year-old) | 0.52 | 0.27 | 1.94 | 0.05 | * |
| SEX (Male) | -0.54 | 0.27 | -2.05 | 0.04 | * |
| YORKSHIRE PARENT(S) (True) | -0.43 | 0.25 | -1.68 | 0.09 | |

**Table 4.4: Logistic mixed effects model to fit results from DL 1 (Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

Figure 4.10 shows that the predicted probability of a correct answer increases between the ages of three and four. The increase in total correct answers between three and four-year-olds shown in the graph of the raw data (Figure 4.11) validates the effect that the model predicts.

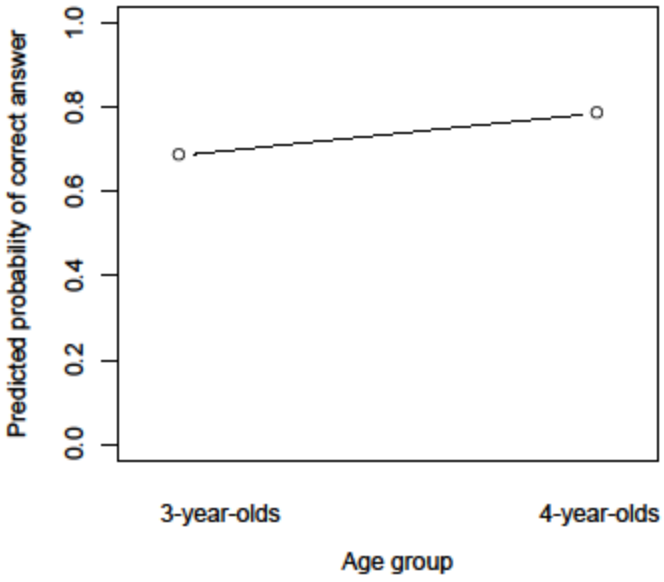



**Figure 4.10: Model prediction showing the effect of age group in DL1**

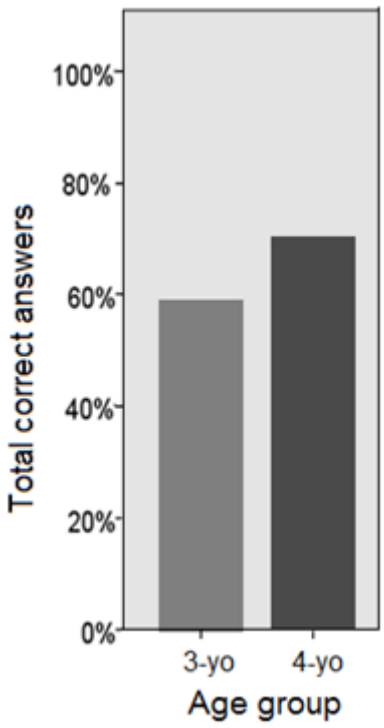



**Figure 4.11: Raw data showing the effect of age group in DL1**

Figure 4.12 shows the model prediction for sex for DL 1. The model predicts a higher probability of a correct answer for the females over the males. The raw data graph in Figure 4.13 confirms the effect of this prediction, showing that the females score a higher mean average of correct responses than the males for this DL.

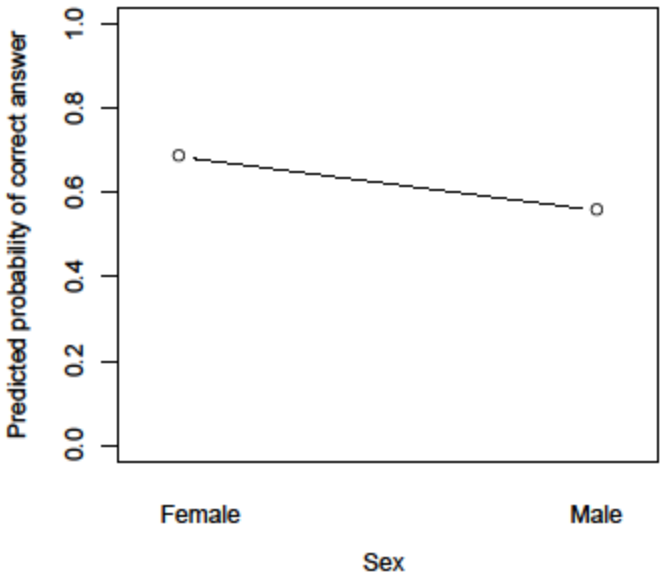



**Figure 4.12: Model prediction showing the effect of sex in DL1**

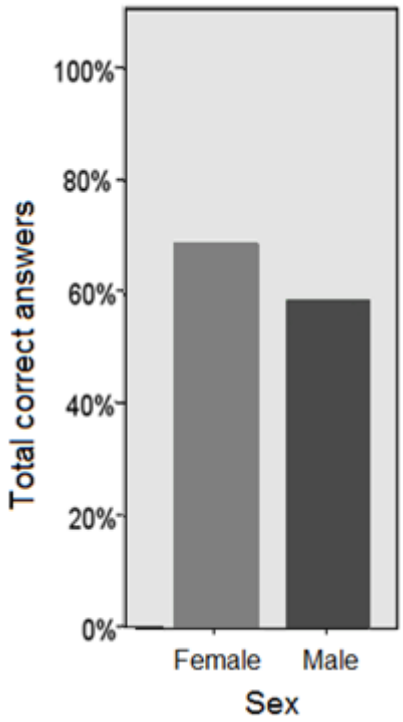



**Figure 4.13: Raw data showing the effect of sex in DL2**

### 4.3.5 DL 2: Same phoneme results

Table 4.5 shows the results of the mixed effects model fit over the results from DL 2 of the experiment. Only the children's age group has a statistically significant effect on their probability of scoring a correct answer.

| | Estimate | Std. Error | z value | Pr(>\|z\|) | Sig. |
|---|---|---|---|---|---|
| (Intercept) | 0.26 | 0.53 | 0.50 | 0.62 | |
| AGE (4-year-old) | 1.31 | 0.55 | 2.36 | 0.02* | * |
| SEX (Male) | -0.53 | 0.50 | -1.06 | 0.29 | |
| YORKSHIRE PARENT(S) (True) | -0.44 | 0.46 | -0.96 | 0.34 | |

**Table 4.5: Logistic mixed effects model to fit results from DL 2**
**(Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

Figure 4.14 and Figure 4.15 present the model prediction and raw data graph for age group for DL 2. Both the figures show an increase in correct answers between the ages of three and four. The model prediction is more robust than for the previous DL, with the predicted probability of a correct answer increasing from around 0.5 for 3-year-olds to over 0.8 for 4-year-olds.

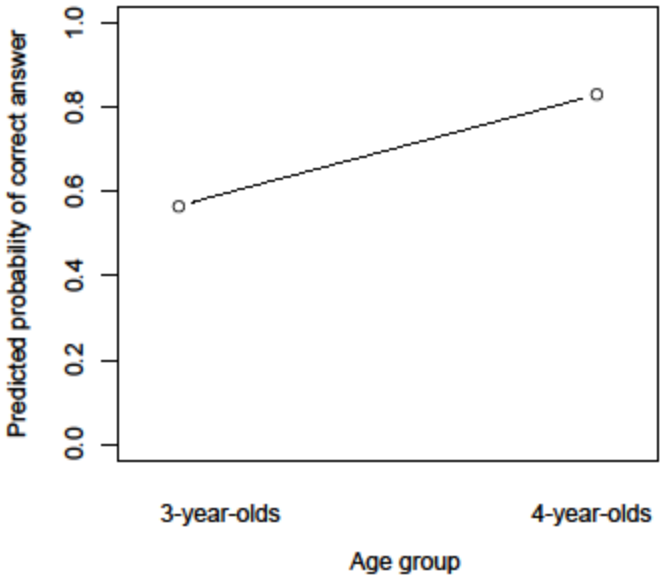



**Figure 4.14: Model prediction showing the effect of age group in DL2**

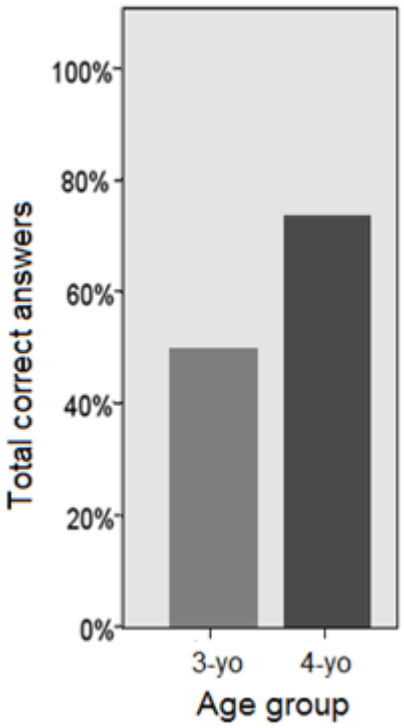



**Figure 4.15: Raw data showing the effect of age group in DL2**

### 4.3.6 Difficulty level 3: Different phoneme results

Table 4.6 shows the results of the mixed effects model fit over the results from DL 3 of the experiment. Only the effect of Yorkshire parentage was found to have a significant effect on the children's probability of scoring a correct answer in this DL.

| | Estimate | Std. Error | z value | Pr(>\|z\|) | Significance |
|---|---|---|---|---|---|
| (Intercept) | 1.61 | 0.45 | 3.58 | 0.0004 | *** |
| YORKSHIRE PARENT(S) (True) | -1.6 | 0.54 | -3.0 | 0.003** | ** |

**Table 4.6: Logistic mixed effects model to fit results from DL 3 (Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

Figure 4.16 and Figure 4.17 show the model prediction and the raw data graphs for age group for DL 3. Both the model prediction and the raw data itself show that the children with no Yorkshire parents perform better than those with parents from Yorkshire. Again, this prediction is fairly robust, with the predicted probability of a correct answer increasing from under 0.6 for those with at least one parent from Yorkshire to over 0.8 for those with no parents from Yorkshire.

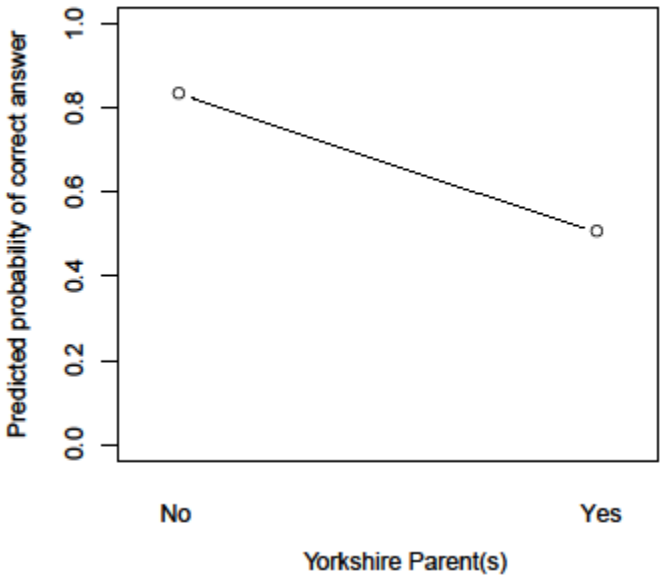



**Figure 4.16: Model prediction showing the effect of Yorkshire parentage in DL3**

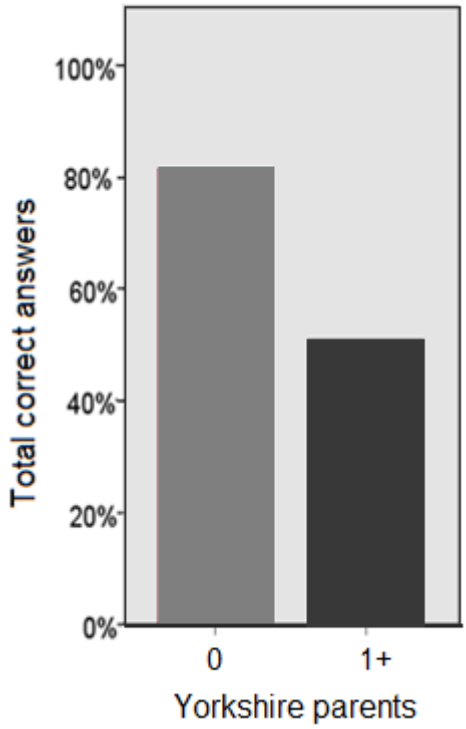



**Figure 4.17: Raw data showing the effect of Yorkshire parentage in DL3**

### 4.3.7 Summary of results across the different DLs

In summary, pre-school children scored above chance level when grouping together speakers based on regionally distributed pronunciation features. This task was easiest when they were asked to group together speakers using the same word and became progressively harder as they were asked to group across different words with the same phoneme and then across different phonemes altogether. Effects of age, sex and input were found relating to results from the different DLs. These effects are discussed further in the next section.

## 4.4 Discussion

### 4.4.1 Age

This Grouping experiment has found that children from 3-years-old score above chance level when grouping together speakers based on regionally distributed pronunciation features. Children of this age have not been tested in this ability previously, and indeed earlier studies have suggested that children under the age of 7 are not able to group speakers according to accent criteria (see section 2.3.4 for a critique of these studies).

Overall, an age improvement was found between the 3- and 4-year-olds. This improvement throughout the pre-school years is in line with other sociolinguistic developments occurring at this time, including adult-like patterns of production (cf. Roberts and Labov, 1995; Foulkes et al., 1999; Smith et al., 2007; Barbu et al., 2013). Furthermore, this improvement with age is consistent with the results from the Recognition Experiment in chapter 3 which found that accent cues help in the recognition of familiar speakers by 2-4-year-olds and that this ability improves throughout the age range.

The probability of improvement with age (calculated by the logistic regression model) was found to be most robust in DL 2 (the same phoneme condition). Whereas the 3-year-olds were predicted by the regression model to perform just above chance level (57% chance of a correct answer), the four-year-olds were predicted to perform well above chance level (83%). This is a much more extensive prediction than the improvement from 69% to79% between the age groups for DL 1 (the same word condition). The task involved in DL2 is undoubtedly harder for the children as they hear different words and have to then extract the similar sounding vowel realisations out of these words in order to be able to group them. The age-related improvement in this process of abstraction shows a development in the understanding of what a phoneme category is and what its variable realisations may be. This development in understanding is a step beyond abstracting across different accent realisations when recognising familiar words, which even younger infants have been found to be able to do (cf. Schmale et al., 2010; Best and Kitamura, 2012). Rather than abstracting across variation, these pre-school children are using the phonetic variation that they hear as criteria for grouping the speakers.

### 4.4.2 Sex

Overall, the girls outperformed the boys in this experiment. This accords with the results from the Recognition experiment in chapter 3, which found that girls were better than boys at recognising a familiar teacher from amongst unfamiliar speakers. The girls were only found to outperform the boys significantly in DL1 (the same word condition), but overall the boys' ability varied a lot more, particularly in DL3 (the different phoneme condition). This difference between the sexes could partly be due to the nature of the task itself, which was focused on the speech of females, using female cartoon pictures and run by a female experimenter. Therefore the boys may not have been as engaged in a task which was

focused on the opposite sex. Support for this interpretation can be found in the results of Cvencek and colleagues (2011), who ran implicit association tests with pre-school boys and girls. They found that girls showed a stronger implicit preference for flowers over insects than boys, indicating the stereotyped association we might expect. Furthermore, the girls were found to show a stronger implicit preference for their own sex than the boys (which supports previous research with adults and primary school age children: cf. Ebert and Steffens, 2008). In order to investigate any task-design bias in the current work, it would be beneficial to run a future study based on a similar experimental design but with a focus on the speech of boys and with a male experimenter. While this research indicates that the present study's task design might positively bias the performance of the girls, another possibility is that the girls are just better at this kind of task than boys. As discussed in chapter 3 (section 3.4), previous research to this effect is inconsistent. Cvencek and colleagues (2011) also dispute this alternative explanation for their results. They report that previous studies have not shown differences between the sexes in similar tasks which have required demands on memory and cognitive processing (Davidson et al., 2006; Cragg and Nation, 2009).

### 4.4.3 Input

Children with parents from outside of Yorkshire had a higher chance of being correct in DL3. Whereas children with at least one parent from Yorkshire were predicted by the regression model to score a correct answer at around chance level (50%), those with both parents from outside of Yorkshire were predicted to score a correct answer well over chance (around 83%). DL3 added an extra level of abstraction to the task for the children as they were being tested on matching different phonemes; either 'Yorkshire sounding' or 'Southern sounding'[5]. Those with outsider parents appear to have made a more distinct category divide between, in this case, 'Yorkshire' and 'Southern' sounding speakers. Therefore, variation in the input seems to be a contributing factor for children's performance in this particular task.

[5] The category labels 'Yorkshire' and 'Southern' are used here for convenience to refer to the distinction between a category of speakers in the children's local area vs. a category of non-local speakers from an area in the South. This is not to say that the children themselves can identify these category labels, but that they are grouping by criteria pertaining to these categorical distinctions.

This finding echoes the results from the Recognition experiment in chapter 3, in which it was found that girls with parents from outside of Yorkshire were better able to recognise a familiar teacher using phonetic accent features. As discussed in section 3.4, the importance of linguistic input is foregrounded in an exemplar account of indexical learning. Children with parents from outside of Yorkshire have had more exposure to speakers with different accents at home and therefore have more stored exemplars of speakers with different accents. These stored exemplars of familiar speakers then form the basis of speaker categories united by regionally distributed features of pronunciation. It is hypothesised, then, that in the Grouping experiment children are able to use their experience with different accents as a basis for grouping the unfamiliar speakers that they hear.

Figure 4.18 shows that three of the top four performers in DL3 (F2, F5, M7) have a Southern parent. They therefore have more experience with a Southern accent as it is presumably something that they hear a lot at home. Their high performance in this task can therefore be explained as a result of their experience with a Southern accent in particular. As the task in the Grouping Experiment involved grouping together speakers with a Yorkshire accent and speakers with an SSBE accent, these children with a Southern parent have the advantage of prolonged experience and, following an exemplar model of language development, an increased store of Southern-sounding vowel exemplars. Therefore, for this relatively difficult task in which they are asked to group together different phonemes, these children have a better idea of which combination of phonetic realisations of these phonemes 'go together' because they have more experience of this combination. As the current findings are based on a relatively small number of children, however, more studies are needed to validate these claims. Further studies could investigate this claim by collecting data about how many contact hours and how much verbal interaction the children have with each parent.

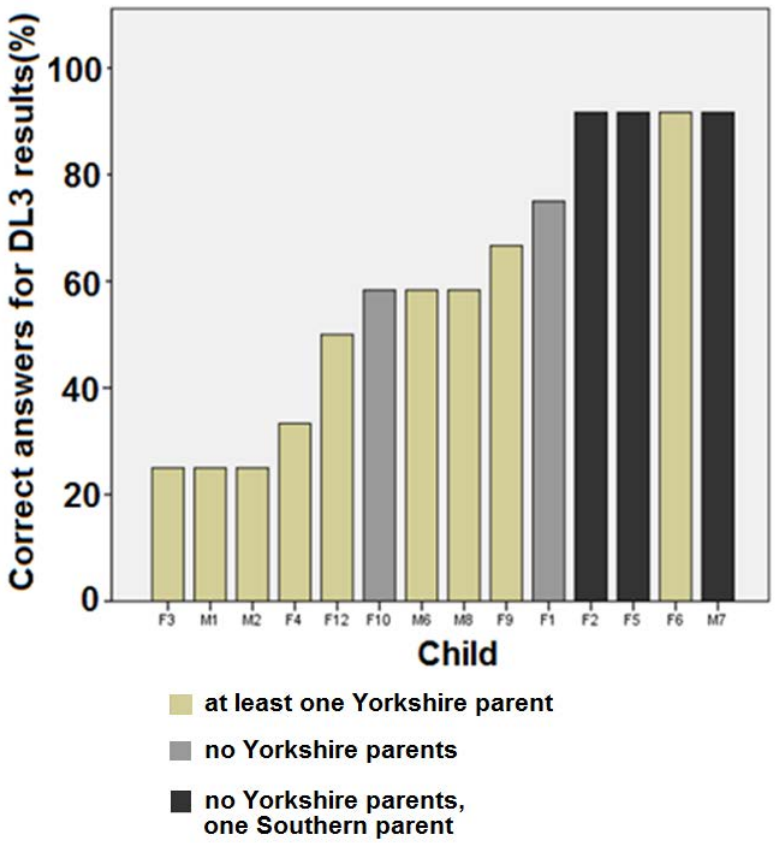



**Figure 4.18: Children's parental background and their performance in DL3**

Other potential factors influencing children's performance in this task should also be considered. The children may have had a lot of contact with other individuals (e.g. grandparents, family friends) from different regions. Therefore their exposure to other accents may have come from frequent contact with individuals other than their parents. Although data to this effect was not collected originally, parents were contacted after the experimental process to ask for details about their children's contact with grandparents and other individuals. Unfortunately only three parents responded to this request and therefore this information could not be used in the analysis. Furthermore, the wide variety of regional accents used in children's television shows and in the media in general means that a high proportion of children's exposure to different regional accents may come from these indirect sources. Stuart-Smith and colleagues (2013) remark on the increased use of different regional accents in the media in recent years, which has meant that many listeners are exposed to accents/dialects that they previously had little experience of. While data

pertaining to children's TV watching habits were not collected for the Grouping Experiment and therefore could not be explored, an attempt at collecting this data was included in the Second Grouping Experiment (see next chapter).

## 4.5 Conclusion

The Grouping Experiment has found that there is development between the ages of 3 and 4 in children's ability to group speakers according to phonological regional variables. This result was found in the DL1 and DL2 parts of the experiment, in which children were asked to group speakers based on hearing the same word featuring an accent difference, or the same phoneme featuring an accent difference. This improvement with age appears to demonstrate development in the children's understanding that variable realisations of phonemes can represent a categorical difference between speakers. In DL3 the children were asked to group speakers based on hearing different words. Therefore, the children were grouping speakers across different realisations of two different phonemes. In this part of the experiment, it was found that varied input helps in the creation of more robust accent categories, as children with parents from outside the Yorkshire region performed better. This finding shows that, beyond perceiving the phonetic differences between variable realisations of the same phoneme, these children were able to group together different phonemes representing an accent distinction. Therefore, the children with parents from outside the region have more robust cognitive accent categories, which make grouping speakers into these categories easier. This finding supports a usage-based model of language acquisition in which speaker categories are based on experienced exemplars which are stored on encounter and then later accessed in speech processing and which are constantly updated throughout the lifetime.

This experiment used an accent guise technique, with stimuli taken from a single speaker in order to focus on children's ability to group by the phonological variables and not other aspects of the speaker's voice. This helped to ensure that the task was establishing children's ability to group together speakers based on their pronunciation rather than their pitch or voice quality, for example. However, this is not a situation that children, or indeed adults will face in real-life. In the real world, we hear speakers with similar accents but different voice qualities and pitches and we hear speakers from the same place but with different strengths of regional accent. Furthermore, while the differences between a

Yorkshire and an SSBE accent are well recognised as diagnostic features of the north and south of England (see section 2.1.3), other regional accent distinctions are not quite so conspicuous and children are perhaps less likely to be able to group speakers accordingly. Chapter 5 will therefore explore these issues further in the Second Grouping experiment, carried out with older children.

# Chapter 5: Primary school children's categorisation of speakers by regional accent

## 5.1 Introduction

The experiments detailed in chapters 3 and 4 demonstrated that children's awareness of regional accent distinctions develops throughout the pre-school years. The Recognition experiment in chapter 3 found that pre-school children utilise accent features to recognise a familiar speaker. Furthermore, the Grouping experiment in chapter 4 found that pre-school children are also able to group speakers by features of their accent. Both of these experiments found an age-related improvement as older children performed better overall. Also, both experiments found an improvement suggestive of the role of children's exposure to variation; children with more exposure to variation were found to perform better overall. These findings indicate that awareness of regional accent is a progressive process and relies upon the accumulation of experience as well as the maturation of cognitive capabilities.

The Grouping experiment used stimuli from one speaker in two accent guises: Yorkshire and SSBE. Therefore, the children's attention in this task was focused on a pervasive accent difference without the need to consider individual speaker differences. Building on these results, the next step is to investigate children's capacity to attend to different speakers as well as other, possibly less familiar accents. Following an exemplar model of indexical learning, it is hypothesised that the grouping of speakers by regional accent follows a developmental process. This process starts with the recognition of familiar speakers and the storing of social information in exemplars of their speech. It then progresses to the grouping together of speakers whose exemplars activate similar social information. The more speakers an individual encounters, the more varied their exemplar store becomes. In turn this means that an individual with a more varied store of exemplars will be able to abstract a larger number of speaker categories from these exemplars. Moreover, a more varied set of voices will be included within each of the categories created by the individual.

Overall, therefore, it is anticipated that as children get older and encounter more speakers and thus more variation, they are better able to analyse and abstract over this variation appropriately in order to categorise speakers by their accent. An exemplar-theoretic account of acquisition hypothesises that cognitive categories of speaker are strengthened each time they are accessed, and with every addition of exemplars from different speakers

added to them. Well established speaker categories are stronger because they have multiple different speakers' exemplars to draw upon, which helps when processing new exemplars. For example, children are able to categorise speakers by their sex/gender from a young age because they have already encountered many different examples of male and female voices. Therefore, incoming exemplars are fairly easy to categorise as belonging to a male or a female because the listener has many different male and female voices to compare the new exemplars to. In the same way, the more exposure a listener has to speakers with different accents, the more substantial their speaker categories based on regional accent distinctions will become. This means that it will be easier to categorise new speakers by their accent as they have multiple speakers in each category to make a comparison with; the commonalities between the incoming exemplars and the stored exemplars will be more apparent.

It is also expected that children who have been exposed to more variation in general will be better at grouping speakers by their regional accent. As the experiments in the previous two chapters have found, children with parents from outside the local region are better at using regional accent features to identify and group speakers. Therefore it is predicted that individuals with more experience and exposure to speakers with different accents will also be better able to differentiate other less familiar accents. The Grouping experiment found that pre-school children can differentiate between accent features indicative of a Yorkshire accent compared to an SSBE accent. These accents are both likely to be frequently encountered by children living in York. The children hear Yorkshire speakers in their local community, and features indicative of SSBE are also present in the speech of many middle-class speakers throughout the country (see section 2.1.2). It is unknown, however, how well children will respond to other Northern accents which are encountered less frequently and yet are more similar to the accent of their local area. Therefore, speakers with accents from two other regions of the north of the UK are included in this experiment: Scottish Standard English speakers and North Eastern English speakers, as well as SSBE speakers. As described in section 2.1.2 and 2.1.3, Scottish Standard English (SSE) has different phonological realisations to SSBE and other English accents. These realisations are not exclusive to SSE however and may be used by Scottish speakers with non-standard accents too. Therefore, in order to represent the broad distinction between speakers from Scotland

more generally and speakers from England, henceforth reference will be made to 'Scottish speakers' and the 'Scottish accent'.

The three regional accent distinctions for this experiment (Scottish, North East and SSBE) were chosen because of their varying differences from the Yorkshire accent. As described in section 2.1.3, the SSBE accent forms a distinctive comparison to the Yorkshire accent, demonstrated by the pronunciation of the BATH, STRUT, FACE and GOAT vowels, which are features of the North/South linguistic divide in England. On the other hand, the Scottish accent is more similar to the Yorkshire accent in terms of vowels, while the feature of rhoticity divides speakers from Scotland and speakers from most of the rest of England (see section 2.1.3). Therefore, the Scottish accent was chosen to represent a different comparison to the Yorkshire accent. This difference was both in regional terms (as the Scottish accent is representative of further north rather than further south) and also in terms of the accent features themselves. The North East accent was chosen as another comparison in order to represent an accent much closer to Yorkshire. Deriving from neighbouring regions in the North of England, these two accents share many similar vowel realisations. Therefore, the differences in the vowel pronunciations of these two accents are more fine-grained (see section 5.3.1) and it is expected that these accents will be much more difficult to tell apart.

Children who have experienced more variation overall from speakers with whom they are in regular contact (such as their parents, family or close family friends ) are hypothesised to have a heightened awareness of accent differences in general and therefore be better able to deal with new kinds of variation. They will be able to compare relatively unknown incoming exemplars to a larger range of stored exemplars and therefore a larger range of speaker categories. Therefore, they will be able to make a more informed decision about how these new exemplars should/could be categorised; whether they belong to any existing speaker categories and/or whether they are different enough to form their own categories.

The extent of young children's awareness of different regional accents has been relatively little explored compared to that of adolescents and adults. While the experiments in chapters 3 and 4 addressed the lack of work investigating pre-school children's awareness, the experiment in the current chapter expands on these experiments by exploring the abilities of primary school children. Although a few studies have been carried out investigating primary school children's awareness of regional accents, these studies have

missed some important lines of investigation. As described in the literature review (see section 2.3.4), Wagner and colleagues' (2014) study found that primary school age children were better at grouping speakers by a second language accent/home accent distinction than a regional accent/home accent distinction. They therefore conclude that children compare accents in terms of how similar they are to their home accent, and also that a regional accent is often familiar enough to be categorised with their home accent. However, this assumption disregards children's individual experiences and the role that their exposure to variation might play. The current experiment (named the Second grouping experiment) aims to address the overlooked effect of accent exposure on children's ability to group different speakers by their accent. This is investigated by collecting information about the regional varieties of the speakers that the children are in regular contact with.

In addition, information about the kinds of television programmes and films that the children regularly watch was collected. This is in order to form an overview of the extent of their exposure to variation via the media. As described in section 2.3.1, in recent years regional and non-standard varieties are being used to a greater extent in different forms of broadcast media. Therefore, nowadays there is greater general awareness of variation because of listeners' exposure to different accents and dialects in the media. Although it is difficult to study this exposure systematically, data collected about the children's favourite programmes will reveal the range of material they watch and therefore indicate potential additional sources of their exposure to variation.

Due to the topic being previously under-researched, the design of the experiment was an important part of the investigation itself. In order to make sure that the children were being tested to their full potential, it was necessary to explore the possibilities of the children's abilities before establishing a final experimental design. Therefore, the pilot designs of the Second grouping experiment and their findings are described here in order to justify the main experiment design.

## 5.2 Research questions

The research questions that the Second grouping experiment addresses are:

(1) Can 5-9 year-olds group **different** speakers by **phonological variables** indexing a **Yorkshire/SSBE** regional accent distinction (Round 1 of the experiment)?

(2) Can they do this with **different** speakers indexing a **Yorkshire/Scottish** regional accent distinction (Round 2 of the experiment)?

(3) Can they do this with **different** speakers indexing a **Yorkshire/North East** regional accent distinction (Round 3 of the experiment)?

(4) Is the children's ability in each round of the experiment affected by whether the phonological variables are embedded in the **same word** (difficulty level 1) or **different words** (difficulty level 2)?

(5) To what extent does the children's ability in each round of the experiment vary with their **age, sex,** their **exposure to different regional accents** and their **parents' level of education**?

(6) How much **individual variation** is there in these children's performances?

## 5.3 Experimental design process

Three different pilot experiment designs were constructed and run before the fourth and final version was carried out. The results from this final design of the experiment will be the main focus of this chapter. However, as the experimental design process was important in revealing both the capabilities of the children and the shortcomings of the experiments themselves, these initial experimental designs and their limitations will be described in this section.

### 5.3.1 Experimental stimuli

A total of 25 speakers were recorded to provide stimuli for the experiments. All of the speakers were young females, aged 18-28. The speakers were chosen for their accents, most of which represented four regional varieties: Yorkshire, SSBE, Scottish Standard English, and North Eastern England. In addition, a young female American speaker and a young female North London speaker (with a non-standard accent) were recorded. These two were recorded to act as a further comparison with the four main regional accents but were not used in the final experiment design. Therefore the use of these speakers will only be briefly discussed in relation to pilot experiment design 3.

The speakers were recorded either in the recording studio of the Department of Language and Linguistic Science at the University of York or in a quiet room using a Zoom H4n

recorder which was set to record at a 16 bit 44.1kHz sampling rate. The speakers were recorded reading a story passage ('Chicken Little'[6]) and a list of sentences (see Appendix 3). These stimuli were created specifically for use in the experiment in order to include words with variables that distinguish the different regional accents. The list of sentences consisted of six sets, with each set featuring a different vowel out of the vowels in the lexical sets: BATH, STRUT, GOAT, FACE, NURSE and lettER. The sentences were designed such that the target word featuring the vowel was always located at the end. The rest of the words in each sentence were chosen carefully in order not to include any other words which contained the variables relevant for these accent differences. The Chicken Little passage also included words featuring these vowels.

Table 5.1 shows the realisations of these vowels by speakers of the four different accents. The BATH, STRUT, FACE and GOAT vowels (features 1-4 in Table 5.1) were recorded in order to differentiate between the Yorkshire and SSBE speakers. These vowels relate to well documented North/South regional distinctions in England (as described in section 4.2.2). The NURSE and lettER vowels (features 5-6 in Table 5.1) differentiate the non-rhotic Yorkshire speakers from the rhotic Scottish speakers who pronounce an /r/ after the vowel sound in these lexical sets. This rhotic/non-rhotic distinction is a well known difference between Scottish and most English accents (Wells, 1982b; Hughes et al., 2012).

The GOAT and FACE vowels (features 7-8 in Table 5.1) also differentiate the Yorkshire speakers from the North East speakers. The 'pan-northern' [e:] (FACE) and [o:] (GOAT) vowels are often cited as representing speakers' pronunciations from both these regions (cf. Haddican et al., 2013). However, while these vowels are similar in terms of vowel quality, acoustic analysis of the formant frequencies reveals differences between the Yorkshire and North East speakers' pronunciations (see the formant plot in Figure 5.1). A lower F2 in the GOAT vowel for the North East speakers (apart from one of the speakers) indicates a back vowel, close to cardinal vowel 7 [o:]. In comparison, the Yorkshire speakers have a more fronted GOAT vowel, represented by [ɵ:]. This fronted version is quite typical of young females in York (Haddican et al., 2013). Lower F1 and higher F2 values indicate closer, more

[6] Adapted from the passage used in Best and Shaw's project: 'You came TO DIE?! Perceptual adaptation to regional accents as a new lens on the puzzle of spoken word recognition', Australian Research Council (ARC Discovery Project DP120104596).

front FACE vowels from the North East speakers, close to cardinal vowel 2 [e:]. On the other hand, the Yorkshire speakers have a slightly more open and back pronunciation of this vowel, closer to cardinal vowel 3 [ɛ:]. In addition, the North East speakers' GOAT and FACE vowels are generally more diphthongal in quality than the Yorkshire speakers' vowels. This is indicated by a larger Euclidean distance between the nucleus and offglide F1 and F2 measurements.

| Accent feature (lexical set) | Yorkshire speakers' realisation | Comparison accent and their realisation |
|---|---|---|
| (1) BATH | [a] | SSBE[ɑ:] |
| (2) STRUT | [ʊ] | SSBE [ʌ] |
| (3) GOAT | [ɵ:] | SSBE [əʊ] |
| (4) FACE | [ɛ:] | SSBE [eɪ] |
| (5) NURSE | [ɜ:] | Scottish [ɜɹ] |
| (6) lettER | [ə] | Scottish [əɹ] |
| (7) GOAT | [ɵ:] | North East [o:] |
| (8) FACE | [ɛ:] | North East [e:] |

**Table 5.1: The realisation of the accent features by speakers of the four regional accents**

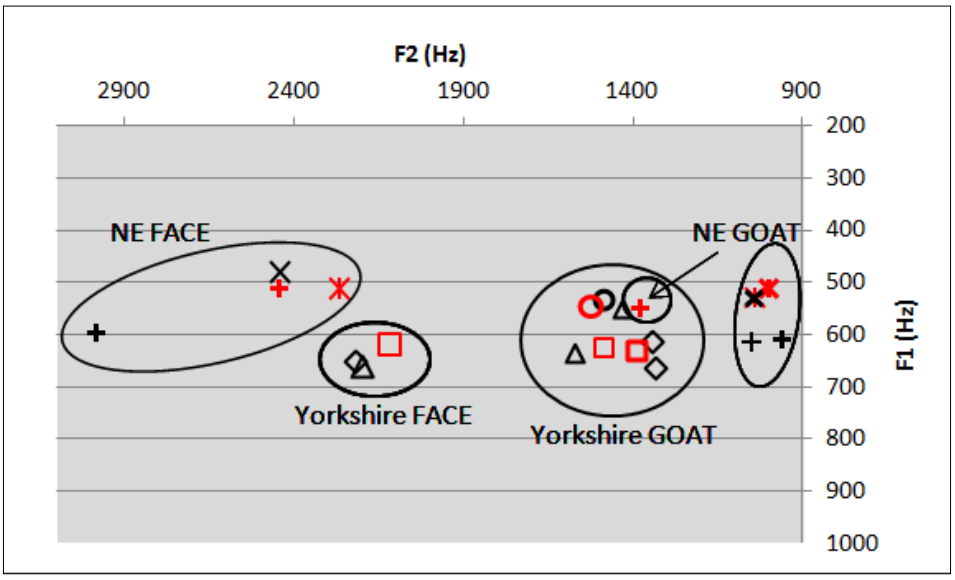



**Figure 5.1: Formant plots for the FACE and GOAT vowels of the Yorkshire and North East speakers (each symbol represents a different speaker's vowel)**

### 5.3.2 Design 1

#### *5.3.2.1 Experiment design and procedure*

The first experiment design was constructed in a similar way to the Grouping experiment (chapter 4). It was designed to be run as a slideshow in Microsoft PowerPoint, with pictures and sound clips. The task was harder than the Grouping experiment, however, as it was designed to be run with older children and to test the assumption that their ability to group speakers by regional accent would be more advanced than that of the younger children. The additional difficulty of this task arose because it involved listening to different speakers (rather than the same speaker, as in the accent-guise design of the previous experiment). In this case, the children had the speakers' different pitches and voice qualities to contend with as well as their different accents. Therefore in order to group speakers by their accent in this task, it was necessary to disregard the additional individual variation. Other aspects that made the task harder included the fact that the children heard multiple accents (Scottish and North East as well as Yorkshire and SSBE) and that there were many more speakers to group in each set. All of these adjustments made the experiment quite challenging. This was in order to test the full potential of the young children's capabilities, before adjusting the design if necessary.

The experiment (or 'game', as it was referred to) was named 'Who sounds like Miss…?' and I carried it out with one boy from York, aged 7.9 years, at his family home. The child was born in York but had non-local parents from the North East (mother) and Warwickshire (father) and regular contact with family members from Derbyshire. The child was asked if he would like to play a game on the computer, listening to some speakers. After agreeing to participate, I sat at the computer with him and gave him instructions on how to play the game. The first part of the game consisted of a training phase in which the child listened to three teachers (Miss A, B and C, represented by different cartoon pictures) reading a short story passage (100 words taken from the 'Chicken Little' passage). Each of the teachers had a different accent: Yorkshire, SSBE and Scottish. As the passage included words featuring the accent distinctions described in the previous section, the aim of the training phase was to familiarise the child with the three different accents and enable them to associate these distinctions with the three different teachers.

For the training phase, the PowerPoint show was started and the child was presented with screen 1 (see Figure 5.2) and told: *'You are going to hear three teachers reading the same story passage. I'd like you to listen carefully to how they sound.'* The PowerPoint was then moved onto screen 2 and he was told: *'First of all, I'd like you to listen to Miss A, by clicking on this speech bubble. Then, when she has finished talking, press the space bar, and Miss B will appear! [screen 3 - indicated by me]. I'd like you to listen to Miss B talk by clicking on her speech bubble. Then, press the space bar again to see Miss C [screen 4 - indicated by me] and click on her speech bubble to hear her talk.'* In order to test whether the child had remembered how each teacher sounded, he was given a task at the end of the familiarisation phase, to match the voice of each speaker to her picture. The instructions were: *'Once you have heard all three teachers talk, press the space bar again and the screen will look like this [screen 5 – indicated by me]. Here you can see the teachers at the top and three smiley faces at the bottom. I'd like you to listen to each of the smiley faces talk and then decide which teacher you think it is. Once you have decided, I'd like you to drag the smiley face into the teacher's box, underneath her' [drag and drop motion indicated by me].* The child was told that he would be given further instructions for the next part of the task after he had completed the first part. I then provided the child with headphones and gave him control of the mouse in order to play the game. I also listened to the experiment through headphones to ensure that the audio clips were working as the child played the game.

After completing the first part of the game, the child was shown screen 6 and told: *'Now, for the next part of the game, there are lots of different speakers here at the bottom of the screen. What I'd like you to do is decide which teacher you think each speaker sounds like the most. Sometimes the speakers will say a whole sentence and sometimes just one word, so listen carefully to how they sound and then decide which teacher they sound like. Click on the number to hear the speaker and then click on the face to drag it into the correct teacher's box. Before you start, click on the speech bubble next to each teacher to hear them speak again and remind you what they sound like.'* The child then clicked on each of the speech bubbles and heard each teacher again, but only reading the first two lines (27 words) of the story passage. He then clicked on each of the smiley face speakers in turn and grouped each of them with one of the teachers. Each of the speakers was heard saying a word or a sentence,

featuring one of the six accent distinctions. There were 20 stimuli altogether, featuring 10 speakers who had Yorkshire, SSBE and Scottish accents.

Once the child had finished grouping all the speakers, he moved on to round 2 of the game which was played in exactly the same way. This time, the child heard three new teachers (Miss D, Miss E and Miss F) all reading the same story passage (the next section of the 'Chicken Little' story). This part of the story also featured distinctions between the accents. Again the teachers all had different accents, but this time there was a Yorkshire speaker, a Scottish speaker and a North East speaker (instead of an SSBE speaker, as in the previous round). The structure of this round of the experiment was the same as the first. After the training phase the child was again tested on whether he could identify each teacher. He then listened to the new speakers and grouped each of them with one of the teachers. For this round, there were 16 stimuli, featuring 7 speakers. Again, each of the speakers was heard saying a word or a sentence, featuring one of the six accent distinctions. For this round, the speakers had Yorkshire, North East and Scottish accents like the teachers.

After completing both rounds, the child was rewarded with a sticker for playing the game.

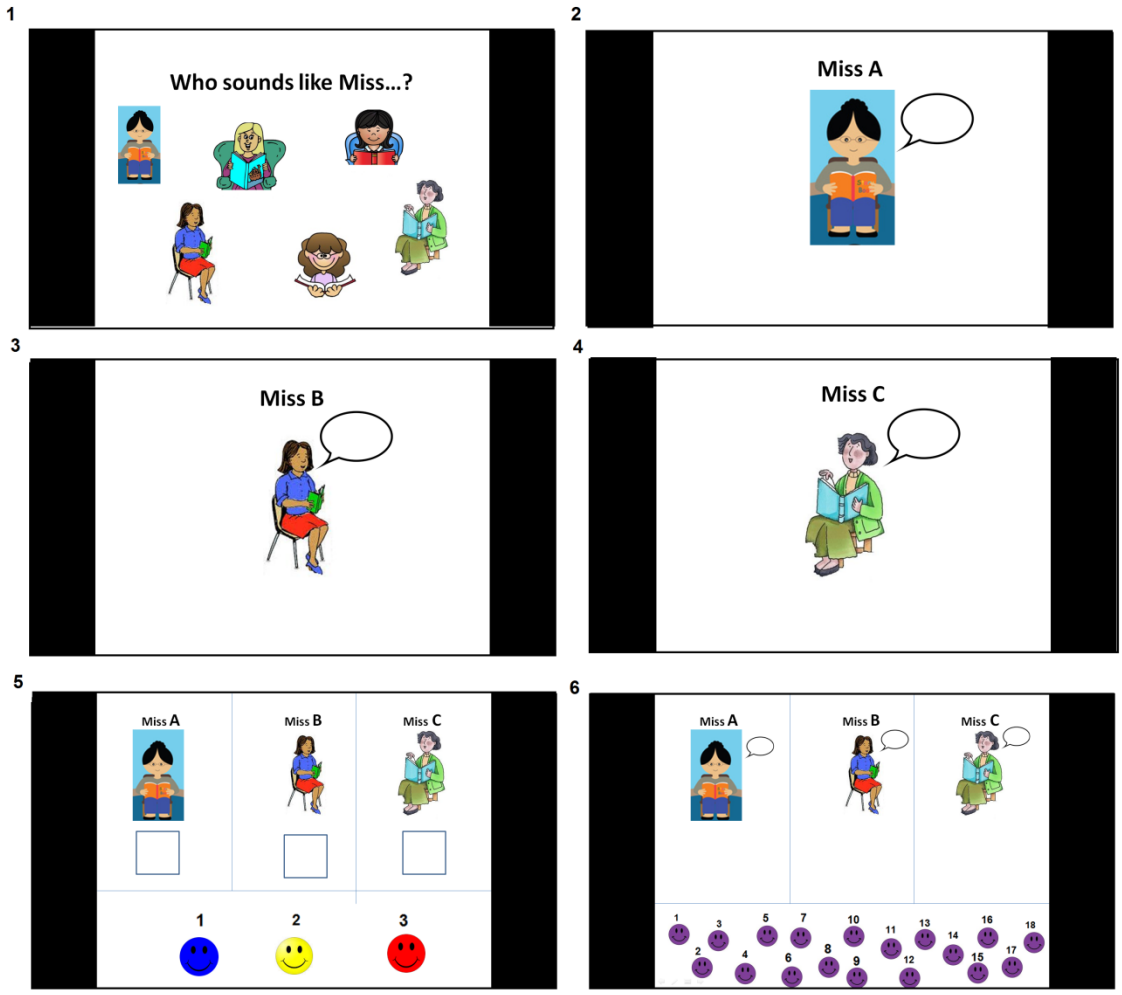



**Figure 5.2: Experiment design 1, screen shots showing the order of the game**

### *5.3.2.2 Results and interpretation*

Altogether, the child who took part in this pilot version of the experiment grouped 18/36 (50%) of the faces with the 'correct' teachers (a correct answer = grouping a Scottish speaker with the Scottish teacher, grouping a Yorkshire speaker with the Yorkshire teacher etc.). While above chance (at 33.3%), this low score indicates that the task involved in the experiment was probably too complex, even for a child of nearly 8 years of age who has had a lot of exposure to family from different regions of the UK. Therefore, this version of the experiment was not run with more children. Instead, it was adjusted in order to address three main problems identified in the design.

First, this experiment relied on the child's ability to hold three voices in short-term memory. Therefore, the cognitive demand involved in this task may have caused the child's

inaccuracies and masked his actual ability to recognise accent differences (see section 2.3.4 for a similar criticism of Floccia et al.'s, 2009 experimental design). Secondly, comments that the child made during the running of the experiment indicated that he was concentrating more on the content of the story, rather than listening to the teachers' voices. This suggested that shorter stimuli might have helped him to focus on how the teachers sound, rather than what they were saying. While it was decided that the teachers' stimuli should be shorter, some of the single-word stimuli of the other speakers were too short to be intelligible. (This was noted by adult participants who also tried out this version of the experiment.) It was therefore decided that only sentence-length stimuli should be used for future versions of the experiment.

Finally, it was decided that there were too many different speakers for the child to group in each set. This made it difficult to determine which speakers the child was comparing each time. This also made the results hard to analyse, as there were too many different variables to consider (the speaker, their accent, the vowel they used and the length of the stimulus).

### 5.3.3 Design 2

#### *5.3.3.1 Experiment design and procedure*

To overcome the difficulties uncovered in the first experiment, the second experiment was constructed in an ABX design (similar to Beck, 2014) in which a listener hears two 'known' speakers (A and B) and then has to decide whether an 'unknown' speaker (X) is more similar to A or B. The experiment was run with two children from York at their family home. These children were sisters, aged 6 and 9. The younger sister was born in York and the older sister had lived in York since the age of 6 months. Their parents were from Wrexham, Wales (mother) and Greater Manchester (father) and the children were reported to have spent a lot of time with their grandparents from Wrexham and Warrington. As with the participant in the first pilot version of the experiment, due to their exposure to speakers from different regions, these participants were tested as a gauge of children's overall understanding of the task. The experiments in the previous two chapters of this thesis have found that children with exposure to variation perform better in accent perception tasks. Therefore, the results from the children taking part in the current version of the pilot experiment should be able to give a good indication of the abilities of the highest performing children overall.

Similarly to the previous design, the experiment was constructed in Microsoft PowerPoint and designed to be run as a game for the children to play, with the same name ('Who sounds like Miss…?').For each set of stimuli, the children heard a speaker A (Miss A), a speaker B (Miss B) and then a speaker X (represented by a smiley face). All speakers were heard reading a sentence. The children then had to decide whether speaker X sounded more like Miss A or Miss B and indicated their decision by dragging the smiley face into either Miss A or Miss B's box (see Figure 5.3).

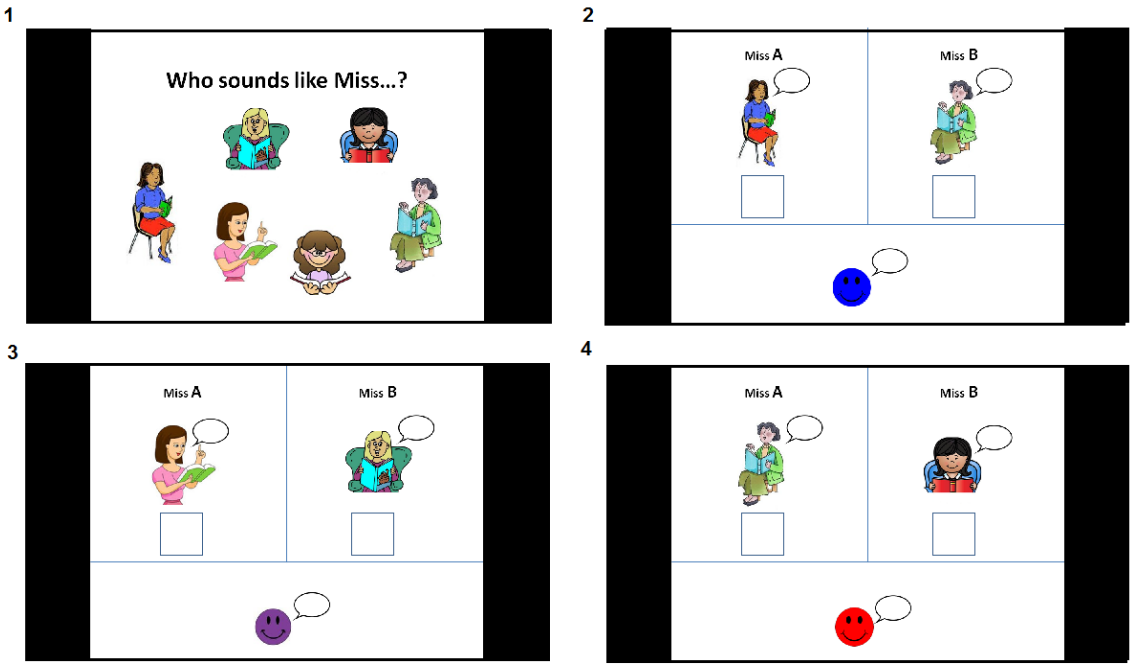



**Figure 5.3: Experiment design 2, screen shots showing the order of the game**

For each set of stimuli, the children heard a new Miss A and Miss B (indicated by the different pictures used to represent the teachers). In each case, Miss A and Miss B had different accents pertaining to one of the three distinctions: Yorkshire and SSBE, Yorkshire and Scottish, Yorkshire and North East. Speaker X had the same accent as either Miss A or Miss B. Altogether, six pictures were used to represent the teachers. These were mixed up so that the children wouldn't be able to associate a particular picture with a particular accent.

The Grouping experiment (chapter 4) found that 4-year-old children were able to match an accent distinction across different words featuring the same vowel. Therefore, the current experiment design assumed this baseline level of ability. For each set of stimuli, each teacher was heard reading a different sentence, featuring a different word. In some cases, both

words featured a vowel from the same lexical set, while in other cases a different lexical set was featured. Speaker X was heard reading a different sentence from both Miss A and Miss B. This sentence also featured a different word. In some cases, the word featured a vowel from the same lexical set as one of the teachers and in other cases the word featured a different lexical set altogether. Therefore there were four different levels of difficulty, depending on (mis)matches between the speakers' vowels. Table 5.2 shows the different difficulty levels, with examples of the vowels (represented by their lexical sets) and words featured at the end of each speaker's sentences. In difficulty level 2, speaker X was heard pronouncing a word featuring the same lexical set as either Miss A or Miss B (the teacher with the matching set was randomized); this was hypothesised to be easier than difficulty level 3 in which the lexical set of the teachers matched but speaker X was heard pronouncing a different lexical set to both of the teachers.

| Difficulty Level | A | B | X | Number of sets |
|---|---|---|---|---|
| **1** | 1<br>GOAT<br>*no* | 1<br>GOAT<br>*road* | 1<br>GOAT<br>*go* | 3 |
| **2** | 1<br>GOAT<br>*go* | 2<br>FACE<br>*day* | 1<br>GOAT<br>*toe* | 5 |
| **3** | 1<br>GOAT<br>*toe* | 1<br>GOAT<br>*road* | 1<br>FACE<br>*day* | 10 |
| **4** | 1<br>GOAT<br>*no* | 2<br>FACE<br>*late* | 3<br>BATH<br>*path* | 9 |

**Table 5.2: Experiment design 2, difficulty levels and examples of the lexical sets**

The experiment included 27 sets of stimuli altogether (9 sets for each accent distinction). A total of 17 speakers from those recorded (described in section 5.3.1) were used for the stimuli. The experiment began with the sets of stimuli from difficulty level 1; only three sets of stimuli were included from this difficulty level as it was expected that the children would find these distinctions relatively easy to make. The rest of the sets of stimuli were randomised so that the accents were mixed up.

### *5.3.3.2 Results and interpretation*

The results reveal that both the children scored around chance level at 50%, and there was no significant difference between the performance of the younger and the older child (see Figure 5.4). Therefore, the children were not able to accurately group speakers according to their accent overall. However, when broken down into the different levels of difficulty (see Figure 5.5), the results show that the children were 100% accurate at grouping speaker X accurately when speaker X's lexical set matched both Miss A and Miss B (difficulty level 1). The children also scored above chance in difficulty level 2, when speaker X's lexical set matched either Miss A or Miss B. However as there were inconsistent numbers of sets of stimuli for each difficult level (see Table 5.2), differences across the difficulty levels should not be considered as definitively reliable. Furthermore, because this experiment design was only piloted on two children and their scores were low overall, these results will not be further analysed. Instead, as we shall see in the next section, the results of this pilot experiment were used to inform the next experiment design which was made easier by focusing on speakers' pronunciations of vowels from the same lexical sets.

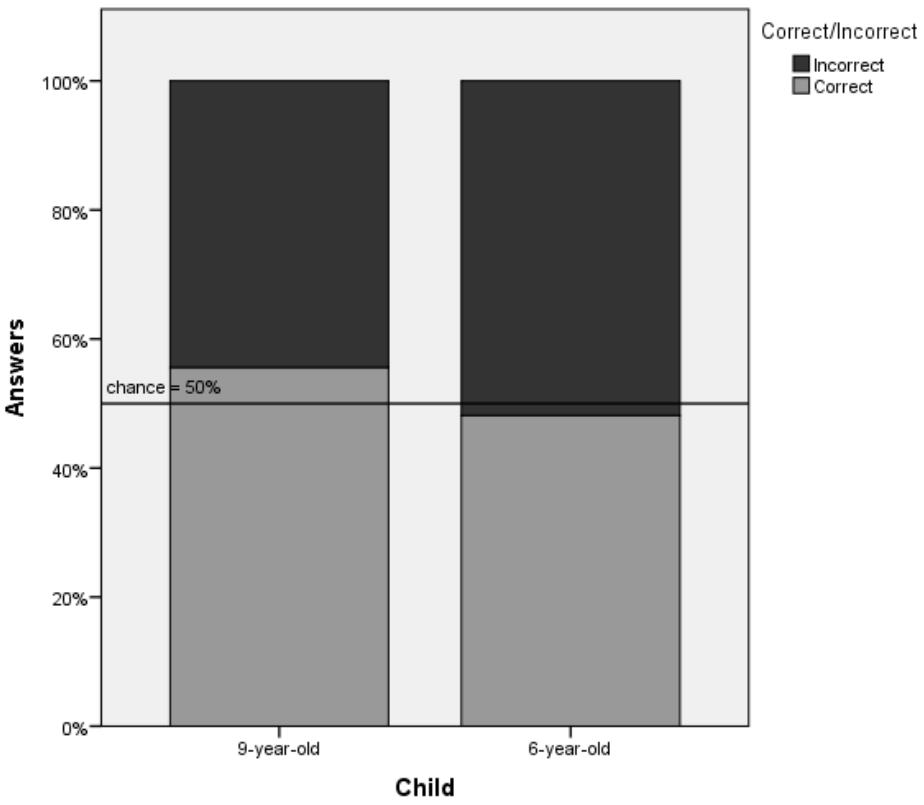



**Figure 5.4: Experiment design 2, children's overall results**

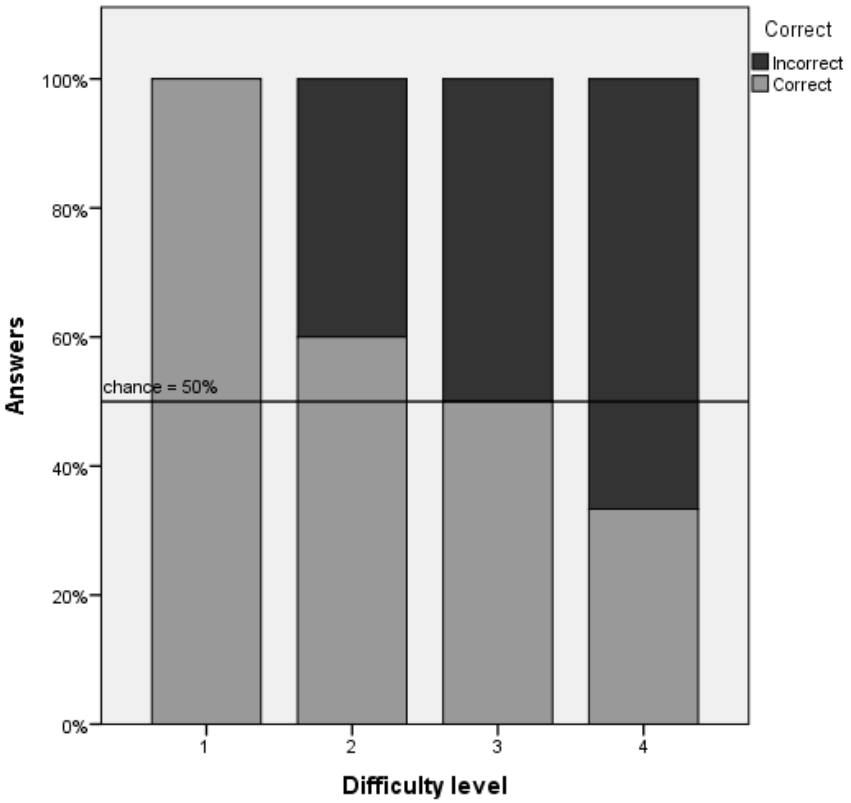



**Figure 5.5: Experiment design 2, results across the different difficulty levels**

### 5.3.4 Design 3

#### *5.3.4.1 Experiment design and procedure*

The results from design 2 indicated that the children found most of the experiment too challenging. Experiment design 3 was therefore made easier, although in a similar ABX construction. For this design it was decided that speakers A and B would always be heard saying the same sentence, featuring the same word and therefore the same lexical set. This would help to ensure that the focus was on the accent difference between the speakers. The experiment was designed with two levels of difficulty. Speaker X was either heard saying a sentence featuring a different word with the same lexical set as speakers A and B (difficulty level 1), or speaker X was heard saying a sentence featuring a different lexical set altogether (difficulty level 2). It was also decided that mixing the stimuli in a random order was too confusing for the children. Therefore the stimuli featuring speakers with the same accents were grouped together. This resulted in three 'rounds' of stimuli, each round featuring a

different accent distinction and two levels of difficulty. These rounds of the experiment were kept in the same order for all of the participants.

In addition, only two lexical set distinctions were included in each round (see Table 5.3). There were 30 sets of stimuli altogether (10 sets in each round). For each set of stimuli, speaker A and speaker B consisted of a randomly ordered pair of speakers with different accents; one speaker had a Yorkshire accent and the other speaker had one of the other accents under consideration. Speaker X's accent then matched either that of speaker A or speaker B. The speakers used in each set of stimuli were not always the same and were presented in various combinations and orders. As an additional way to randomize the order of the stimuli amongst the children, two versions of the experiment were created, with speakers A and B in a different order.

A total of 18 different speakers were used for the stimuli representing the four accents under consideration: 6 Yorkshire, 5 SSBE, 4 Scottish and 2 North Eastern. In addition, one stimulus featuring a non-standard North London speaker was included in round 1 and two stimuli featuring an American speaker were included in round 2. The North London speaker was included in order to find out if she would form a robust comparison with an SSBE speaker; while they both pronounced the GOAT vowel with a diphthong, the North London speaker used a much wider, fronted diphthong in her pronunciation ([ʌʏ] rather than SSBE [əʊ]); a pronunciation typical of working class London speech. The American speaker was included in order to test whether the children would group the American and Scottish speakers together on the basis of their rhotic pronunciation of the NURSE vowel.

| Round | Speakers | Difficulty Level | A | B | X |
|---|---|---|---|---|---|
| 1 | Yorkshire and SSBE | 1 | 1<br>GOAT<br>*toe* | 1<br>GOAT<br>*toe* | 1<br>GOAT<br>*go* |
| | | 2 | 1<br>GOAT<br>*no* | 1<br>GOAT<br>*no* | 2<br>FACE<br>*wait* |
| 2 | Yorkshire and Scottish | 1 | 1<br>NURSE<br>*hurt* | 1<br>NURSE<br>*hurt* | 1<br>NURSE<br>*burst* |
| | | 2 | 1<br>NURSE<br>*word* | 1<br>NURSE<br>*word* | 2<br>lettER<br>*sugar* |
| 3 | Yorkshire and North East | 1 | 1<br>GOAT<br>*goat* | 1<br>GOAT<br>*goat* | 1<br>GOAT<br>*no* |
| | | 2 | 1<br>GOAT<br>*go* | 1<br>GOAT<br>*go* | 2<br>FACE<br>*late* |

**Table 5.3: Experiment design 3, the accents and features used in each round**

This version of the experiment was run with 19 children, from two schools (10 from one and 9 from another). These schools were recruited via my personal contacts with the teachers. Both schools were based in Yorkshire, one in Pontefract (West Yorkshire) and one in York (North Yorkshire). The children were all in years 3-4 at school and their ages ranged from 7.2 to 8.9 years. All but two of the children were born in Yorkshire; one child was originally from Birmingham and another originally from London.

Background information was also collected from children's parents (see Appendix 1). This information asked about the regional background of the child, their parents and whether the child had regular contact with anyone else from another region. It also asked about the parents' education and occupation.

The experiment was run in a similar way to design 2. However, for this version the children were told that they were going to play three rounds of a game, and a title screen was used to indicate when the child was beginning each round (see Figure 5.6). This also helped to break the game up rather than presenting it as one long, continuous task.

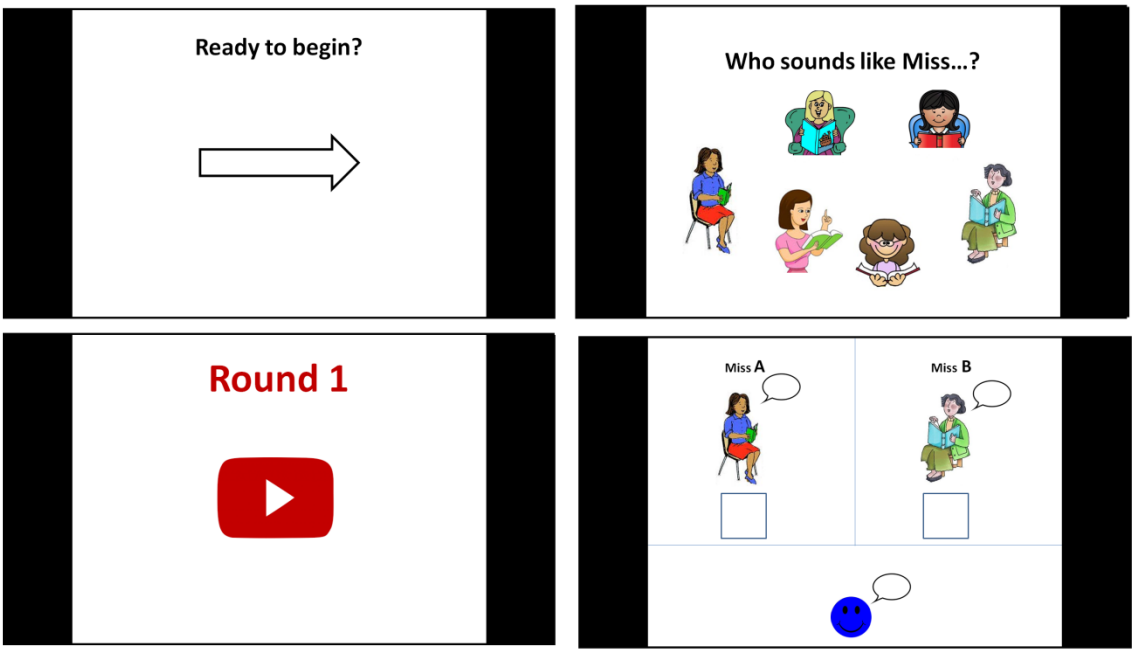



**Figure 5.6: Experiment design 3, screen shots of the game**

### *5.3.4.2 Results and interpretation*

The background information collected from the children's parents revealed that there were prevailing differences between the parents of children from the different schools which could impact their results. All of the children from school 1 had at least one Yorkshire parent; in fact 8/10 of the children had both parents from the local area. On the other hand, 4/9 children from school 2 had both parents from outside of Yorkshire. Therefore the children from school 2 with non-local parents are more likely to have been exposed to more linguistic variation at home. In line with the results from the experiments in chapters 3 and 4, this is predicted to result in a better performance by these children. Furthermore, there is a difference between the educational levels of the children's parents from each of the schools, which indicates a socioeconomic class division. While 8/10 children from school 1 had parents who left education after high school, 4/9 children from school 2 had parents with a postgraduate education, and a further 4 of these children's parents had an undergraduate education. This indicates that the children from school 2 come from a higher social class background which in turn suggests that these children will have had more exposure to standard forms at home (see section 5.6.4 for a further discussion of the link between class and exposure to standard forms). Therefore these children are more likely to

be more familiar with SSBE forms and better able to differentiate between SSBE and Yorkshire accent differences.

Due to these differences between the children's backgrounds at each of the schools, the results for each school will be presented individually. This will enable a comparison between the performances of the children across the social dimensions that these schools represent.

Table 5.4 shows the overall results. No significant difference was found between the two versions of the experiment (with Miss A and B in a different order) and therefore results for both versions are presented together. A significant difference was found between the overall performance of the two schools ($t = 2.96$, $df = 217$, two-tailed-$p = 0.009$); only the children from school 2 scored above chance level overall. This difference in performance between the two schools could be related to the different social backgrounds of the children (as described above). The children at school 2 came from higher class and more regionally diverse backgrounds. They were therefore more likely to be exposed both to more standard forms and more regional variation, which may have contributed to their better performance in the experiment.

| School | Overall Mean % (SD) |
|---|---|
| **Both 1&2** | 53.16% (8.14) |
| **1 (Pontefract)** | 48.67% (6.89) |
| **2 (York)** | 58.15% (7.1) |

**Table 5.4: Experiment design 3, overall results for both schools**

Figure 5.7 shows a box plot of the performance of each school for each of the rounds. In both schools, children performed best in round 1. The children from school 2 scored particularly well in this round with a mean score of 71%. This indicates that the children found it easier to group speakers by a Yorkshire/SSBE accent distinction than a Yorkshire/Scottish or a Yorkshire/North East distinction.

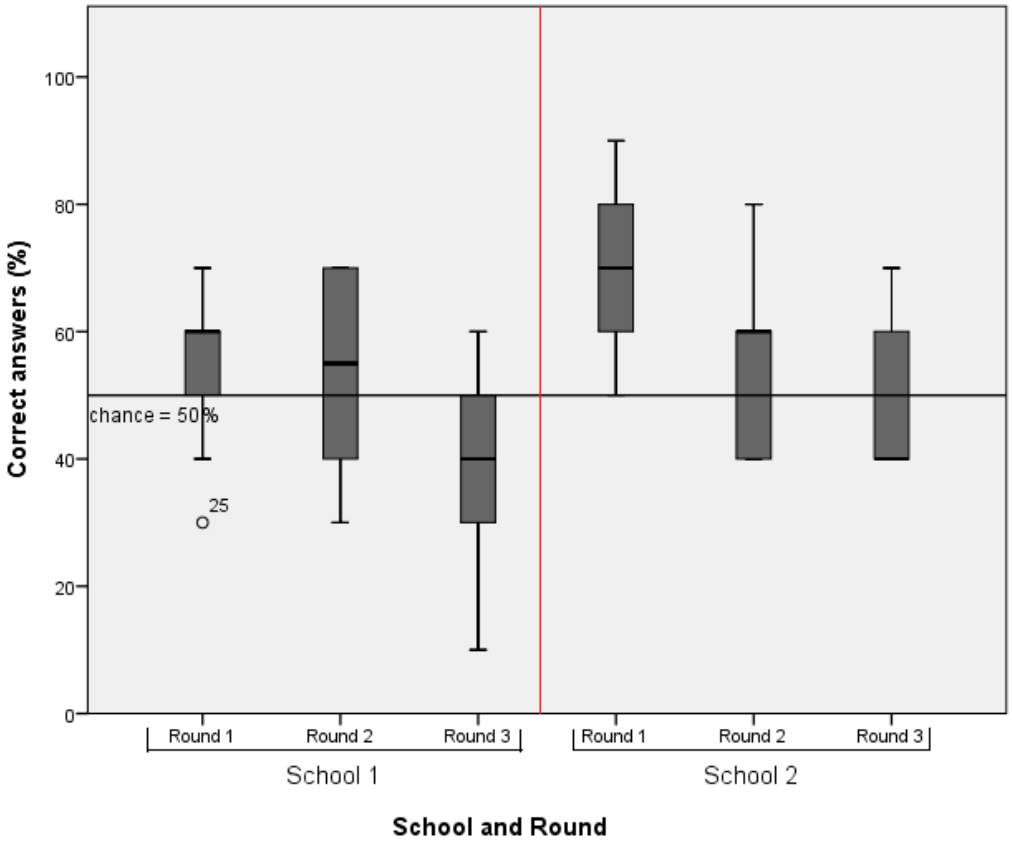



**Figure 5.7: Experiment design 3, results of each round for each school**

No significant difference was found between the performances of the two sexes. Furthermore, as Figure 5.8 shows, no significant difference was found between the performances in difficulty levels 1 and 2 for each round. Thus the children did not perform better when speaker X's lexical set matched the lexical set of speakers A and B.

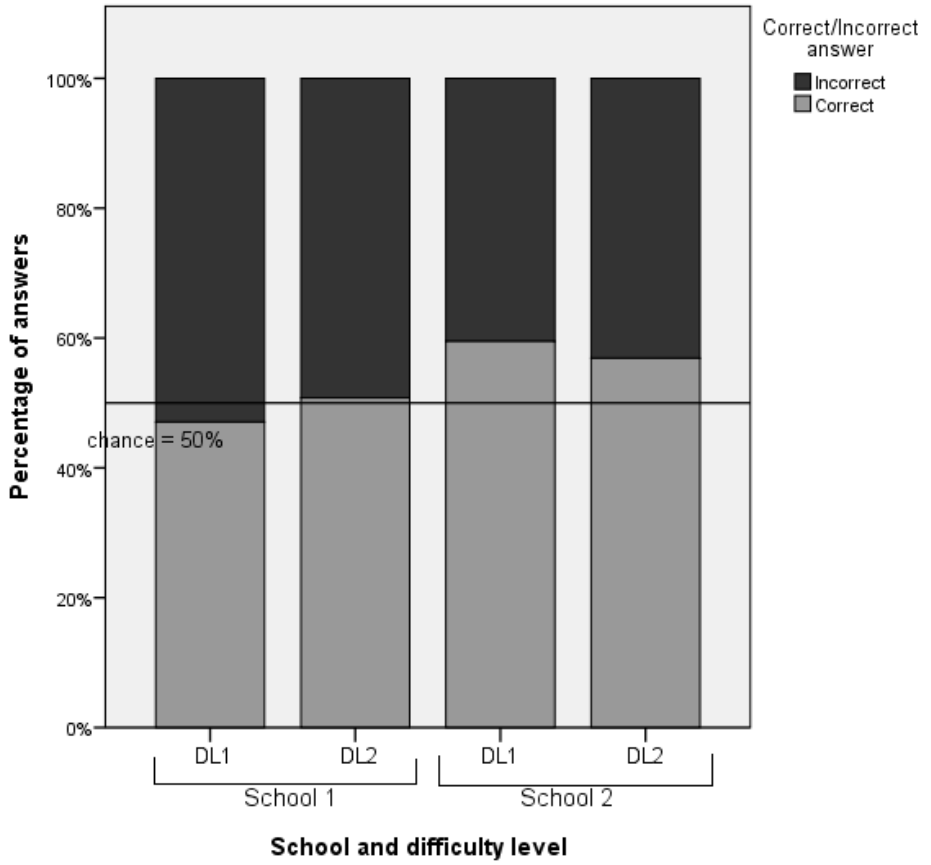



**Figure 5.8: Experiment design 3, results across the difficulty levels for each school**

The children's responses to individual stimuli were further investigated in order to determine whether any of the stimuli proved particularly difficult. Figure 5.9 shows the overall percentage of correct answers for each set of stimuli, with all 19 children combined. For some of the stimuli sets, most of the children grouped the speaker incorrectly. For example, for stimulus set 11 (see the blue bar in Figure 5.9), over 90% of the children grouped the speaker incorrectly. This was the first set of stimuli in round 2 of the experiment and featured a Yorkshire/Scottish distinction of the NURSE vowel. The low performance for this particular set could be due to the fact that one of the speakers has a much lower pitched voice than the other two. Therefore the children may have grouped the other two speakers together based on their more similar pitch. This explanation is supported by results from the Identification experiment (in chapter 3), which showed that a speaker's pitch can be an important criterion for identifying a speaker. Furthermore, the children were only told to 'decide whether the speaker sounds more like Miss A or more like Miss B'. Therefore it is difficult to know which criteria they were using when grouping each of the speakers.

Despite a couple of low performances for individual stimuli, the results are fairly uniform overall: the mean score is 53% correct, with a high and low of 67% and 40%. Therefore, although these scores are relatively low overall, they are consistent, and indicate that the children faced the same difficulties with the experiment. The children's performance in the stimulus sets featuring the American and North London speakers are consistent with the rest of the results (see the red bars in Figure 5.9). They performed above chance level when grouping an SSBE speaker with another SSBE speaker, rather than the non-standard London speaker (stimulus set 6). They also performed above chance level when grouping the rhotic American speaker with the rhotic Scottish speaker, rather than the non-rhotic Yorkshire speaker (stimulus set 19). They performed around chance level in the alternate situation however (stimulus set 15), when grouping a rhotic Scottish speaker with the rhotic American speaker rather than the non-rhotic Yorkshire speaker. This difference highlights a shortcoming of the experiment design, as discussed below.

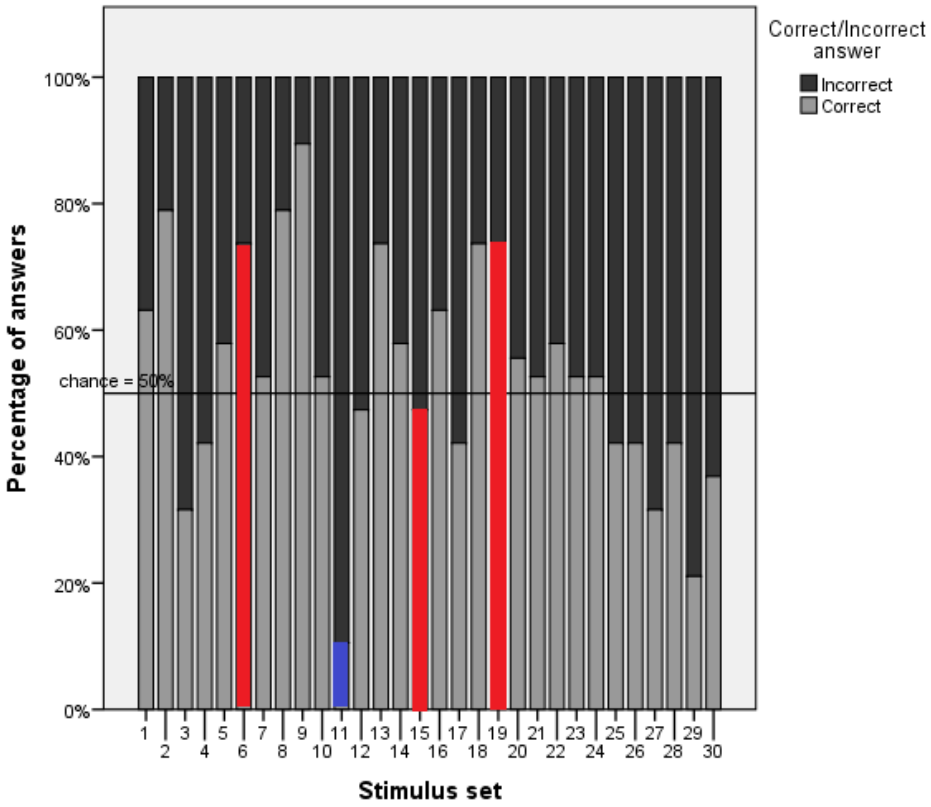



**Figure 5.9: Experiment design 3, percentage of correct answers for each stimulus set**

The reliability of the results is limited because only one speaker (X) is grouped with another speaker (A or B) in each stimulus set. Therefore, it is difficult to conclude that the speaker in each set was grouped solely on the basis of accent; there are no other speakers with the same accent who are also grouped with speaker A or B, alongside speaker X to support this analysis. Strictly speaking the ABX design is a discrimination task rather than a grouping

task. In order to investigate children's categorisation of accents more fully, a task design based on grouping more than one speaker is more suitable. Having a range of speakers to group in each stimulus set will help to avoid the results being affected by the children grouping these speakers on other aspects of their voice, such as their pitch or voice quality. The criteria that the child listeners use to group speakers can never be definitely known; this is true of studies with adults as well (such as those described in section 2.3.3). Furthermore, discrimination between unfamiliar speakers is likely to rely on the processing of a range of features of the voice, not just one (as described in section 2.5.1). However, if groups of speakers created by the children are found to correspond with accent distinctions, we can interpret this as an indication of the important role of the speakers' accents for these groupings. Due to these considerations, the final experiment is based on a grouping design, as described in the following section.

## 5.4 Final design: methodology

The final experiment design was constructed in consideration of the deficiencies discovered in the three pilot designs. This final version of the Second grouping experiment was therefore simpler in key respects than previous designs of the experiment, featuring fewer lexical set distinctions and fewer levels of difficulty. This was in order to ensure that the experiment captured the children's best ability in this task and that their performance was not impeded by the task design.

Additionally, in this final experiment design, samples of the children's productions were recorded in order to investigate whether the children's own productions might affect their perception of vowels indicative of a regional distinction. As emphasised by Vihman (2014), the link between children's perception and production is an important line of investigation in accounting for children's phonological acquisition, As children start to talk, their own output (vocal productions) form part of their input (and therefore their perceptual experience). In particular, studies of infant word-learning have found that infants' listening to speech is affected by their own vocal productions, showing early evidence of the perception-production link (cf. DePaolis et al., 2011; 2013; Majorano et al., 2014). With regards to the current study, the children's productions were recorded as an additional indication of their exposure to regional variation from their own speech.

The final version of the Second grouping experiment addresses the research questions described in section 5.2.

### 5.4.1 Participants

Thirty four primary-school children (22 girls and 12 boys), covering an age range of between 5.6 and 9.8 years old took part in the final experiment. Twenty eight of these children were born in York. Out of the remaining six children, three moved to York under the age of six months and the other three moved from elsewhere (two from London who moved to York at the age of 3.5 and 5.5, and one from Malaysia who moved to York at the age of 6). The regional backgrounds of the children are considered alongside their parents' regional backgrounds in the analysis of the results. Due to difficulty in recruiting schools to participate, these children were unevenly distributed between three schools in York. Sixteen children were recruited from school 1, seven from school 2 and eleven from school 3. The three schools were quite different in terms of their pupil populations and their overall academic performance. Information from their most recent Ofsted reports indicates that schools 1 and 3 have a good/outstanding overall performance and a lower than average proportion of pupils eligible for the pupil premium (additional funding provided by the government for disadvantaged pupils, often from lower class social backgrounds). In contrast, school 2 is reported as 'requiring improvement' and has an above average proportion of pupils eligible for the pupil premium. This indicates that school 2 has a higher proportion of children from lower social class backgrounds than schools 1 and 3. The social class backgrounds of children from the different schools will be considered in the analysis of the results from this experiment, although measured through their parents' education instead (see section 5.4.7).

Consent forms and background information sheets were collected from children's parents (see Appendix 1). This information asked about the regional background of the child and their parents, and whether the child had regular contact with anyone else from another region. It also asked about the parents' education and occupation and the children's favourite television programmes/films (see sections 5.4.6 and 5.4.7 for a description of what this information was used for).

### 5.4.2 Experiment design and stimuli

The experiment was constructed in a grouping design but still based on hearing two teachers (Miss A and Miss B) and grouping new speakers with the teachers. In order to simplify the task from experiment design 3, the number of speakers to group was limited to five per set of stimuli. As in experiment design 3 there were three 'rounds' to the game, each featuring a different accent distinction. Again there were two levels of difficulty within each round. However, as experiment design 3 had still proved to be quite difficult for many of the children, the difficulty levels were made easier. For difficulty level 1, the five grouping speakers, as well as Miss A and Miss B, were all heard reading the same sentence, featuring the same lexical set. For difficulty level 2, whereas the teachers were all heard reading the same sentence, the 5 grouping speakers were heard reading another sentence but featuring the same lexical set. These two difficulty levels correspond to the 'same word' and 'same phoneme' difficulty levels used in the Grouping experiment.

Each round followed the same order and consisted of two difficulty level 1 stimulus sets first, followed by one difficulty level 2 stimulus set. Each round focused on two variables that differentiate the accents. For each stimulus set, Miss A and Miss B had different accents and the other speakers had an accent matching either Miss A or Miss B. The orders of the accents heard were randomised within each set. In each round, a total of four Yorkshire speakers and four speakers from the other accent (SSBE, Scottish or North East) were heard (apart from round 3, in which five Yorkshire speakers were heard). Therefore a total of 17 different speakers were heard altogether. Table 5.5 presents a summary of the experiment design (named the Second grouping experiment), along with the lexical sets and words featured in each stimulus set.

| Round | Speakers | Difficulty Level | Miss A | Miss B | Grouping speakers |
|---|---|---|---|---|---|
| 1 | Yorkshire and SSBE | 1 | 1<br>GOAT<br>*no* | 1<br>GOAT<br>*no* | 1<br>GOAT<br>*no* |
| | | 1 | 1<br>FACE<br>*shade* | 1<br>FACE<br>*shade* | 1<br>FACE<br>*shade* |
| | | 2 | 1<br>GOAT<br>*go* | 1<br>GOAT<br>*go* | 2<br>GOAT<br>*toe* |
| 2 | Yorkshire and Scottish | 1 | 1<br>NURSE<br>*church* | 1<br>NURSE<br>*church* | 1<br>NURSE<br>*church* |
| | | 1 | 1<br>lettER<br>*sugar* | 1<br>lettER<br>*sugar* | 1<br>lettER<br>*sugar* |
| | | 2 | 1<br>NURSE<br>*word* | 1<br>NURSE<br>*word* | 2<br>NURSE<br>*burst* |
| 3 | Yorkshire and North East | 1 | 1<br>GOAT<br>*goat* | 1<br>GOAT<br>*goat* | 1<br>GOAT<br>*goat* |
| | | 1 | 1<br>FACE<br>*late* | 1<br>FACE<br>*late* | 1<br>FACE<br>*late* |
| | | 2 | 1<br>GOAT<br>*toe* | 1<br>GOAT<br>*toe* | 2<br>GOAT<br>*go* |

**Table 5.5: Design of the Second grouping experiment, with the lexical sets and words featured in each set of stimuli**

### 5.4.3 Experimental procedure

I ran the experiment with each child individually, either in a quiet space at their school or at the child's home. The children were invited to play a game on the computer with me. Before taking part in the listening game (again entitled *'Who sounds like Miss…?*'), they were recorded describing some images on screen. The images were chosen to represent objects featuring the BATH, STRUT, FACE and GOAT vowels (see screen 1 in Figure 5.10) to elicit the children's own production of these regionally variable vowels.

The children were told: *"For this game, you are going to listen to some speakers but before you start I would like you to have a look at the screen and describe the pictures you see.*' The children were then recorded describing the pictures using a Zoom H4n recorder (set to record at a 16 bit 44.1 kHz sampling rate). Once the children had described the pictures

eliciting their GOAT, FACE, BATH and STRUT vowels, they were told: *"Great, now we know that your eyes are working! Let's move on to the listening part of the game."* I then moved the PowerPoint show on to slide 2 (in Figure 5.10) and gave further instructions: *"The game is called 'Who sounds like Miss…?' and there are three rounds to the game. Let's start round 1 [slide 3 displayed] and I'll explain what you need to do. On the screen you see two teachers [slide 4 displayed], Miss A and Miss B. What I'd like you to do is click on Miss A and listen to her read a sentence. Then click on Miss B and listen to her read a sentence. Once you have heard both Miss A and Miss B speak, I'd like you to listen to these other speakers at the top of the screen, represented by smiley faces. To hear the smiley face speak, you need to click on the speech bubble [indicated by me]. After listening to the smiley face speak, I'd like you to decide whether you think they sound like they come from the same place as Miss A or the same place as Miss B. Once you have decided, you can click on the smiley face to drag it either into Miss A's box or into Miss B's box [indicated by me]. Then, I'd like you to do the same with each of the other smiley faces. So you click on the speech bubble to hear each smiley face and then click on the face to drag it either into Miss A or Miss B's box. If you forget what Miss A or Miss B sound like you can always click on them again to listen to them again."* Due to the results from design 3, the instructions for this final version were slightly altered. The children were asked to group the speakers according to 'the place' they came from rather than if they 'sounded similar'. This was to try and avoid the children making their decision based on other non-regional aspects of the speakers' voices, such as pitch.

After the instructions were given and the child confirmed that they understood the task, the child was then provided with headphones and given control of the mouse to start the game. I also wore headphones in order to monitor the game and check that all the audio clips worked. After the three stimulus sets of round 1, the child was presented with a screen indicating that they were about to start round 2, and they were asked if they would like to continue playing the game. Rounds 2 and 3 were played in exactly the same way as round 1. At the end of the game, children were rewarded with a sticker.

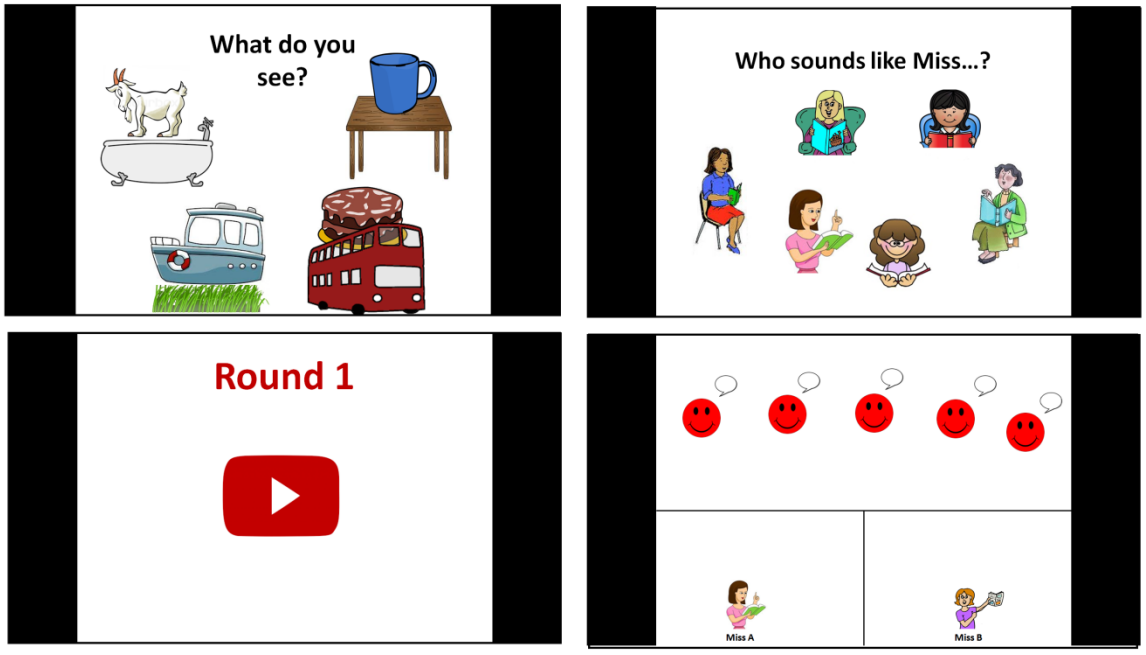



**Figure 5.10: Second grouping experiment, screen shots of the game**

### 5.4.4 Running of the experiment

All 34 children completed the whole experiment. The children showed a good understanding of the task and most of them finished it without needing to ask any questions. There were a few occasions where the children seemed puzzled at the start of the first set of difficulty level 2 sentences (when the sentences heard by the teachers and the smiley faces were different) and looked towards me for help. In these cases, the children were reminded that they were listening to where the speakers came from based on how they sounded and therefore it did not matter if they were saying something different. This satisfied their confusion and they carried on playing the game. The only other problem encountered during the playing of the game was the ability of some of the younger children to control the mouse. Although all of the children knew how the mouse worked, some of the children with smaller hands struggled to move the mouse very quickly. However, all of the children persevered and were able to control the mouse after some practice. In the instances in which the children took a longer time to manoeuvre the curser, they were reminded that they could listen to the teachers speaking again, in case they had forgotten how they sounded.

### 5.4.5 The children's productions

As described in the procedure section, the children's productions were recorded with the intention of analysing this data alongside the results of the perceptual experiment. However, as the majority of the children pronounced the vowels in the same way, with northern pronunciations of the BATH and STRUT vowels ([a] and [ʊ]) and diphthongal pronunciations of the GOAT and FACE vowels ([əʊ] and [eɪ] - reflecting a current change in progress in York, described in section 2.1.3), this data was not further analysed. Moreover, in order to draw conclusions from the results based on a production-perception link, a detailed analysis of the children's own productions would need to take into account their productions in spontaneous speech in their home environment. While further analysis of production data is outside the main theoretical considerations of the current thesis, the perception-production link is an important line of investigation and worthy of future study in relation to children's perception of regional accents.

### 5.4.6 Children's media exposure

Information about the children's favourite television programmes/films was collected in order to consider how children's exposure to the media might contribute to their exposure to variation overall. The children's parents reported a wide variety of shows that their children watched; 73 separate programmes were listed altogether as well as general references to channels and their programmes such as *'CBBC', 'CBeebies', 'CITV'* and *'The Disney Channel'* . This large number of different programmes meant that it was not possible to analyse this data alongside the children's results systematically. However, the range of shows reported and the kinds of variation they include indicate that, overall, the children are exposed to a variety of accents. American TV programmes and films were reported as the most popular amongst the children overall, which suggests that most of the children will have had a lot of exposure to American accents. Furthermore, some of the parents specifically commented on their children watching *'American programmes'* in general and one parent even made a point of saying that their child *'puts on an American accent when playing with their toys'.*

However, there are also indications of the children hearing a variety of other accents, for example on the CBBC/CBeebies television channels on which there are presenters and

characters in programmes with a very wide variety of accents including Irish, Scottish and Manchester varieties. As explained in the literature review (section 2.3.1), the range of accents now heard in the British broadcast media has heightened listeners' exposure to variation outside their own region. While this observation has generally only been reported with regard to adolescent and adult listeners, it appears to be true for children as well. Although beyond the scope of the current project, further systematic investigation into the connection between children's exposure to variation in the media and their perception of this variation would be beneficial in adding to an account of children's developing accent awareness.

### 5.4.7 Children's background information

Background information was collected from the children's parents in order to interpret the statistical effects of these independent variables on the children's results (addressing research question 5 – see section 5.2). Five variables based on the children's backgrounds were used in the statistical analysis. In line with the previous two experiments, three of these variables were: the children's age (entered into the statistical models as a continuous variable), the children's sex (male/female), and whether they had any Yorkshire parents (yes/no).

The other two variables were: the parents' highest level of education, and whether the children had regular exposure to speakers of other UK regional varieties. The parents' education level was used as a proxy for indicating the children's social class background. As described in section 5.4.1, the children were recruited from three schools which represented children from different social class backgrounds. Due to the uneven distribution of children amongst the different schools however, it was decided that the children's parents' level of education would form the basis for a fairer comparison of their social class. Furthermore, parental education as an indication of social class has been used by other studies looking at the different features of child-directed speech amongst higher and lower class parents (Huttenlocher et al., 2007; Rowe, 2008;), which has implications for the children's exposure to variation (see section 5.6.4.). For the purposes of the statistical analysis, a binary distinction was made on the basis of whether the children had a parent with a postgraduate degree (yes/no). This measurement formed a way of representing the differences between the schools, as well as creating the most even numerical division of the

sample. While school 1 had 50% of pupils with postgraduate parents, school 2 had 82% and school 3 just 29%.

The children's exposure to other regional varieties was measured based on their parents' answer to question (8) on the background information sheet .This question asked the parents to list any people from a different region who their child frequently spends time with. The answers to this question revealed that, overall, the children had experience with a range of speakers from a variety of regions. Therefore it was not possible to enter the exposure to individual accents into the statistical model. Instead, this information was entered into the model as a binary variable and the children were split into a two-way distinction according to whether they had regular exposure to non-Yorkshire speakers of English (yes/no). The children's exposure to the particular accent of each round will be also be further explored in the separate analysis of each round.

## 5.5 Results

In this results section, the overall results are outlined (section 5.5.1), before being explored statistically (section 5.5.2 onwards) in line with the research questions (see section 5.2).

### 5.5.1 Overall results

Table 5.6 presents each child's overall results as well as their results for each round of the experiment (presented in rank order of the children's overall percentage correct).

| Child | Age | Sex | School | Round 1 (SSBE) (% correct) | Round 2 (Scottish) (% correct) | Round 3 (N East) (% correct) | Overall (% correct) |
|---|---|---|---|---|---|---|---|
| 1 | 9.08 | F | 1 | 86.67 | 93.33 | 73.33 | 84.44 |
| 2 | 6.37 | F | 1 | 100.00 | 60.00 | 80.00 | 80.00 |
| 3 | 6.18 | F | 3 | 100.00 | 66.67 | 66.67 | 77.78 |
| 4 | 8.7 | F | 1 | 66.67 | 86.67 | 73.33 | 75.56 |
| 5 | 6.37 | F | 1 | 100.00 | 60.00 | 60.00 | 73.33 |
| 6 | 7.18 | F | 1 | 73.33 | 80.00 | 66.67 | 73.33 |
| 7 | 8.55 | M | 1 | 93.33 | 86.67 | 40.00 | 73.33 |
| 8 | 5.9 | F | 3 | 86.67 | 60.00 | 73.33 | 73.33 |
| 9 | 8.13 | F | 3 | 100.00 | 40.00 | 73.33 | 71.11 |
| 10 | ? | M | 1 | 93.33 | 33.33 | 80.00 | 68.89 |
| 11 | 8.31 | M | 1 | 93.33 | 66.67 | 46.67 | 68.89 |
| 12 | 5.66 | F | 3 | 60.00 | 80.00 | 66.67 | 68.89 |
| 13 | 6.12 | F | 3 | 80.00 | 46.67 | 80.00 | 68.89 |
| 14 | 6.1 | F | 3 | 100.00 | 53.33 | 53.33 | 68.89 |
| 15 | 9.1 | F | 1 | 73.33 | 53.33 | 73.33 | 66.67 |
| 16 | 6.49 | F | 2 | 86.67 | 46.67 | 66.67 | 66.67 |
| 17 | 6.82 | M | 2 | 60.00 | 66.67 | 73.33 | 66.67 |
| 18 | 7.67 | F | 3 | 73.33 | 80.00 | 46.67 | 66.67 |
| 19 | 6.19 | F | 3 | 100.00 | 46.67 | 53.33 | 66.67 |
| 20 | 6.09 | M | 3 | 93.33 | 60.00 | 46.67 | 66.67 |
| 21 | 6.91 | M | 2 | 53.33 | 73.33 | 60.00 | 62.22 |
| 22 | 6.33 | F | 1 | 100.00 | 40.00 | 40.00 | 60.00 |
| 23 | 9.78 | M | 1 | 100.00 | 40.00 | 40.00 | 60.00 |
| 24 | 8.59 | F | 2 | 60.00 | 66.67 | 53.33 | 60.00 |
| 25 | 8.78 | M | 1 | 86.67 | 33.33 | 53.33 | 57.78 |
| 26 | 6.72 | F | 2 | 66.67 | 46.67 | 60.00 | 57.78 |
| 27 | 6 | M | 3 | 60.00 | 40.00 | 73.33 | 57.78 |
| 28 | 5.63 | F | 3 | 66.67 | 60.00 | 46.67 | 57.78 |
| 29 | 9.01 | F | 2 | 40.00 | 66.67 | 60.00 | 55.56 |
| 30 | 8.83 | F | 1 | 66.67 | 46.67 | 40.00 | 51.11 |
| 31 | 6.95 | M | 1 | 66.67 | 46.67 | 33.33 | 48.89 |
| 32 | 7.09 | M | 1 | 53.33 | 26.67 | 60.00 | 46.67 |
| 33 | 7.22 | F | 2 | 40.00 | 26.67 | 60.00 | 42.22 |
| 34 | 6.85 | M | 1 | 53.33 | 20.00 | 26.67 | 33.33 |
| **Average** | **7.26** | **-** | **-** | **76.67** | **56.67** | **60.00** | **66.67** |

**Table 5.6: Second grouping experiment, results for each child (in rank order based on overall score), F=Female, M=Male**

Figure 5.11 shows the mean percentage correct for the results overall as well as for each round individually. Overall, the children performed above chance level (50%). Only four

children scored below chance. The children scored best in round 1, with an average of 77% correct and 8 children scoring 100% in this round.

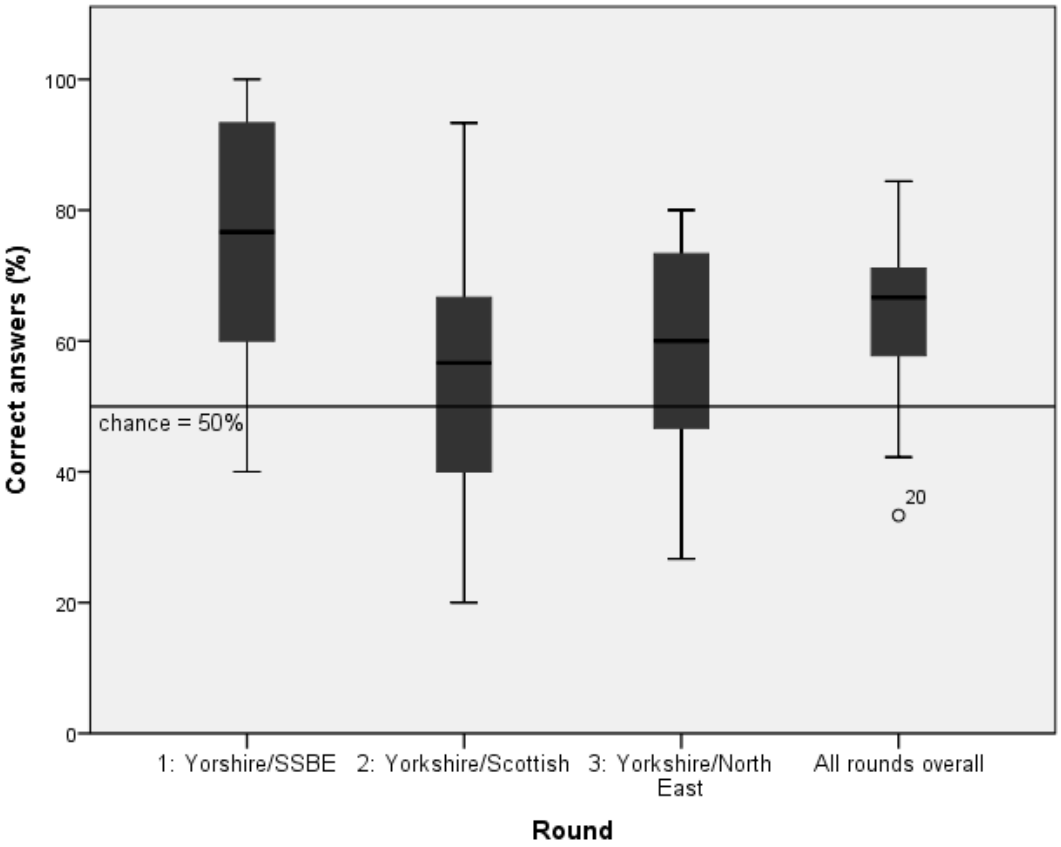



**Figure 5.11: Second grouping experiment, mean correct results for each round**

The majority of children (20/34) scored better in round 1 than rounds 2 or 3, another indication that this was, as predicted, the easiest round. In order to decipher whether there was a difference in these children's performance between rounds 2 and 3, Figure 5.12 shows the distribution of these 20 children's scores across the three rounds. Out of these children, six scored second best in round 2, nine scored second best in round 3 and five had the same score for rounds 2 and 3. This indicates that children's scores were quite varied between rounds 2 and 3 and that neither round was easier overall. The results of children who scored particularly well in rounds 2 and 3 will be investigated in more detail when the results for each round are analysed individually.

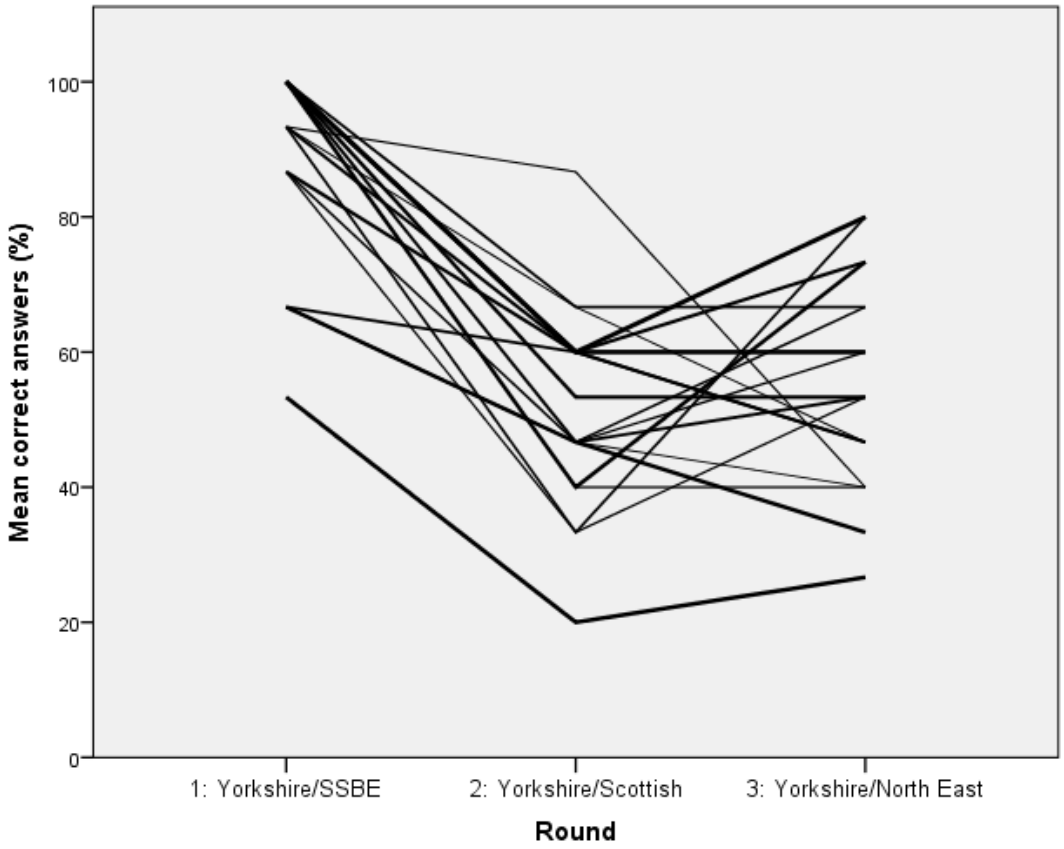



**Figure 5.12: Performances across the rounds for those who performed best in round 1**

### 5.5.2 Statistical analysis overall

Analogous to the experiments in the previous chapters, binary mixed effects logistic models were run in R (R Core Team, 2013) using a stepwise backward regression method. An initial model was run on the overall results in order to investigate whether the children's performance in each round was significantly different. Therefore, round (1-3) was entered into the model as a continuous variable alongside the binary variable of difficulty level (1/2) and the five background variables described above: age (continuous), sex (male/female), Yorkshire parent (yes/no), parent with postgraduate education (yes/no) and exposure to other regional varieties (yes/no). This model was used to explore the effects of the independent variables and in particular to test for a significant difference between the rounds. Therefore interactions were not included in this model. For the difficulty level predictor, the default was set to 1; for the sex predictor, the default was set to female; for the parents' education predictor, the default was set to no postgraduate education; for the

Yorkshire parent predictor, the default was set to no Yorkshire parents; finally for the exposure predictor, the default was set to no regular exposure to speakers of other regional varieties. The same defaults were set for the all of the following statistical models.

As Table 5.7 shows, the variable 'round' was a highly significant predictor of whether the children scored a correct answer and the by far the biggest predictor overall. The negative figure in the estimate column shows that in the higher rounds (2 and 3), the children were much less likely to score a correct answer. Due to the large effect that round had on the results, separate statistical models were run on each round individually in order to investigate the effects of the other variables on the children's performance in each round.

| | **Estimate** | **Std. Error** | **z value** | **Pr(>\|z\|)** | **Sig.** |
|---|---|---|---|---|---|
| (Intercept) | 0.86 | 0.54 | 1.59 | 0.11 | |
| ROUND (1-3) | -0.44 | 0.07 | -6.33 | 2.4e-10 | *** |
| DIFFICULTY LEVEL (2) | -0.34 | 0.12 | -2.78 | 0.005 | ** |
| SEX (Male) | -0.32 | 0.15 | -2.2 | 0.03 | * |
| AGE (Years) | 0.09 | 0.06 | 1.46 | 0.14 | |
| PARENT(S) WITH POSTGRADUATE EDUCATION (True) | 0.21 | 0.18 | 1.19 | 0.23 | |
| YORKSHIRE PARENT(S) (True) | -0.05 | 0.16 | -0.30 | 0.76 | |
| REGULAR EXPOSURE TO SPEAKERS OF OTHER REGIONAL VARIETIES (True) | 0.36 | 0.18 | 1.99 | 0.05 | * |

**Table 5.7: Logistic mixed effects model to fit results across all rounds of the experiment (Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

### 5.5.3 Statistical analysis of Round 1: Yorkshire vs. SSBE results

Table 5.8 shows the results of the mixed effects model fit over the results from Round 1 of the experiment. Excluding 'round', the same six independent variables as the previous model were included in the model. In line with the previous experiments, a two-way interaction between age and sex was included in the model. A two-way interaction between postgraduate education and Yorkshire parents was also initially included in the model, in order to test whether one of these variables was more of a driving force (and consequently affecting the impact of the other variables). However, the interaction was not found to be significant and therefore was removed from the final model reported in Table 5.8.

In this model, three of the variables were found to significantly predict the probability of the children scoring a correct answer: the difficulty level, the interaction between age and sex, and whether the children had regular exposure to speakers of other regional varieties. Each of these predictor variables is described in turn in the following sections.

| | **Estimate** | **Std. Error** | **z value** | **Pr(>\|z\|)** | **Sig.** |
|---|---|---|---|---|---|
| (Intercept) | 1.35 | 1.77 | 0.77 | 0.44 | |
| DIFFICULTY LEVEL (2) | -1.02 | 0.28 | -3.65 | 0.0003 | *** |
| SEX (Male) | -7.63 | 2.85 | -2.68 | 0.007 | ** |
| AGE (Years) | -0.05 | 0.20 | -0.26 | 0.79 | |
| PARENT(S) WITH POSTGRADUATE EDUCATION (True) | 0.75 | 0.44 | 1.68 | 0.09 | |
| YORKSHIRE PARENTS (True) | 0.32 | 0.43 | 0.73 | 0.46 | |
| REGULAR EXPOSURE TO SPEAKERS OF OTHER REGIONAL VARIETIES (True) | 1.11 | 0.48 | 2.31 | 0.02 | * |
| AGE (Years):SEX (Male) | 1.04 | 0.39 | 2.63 | 0.008 | ** |

**Table 5.8: Logistic mixed effects model to fit results across round 1 of the experiment (Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

#### *5.5.3.1 Difficulty level*

Unexpectedly, Figure 5.13 shows that the predicted probability of a correct answer is higher for the stimuli in difficulty level 2 (when the teachers were heard saying a different sentence from the other speakers). The predicted probability of a correct answer declines from around 0.72 for DL2 to around 0.5 for DL1. This result is also reflected in Figure 5.14 which displays the raw data.

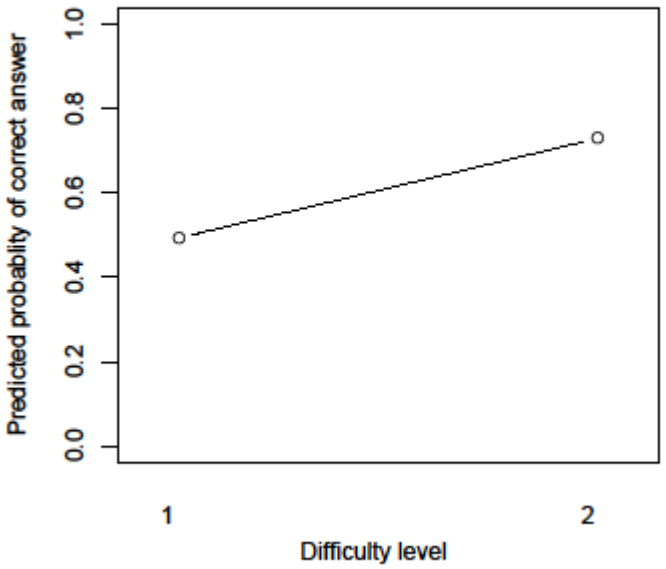



**Figure 5.13: Model prediction showing the effect of difficulty level in round 1**

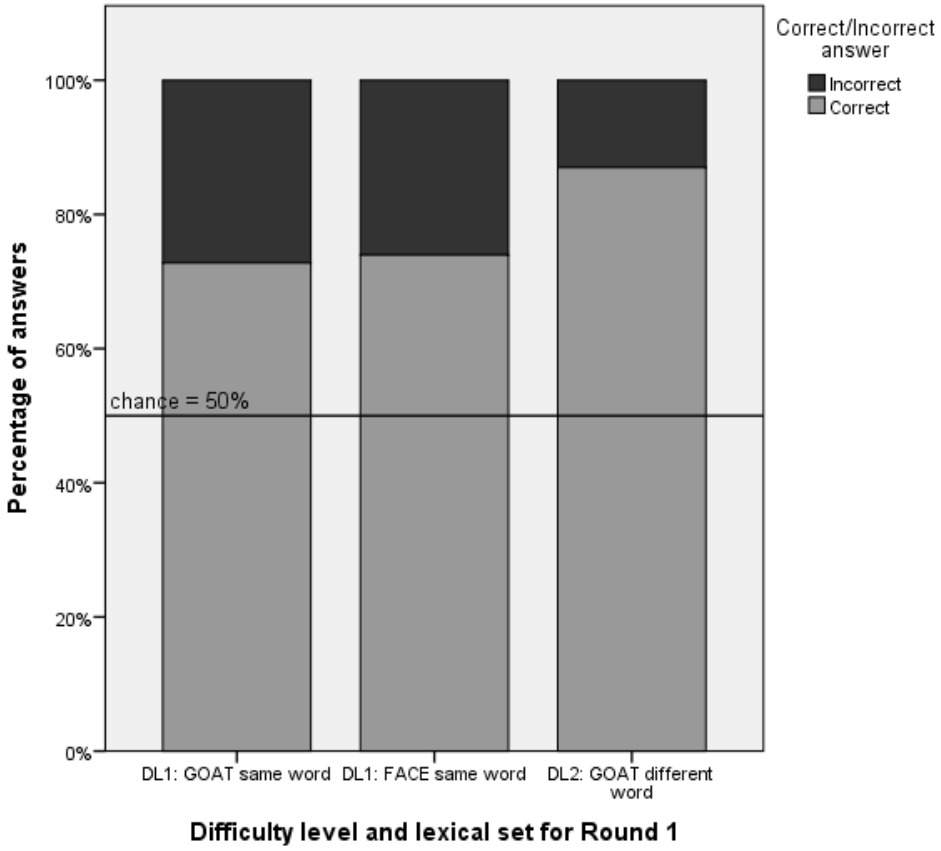



**Figure 5.14: Raw data showing the effect of difficulty level in round 1**

Figure 5.15 shows the percentage of correct answers for each of the individual stimuli in round 1 of the experiment. There were consistently higher scores in DL2 (over 80%) compared to the two sets of DL1. This indicates that the children found this set of stimuli easier to group in general, rather than low scoring individual stimuli in the two DL1 sets driving this effect. Explanations for this unpredicted difference are explored in the discussion section (section 5.6.2).

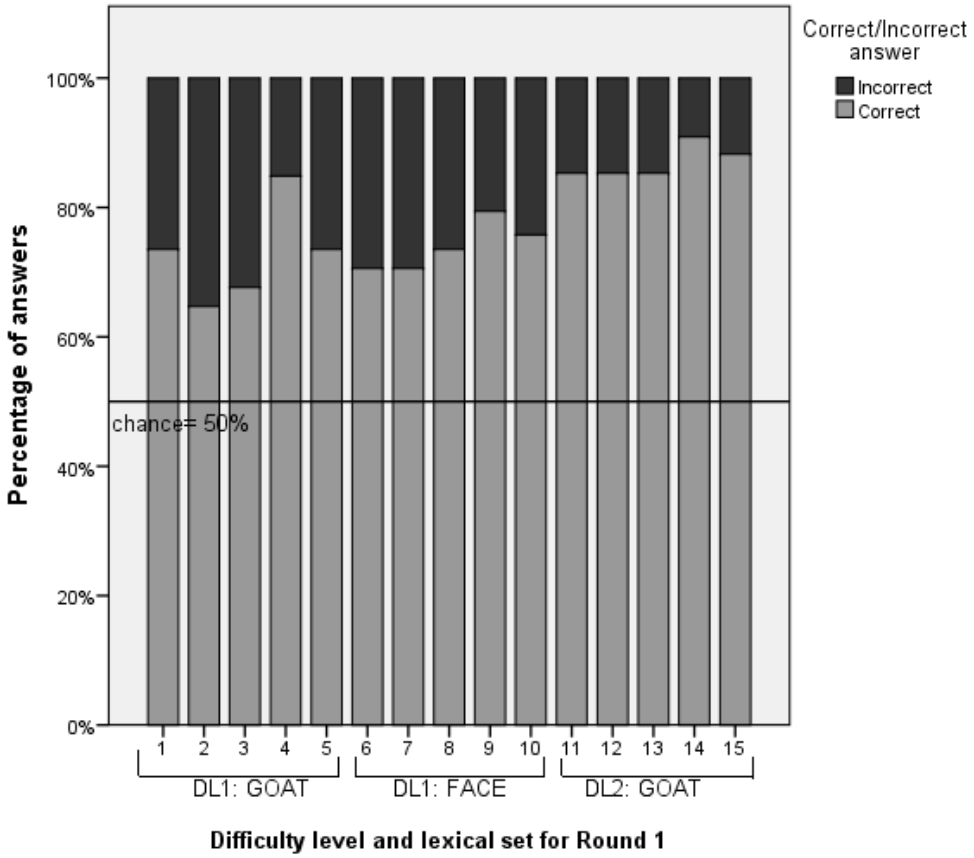



**Figure 5.15: Results for the individual stimuli in round 1**

### *5.5.3.2 Interaction of age and sex*

Figure 5.16 shows that the predicted probability of a correct answer increases for the boys as they get older. This is a large increase, from below chance for 6-year-olds to well over chance (above 0.8) for 9-year-olds. The same is not true for the girls, however, who seem to have reached a ceiling level of performance (around 0.8) at the age of 6. Although the plot

shows a slight (but non-significant) decline with age for the girls, this is likely to be due to a couple of low scoring individuals, who were the only two participants to score below 50% at all (see the raw data in Figure 5.17).

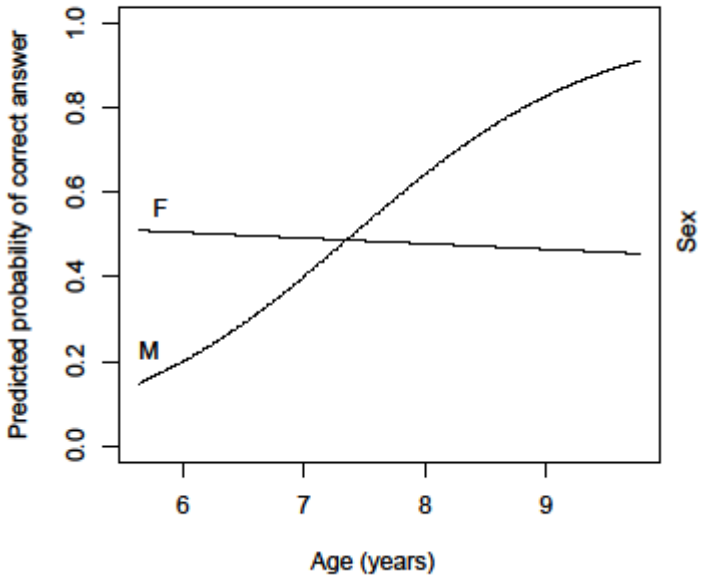



**Figure 5.16: Model prediction showing the interacting effects of age and sex in round 1 (F=female, M=male)**

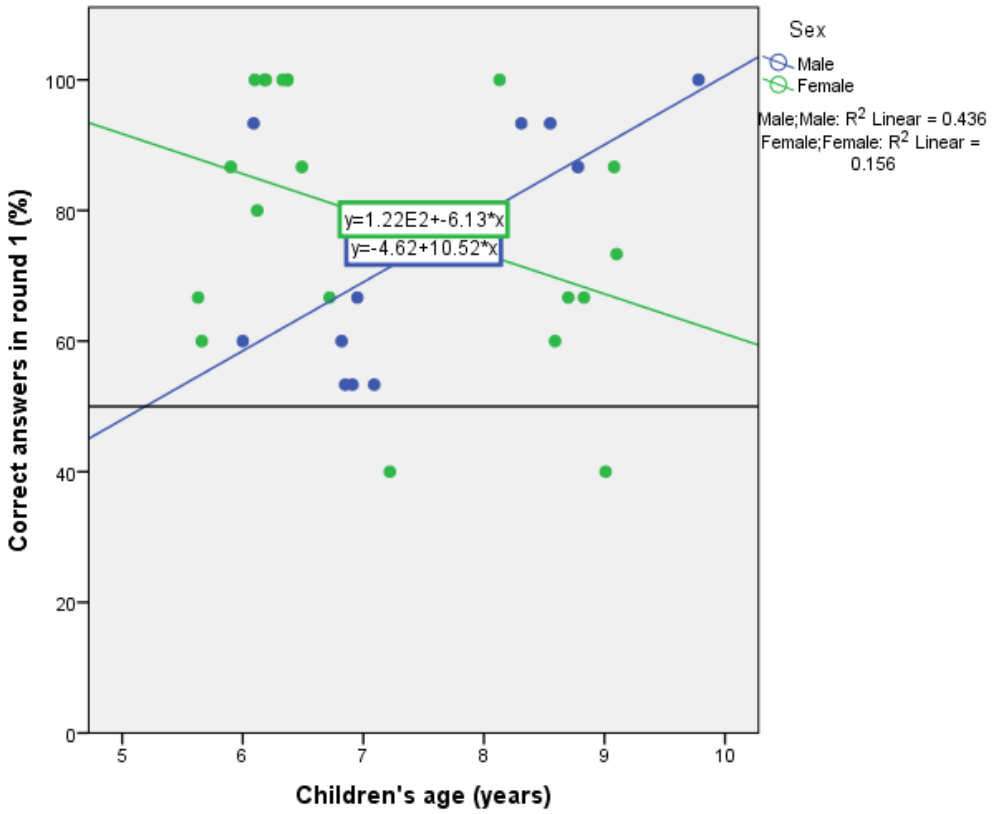



**Figure 5.17: Raw data showing the effects of age and sex in round 1**

### *5.5.3.3 Exposure to speakers of other regional varieties*

Figure 5.18 and Figure 5.19 show that the predicted probability of a correct answer increases for children who are reported as having regular exposure to non-Yorkshire English speaking family or friends. This finding was anticipated in line with results from the previous experiments in which the children's performance was predicted by the Yorkshire/non-Yorkshire background of their parents. Although it is not their parents' backgrounds which show a significant effect in this case, the prediction shown in Figure 5.18 is a representation of children's wider exposure to variation. This indicates that perhaps a more general exposure to variation plays a bigger role in children's perception as they get older. In accordance with research question (6), the next section investigates individual variation amongst the children in this round.

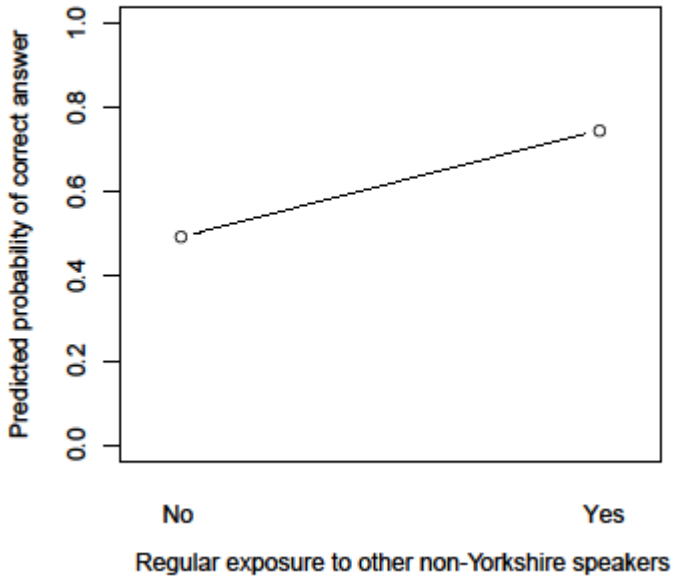



**Figure 5.18: Model prediction showing the effect of children's regular exposure to other regional varieties in round 1**

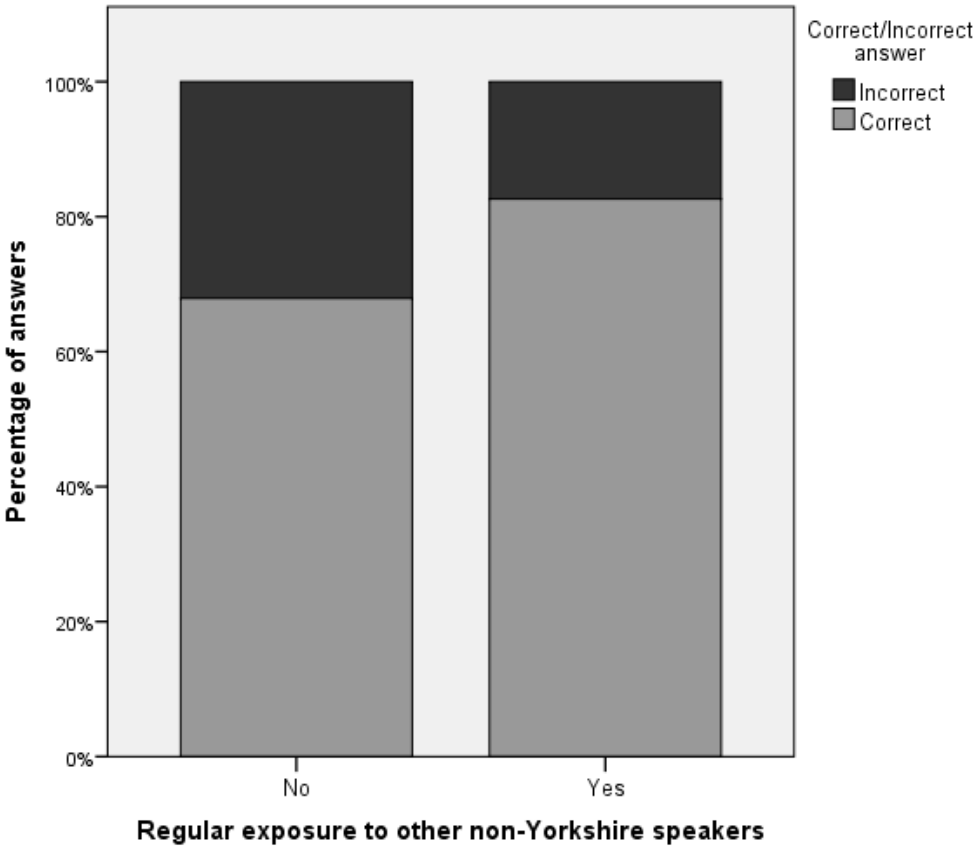



**Figure 5.19: Raw data showing the effect of children's exposure to other regional varieties in round 1**

### 5.5.4 Individual variation in Round 1

As noted in section 5.5.1, eight children scored 100% correct answers in this round. Table 5.9 shows the background details of these children. All of these top scoring children are from school 1 or school 3, with none from school 2. This reflects the lower performance of the children from school 2 overall, which is evidenced in Table 5.6. Table 5.6 shows that even the highest scoring individuals from school 2 (children 16 and 17) scored at the average level overall (67%). This is possibly a reflection of the school's difference in terms of their performance and their proportion of pupils from disadvantaged backgrounds (as described in section 5.4.1). However, because of the uneven distribution of children across the schools (only 7 from school 2), differences in performance between the schools will not be further investigated. Furthermore, as the children's backgrounds (their parents' education and their Yorkshire/non-Yorkshire regional origin) were found to be a good reflection of the differences between the children at these schools, the addition of school in the analysis would be redundant.

An intriguing finding from Table 5.9 is that six of the eight children are pairs of siblings (two of them twins). This supports the interpretation that children's exposure to variation plays a role in their perceptual ability, as these pairs of siblings are likely to have experienced a very similar linguistic input from their parents, family and family friends. Out of these eight children, four have a Southern parent (in bold in the table) and therefore their exposure to a Southern accent at home could be a reason for their high performance in this round of the experiment. There are six other participants in the experiment who have a Southern parent. Out of these six children, three also score very highly in this round (87-93 % correct answers). Two of the other children do not score so highly (60-67% correct answers; their scores are above chance but below the overall average). Only one child with a Southern parent scores below chance (see child 33 in Table 5.6).

The remaining four children in Table 5.9, who do not have Southern parents, all have Yorkshire parents and no reported regular exposure to Southern speakers. However, they do all have family from other regions of the UK whom they reportedly see regularly and therefore they have prevalent exposure to varieties other than Yorkshire English.

All of these top scoring children have at least one parent with a postgraduate level of education. As described in section 5.4.7, this measurement was taken as an indication of the

children's social class and therefore suggests that these children come from higher class backgrounds. In turn, this has implications for the quality and quantity of the linguistic input that the children may experienced, including the kind of linguistic variation that they are likely to have been exposed to (see discussion section 5.6.4 for further consideration of this finding).

| Child | Age | Sex | School | Parents from | Parents' highest education | Other exposure to regional varieties |
|---|---|---|---|---|---|---|
| 2 (twins) | 6.37 | F | 1 | Leeds, Doncaster | Postgrad degree | Grandmother from Newcastle, grandfather from Wales |
| 5 (twins) | 6.37 | F | 1 | | | |
| 9 (siblings) | 8.13 | F | 3 | **London**, Merseyside | Postgrad degree | Childminder from London |
| 14 (siblings) | 6.1 | F | 3 | | | |
| 22 (siblings) | 6.33 | F | 1 | York | Postgrad degree | Uncle and cousin from Newcastle |
| 23 (siblings) | 9.78 | M | 1 | | | |
| 3 | 6.18 | F | 3 | **Northampton,** Germany | Postgrad degree | |
| 19 | 6.19 | F | 3 | **London**, Sheffield | Postgrad degree | Grandmother from Doncaster, grandfather from Portsmouth |

**Table 5.9: Background details of the top scoring children in round 1**

Table 5.10 presents the five lowest scoring children in round 1 (the only children to score at chance level or lower). Three of these children have both parents from York (in bold in the table) and only one child is reported as having regular exposure to a speaker from another region of the UK. Child 29 was born in Malaysia and has been in York for 3 years; her main exposure to English has therefore been in the local York area. Child 33 is the only one with a Southern parent. Overall, the children's fairly limited exposure to other regional varieties could explain their low scores in this round. In addition, these children's parents generally have a lower level of education than the highest performing children. Again this could indicate the lower social class backgrounds of these children and therefore their different exposure to linguistic variation in the input (see section 5.6.4). There are only three other participants in the experiment who have both parents from York and whose parents have a

lower than postgraduate level of education; these children score above chance in this round (60-67% correct answers) but not above the children's overall average (77%).

| Child | Score Round 1 (%) | Age | Sex | School | Parents from | Parents' highest education | Other exposure to regional varieties |
|---|---|---|---|---|---|---|---|
| 29 | 40.00 | 9.01 | F | 2 | Malaysia | Postgrad degree | |
| 33 | 40.00 | 7.22 | F | 2 | York, Oxford | High school up to 18 | |
| 21 | 53.33 | 6.91 | M | 2 | **York** | High school up to 18 | |
| 32 | 53.33 | 7.09 | M | 1 | **York** | Undergrad degree | Uncle from Essex |
| 34 | 53.33 | 6.85 | M | 1 | **York** | High school up to 18 | |

**Table 5.10: Background details of the lowest scoring children in round 1**

### 5.5.5 Statistical analysis of Round 2: Yorkshire vs. Scottish results

The same variables and interactions were included in the statistical model for round 2. Using the stepwise backward regression method, non-significant variables were pruned from the model one at a time, until the predicting power of the model could not be improved.

| | Estimate | Std. Error | z value | Pr(>\|z\|) | Sig. |
|---|---|---|---|---|---|
| (Intercept) | -1.005 | 0.81 | -1.24 | 0.22 | |
| SEX (Male) | -0.42 | 0.26 | -1.6 | 0.11 | |
| AGE (Years) | 0.16 | 0.11 | 1.52 | 0.13 | |
| PARENT(S) WITH POSTGRADUATE EDUCATION (True) | 0.47 | 0.25 | 1.86 | 0.06 | |

**Table 5.11: Logistic mixed effects model to fit results across round 2 of the experiment (Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

As Table 5.11 shows, no variables were significant in the final model. However, postgraduate education was close to significance at the 0.05 level, indicating that children from a higher social class background were more likely to perform well in this round. The next section will explore this near significant effect, along with the significant effects from round 1 of the experiment in order to investigate whether these trends remain in round 2.

### 5.5.6 Exploration of the effects of independent variables/ individual variation in Round 2

Although no variables were found to be significant in the statistical model for round 2, this section explores the effect of the near significant variable (parents' level of education) on the children's results in this round. In addition, this section explores those variables found to be significant in round 1 which also show a correlation with the results in this round This is in order to investigate possible trends in the data which were not captured by the statistical model. Subsequently, children's individual performances are then considered in more detail in section 5.5.6.5.

#### *5.5.6.1 Parents' postgraduate education*

Figure 5.20 shows that, overall, children who have at least one parent with a postgraduate level of education perform above chance level (just over 60%) in this round of the experiment. In comparison, children whose parents do not have a postgraduate level of education perform around chance level.

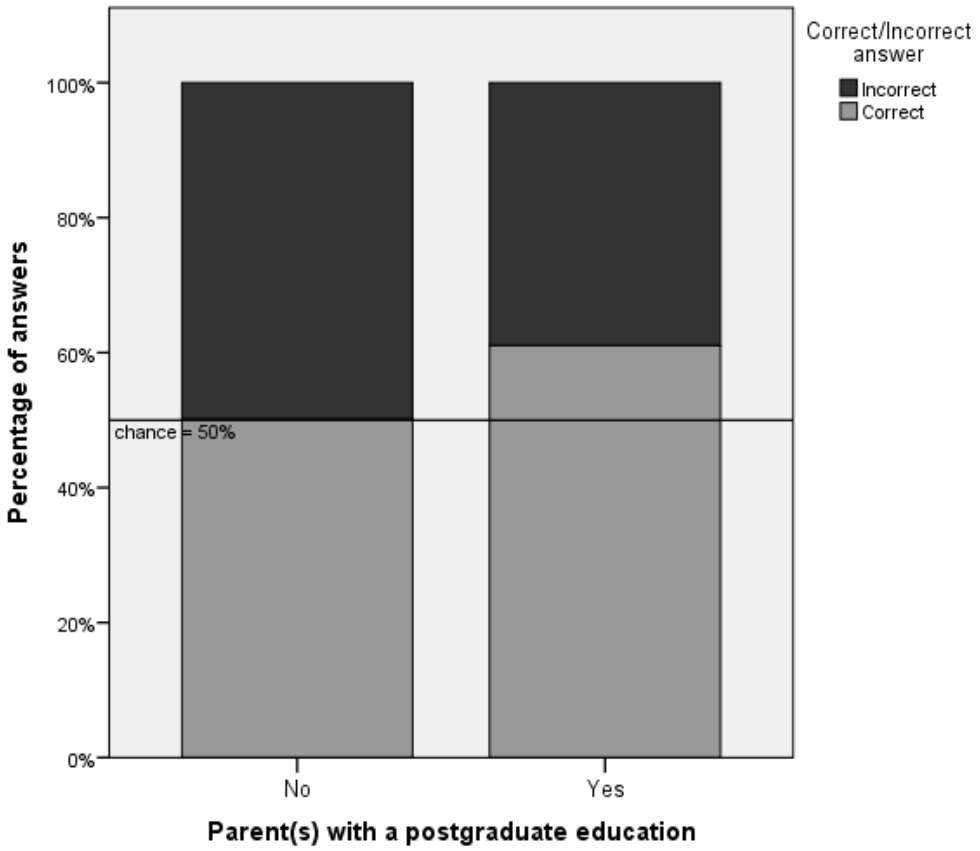



**Figure 5.20: Raw data showing the effect of the children's parents' level of education in round 2**

### *5.5.6.2 Difficulty level*

Although difficulty level was not significant, as Figure 5.21 shows, there was a difference between the scores for each set of stimuli in this round. Children performed better for the set of stimuli featuring words of the NURSE lexical set and when the teachers were heard saying the same sentences as the other speakers. This indicates that the difference in the pronunciation of NURSE is a more conspicuous distinction for the children than the lettER vowel. This may be because the Scottish speakers' post-vocalic /ɹ/ was in a stressed syllable in NURSE words, as opposed to in an unstressed syllable in lettER words. Therefore, the difference between the rhotic/non-rhotic pronunciations of the lettER words was perhaps not as noticeable as the difference between the rhotic/non-rhotic pronunciations of the NURSE words. The difference in salience between stressed and unstressed syllables is evidenced in the word-learning of young infants. As Vihman and colleagues (2004) found in their study of 11-month-olds, while the infants' word recognition was blocked by changes to the onset consonant of a stressed syllable, word recognition was not blocked by changes to

the onset consonant of an unstressed syllable. The authors conclude that this result demonstrates the importance of stressed syllables for infants' word representations.

The children in the current experiment performed worse when they grouped together speakers saying the NURSE vowel in DL2, however. This suggests that they struggled to abstract across the pronunciations at the level of the phoneme in order to group the speakers by their pronunciations of the NURSE vowel.

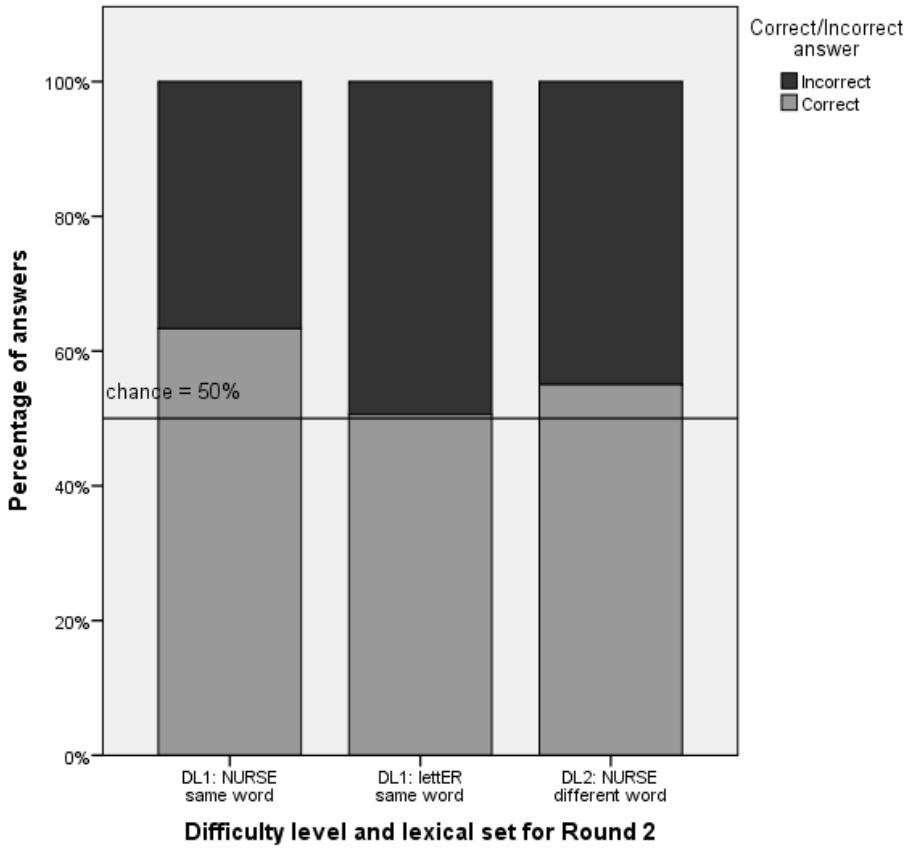



**Figure 5.21: Raw data showing performance across the difficulty levels in round 2**

#### *5.5.6.3 Sex*

Figure 5.22 shows that the females perform better than the males in round 2, with an average of just over 60% correct answers, compared to just under 50% (around chance level) for the males. This trend in the results of round 2 is in line with the better overall performance of the females in round 1 (as well as round 3 – see section 5.5.7.2).

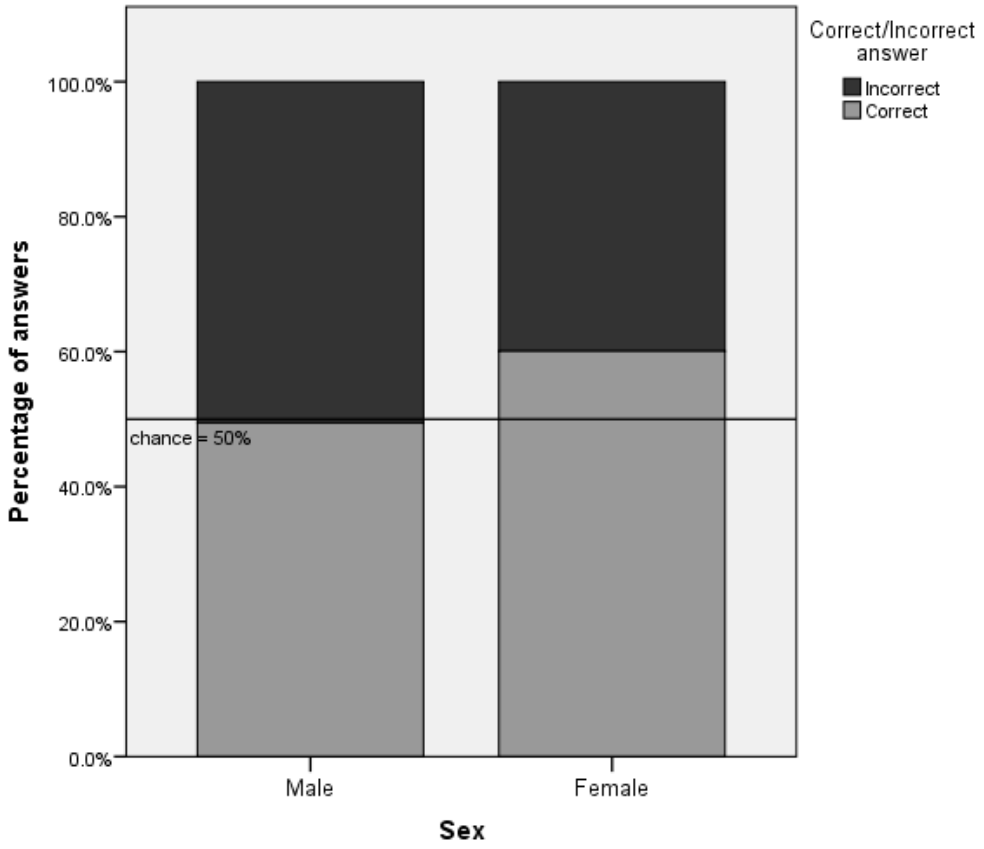



**Figure 5.22: Raw data showing the effects of age and sex in round 2**

#### *5.5.6.4 Exposure to speakers of other regional varieties*

Figure 5.23 again shows a trend in this round of the experiment consistent with a significant effect found in round 1; children who have regular exposure to speakers from other regions of the UK perform better than those who do not have this exposure (although only by a small percentage - 59% vs. 53% correct answers).

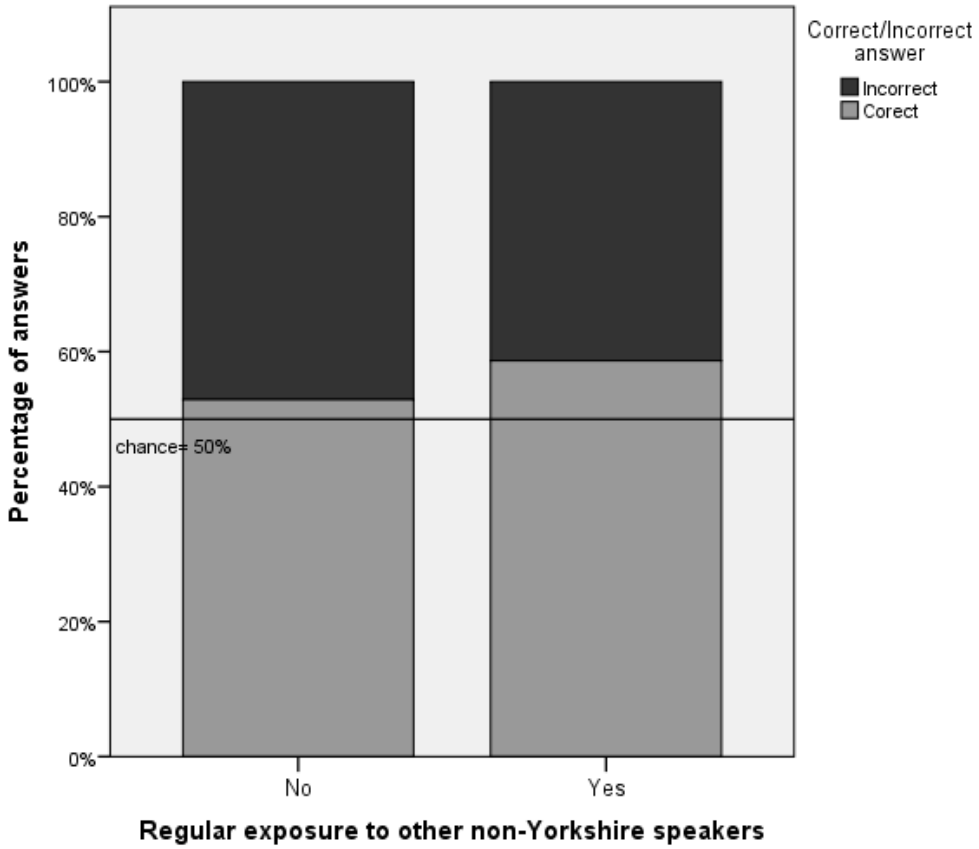



**Figure 5.23: Raw data showing the effect of children's exposure to other regional varieties in round 2**

#### *5.5.6.5 Individual variation in Round 2*

Unlike round 1, no children scored 100% in round 2. Furthermore, as there was only one child with the top score in this round (93%), Table 5.12 shows the background information of the children with the highest and the second highest scores. These three children scored 13 or 14 out of 15 correct in this round. An interesting finding is that two of these children have Scottish family (in bold in the table), which could explain their high scores in this round. This explanation is further justified by the fact that none of the other participants in the experiment are reported as having any Scottish family or regular exposure to Scottish speakers.

| Child | Score Round 2 (%) | Age | Sex | School | Parents from | Parents' highest education | Other exposure to regional varieties |
|---|---|---|---|---|---|---|---|
| 1 | 93.33 | 9.08 | F | 1 | Wiltshire, Newcastle | Postgrad degree | Aunt, uncle and cousins in **Scotland**, close family friend from Wales |
| 7 | 86.67 | 8.55 | M | 1 | Hull, Tyne and Wear | Postgrad degree | Grandparents in Tyne and Wear, grandmother in **Scotland**, aunt and cousins in Hull |
| 4 | 86.67 | 8.7 | F | 1 | Darlington (mother) | Undergrad degree | |

**Table 5.12: Background details of the highest scoring children in round 2**

The background details of the three lowest scoring children in round 2 are shown in Table 5.13. These children scored 3 or 4 correct out of 15. Although two of these children have parents from York, indicating a lack of regional variation in their input, there are other participants with both parents from York who perform better in this round and above the average (57%), for example Child 21 who scored 73% and Child 24 who scored 67%. The three lowest scoring children in this round are also the lowest scoring individuals in round 1 (see Table 5.6) and the lowest scoring individuals overall. This suggests that these three children are low performers in general and this is not necessarily due to their background. Therefore, while general exposure to regional variation does not appear to play much of a role in this round of the experiment, exposure to Scottish speakers in particular may have more of an effect on the children's performance.

| Child | Score Round 2 (%) | Age | Sex | School | Parents from | Parents' highest education | Other exposure to regional varieties |
|---|---|---|---|---|---|---|---|
| 34 | 20.00 | 6.85 | M | 1 | York | High school up to 18 | |
| 32 | 26.67 | 7.09 | M | 1 | York | Undergrad degree | Uncle from Essex |
| 33 | 26.67 | 7.22 | F | 2 | York, Oxford | High school up to 18 | |

**Table 5.13: Background details of the lowest scoring children in round 2**

### 5.5.7 Statistical analysis of Round 3: Yorkshire vs. North East results

Table 5.14 shows the resulting model for round 3. The same variables and interactions as the previous two models were initially included in the model and then the model was pruned of non-significant predictors. Two variables were significant: difficulty level and sex.

| | Estimate | Std. Error | z value | Pr(>\|z\|) | Sig. |
|---|---|---|---|---|---|
| (Intercept) | 1.15 | 0.59 | 1.95 | 0.05 | * |
| DIFFICULTY LEVEL | -0.39 | 0.2 | -1.94 | 0.05 | * |
| SEX (Male) | -0.44 | 0.2 | -2.2 | 0.03 | * |
| AGE (Years) | -0.04 | 0.08 | -0.48 | 0.63 | |
| YORKSHIRE PARENT(S) (True) | -0.22 | 0.2 | -1.13 | 0.26 | |

**Table 5.14: Logistic mixed effects model to fit results across round 3 of the experiment (Significance level: '*'= 0.05, '**' = 0.01, '***' = 0.001)**

#### *5.5.7.1 Difficulty level*

Similar to round 1, Figure 5.24 and Figure 5.25 show that the predicted probability of a correct answer is higher than for the stimuli in difficulty level 2. However this prediction is not as robust as for round 1; the predicted probability of a correct answer declines from just over 0.6 for DL2 to just over 0.5 for DL1.

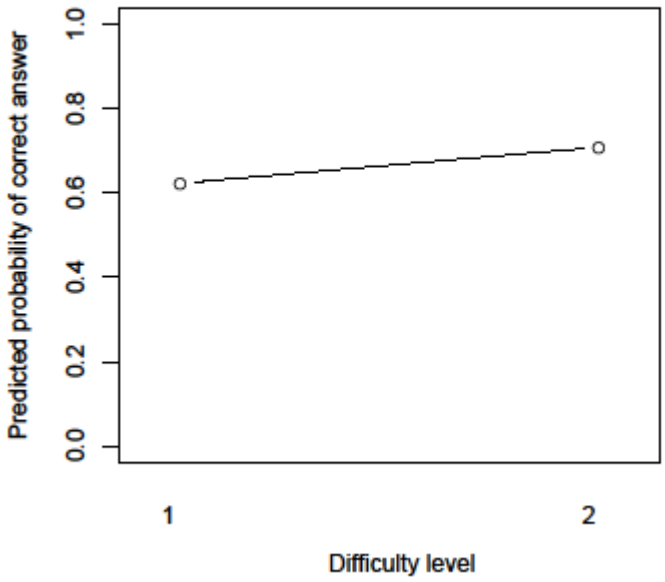



**Figure 5.24: Model prediction showing the effect of difficulty level in round 3**

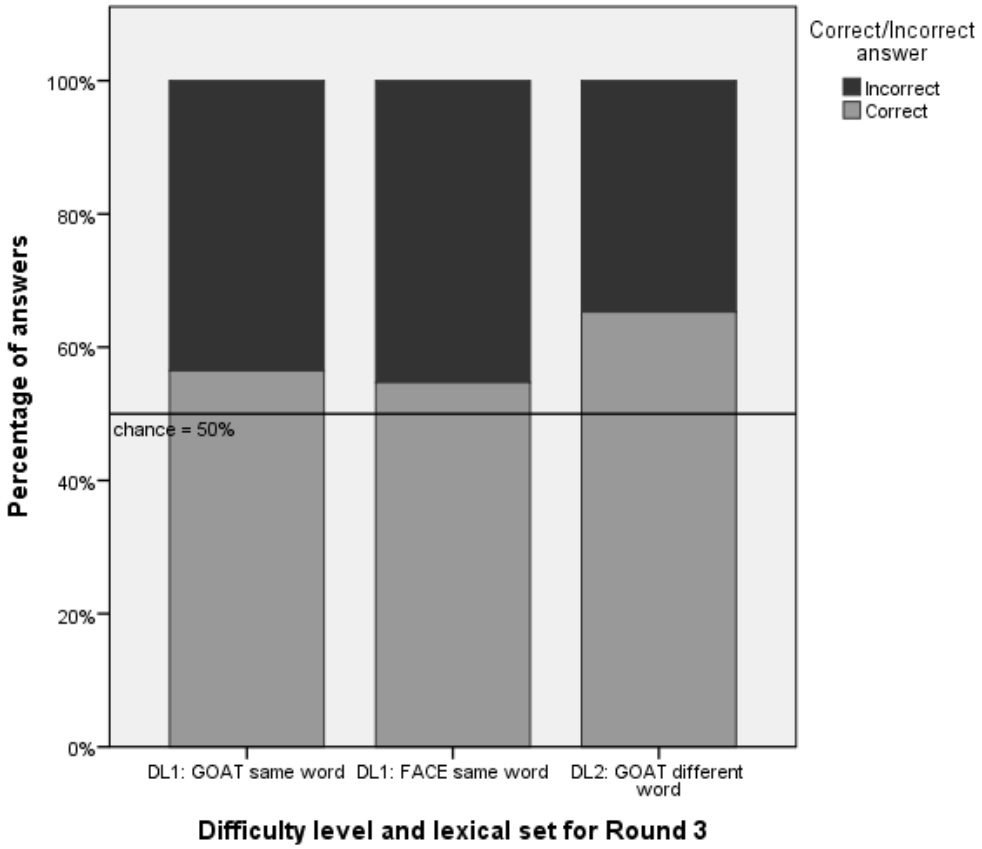



**Figure 5.25: Raw data showing the effect of difficulty level in round 3**

Figure 5.26 shows the percentage of correct answers for each of the individual stimuli in round 3. Whereas the percentage of correct answers for DL2 was consistently higher than chance, the scores for the two sets of DL1 were much more varied, with some above and some below chance.

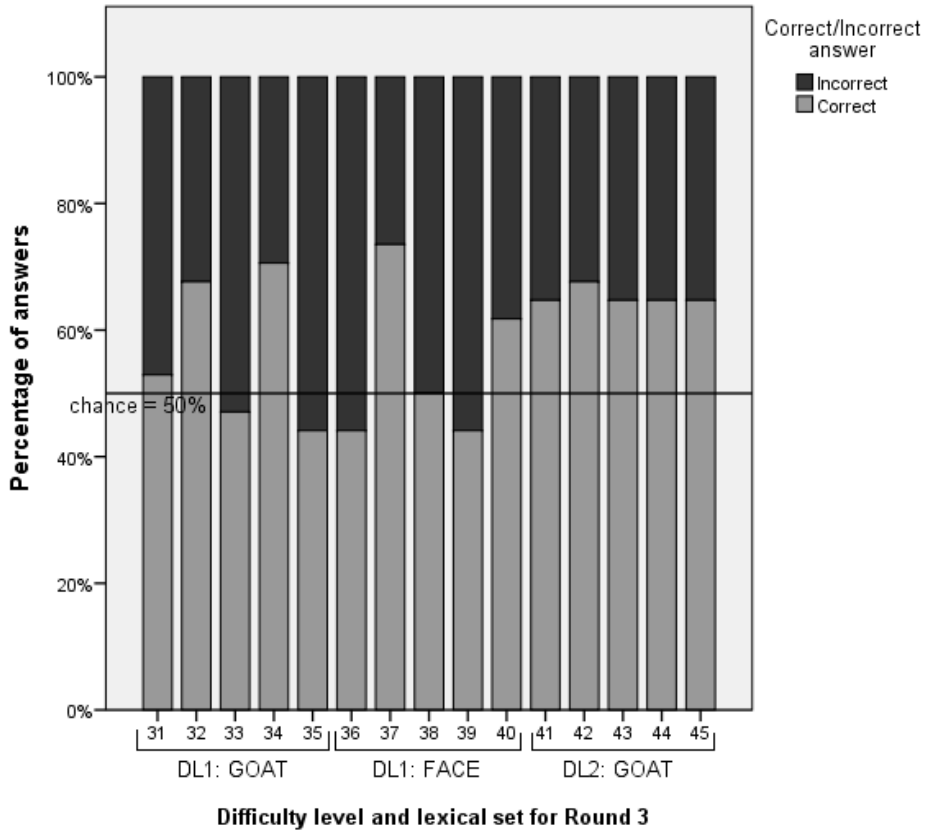



**Figure 5.26: Results for the individual stimuli in round 3**

This unexpected contrast in performance between the difficulty levels might be explained by the degrees of difference between the vowels in each set of stimuli. As described in the methodology (see section 5.3.1), overall the GOAT and FACE vowels were pronounced very similarly by both the Yorkshire and the North East speakers, with only small differences in their height and frontness. However there was a less definite distinction between the Yorkshire and North East speakers' vowel pronunciations in the first two sets of stimuli in this round of the experiment. In the first set of stimuli, one of the North East speaker's GOAT vowels was much closer to the Yorkshire speakers' GOAT vowels (see circled speaker in Figure 5.1). This was reflected in the results, as this speaker was only categorised with the teacher from the same region by 44% of the children. In the second set of stimuli, there was

some overlap between the degree of both the height and frontness of the Yorkshire/North East speakers' FACE vowels (again see Figure 5.1). Therefore, the vowels of the two groups of speakers in this set were not completely distinct. In the third set of stimuli, however, there was a more definite distinction between the Yorkshire and North East speakers' GOAT vowel pronunciations. As a result, despite this set of stimuli being from DL2, the children perhaps performed best in this set of stimuli because they were aided by the speakers' more categorical vowel pronunciation differences.

#### *5.5.7.2 Sex*

Similar to the results from the Grouping experiment in chapter 4, Figure 5.27 shows the predicted probability of a correct answer is higher for the females than the males in this round of the experiment. The raw data graph in Figure 5.28 also illustrates the effect of this prediction.

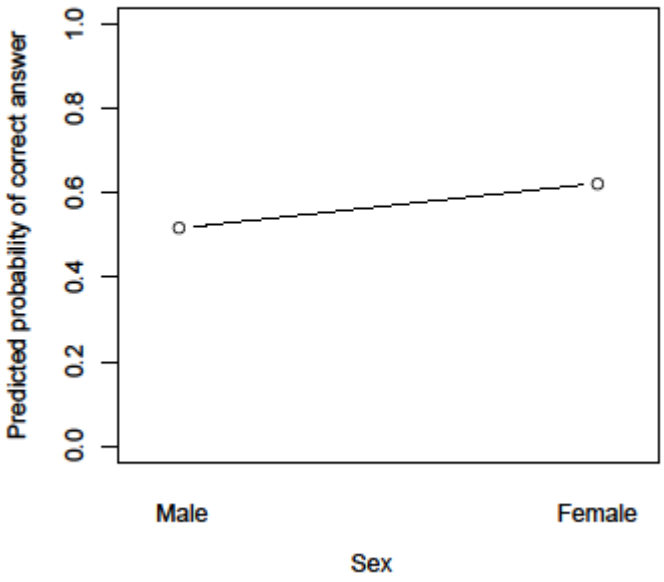



**Figure 5.27: Model prediction showing the effect of sex in round 3**

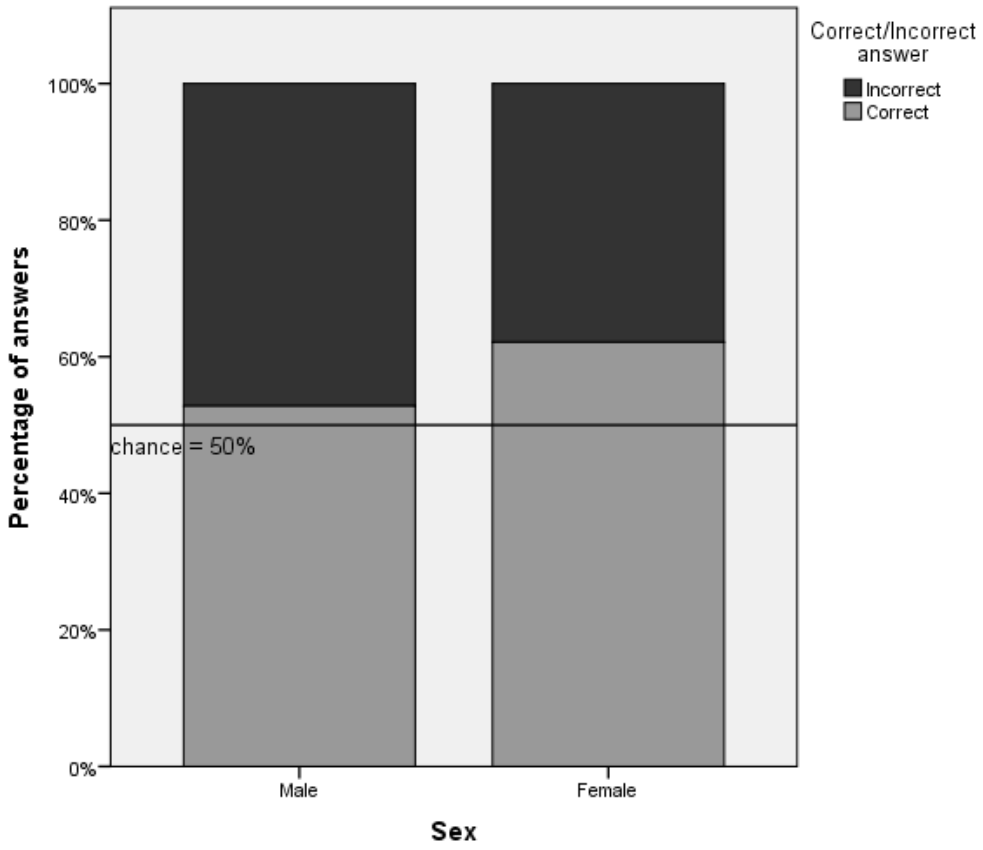



**Figure 5.28: Raw data showing the effect of sex in round 3**

### 5.5.8 Exploration of the effects of other independent variables/ individual variation in Round 3

This section investigates correlations based on the independent variables explored in the other rounds of the experiment. Only one of these variables was found to correlate with children's performance in this round of the experiment (and only to a small degree- see section 5.5.8.1). Section 5.5.8.2 then analyses children's individual performances in more detail.

#### *5.5.8.1 Exposure to speakers of other regional varieties*

Figure 5.29 shows a slight difference in performance between children who have regular exposure to speakers from other regional areas and those who do not. Although this is in line with the results from rounds 1 and 2, this difference is only small (just over 3%). Furthermore, this difference is not supported by the investigation into children's individual performances in this round, as described in the following section.

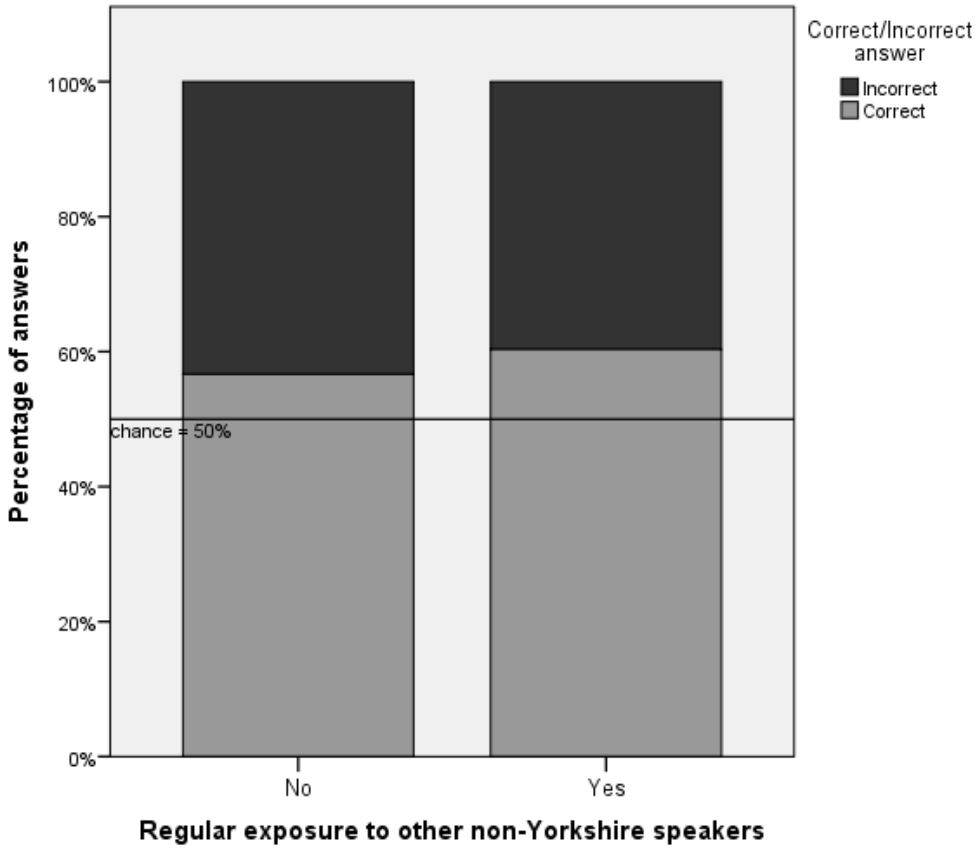



**Figure 5.29: Raw data showing the effect of children's exposure to other regional varieties in round 3**

#### *5.5.8.2 Individual variation in Round 3*

Despite being the round with the smallest range in scores, in round 3 there was still a fairly large range (from 27% to 80%), indicating a lot of individual variation. Consistent with the analysis of the results in round 2, Table 5.15 shows the background information for the children who had the highest and second highest scores in round 3. These ten children scored 12/15 (80%) or 11/15 (73%) correct answers. As shown in the table (in bold), four of the children were regularly exposed to speakers from the North East, which may explain

their high scores in this round. However, the other children come from a mixture of backgrounds as some have parents from other regions of the UK whereas others have parents from Yorkshire and no other regular exposure to regional varieties. Furthermore, three low scoring individuals also have family from the North East (see Table 5.15) and the only other child with family from the North East scored 60%. Overall, this reveals the mixture of results in this round and indicates that neither exposure to the North East variety itself, nor exposure to regional variation in general, appears to correlate with children's performance in this round of the experiment. This finding is further supported by investigating the lowest performing individuals in this round (see Table 5.15).

| Child | Score Round 3 (%) | Age | Sex | School | Parents from | Parents' highest education | Other exposure to regional varieties |
|---|---|---|---|---|---|---|---|
| 2 | 80 | 6.37 | F | 1 | Leeds, Doncaster | Postgrad degree | Grandmother from **Newcastle,** grandfather from Wales |
| 13 | 80 | 6.12 | F | 3 | Lebanon, Derby | Postgrad degree | Childminder from London |
| 10 | 80 | ? | M | 1 | York, Scarborough | Undergrad degree | |
| 1 | 73.33 | 9.08 | F | 1 | Wiltshire, **Newcastle** | Postgrad degree | Aunt, uncle and cousins in Scotland, close family friend from Wales |
| 4 | 73.33 | 8.7 | F | 1 | **Darlington** (mother) | Undergrad degree | |
| 17 | 73.33 | 6.82 | M | 2 | York | Postgrad degree | |
| 8 | 73.33 | 5.9 | F | 3 | York, Grimsby | Undergrad degree | |
| 15 | 73.33 | 9.1 | F | 1 | York, Wakefield | Undergrad degree | |
| 9 | 73.33 | 8.13 | F | 3 | London, Merseyside | Postgrad degree | Childminder from London |
| 27 | 73.33 | 6 | M | 3 | Essex, **Middlesbrough** | Postgrad degree | Aunt from Dublin, grandmother from **Hartlepool**, grandmother from Essex |

**Table 5.15: Background information of the highest scoring children in round 3**

Table 5.16 shows the children with the lowest scores in round 3. These six children grouped 4-6 out of 15 speakers correctly (27-40%). They come from a mixture of backgrounds, and three have parents or family from the North East. Two of these children are siblings (child 22 and child 23), which suggests that they have had similar experience of the North East accent through interactions with their family members from the North East. However, their exposure to the North East variety does not appear to have helped them in this round. The differences between how these children performed in round 1 highlight the inconsistencies with their performance in round 3. Whilst children 7, 22 and 23 scored very highly in round

1, child 34 scored around chance level and children 30 and 31 did not score above average. Overall, this further indicates the children's high level of individual variability. Despite the variation in the children's background overall, there are two pairs of siblings who have a similarly low score in this round. This echoes the correspondent sibling performances of round 1 and indicates a potential link between the children's background and their performance.

| Child | Score Round 3 (%) | Score Round 1 (%) | Age | Sex | Parents from | Parents' highest education | Other exposure to regional varieties |
|---|---|---|---|---|---|---|---|
| 34 | 26.67 | 53.33 | 6.85 | M | York | High school up to 18 | |
| 31 | 33.33 | 66.67 | 6.95 | M | Cheshire, Derbyshire | Postgrad degree | |
| 7 | 40.00 | 93.33 | 8.55 | | Hull, **Tyne & Wear** | Postgrad degree | Grandparents in **Tyne & Wear**, grandmother in Scotland, aunt and cousins in Hull |
| 30 | 40.00 | 66.67 | 8.83 | F | York | High school up to 18 | |
| 22 (siblings) | 40.00 | 100.00 | 6.33 | F | York | Postgrad degree | Uncle and cousin from **Newcastle** |
| 23 | 40.00 | 100.00 | 9.78 | M | | | |

**Table 5.16: Background information of the lowest scoring children in round 3**

### 5.5.9 Summary of results across the different rounds

Overall, primary school children scored above chance level when grouping together different speakers based on regionally distributed pronunciation features. The children performed best in round 1, when grouping together speakers based on a Yorkshire/SSBE accent distinction. They performed less well in rounds 2 and 3, when grouping speakers based on a Yorkshire/Scottish and then a Yorkshire/North East distinction. Effects were found relating to the children's age, sex, and their exposure to other regional varieties, as well as the difficulty level of the set of stimuli. In the following section, these results will be discussed in line with the research questions.

## 5.6 Discussion

The Second grouping experiment has found that primary school children, between the ages of 5 and 9, score above chance at grouping different speakers on the basis of their regional accent. This contradicts findings from Floccia and colleagues (2009) and Wagner and colleagues (2014), who concluded that children under the age of 7 were not able to categorise speakers by a regional accent distinction. The experiment design process was crucial in uncovering this ability and revealed the importance of constructing an experiment suitable for children of this young age and in accordance with their cognitive capabilities.

The results from this experiment build on the findings from the previous experiments outlined in chapters 3 and 4. These earlier experiments found that children can use regional accent features when identifying a familiar speaker, and that they can group together speaker guises on the basis of a regional accent distinction. In this Second grouping experiment, children have shown the ability to group different speakers by a regional accent distinction. This is regarded as a harder task as it involves overlooking individual speaker differences and drawing on more abstract categories of speakers, based on similarities between their pronunciations. This progression in abilities is consistent with an exemplar model of indexical learning which posits that categories of speakers develop initially from an individual's encounters with familiar speakers (for further discussion see general discussion section 6.4).

The following sections will further discuss the results in line with the research questions (set out in section 5.2).

### 5.6.1 Differences across the rounds

As predicted, children were found to perform best in round 1, when grouping speakers based on a Yorkshire vs. SSBE accent distinction, rather than a Yorkshire vs. Scottish or a Yorkshire vs. North East distinction. In round 1, the children's mean score was 77% correct answers, with only two children scoring below chance at 50%. Furthermore, eight children scored 100% correct answers in this round.

These results indicate that out of the accents tested in the experiment, the differences between Yorkshire and SSBE accented speakers were the most conspicuous for the children. As described in the introduction, features of SSBE are ubiquitous in the speech of middle

class people throughout the country. Therefore, the children's prevalent exposure to features of SSBE, alongside the features of their local Yorkshire accent, is likely to have helped in their ability to perceive differences between the accents. This finding also builds on the results of the Grouping experiment in chapter 4, which found that pre-school children were able to differentiate speaker guises on the basis of accent features pertaining to a Yorkshire/SSBE distinction. The Second grouping experiment has expanded on these results and found that older primary school children can use this distinction to group different speakers. Therefore these older children are able to disregard the individual differences of the speakers in favour of common properties of the accents.

In round 2, children scored lower than round 1; their mean score was 57% correct answers and the highest score of 93% was achieved by only one child. In round 3, children scored a mean of 60% correct, and three children scored the highest at 80%. Children's lower performances in these rounds indicate that the Yorkshire vs. Scottish/North East accent distinctions were much harder for them to differentiate than the Yorkshire vs. SSBE distinction. This may partly be explained by the fact that features of the Scottish and North East accents are less pervasive than SSBE in general; they do not hold the same level of prestige and therefore features of these accents have not been adopted by middle class speakers as they have for SSBE features. Furthermore, whereas the FACE and GOAT vowels of round 1 represent stereotypical distinctions between the north and the south of England (see section 2.1.3), the variables of rounds 2 and 3 do not present such salient contrasts.

A rhotic/non-rhotic distinction of the NURSE and lettER vowels differentiates the Scottish speakers from the Yorkshire speakers in round 2. While rhoticity is generally regarded as a distinguishing feature of Scottish accents in comparison to English accents, (as described in section 2.1.3), this distinction does not appear to be discernible for many of the children in this experiment. This could be due to the children's lack of experience with the Scottish accent in general (see section 5.6.4). However, as described in section 5.4.6, the background information collected from parents revealed that, overall, the children watched many American television programmes and films. Some parents even commented directly (on the background information sheet or through personal communication) that their children sometimes spoke with a pretend American accent when playing with their toys, for example. One of the most salient features of an American accent for English English speakers is the

use of postvocalic /r/; this is often one of the features pronounced by speakers when imitating an American accent (see Trudgill, 1986:12-13). Therefore, it would perhaps be expected that the children's exposure to rhoticity through American television (and their attempts to imitate the accent) would help them to differentiate between rhotic/non-rhotic pronunciations in round 2 of this experiment.

Despite the rhotic similarity in these varieties of English, however, there is 'complex articulatory variability' (Lawson et al., 2011:260) in the realisation of /r/ in both Scottish (Stuart-Smith, 2003; 2007) and American English (Zhou et al., 2008). Therefore in the current experiment the Scottish speakers' rhotic realisations may not have been strongly or consistently rhotic enough for the children to perceive them as rhotic. The variability inherent in rhoticity is also shown by the fact that it is difficult to measure acoustically; there are conflicting interpretations regarding the role of F2 and F3 in the perceptual salience of rhotic segments (Heselwood and Plug, 2011). In their study using ultrasound tongue imaging, thus focusing on the production of rhotic segments, Lawson and colleagues (2011) found a social differentiation between Scottish middle-class speakers' more strongly rhotic variants (using a bunched tongue) and Scottish working-class speakers' less rhotic variants (using a raised tongue-tip/front). The social indexicality of Scottish speakers' rhoticity further complicates its relationship to American rhoticity; American rhoticity has been found to be speaker-specific (Mielke et al., 2010) but not socially stratified. Therefore, because rhoticity in Scottish English and American English are signalling different levels of variation, it is perhaps too much of a simplification to equate exposure to rhoticity from American speakers with exposure to rhoticity from Scottish speakers. In terms of the current experiment, therefore, it is not surprising that the children's exposure to American television does not help them to differentiate between Scottish and Yorkshire accents on the basis of rhoticity.

The lack of effect from exposure to American rhoticity in the media could also suggest that this more indirect form of exposure does not contribute to perceptual awareness of regional accent distinctions in the same way, or to the same extent, as direct exposure (from family/friends/speakers in the community etc.). Therefore, because of the children's overall lack of direct experience with the Scottish accent specifically, and rhoticity more generally, the difference between the Yorkshire/Scottish speakers was not as discernible to them as

the difference between the Yorkshire/SSBE speakers; consequently they performed worse in round 2. As described in section 2.3.1, although exposure to certain non-standard variants in the media has been found to play some role in helping to spread linguistic change in production, no studies appear to have investigated whether media exposure plays a role in accent perception. Therefore research into the different effects of direct and indirect exposure would be useful to investigate this further (see section 6.5). This is also an important factor to consider under an exemplar theoretic framework, in which the storing of exemplars depends upon their importance at the time of encounter (see section 2.4.2). As suggested in section 6.5, it would be fruitful to further investigate children's media watching habits, including their psychological attachment and investment in what they watch, and its effect on their perception of accent differences.

In round 3, the children also scored lower than round 1; their mean score was 60% correct answers and the highest score of 80% was only achieved by three children. Children's difficulty in discriminating between the Yorkshire and North East speakers' pronunciations of the GOAT and FACE vowels is unsurprising as these vowels do not represent a distinction on the level of a stereotype. They were pronounced very similarly by both sets of speakers, with only a small average difference in height/frontness differentiating them and some overlap (see section 5.3.1). Therefore the acoustic similarity of these vowels translated into only a small audible difference between the speakers with a Yorkshire accent and the speakers with a North East accent.

### 5.6.2 Differences across the difficulty levels

It was expected that the children would find it easier to group together the different speakers when they were heard saying the same sentence, featuring the same word as the two teachers (difficulty level 1). However, in both rounds 1 and 3, the children were found to perform significantly better in difficulty level 2, when the speakers they were grouping were heard pronouncing a different sentence, featuring the same phoneme as the teachers' sentences. This is contrary to the results of the Grouping experiment (chapter 4) in which it was found that children performed better at grouping together speaker guises when they were heard saying the same sentence, featuring the same word.

As shown in section 5.5.3.1, in round 1 the individual results for the first two sets of stimuli (difficulty level 1) were all lower than the results for the third set of stimuli (difficulty level 2). This suggests that no individual stimuli in particular caused the low overall results in these sets and that neither one of the FACE or the GOAT vowel distinctions was more conspicuous to the children. Furthermore, as the children performed well above chance for each of the individual stimuli, it is important to acknowledge that the difference in performance between the difficulty levels is a difference between high overall scores.

The consistently higher performance for the third set of stimuli in round 1 could indicate a practice effect; it might indicate that the children performed better in the third set of stimuli because they had become accustomed to the experiment design and therefore were more confident about how to complete the task, as well as more familiar with the accent differences of the speakers. A different (but compatible) interpretation is that, in DL1, the children might have been more confused as to how to group the speakers because they heard the same sentence featuring the same word from each speaker. Therefore, the children might have paid more attention to the fact that the speakers were all saying the same thing and not have noticed any differences between the speakers. In DL2, on the other hand, the children heard different sentences featuring different words. Therefore, they already had a stronger initial indication of a difference between the speakers and this may have primed them to listen out for more differences.

A different explanation is more likely for the higher performance in DL2 of round 3. While the children's results for the DL2 stimuli in round 3 were fairly uniform, their results in the two DL1 sets of stimuli were more varied (see section 5.5.7.1). Therefore the children did not perform consistently higher for the DL2 stimuli. As explained in section 5.5.7.1, the difference between the children's scores in the different sets of round 3 is likely to be because in the third set of stimuli, the speakers' vowels formed two more distinct categories pertaining to a Yorkshire/North East speaker distinction. The difference between these sets of speakers was much more clear-cut than the difference between the speakers in the first two sets of stimuli. The range of results across the individual stimuli in the DL1 sets therefore reflects the varying degrees of difference between the vowel pronunciations of the two groups of speakers in these sets.

### 5.6.3 Effects of age and sex

Overall, the girls performed better than the boys in the Second grouping experiment. This is evidenced in the logistic regression models run for both rounds 1 and 3 of the experiment. In round 3, the girls were predicted to score better than the boys regardless of other factors. Although not a significant effect, this is supported by the results of round 2, which found that the girls performed better than the boys overall. In round 1 however, an interaction between sex and age was found. The girls' performance was predicted to stay high across the age range and although the model prediction shows a slight decline with age, this result is probably due to a couple of low scoring individual girls (see child 29 and 33 in Table 5.6) who were the only children to score below chance in this round. Therefore, the overall high performance indicates that the girls have already reached their peak performance in this task around the age of 6. On the other hand, a development effect was found for the boys as the older boys were predicted to perform better than the younger boys in round 1.

These results support the results of the previous Grouping experiment (chapter 4) in which the girls were found to significantly outperform the boys. Therefore, as the Second grouping experiment is a progression from the Grouping experiment, the distinction between the scores for the different sexes in both these experiments is consistent with the explanation that the boys' abilities are initially behind the girls, but that they are catching up as they get older. As suggested in the interpretation of the Grouping experiment (see section 4.4.2), one possible explanation for boys' lower performance in general could be that the task design itself is more orientated towards girls, as the stimuli featured all female speakers and the task was run by a female experimenter.

### 5.6.4 Effect of the children's exposure to variation, social class background and their individual variation in performance

In round 1 of the experiment the logistic regression model found that children's exposure to speakers of other regional varieties predicted a better performance. Those children who were reported to have regular contact with non-Yorkshire speakers from elsewhere in the UK performed better overall. This result endorses the general hypothesis of this thesis: that variation in their linguistic input helps children to interpret the variation that they encounter. Furthermore, six of the eight children who scored 100% in round 1 were pairs of

siblings. This also supports an input-based explanation as each sibling pair is likely to have experienced a similar linguistic environment. Therefore, these siblings are likely to have similar stores of exemplars which they can utilise in perception when grouping incoming exemplars with those already stored.

Although the children's parents' level of education was not found to be a statistically significant predictive factor, there was a clear difference between the top- and bottom-scoring individuals in round 1 of the experiment. The top-scoring children all have at least one parent with a postgraduate education whereas the bottom-scoring children mainly have parents who left education after high school. The role of the parents' education can be explained by considering what this may mean for the children's exposure to linguistic input from their parents. Studies investigating children's lexical development have found that children from higher class backgrounds, as measured by their parents' education, are exposed to more linguistic input because their parents tend to spend more time interacting and talking to them (Hart and Risley, 1995; 2003; Huttenlocher et al., 2007). As a result, children with more educated parents are found to have an advantage in the learning of vocabulary as well as other aspects of language development such as syntactic and semantic understanding (Hart and Risley, 2003). Exposure to more linguistic input could also play a role in children's development of accent awareness, as they are exposed to more linguistic variation as part of this input. However, any such effect would be fairly indirect and it is important to consider the nature of the differences in the input itself, rather than simply the quantity of the input.

Sociolinguistic studies of child-directed speech have found that parents and caregivers often use different patterns of variation when speaking to their children, compared to when they speak to other adults (Foulkes et al., 2005; Smith et al., 2013). This variation often relates to the use of standard vs. vernacular forms; parents often use more standard forms in speech to their children, in linguistic contexts where they would almost exclusively use the local, vernacular form with other adults. In this way, children are exposed to more variation in their input as they experience more variability with regards to these standard/vernacular forms (Foulkes et al., 2005:197). Moreover, the linguistic input that children receive inevitably reflects their parents own education and social class (see section 2.1.2). Children from higher class families are likely to be exposed to more standard forms at home, which

contrast with some of the accent features of their local community, and therefore they are likely to be exposed to more variation overall.

Another related aspect of the children's exposure to variation is in terms of their exposure to diphthongal pronunciations of the FACE and GOAT vowels in their local community. As described in section 2.1.3, there is a current change in progress whereby standard Southern diphthongal forms of FACE and GOAT are spreading to northern areas such as York (as described by Haddican et al., 2013). This is reflected in many of the children's own productions of the FACE and GOAT vowels (recorded before they took part in the experiment) which were found to be predominantly diphthongal. Only three children pronounced these vowels as monophthongs, more typical of a Yorkshire accent (as described in section 5.4.5). This reflects the imbalance in the children's backgrounds, as overall there were more children from the schools with a larger percentage of residents from a higher social class background (see section 5.4.1). The children were likely to be paying close attention to their speech while listing items on the screen and therefore their pronunciations were possibly susceptible to style-shifting. However, this is still an indication that the children's exposure to diphthongal pronunciations of these vowels derives from their own and their peers' productions as well as from their parents/ other middle class speakers in their local community.

In round 2, although not significant, a trend in the results was found relating to children's exposure to variation; children who had regular exposure to non-Yorkshire speakers were found to perform slightly better overall. In addition, the children's parents' level of education was a near significant predictor of their results in this round. Children who have at least one parent with a postgraduate level of education performed better overall. As described above, the children whose parents have a higher level of education are likely to be exposed to more variation in general. This variation in the children's input could therefore have an impact on their general ability to distinguish accent variation; it may help them to interpret other accent variation (i.e. between Yorkshire and Scottish accents) more readily.

The only two children with Scottish family were two of the three highest scoring children in round 2 (93% and 87% correct answers). This suggests that their exposure to Scottish accents in their family might have helped them to group speakers on the basis of a Yorkshire/Scottish accent distinction in this round of the experiment. As there were only

two children with Scottish family altogether, however, more children would need to be tested to substantiate this interpretation.

In round 3 of the experiment, no correlations were found between children's exposure to variation and their performance. Generally, the children's performances in this round were varied and not consistently connected to their overall performance. As Table 5.16 shows, while some of the children who scored the lowest in round 3 also scored low in round 1, others had very high scores in round 1. Overall, the variability in the children's results and the fact that only three children score the highest result of 80% demonstrates the general difficulty of this round. In turn, this suggests that the differences between the GOAT and FACE vowels of the Yorkshire and North East speakers were not conspicuous enough for most of the children to make a distinction between them (for further discussion see section 6.3.2.1).

## 5.7 Conclusion

The Second grouping experiment has found that primary school children are able to group different speakers according to their regional accent. Children were better at grouping speakers by a Yorkshire/SSBE regional accent distinction than a Yorkshire/Scottish or a Yorkshire/North East distinction. This indicates that some features diagnostic of regional accent distinctions are easier to differentiate than others and therefore that conclusions about children's abilities should not be made on the basis of one comparison. Furthermore, children's exposure to variation predicted their performance in the Second grouping experiment. This supports the results of the previous experiments in chapters 3 and 4 and demonstrates the importance of considering children's individual experience when analysing this kind of data.

# Chapter 6: General discussion

## 6.1 Overview of the investigation

This thesis set out to explore children's developing awareness of linguistic differences relating to regional accents and how speakers are grouped together according to these differences. The investigation was based on the hypothesis that the ability to group speakers by their accent arises from the initial ability to recognise individual familiar speakers. Therefore, the focus of this sequence of experiments was on tracking this development, from familiar speaker recognition, through to unfamiliar speaker categorisation based on different accent distinctions.

Each of the experiments carried out in this thesis has attempted to establish a stage in children's development of accent awareness. In turn, these developments have been analysed in an exemplar theoretic framework, in which it is proposed that accent awareness develops from the storing of exemplars from individual familiar speakers. The findings from the experiments and interpretations of their results will be expanded on in the following sections. Section 6.2 addresses the question of children's development in regional accent awareness by summarising the results of each of the experiments consecutively. Section 6.3 then moves on to summarise the effects of the independent variables found to play a role in the children's performance across the experiments. Section 6.4 proposes a theoretical account of the findings from the experiments in this thesis, incorporating exemplar theory (ExT) with a model of speaker recognition based on studies in voice perception. Section 6.5 discusses the limitations of the current thesis, moving on to consider possible future directions. The section ends with an overall conclusion of the aims and contributions of the thesis.

## 6.2 Development in the pre-school/primary school years

Underlying the experiments is the hypothesis that speaker groups arise from the initial ability to recognise individual familiar speakers. Therefore, as little work on children's recognition of familiar speakers has been carried out, this was the line of investigation for the first two experiments in chapter 3. The Identification experiment tested children on their ability to identify familiar speakers and found that 2-4-year-old children performed above chance level when identifying six familiar nursery teachers. It was found that the

speaker with the most distinctive pitch and voice quality was the most easily identified teacher. The Recognition experiment went on to investigate whether the same children relied upon features of a familiar speaker's accent in order to identify her. This experiment found (1) that the familiar speaker was less likely to be identified if her accent was disguised via regionally variable segment realisations; and (2) that unfamiliar speakers were more likely to be misidentified as the familiar speaker if they had an accent similar to hers. Having established that pre-school children were able to identify familiar speakers, and that the speaker's accent (as well as pitch and voice quality) played a role in this process, chapter 4 moved on to investigate whether children of a similar age could group together unfamiliar speaker accent guises based on their accent. The Grouping experiment found that 3-4-year-olds performed above chance level at grouping speaker guises based on a Yorkshire/SSBE distinction.

Following on from these results from pre-school children, the next step in the investigation was to uncover the capabilities of slightly older children, attending primary school. This task was made harder in two main ways: the children listened to audio stimuli from a range of different unfamiliar speakers (rather than speaker accent guises) and the children heard three different accent distinctions. Therefore, the children's task was to overlook differences in the pitch and voice quality of individual speakers in order to group them by similar features of their accent. Due to the unreliability of previous results testing children in this ability (cf. Floccia et al.'s 2009 study, described in section 2.3.4), the extent to which children would be competent in the task was unknown. Therefore, an exploratory experiment design process was undertaken so that their abilities in this task were not underestimated. The final experimental design of the Second grouping experiment found that children performed well at grouping together unfamiliar speakers based on a Yorkshire/SSBE distinction. The children did not perform as well when grouping together speakers based on a Yorkshire/Scottish or a Yorkshire/North East distinction however (see section 6.3.2.1 for further discussion on this).

Overall, therefore, these four experiments show a progression in young children's awareness, from the use of accent features to help identify familiar speakers, to the use of accent features as grouping criteria for (1) unfamiliar speaker guises and (2) different unfamiliar speakers' voices. Furthermore, these experiments contradict the results of

previous experiments (Floccia et al., 2009; Wagner et al., 2014) which found that children under the age of 7 were not able to categorise speakers by a regional accent distinction.

Throughout these experiments, independent variables reflecting the children's backgrounds were investigated in line with the results. The next section summarises the variables found to play a role in children's performance and discusses their implications.

## 6.3 The role of independent variables

### 6.3.1 Child maturation

Due to the overall focus of the thesis being on the development of children's accent awareness, child maturation was an important consideration in both designing this series of experiments and in analysing the results. The children's age and sex were found to have a combined impact on their ability throughout the tasks; generally the older children performed better than the younger children throughout the experiments and the girls performed better than the boys. While it is impossible to completely separate maturational ability from the effects of an individual's experience (see section 6.3.2), separate effects for these independent variables were found in the statistical analyses. Age effects amongst young children should be treated with caution, however, as other factors such as vocabulary size have been found to have more of an effect on children's processing speed than age alone. For example, Fernald et al. (2001) found that vocabulary size, rather than age, was a better predictor of children's ability to recognise familiar words from hearing only part of the word. Their experiment found that amongst 18 month and 21 month-old children, those with a larger lexicon (over 100 words in production) were more accurate in this task. Although this experiment highlights the important role of other factors alongside age, consideration of these factors was beyond the scope of the current study. Therefore, for the present purposes, age is used as a rough approximation of general maturation; further investigation of the impact of other related factors, such as vocabulary size, are worthy of future study.

The girls performed better than the boys in the Recognition experiment, the Grouping experiment and the Second grouping experiment. Their consistently higher performance across these experiments, which progress in difficulty, suggests that they are generally more advanced in the abilities that these tasks are testing. This is supported by the results

showing an improvement in the boys' performance both in the Recognition experiment and the Second grouping experiment as they get older, indicating a developmental change in their ability.

There is some support for the more advanced ability of girls in this task being due to a biological, sex-related difference. Infant girls are generally found to mature physically at a faster rate than boys (Bornstein et al., 2004) and therefore they may have an early advantage in language processing tasks due to earlier brain maturation. Furthermore, there is some evidence to suggest that girls are better at language-related tasks. For example, studies have found that girls are better at tasks that require the accessing of phonological information in long-term memory and quick processing speed in perception (Halpern, 1997; Sternberg, 2004; Lundberg et al., 2012). Also, girls have been found to advance more quickly in vocabulary learning and word development (Huttenlocher et al., 1991; Fenson et al., 1994; Naigles, 1996). However, other studies have found no difference between cognitive processes, such as the working memory and cognitive flexibility of young boys and girls (Davidson et al., 2006; Cragg and Nation, 2009) and none of the previous experiments in infant/child accent awareness (such as those discussed in section 2.3.2 and 2.3.4) have reported differences in performance between the sexes.

Furthermore, it is controversial whether any of these language-related differences between the sexes are due to biological rather than social factors. There are studies showing that girls and boys receive different linguistic input from their parents, which reflects the use of different sociolinguistic variants by males and females (Foulkes et al., 2005). Therefore, it is also possible that this form of children's linguistic socialisation affects their overall perception of linguistic variation. For example, in Foulkes and colleagues' (2005) study, child directed speech (CDS) from mothers to their sons was found to contain a higher proportion of glottals for /t/than CDS from mothers to their daughters, which contained more standard [t]. The authors propose that "mothers are tuning their phonological performance in line with their child's developing gender identity" (Foulkes et al., 2005:198); they use the [t] variant more when speaking to their daughters, perhaps in order to encourage their daughters to use this more positively evaluated form themselves. This evidence for boys and girls being exposed to different levels of linguistic variation in their input is therefore a possible additional explanation for the difference in performance

between the two sexes in the current set of experiments. If girls are exposed to more standard variants in speech from their parents, there is more of a contrast between the linguistic input they receive at home and the localised forms they experience in their local community more generally. Therefore, this extra exposure to variation may help them to perceive phonetic variation, such as in these experimental tasks. As Foulkes and colleagues (2005:180) point out, there are very few other studies which have addressed the segmental features of CDS and, in particular, how the parents'/children's gender might play a role. Consequently, this explanation is fairly speculative and further evidence for the use of different levels of variation in CDS between boys and girls is needed to substantiate this account.

In terms of the current set of experiments, it is also important to consider whether the difference in performance between the sexes was also partly due to the design of the tasks themselves, which were possibly more engaging for the girls than the boys. As described in section 4.4.2, measures of children's attitudes towards sex differences have found that they show both explicit and implicit preferences for their own sex and its stereotyped associations at 4-years-old (Cvencek et al., 2011). Therefore, the girls may have performed better in the experiments, even from a younger age, partly because they had a more positive attitude towards the female speakers and the female pictures. This means that they were more likely to enjoy the tasks and therefore be more engaged in the experiments than the boys. An explanation due to engagement with the task is also supported by the difference in learning approaches which have been found between boys and girls. Studies of young children's classroom behaviour have found girls' more attentive and persistent approach to tasks in the classroom supports their literacy learning (Chiu, 2001; Ready et al., 2005).

The better performance of the girls throughout the experiments may therefore be due to a combination of biological/maturational, social and task-related factors. Alongside this difference between the sexes, however, there is a general age-related improvement in most of the tasks reported here. The youngest children (2-4-year-olds) were tested in the Identification and Recognition experiments reported on in chapter 3. In the Identification experiment, older children who had been attending the nursery for longer performed better overall. An age improvement was also found in the Recognition experiment, although only for the boys. Children's improvement in performance in these experiments throughout the

pre-school years is consistent with the results from Spence and colleagues (2002), who found that children's recognition of familiar cartoon voices improved between the ages of 3 and 5. As described in section 3.4, the process of identifying a familiar speaker can be explained through a cognitive model of recognition, based on the retrieving of stored instances as well as the accessing of more abstract information (Diana et al., 2006). In such a model, when a listener hears a speaker, previously stored instances (or exemplars) of that speaker are activated, as well as indexical information relating to the speaker (their name, for example). Explicit memory, which is used to remember particular episodes, is strongly age dependent and has been found to improve in the pre-school years (e.g. Drummey and Newcombe, 1995). This helps to explain children's improvement in the Identification and Recognition experiments throughout the pre-school years; in these early years, children improve in their ability to store and access exemplars encoded with speaker-specific information.

An age improvement was also found between 3 and 4 year olds in the Grouping experiment, which tested children's ability to group speaker guises based on their accent. This task was focused on the children's familiarity with accent distinctions rather than individual speakers. Therefore, the results demonstrate a progression in children's awareness of accent differences between the ages of 3 and 4. The most marked improvement across the age range was the performance in difficulty level (DL2) of this experiment. In DL2, children grouped speaker guises across different pronunciations of the same phoneme, featured in different words. Therefore, the improvement with age in this particular task reflects a development in the children's understanding of phonemic categories and their different phonetic realisations. This shows a progression in the children's abilities from younger infancy, when they have been found to be able to abstract across different accent realisations in order to recognise familiar words (cf. Schmale et al., 2010; Best and Kitamura, 2012). Further to abstracting from phonetic variation in order to generalise across words, the pre-school children in the Grouping experiment demonstrated the ability to use this variation in order to group the speakers. This result somewhat contradicts previous studies (see section 2.3.4), which found that children under the age of 7 could not abstract across different accent realisations when grouping speakers (Floccia et al., 2009) or reproducing

pronunciations (Nathan et al., 1998). The results from the Grouping experiment indicate that rather being determined by age, the ability is more likely to be experience-related.

The progression from pre-school age children in the Grouping experiment to primary school children in the Second grouping experiment marks an important stage in children's development. When they start school, children's acquisition of the phonology of their local accent often accelerates as they interact with more children from the local area (as found by Tagliamonte and Molfenter, 2007). In round 1 of the Second grouping experiment, the children were tested on their ability to categorise different speakers based on a Yorkshire/SSBE distinction. In this round of the experiment, the boys were found to improve between the ages of 5 and 9. This affirms the girls' overall higher performance throughout the experiments and the boys' slightly later developing ability in these tasks.

The absence of an improvement with age in the other rounds of Second grouping experiment reflects the limit of an account of accent awareness based on age alone. As described in section 2.3.3, adults are very variable in their ability to differentiate between accents, and can generally only group speakers into categories based on very broad dialect areas (Clopper and Pisoni, 2004a). Variable performance in adulthood indicates that individual variation is based on more than just a listener's age. Although maturational, age-related developments play a role in childhood, it is hard to disentangle these developments from the more individualistic aspect of the children's overall experience with linguistic variation; ultimately as children get older they encounter more speech and therefore more linguistic variation.

A child's phonological systems may not have reached an adult-like level until the age of 7 (cf. Sander, 1972). A phonological account based on this premise was proposed by Nathan and colleagues (1998). As described in section 2.3.4, their study found that 4-year-olds were more likely to phonetically reproduce words heard in a different (Glaswegian) accent, whereas 7-year-olds were more likely to pronounce these words in their own (London) accent. The authors explain this difference as due to the 7-year-olds having more developed, phonologically abstract representations than the 4-year-olds, who had not experienced as much accent variation in their linguistic input and therefore were not as adept at abstracting across unfamiliar phonetic variation in line with their phonological representations. Therefore, despite this difference being age-related, it is also inevitably experience related.

While the Grouping experiment reported here found that 3-4-year-olds were able to abstract across phonetic variation to group speakers at the level of phonemes, this might be limited to the particular phonetic realisations encountered in the experiment and not necessarily be widened out to account for less familiar phonetic variation. In the Grouping experiment, the Yorkshire/SSBE pronunciation differences between the BATH and FACE vowels are likely to have been experienced by most of the children. However, in Nathan and colleagues' study, the Glaswegian accent was specifically chosen as being one that was unfamiliar to the children participating in their experiment. Therefore, once children have reached the point at which they can categorise sounds into phonological categories, much of the rest of their success in accent perception is likely to be on account of their overall experience with variation.

### 6.3.2 Input/Exposure

The role of experience in a listener's ability to perceive accent differences is multifaceted. Fundamentally, a listener's experience relates to their quantity of exposure to variation, which in turn relates to their age and their social background. However, there is also the issue of how a listener personally relates to the variation they encounter. An important consideration is how much attention a listener pays to their linguistic input and how this is mediated through their personal/social group attitudes. As described in section 2.4.2, Smith and Zárate's (1992) exemplar model of social judgment proposes that our social judgments are influenced by specific representations we have stored in memory, which arise from individual encounters. The social information that we store as part of these specific representations is then accessed when similar stimuli are encountered. The particular representations and social information that are stored depend on the attention we pay the exemplar stimuli we encounter. In turn, this attention depends on our personal/social group attitudes and motivations. For example, we are more likely to pay more attention to attributes of a stimulus that we personally relate to and we are also more likely to pay attention to individual attributes of those in our social 'in-group' compared to group attributes of those in a social 'out-group'. These, along with many other potential social motivations, have an influence on how we then interpret future encounters with similar stimuli.

An analogous process is presented in a sociolinguistic exemplar-based account, in which it is proposed that our cognitive development of speaker groups is based upon the storing of individual linguistic exemplars (see section 6.4). These linguistic exemplars are encoded with social information, pertinent at the time of encounter and what makes this information pertinent depends upon many different aspects of our social motivation (as described above).

More work is needed to further investigate this link, between what kind of social motivations cause us to pay more attention to certain stimuli and how the accessing of these stimuli in perception can influence our interpretation of newly encountered stimuli. Work in this area would be of interest to sociolinguistics as it would encourage exploration into how our language attitudes relate to our memory processes. This, in turn, could tell us more about how language attitudes, such as linguistic stereotypes, develop as they build up from the social information we store and access as part of our linguistic exemplars (see section 7.4 for possible future directions for this research).

An additional issue in the investigation of measuring exposure to variation is the fact that the variation itself is multifaceted. As described in section 2.2.1, studies have found that pre-school age children pick up on adult-like patterns of sociolinguistic production (Roberts and Labov, 1995; Foulkes et al., 1999; Smith et al., 2007; Barbu et al., 2013), learning how linguistic variation relates to social factors such as the relationship between local/vernacular and standard variants, norms of gender use and stylistic shifting.

These studies are an indication of the many different ways in which variation is encountered and therefore the difficulty in accounting for exposure to this variation as an overall measure. Therefore, in line with second dialect acquisition studies (see section 2.2.2), as well as Beck (2014) and van der Feest and Johnson (2015), who were investigating similar perceptual abilities, the experiments in the current thesis used the outside/insider status of children's parents as a criterion for differentiating between the children's different levels of experience with variation. The performance of children with both parents from outside of the local, Yorkshire region was therefore compared to the performance of children with at least one parent from the local region. Underlying this division was an assumption that children with parents from elsewhere were more likely to experience more variation pertaining to regional accent distinctions, through being exposed to both speech

from their local community and speech from their parents, who were not from the local community. As the parents' speech was not recorded, it is not possible to know (1) how different their speech was in comparison to the local community and therefore (2) the exact levels of variation that the children were exposed to. However, in line with previous studies, this division was used as a proxy for variation. Due to the limited number of participants who had parents from a wide range of regions, this binary distinction was made as categorising the children by more specific regional distinctions would have resulted in very few/only single individuals in each category. In the Second Grouping experiment, a further measure of exposure was analysed; this measure took into account parents' self-reported details of their children's regular exposure to regional variation from other speakers who the children regularly encountered. Children were also categorised along a binary distinction with regards to this distinction (regular exposure to other non-Yorkshire speakers/not). Again, as this was an indirect measurement of the children's exposure, the results must be treated with some caution. Nevertheless, it is hoped that these measurements and their resulting significant effects can form the basis for an account drawing attention to children's exposure to variation in the input.

#### *6.3.2.1 Exposure to regional variation from family and friends*

The accumulation of the results from this thesis found that, in general, children with non-local parents and/or regular exposure to other non-local speakers performed better in tasks which required identifying speakers and grouping speakers according to their regional accent. While the children's parents' local/non-local status also linked to other factors related to the children's social class background (such as their parents' level of education and the school they attended – see section 6.3.2.2), these results are interpreted as an indication that children who are exposed to more variation in their input are better at distinguishing between speakers based on their regional varieties.

The finding that exposure to variation plays a role in children's developing awareness of regional accents supports an ExT account of the development of social-indexical knowledge. As described above, at the heart of an exemplar model of memory is the idea that individually encountered stimuli are stored with social detail. Therefore, our experience with individual stimuli, in this case in the form of linguistic exposure, is central to such a model. In such an ExT account, individual exemplars and the social details they contain are

then grouped together according to these details, at higher levels of abstraction, to form speaker categories (see section 6.4 for further discussion).

In order to account for the development of social-indexical knowledge alongside linguistic knowledge, this thesis advocates a usage-based account of language acquisition (as described in section 2.4.1). Thus, the role of linguistic input and exposure to variation has been one of the main points of focus in analysing the results. In the Recognition experiment, girls with non-local parents were significantly better at recognising a familiar speaker from one-word utterances featuring regional accent variables than girls with local parents (the boys performed worse overall, see previous section). Furthermore, in the Grouping experiment, both the boys and girls with both parents from outside Yorkshire were found to perform better in DL3. In DL3, the children's task was to group Yorkshire/SSBE accented speaker guises across their pronunciation of different phonemes. Finally, in round 1 of the Second grouping experiment, children who had regular contact with non-Yorkshire friends and family from different regions of the UK performed better at grouping together different speakers based on a Yorkshire/SSBE accent distinction.

Results from previous studies have not been consistent with regards to the impact of variation in children's input on their ability to perceive accent differences. In particular, studies with 20-24-month olds have disagreed on whether the local variety overrides the mixed input that children with non-local parents receive; there is disagreement as to whether a mixed input helps or hinders children to recognise words heard in different accents. As discussed in section 2.3.2, the results of van der Feest and Johnson's (2015) study are the most convincing and also the most comparable with the results of the experiments in this thesis. This is because their study categorises children with a mixed input in a similar way (children with both non-local parents as opposed to children with one or more non-local parents). Their study found that 24-month-old Dutch children who received a mixed input (parents who spoke a different variety of Dutch compared to their local variety) were able to adapt how they processed the speech of speakers with a local accent (in which all fricatives are devoiced) compared to speakers with their parents' non-local accent (in which there is a contrast between voiced/devoiced fricatives). Children with a mixed input were able to detect mispronunciations by speakers with a non-local accent whereas children with a more uniform input (i.e. who had local parents) were not able to

detect these mispronunciations. Therefore, this study demonstrates an advantage for 24-month-olds who have had more exposure to regional variation; they can use their experience of accent variation to inform their interpretation of the speech that they hear. Similarly, in the experiments of this thesis, children who have had more exposure to regional variation may be using this experience to interpret the accent differences that they hear in testing. In ExT terms, we can infer that the children have more exemplars of speakers with different accents stored in memory. The children may therefore be better able to contrast the accents of the speakers that they hear in the experiments because they can draw upon the social-indexical information embedded in these exemplars. The disadvantage for children who have a more uniform input can therefore be interpreted as their having a more specific and less flexible perceptual phonetic space (in Evans and Iverson's, 2007 terms) built up from their more homogenous store of exemplars. Their perception of variables outside of this space may be somewhat hindered, as they have fewer stored exemplars with which to compare the incoming exemplars.

Contrary to the findings in this thesis, Beck (2014) did not find that 5-6-year-olds children's ability to differentiate between regional accents in an ABX discrimination task was affected by their exposure to regional variation. In her experiment, children listened to six different vowel distinctions in two regional accents (Philadelphian and General Southern U.S. English). Beck did find a difference in the children's performance depending on the vowel that they heard and therefore is it possible that she may have found an exposure effect for the vowels that the children performed best with. Despite the lack of an exposure effect, however, Beck did find a correlation between the children's awareness of regional variation and their performance in this task. Children who were able to correctly label a speaker as being 'local' performed better in the ABX discrimination task, which relied upon them differentiating speakers based on a local/non-local accent distinction. On the other hand, there was no correlation between children correctly labelling 'non-local' speakers and performing better in the ABX task. Beck therefore concludes that, overall, the children had conceptualised a local/familiar speaker category but not necessarily a non-local/unfamiliar one. She supports this analysis with results from another part of the experiment in which each child was asked to listen to two different people say the same word and choose the speaker who sounded most like them. In each trial, the children heard one local and one

non-local speaker. Beck found that children with non-local parents had a 'confused' local category and were therefore less likely to identify local speakers as sounding like themselves. This highlights the fact that, for children with non-local parents, there is no necessary connection between familiar speakers and local speakers; their familiar speaker category is made of both local and non-local speakers.

There is evidence from studies in second dialect acquisition that children with non-local parents are not overtly aware of differences between the speech of their parents and the community until early adolescence at least (e.g. Chambers, 2002; Hazen, 2002). However, this overt awareness does not seem to be important for children's overall perception of variation. As the experiments in this thesis have shown, even from a pre-school age, children's increased exposure to variation can help them in their ability to recognise and group speakers based on regional accent features. Furthermore, the results from the experiments in this thesis provide some evidence that exposure to a particular non-local accent helps in the ability to group speakers with this accent. For example, children with Southern parents were amongst the highest performers in both the Grouping experiment and the Second grouping experiment when grouping together speakers based on a Yorkshire/Southern speakers. Also, in the Second grouping experiment, the two children with Scottish family performed best overall when grouping speakers based on a Yorkshire/Scottish distinction. It is possible to explain these findings as an extension of what Stevenage and colleagues (2012) refer to as the 'other-accent' effect. The 'other-accent' effect describes the fact that listeners are better at identifying speakers who have the same accent as them (see section 2.5.3). In modifying the description of this effect to account for the results from this thesis, it is proposed that 'other' could, instead, exemplify 'unfamiliar' and that 'identifying speakers' could be adapted to 'grouping speakers'. Therefore, as well as helping in the identification of speakers, the familiarity of an accent can help in the grouping of speakers on the basis of this accent.

This proposition echoes Beck's (2014) description of a 'familiar speakers' category, which may include both local and non-local speakers if a child has non-local parents. In an ExT account (see section 6.4), it is suggested that a 'familiar speakers' category contains sub-categories pertaining to the individual speakers themselves; each familiar speaker forms their own category of exemplars, while a broader category of 'familiar speakers' emerges

from an abstraction across these subcategories. Therefore, in this approach, rather than having a 'confused' local speaker category (as Beck proposes), children with non-local parents have different subcategories as part of their overall 'familiar speakers' category.

Despite the overarching evidence that variation in the input helps children to perceive accent differences, no exposure effects were found in round 3 of the Second grouping experiment. In this round of the experiment, the children grouped different speakers on the basis of a Yorkshire/North East accent distinction. Neither children's overall exposure to variation, nor their exposure to North East varieties in particular, seemed to play a role in their ability to group speakers based on this accent distinction. The difference between these two groups of speakers' FACE and GOAT vowels were much more fine-grained than the features pertaining to the other accent distinctions. As the pronunciation of these vowels only differed by small degrees of height and frontness, it is likely that these were insufficient differences for many of the children to perceive.

There were also no effects of general exposure to regional variation on the children's ability to group the speakers based on a Yorkshire/Scottish distinction in round 2 of the experiment. As suggested in section 5.6.1, this could be due to the characteristic variability of rhoticity found amongst Scottish speakers (Lawson et al., 2011); the Scottish speakers in the stimuli may not have pronounced the stimuli with a consistently substantial level of rhoticity for the children to be able differentiate them from the Yorkshire speakers. As mentioned above, the two children with Scottish family members performed best in this round of the experiment and therefore it is possible that these children's direct exposure to Scottish speakers helped them to group speakers based on a Yorkshire/Scottish distinction. Although many of the children are regularly exposed to rhoticity through American accents in the media (see next section), this more indirect form of exposure perhaps does not contribute to their perceptual awareness as direct exposure to speakers that they interact with. Furthermore, as described in section 5.6.1, the differences between rhotic variability in American compared to Scottish speakers may mean that is too complicated to equate the two types of rhoticity; whereas levels of rhoticity are socially stratified in for Scottish speakers, this is not the same for American speakers where levels of rhoticity have been found to be more speaker-specific (cf. Mielke et al., 2010)

#### *6.3.2.2 Other forms of exposure to variation*

Other factors connected to the children's exposure to variation were explored in relation to the results of the Second grouping experiment. These factors related to the children's school, their parents' education and their exposure to variation in the media. While these forms of exposure to variation were not found to be statistically significant, they did contribute to an explanation of the trends and some patterns of individual variation. Insufficient background information was collected from the parents of the children who took part in the other experiments, and therefore these variables could not be considered systematically in the analysis of their results.

Children from three different schools took part in the Second grouping experiment. The children who attended the school with the lowest academic rating, with the most disadvantaged pupils, performed worst in the experiment overall. However, as explained in section 5.5.4, due to the uneven distribution of children from each of the schools, this variable was not investigated further. Rather, other factors which were a good reflection of these differences between children at the different schools were investigated instead. As well as the regional origin of their parents (described above), the education level of their parents was considered in the analysis. Although not a significant factor, in both rounds 1 and 2 of the Second grouping experiment, most of the top performing children had at least one parent with a postgraduate degree. As described in section 5.6.4, this better performance can be attributed to the likelihood that children from higher social class backgrounds (measured through means such as their parents' education) are more likely to be exposed to more variation in child-directed speech. This is because, as Foulkes and colleagues (2005) found, parents tend to use more standard forms in CDS compared to the higher proportion of vernacular forms they use with other adults. Therefore, children with higher class parents who use more standard forms in their CDS, are likely to experience more variation in their linguistic input. This is because their parents' use of standard forms contrasts with the vernacular forms of their local community.

The final proxy for variation, which was only investigated informally, was children's exposure to regional variation in the media. Background information collected from the parents of participants in the Second grouping experiment found that all of the children watched television programmes featuring speakers of different varieties of English (see

section 5.4.6). Many of these programmes were American, indicating a considerable amount of exposure to American accents amongst the children. However, there were also many programmes which featured other regional varieties of British English. While a systematic study of the impact of this variation is beyond the capacity of this thesis (see section 6.5 for potential future directions on this), the collection of this data at least indicates the variety of programmes young children are watching and the potential for the accent varieties heard in these programmes to affect children's perception of this variation (as Lippi-Green, 2012 has suggested – see section 2.3.1).

## 6.4 A theoretical account

Overall, the results from the experiments in this thesis have revealed two main findings. Firstly, there is a progression in children's accent awareness throughout the pre-school and primary school years. This progression can be tracked from the use of accent features in the recognition of familiar speakers, to the grouping of unfamiliar speakers based on features of their accent. Secondly, children's accent awareness is affected by different factors which, above all, relate to their exposure to linguistic variation. This exposure comes in different forms; the children's age and sex, their social class background and most notably their exposure to regional variation from their parents and close family/friends.

These two findings can be combined in an explanation couched in an ExT framework, incorporating findings from familiar speaker recognition. Here, it is being proposed that research in voice perception, which has investigated the processes of familiar speaker recognition and unfamiliar speaker discrimination, can inform and be integrated into an ExT account of the development of social-indexical knowledge. ExT is increasingly being used by different areas of linguistic research in order to account for the link between social and linguistic information (as described in section 2.4.2). ExT gives us an insight into the developing link between social and linguistic information in a way that many models of language acquisition do not appear to do. As described in section 2.4.1, formalist models of language acquisition traditionally focus on describing children's learning of the abstract patterns of their native language(s) and are not concerned with also explaining their acquisition of social-indexical knowledge alongside the linguistic. Therefore, a usage-based account, and in particular ExT with its emphasis on the importance of individual instances, is proposed as the best way to describe this relationship. This is not to say that

individualistic episodic memories can solely account for this relationship, however; the role of abstraction is still important. As described in section 2.4.2, amongst investigators who promote exemplar theory, there is general consent for a hybrid model of language development, which advocates a role for both episodic memories and abstractions across these episodes. In relation to the current investigation, in such a model it is the storing and abstraction across speakers' individual exemplars that leads to the development of perceptual accent categories.

Voice perception studies have investigated the different neural processes involved in familiar speaker recognition compared to unfamiliar speaker discrimination. As described in section 2.5.1, Kreiman and Sidtis (2011) propose a framework describing the differences but also the links between these two processes. Whereas when we process a familiar speaker's voice we tend to recognise features of their voice in an overall pattern (using the right hemisphere of the brain), when we hear an unfamiliar speaker, we rely more upon analysing particular features of their voice (using the left hemisphere of the brain). Kreiman and Sidtis propose that these two neural processes (pattern recognition vs. featural analysis) are used to different extents when processing familiar vs. unfamiliar speakers' voices. Therefore, in this framework, unfamiliarity-familiarity lie on a continuum along a scale of homogeneity-heterogeneity (see Figure 2.3); in this respect an unfamiliar set of voices are more homogenous to a listener than a familiar set of voices.

This framework of familiarity-unfamiliarity from work in voice perception can be associated with an exemplar-based account of memory which suggests that when making social judgments, we tend to pay more attention to individually distinguishing features of those we know better (members of our social in-group), whereas we pay more attention to group level attributes of those whom we do not know (members of out-groups) (Smith and Zárate, 1992:14). Therefore, by integrating these accounts, we can suggest that, when processing a speaker's exemplars, the more familiar they are or the more similar they are to speakers in our social in-group, the more likely we are to attend to both their distinctive vocal features and their distinctive social attributes.

This interpretation can be further integrated into an ExT account of the building of speaker groups from abstractions across individual instances over time. Just as Pierrehumbert (2003) suggests that phonology is built from the categorisation of phonetically similar

exemplars through statistical learning (as described in section 2.4.2), in a similar way, social categories develop as abstractions over stores of exemplars which are also encoded with social detail. In Smith and Zárate's (1992) exemplar-based account of social judgment, they describe some of the possible motivations for attending to particular social attributes of exemplars; the attention we pay to stimuli depends on influences relating to the preservation of the link between our personal and group-level social relationship (see section 2.4.2). Therefore, it is proposed that categories of speaker develop over time to reflect relevant social and/or linguistic distinctions. These categories can develop at different levels of abstraction; for example a 'familiar female speakers' category might develop but also include the subcategory, 'familiar female speakers from work(/school/nursery)'. Foulkes (2010) proposes that phonetically transparent categories (such as speaker sex) develop first, before less transparent categories (such as speaker accent) – see section 2.4.2.

Furthermore, an ExT account of the building of speaker categories (such as Foulkes 2010), suggests that these categories initially develop from the storing of familiar speakers' exemplars. Therefore, a speaker category pertaining to a regional accent will build up from the storing of exemplars from speakers with that accent, the most likely of which to be stored are familiar speakers. The recognition of a familiar speaker relies upon the recognition of the many different features which make up a speaker's voice. These features (such as voice quality, pitch and regional accent) can be biologically or socially determined to different extents (see section 2.5.1). As described above, studies in voice perception (e.g. Kreiman and Sidtis, 2011) suggest that familiar speakers' voices are rapidly stored in memory as fairly holistic patterns of features and therefore that a process of general pattern recognition is employed when a familiar speaker's voice is recognised.

ExT can explain the pattern recognition process involved in familiar speaker recognition as well as how familiar speakers come to be recognised quickly. In an ExT account, exemplars from familiar speakers are more often stored as they are encountered more frequently. Additionally, as described above, individual social attributes of a familiar speaker are more likely to be stored and accessed in their linguistic exemplars, as these speakers form part of a perceiver's social in-group (an in-group of familiarity). In order to fit into our social in-group, we are invested in remembering social attributes of its members, and therefore we

are more likely to pay attention to individual attributes of familiar speakers. Therefore, in many circumstances, the more familiar a speaker is, the more likely their linguistic exemplars are to be stored with social information, and a category of exemplars, in the form of specific memories of their speech, can build up very quickly. When recognising a familiar speaker, it is the category of exemplars pertaining to specific memories of their speech which is accessed. This corresponds to the 'pattern recognition process' suggested by voice perception studies (mentioned above); the overall pattern that we recognise comes from accessing a familiar speaker's store of exemplars.

When it comes to discriminating between and grouping together unfamiliar speakers, there are no exemplar categories of these speakers stored in memory. Therefore, the listener compares an unfamiliar speaker's exemplars with the exemplars of other speakers that they do have stored in memory. As described above, voice perception studies propose that unfamiliar speaker discrimination relies more upon comparing particular features of the voice, rather than the holistic process of pattern recognition employed in familiar speaker recognition. Therefore, incorporating this evidence into an exemplar account, we can propose that a listener compares particular features of an unfamiliar speaker's exemplars with their stored exemplars and exemplar categories. Which particular features are compared will depend on the exemplar categories that the listener has already stored in memory as well as which particular social dimensions they are paying attention to at any particular time. As Smith and Zárate (1992) indicate, social judgements are numerous and fluctuating and therefore the processes of exemplar storage, categorisation and retrieval are very much dependent on the context and circumstances of an individual's encounter with each exemplar.

In general terms, however, an ExT account would propose that the more exemplar categories (containing both linguistic and social information) that a listener has stored in memory, the more categories they have to compare unfamiliar speakers' exemplars with. Furthermore, if a listener already has stored exemplars pertaining to the distinctions between a group of unfamiliar speakers, this will help the listener to categorise members of this group, based on these distinctions. Therefore, if a listener has a set of stored exemplars from speakers who can be categorised according to a social distinction such as their regional

accent, this will help the listener to categorise incoming exemplars on the basis of their similarity to these stored exemplars.

The experiments in this thesis have demonstrated different stages in the account of speaker recognition, discrimination and categorisation described here. The Identification experiment established pre-school children's ability to access their stores of individual speaker exemplar categories in order to identify familiar speakers. The Recognition experiment then found that a familiar speaker's segmental accent features formed part of this identification process. This experiment found that when the familiar speaker's accent was disguised the children were less likely to recognise her, which implies that the segmental phonetic information in her voice did not closely match that of the children's stored exemplars of her voice. Therefore, this caused impairment in the children's ability to access their stored exemplars of this speaker and identify her from the social information stored as part of these exemplars. Furthermore, an unfamiliar speaker was misidentified as the familiar speaker if their pronunciation of vowels featuring a regional accent distinction were phonetically similar enough to the vowels of the familiar speaker to activate stored exemplars of the familiar speaker's voice. In the Grouping experiment, pre-school children were found to be able to group together speaker guises on the basis of accent distinctions. The Second grouping experiment found that slightly older children were able to group together different speakers according to their accent. Children with more exposure to regional variation were found to perform better in both the Grouping and the Second grouping experiments. These children were able to compare the unfamiliar speakers' exemplars with more accent exemplar categories that they already had stored in memory. This helped them to categorise the unfamiliar speakers based on these accent distinctions; the unfamiliar speakers' exemplars were compared to exemplar categories created by familiar speakers.

Overall, therefore, we can account for the children's progressive abilities throughout the experiments in this thesis by proposing the continuous building and adapting of exemplar categories which are used as measures of comparison in perception.

## 6.5 Limitations of the present research

Due to the scarcity of findings from previous investigations into children's awareness of regional accents, the experiments in this thesis were exploratory in nature and therefore there were several instances of experimental re-design (particularly for the Second grouping experiment). As the overall focus was on children's emergent awareness of regional accents, it was necessary to scale the experiment designs in order to capture these developments. Inevitably, this meant that the experiments were limited in their scope as they were focused on basic-level abilities. Furthermore, the limited attention span of young children meant that the extent of the experiments' findings were additionally restricted as the experiments needed to be kept short. Therefore, although the Grouping experiment established a progression in the ability of pre-school children to group Yorkshire/SSBE speaker guises on the basis of their accent, it was not possible to further investigate whether particular accent features were more or less effective in enabling the children to make this distinction. Future work on a systematic investigation into the relative strength of association of different accent features with particular accents would therefore be worthy of investigation. In similar vein, investigating the effect of different accent features on the results of the Second grouping experiment would be valuable in order to see if there are discernible differences in how well these features characterise the accent distinctions for the children.

The experiments were also limited in their measurement of the independent variables analysed in the results. In the Recognition and Grouping experiments, only information pertaining to the parents' regional background was collected. In the Second grouping experiment, however, additional information was gathered about the children's exposure to other regional varieties through their contact with friends and family from other regions of the UK. While this information was found to be an important predictor of the children's results, it was still only a simplified representation of the children's exposure to regional variation. In order to validate the findings relating to children's exposure, it would therefore be beneficial to formulate a more comprehensive way of measuring their exposure to regional varieties. This would be a more labour-intensive process and require more work on the part of the children's parents (for example, by keeping a detailed account of their children's interactions). Similarly, as mentioned in section 5.4.6, a more extensive

exploration of children's exposure to regional variation in the media would be a valuable addition to an overall measurement of variation in children's input. Again, such an investigation would be demanding both for the researcher and the parents as the children's television-watching habits would need to be tracked in considerable detail. Furthermore, the children's engagement with, as well as their exposure to, different forms of media entertainment would need to be explored in such an investigation. As Stuart-Smith and colleagues (2013) found in their investigation of language change in Glasgow, a strong psychological attachment to a television drama was one significant factor influencing this change.

Another limitation of the results of these experiments is their lack of consideration of children's overt awareness of regional accent differences. Although the experiments in this thesis have discovered a certain level of awareness amongst children towards regional accent differences, they did not investigate whether children are overtly aware of these differences. Anecdotally, some of the older children expressed an awareness that these differences had something to do with where the speakers were from, but empirically investigating children's explicit knowledge about regional accents was beyond the scope of the current study. Beck (2014) found a correlation between children's awareness of local/non-local speakers and their performance in an accent discrimination task. Therefore it would be interesting to track the development of children's metalinguistic awareness of how a speaker's accent relates to where they are from; it could then be investigated whether this explicit awareness helps in children's general perceptual awareness of these accent differences overall. Children's explicit awareness of regional accents also links to a proposed future direction for the present research (see section 7.4).

## 6.6 Summary

This chapter has discussed the overall findings from the results of the four experiments carried out for this study. Overall, these experiments found that children's awareness of regional accent develops from a younger age than previously investigated and that this ability is affected by maturational development such as their age and sex, as well as effects of their exposure to linguistic variation. The following chapter will conclude the study by presenting a final overview of the research findings in the context of previous research, as well as suggesting future directions for further research in this area.

# Chapter 7: Conclusion

## 7.1 Scope of the thesis

This thesis has attempted to address an imbalance in the research relating to children's sociolinguistic development in the pre-school and early primary school years. Sociolinguistic development, here, refers to children's establishing of the social principles of variation - their learning of the indexical properties of linguistic variants, including deciphering who is more likely to use particular variants and in which particular contexts. While there is a growing body of research demonstrating children's acquisition of regional variation in their own production (Roberts and Labov, 1995; Foulkes et al., 1999; Smith et al., 2007; Barbu et al., 2013), there is a lack of research into how this relates to their perception of such variation.

The few studies that have attempted to address the question of young children's awareness of regional variation have come to conflicting conclusions, perhaps largely because they have used quite different methodologies. Whereas Floccia and colleagues (2009) found that 5-year-olds were not able to group speakers by a regional dialect distinction, Beck (2014) found that children of 5 years show the more basic ability of being able to match together speakers based on one-word stimuli featuring regional accent features. The design and the analysis of Floccia and colleagues 's study, however, shows a lack of focus on the linguistic detail; the children listened to different stimuli to each other and therefore there was no control over which particular accent features they heard. This casts doubt on the reliability of the results, as there is no way of knowing which accent features the children may have attended to when they were grouping the speakers. Furthermore, there is an absence of research focusing on younger, pre-school children's accent perception.

While studies on accent perception have been carried out with infants, this work has mainly focused on infants' word learning. It has been found that infants are initially sensitive to accent differences and unable to recognise familiar words in an unfamiliar accent. However, by the age of about 18-months, children are able to generalise across accent differences and understand familiar words heard in an unfamiliar accent (cf. Schmale et al., 2010; Best and Kitamura, 2012). Beyond these word-learning abilities in infancy, the perception of regional variation by pre-school children before the age of around 5 years does not appear to have

been considered. Therefore, in an attempt to fill this gap in knowledge, this research began the investigation of children's awareness of regional accent by exploring the abilities of pre-school children from 2-years-old.

## 7.2 Overall findings

In relation to the research questions set out in the introduction (see section 1.3), this research presents four main findings:

(1) There is development in children's awareness of regional accent in the pre-school and primary school years.
(2) Independent variables, relating to children's maturation and their exposure to regional accent variation, contribute to their developing awareness of regional accent.
(3) This research supports evidence for an important link between social/speaker-specific information and linguistic information, which are proposed as stored together in memory and then processed together in perception.
(4) The cognitive storage of this social/speaker-specific information is best explained in a theoretical model which relies upon the development of speaker categories building up from individual exemplars.

Overall, the experiments found that children's awareness of regional accent distinctions emerge at a younger age than established in previous studies. In particular, it was found that 3-4-year-old pre-school children were able to group speakers according to phonological regional variables, an age group which had not been previously tested for this ability before to my knowledge.

Results from across the experiments found significant effects relating to the children's age and sex; in general, the older children performed better than younger children and girls performed better than boys throughout the experiments. The most revealing finding, however, was the effect of children's linguistic input on their performance in these tasks; children with more exposure to regional variation were found to perform better overall.

The outcomes of these experiments have been found to be best accounted for in an exemplar theoretic framework, in which linguistic instances and social information are stored in memory together, and in which categories of speaker emerge from abstractions across these

stored exemplars. Furthermore, findings from voice perception studies in speaker recognition and discrimination have also been incorporated into an ExT account of this development. It is hoped that this research will form the foundation for future research into important considerations such as developing social attitudes towards members of speaker groups (see section 7.4).

## 7.3 Contribution of this thesis to wider research fields

This research has contributed to the broad fields of sociolinguistics and speech perception by investigating children's developing awareness of sociophonetic variation in relation to regional accent distinctions. As described in the introduction, while work in child language development has not generally focused on the acquisition of variation, studies in sociolinguistics and speech perception have generally focused on studying adolescents or adults. Therefore, the current research adds to the growing body of socioperceptual research by providing results from children, who are under-represented in this interdisciplinary research area. In particular, this research has uncovered new information regarding children's ability to group speakers according to phonological regional variables; it has been found that children can do this from a younger age than previously established and that while younger children can group different speaker guises, older children are able to group different, unfamiliar speakers according to these variables.

The methodological design of each of the experiments was carefully considered, to be sure to create experiments of an appropriate difficulty level for the children taking part in this research. Therefore, the experimental designs were refined through a series of pilots and adaptations and it is hoped that these experiments provide an innovative contribution to the repository of experimental designs for use with young children.

Finally, the research from this thesis aims to form a foundation for future work exploring the development of social judgments and stereotypes based on speakers' accents (see next section).

## 7.4 Future directions

This thesis has found that children learn to categorise speakers based on their accent from an early age. As categorisation is a cognitive strategy which we use in order to understand our social world, investigating the formation of these categories is a first step to

understanding how we develop social judgments about category members. Therefore, a proposal for a future project based on the current research is to investigate the development of both implicit and explicit social evaluations children make about speakers with different regional accents. This would provide new information about when and how we develop a range of positive and negative social judgments about members of accent categories based on the way they speak. It is important to understand the development of these social judgments amongst children so that social and educational policy and practice can focus on minimising the potential for accent discrimination. This is relevant for TV and film production, where assumed perceptual links between a speaker's accent and their personality is often used as a shortcut to characterisation. Further investigation is needed to explore Lippi-Green's (2012) claim that Disney's use of non-native accents to represent 'bad' characters teaches children to discriminate.

As described in chapter 2, sociolinguistic studies have investigated children's acquisition of community norms of pronunciation and their perception of standard/non-standard accent features in relation to their exposure to these forms (e.g. Foulkes et al., 1999; Barbu et al., 2013). Meanwhile, studies in developmental psychology have looked at the psychological beginnings of social evaluations, e.g. the development of a same-accent bias. Kinzler and DeJesus (2013) found that 5-6-year-olds in the U.S. made explicit, positive judgments about speakers with their local accent whereas 9-10-year-olds made judgments akin to widely held adult stereotypes (e.g. Northerners sound intelligent, Southerners sound friendly). The authors hypothesise that young children have a social preference for the familiar before stereotype-driven preferences emerge at 9-10-years. Crucially, however, there is no linguistic analysis of the accent-specific variables used in the stimuli. Furthermore, they did not investigate the potential effect of the children's exposure to other accents, which this thesis has found to play a significant role in accent awareness. Accordingly, there does not appear to be any research addressing an underlying link between the research in sociolinguistics and the research in developmental psychology. This link is between how children's developing awareness of regional accent groups and their experience of and exposure to linguistic variation in early life influences their social evaluation of speakers with different accents. Therefore it would be beneficial to investigate this link in a future

project in order to better understand children's social judgments about speakers with different accents.

Research in social psychology suggests that an individual's social judgments are made of both explicit and implicit attitudes (Greenwald and Banaji, 1995; Greenwald et al. 1998). Therefore, in a future project investigating children's social judgements, it would be beneficial to investigate both their implicit and explicit evaluations. While a few studies have investigated children's explicit evaluations of different accents (e.g. Giles et al., 1983; Kinzler and DeJesus, 2013), none appears to have investigated their implicit associations. The Implicit Association Test (IAT) and the Matched Guise Test (MGT) are used by linguists and psychologists to measure implicit social attitudes. The IAT measures how quickly participants sort a target concept (e.g. 'Yorkshire accent') when it is categorised with a positive compared to a negative attribute (e.g. 'friendly' vs. 'unfriendly'), in order to uncover which pairing is more embedded in memory. The adult IAT has only recently been adapted for use with children, mainly to investigate evaluations of race (cf. Baron and Banaji, 2006). Originally designed by Lambert (1967), the MGT has been used more extensively for accent evaluations, where listeners rate speakers with different accents on social attributes such as friendliness. However, the listeners hear only one speaker using different 'accent guises', thus revealing attitudes about accents rather than individuals. A suitable measure of children's implicit attitudes towards regional accents could therefore be developed in a future project, by adapting one or both of these tests. This could be run along with an appropriate method of measuring the children's explicit attitudes towards regional accents, building on previous studies (such as Kinzler and DeJesus, 2013). Both of these measures would then be analysed alongside a linguistic analysis of the children's ability to categorise speakers by accent.

Findings from such a project would provide new understanding of how social opinions about speakers are linked to a more general awareness of regional accents and exposure to linguistic variation. Important implications of this research include unravelling the roots of 'linguistic profiling', which can lead to accent discrimination, and expanding our knowledge about how characterisation based on accent (such as in children's TV) can be used for positive, educational purposes rather than reinforcing damaging stereotypes.

# Appendices

## Appendix 1: Experiment information sheets and consent forms

THE UNIVERSITY of York.

DEPARTMENT OF LANGUAGE AND LINGUISTIC SCIENCE

Heslington, York, YO10 5DD, UK
Phone number XXXXXXXXXX
Email erwj500@york.ac.uk

### Familiar Voices Game
### Information sheet

Researcher: Ella Jeffries

*Please read and keep this information sheet for you records. If you have any questions please feel free to contact Ella (contact details at the end).*

**About the game**

The game is a 10 minute task run on a computer. The children will listen to some sound clips through headphones and be asked to guess 'Who is talking?' by pointing at the picture of the nursery worker they think they can hear speaking. They will then be asked to identify whether they hear one of the teachers talking or whether it is a speaker they don't recognise. The game will be played with the researcher (Ella) in a quiet corner of the nursery with other nursery staff present in the room.

**About the research**

The game is being carried out as part of Ella Jeffries' PhD research at the University of York which is looking into children's ability to recognise familiar speakers and how they might identify a familiar speaker based on his/her accent. This PhD project is being supervised by Professor Paul Foulkes and Dr Carmen Llamas. Contact details for all the researchers are at the end of this sheet.

**Benefits**

The game will be stimulating for the children as it is an interactive and fun activity. Results of the research may be beneficial in revealing new information about the abilities of pre-school age children and what this may mean for their development of language and language learning. Information about the study and the implications of any results can potentially be part of and enhance the nursery's communication and language curriculum.

Every effort will be made to ensure that the children enjoy taking part, but if they become bored or don't seem to like playing the game, they are free to stop playing at any time. All information is strictly confidential and no real names will be used in any presentations, publications or theses. All data and information will be stored securely in the Department of Language and Linguistic Science at the University of York. If you decide to withdraw from the study, your child's data information will be permanently deleted and not used in the research project.

Screen shot of the game

**Results of the study**

The data you and your child provide will be used alongside the data of other participants to find out the level of comprehension and ability of pre-school children in recognising familiar voices. A debriefing letter and feedback information from the study will be produced when the results have been collected and analysed. This will include a summary of relevant findings and implications of the results. *[Please fill in your contact details on the Consent form if you wish to receive a summary of the results. A summary will also be provided for the nursery.]*

*This study has been reviewed and approved by the Departmental Ethics Committee of the Department of Language and Linguistic Science at the University of York. If you have any questions regarding this, you can contact the chair of the L&LS Ethics Committee, Dr Dominic Watt (**email:** dominic.watt@york.ac.uk; **Tel:** (01904) 322671).*


Game Title Screen

If you have further questions regarding this study, please feel free to contact:

**Researcher: Ella Jeffries**
Department of Language and Linguistic Science
University of York, Heslington, York, YO10 5DD
**tel:** XXXXXXXXX
**email:** erwj500@york.ac.uk

Supervisors: Professor Paul Foulkes
**email:** paul.foulkes@york.ac.uk

Dr Carmen Llamas
**email:** carmen.llamas@york.ac.uk

THE UNIVERSITY of York.

DEPARTMENT OF LANGUAGE AND LINGUISTIC SCIENCE

Heslington, York, YO10 5DD, UK
Phone number 07912690196
Email erwj500@york.ac.uk

# Grouping Voices Game
# Information sheet

Researcher: Ella Jeffries

*Please read and keep this information sheet for you records. If you have any questions please feel free to contact Ella (contact details at the end).*

## About the game

The game is a 10 minute task run on a computer. The children will listen to some sound clips of mothers and daughters through headphones and be asked to choose which mother each daughter should be put with. The game will be played with the researcher (Ella) in a quiet corner of the nursery with other nursery staff present in the room.

Ella playing the game with a pre-schooler

## About the research

The game is being carried out as part of Ella Jeffries' PhD research at the University of York which is looking into children's ability to group a speaker based on his/her accent. This PhD project is being supervised by Professor Paul Foulkes and Dr Carmen Llamas. Contact details for all the researchers are at the end of this sheet.

## Benefits

The game will be stimulating for the children as it is an interactive and fun activity. Results of the research may be beneficial in revealing new information about the abilities of pre-school age children and what this may mean for their development of language and language learning. Information about the study and the implications of any

Screen shot of the game 1

nursery's communication and language curriculum.

Every effort will be made to ensure that the children enjoy taking part, but if they become bored or don't seem to like playing the game, they are free to stop playing at any time. All information is strictly confidential and no real names will be used in any presentations, publications or theses. All data and information will be stored securely in the Department of Language and Linguistic Science at the University of York. If you decide to withdraw from the study, your child's data information will be permanently deleted and not used in the research project.

**Results of the study**

The data you and your child provide will be used alongside the data of other participants to find out the level of comprehension and ability of pre-school children in grouping speakers according to their accents. A debriefing letter and feedback information from the study will be produced when the results have been collected and analysed. This will include a summary of relevant findings and implications of the results. *[Please fill in your contact details on the Consent form if you wish to receive a summary of the results. A summary will also be provided for the nursery.]*

*This study has been reviewed and approved by the Departmental Ethics Committee of the Department of Language and Linguistic Science at the University of York. If you have any questions regarding this, you can contact the chair of the L&LS Ethics Committee, Dr Dominic Watt (**email:** dominic.watt@york.ac.uk; **Tel:** (01904) 322671).*

If you have further questions regarding this study, please feel free to contact:

**Researcher: Ella Jeffries**
Department of Language and Linguistic Science
University of York, Heslington, York, YO10 5DD
**tel:** 07912690196
**email:** erwj500@york.ac.uk

Supervisors: Professor Paul Foulkes
**tel:** (0)1904 322669
**email:** paul.foulkes@york.ac.uk

Dr Carmen Llamas
**tel:** (0)1904 322618
**email:** carmen.llamas@york.ac.uk

Screen shot of the game 2

## Familiar Voices/ Grouping Game
## Background Information

Child's name _____

Child's Date of Birth _____

Was your child born in York? Yes ❒ No ❒

If no, how long have they lived in York?_____

Are the child's parents/carers originally from York? Yes ❒ No ❒

If no, please indicate where they are from and how long they have lived in York

_____

_____

Is your child exposed to another language other than English at home? Yes ❒ No ❒

If yes, what language(s)?_____

_____

Which days of the week does your child attend nursery? (Please circle below)

Mon Tues Wed Thurs Fri

Roughly how many hours a week does he/she attend the nursery?

_____

**Thank you!**

# Grouping Speakers Game: *Who sounds like Miss..?*

## Information sheet

Researcher: Ella Jeffries

Your child is invited to participate in a game to be played at school as part of a research project being carried out at the University of York.

*Please read and keep this information sheet for you records. If you have any questions please feel free to contact Ella (contact details at the end).*

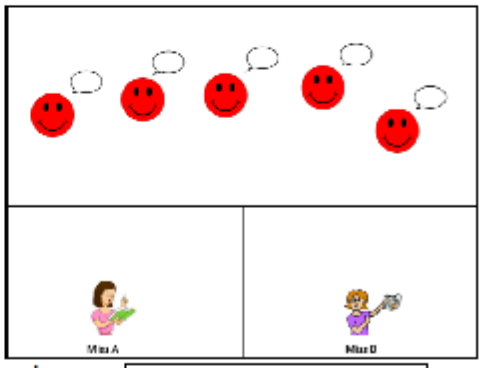
### About the research

The game is being carried out as part of Ella Jeffries' PhD research at the University of York which is looking into children's ability to group a speaker based on his/her accent. This PhD project is being supervised by Professor Paul Foulkes and Dr Carmen Llamas. Contact details for all the researchers are at the end of this sheet.

Previous work with pre-school children has found that 3-4-year-olds are able to group speakers using accent information. The current study looks at how these groups of speakers develop and whether older children are able to create more groups of speakers based on a wider range of regional accent differences. The individual child's experience and exposure to speakers with different accents as well as their attitudes towards speakers and their accents will also be investigated. Two similar activities have already been carried out by the researcher with pre-school children, who enjoyed playing the games and responded very positively to them.

### About the game

The game will last about 15 minutes and will involve two tasks run on a computer. First, the children will be asked to describe some pictures that they see on screen (e.g. a picture of a cake on a table) and their responses will be audio recorded. Second, the children will be asked to listen to two cartoon teachers reading some sentences and then listen to sentences spoken by some other speakers. They will be asked to group each speaker with the teacher who they most sound like, by using the computer's mouse to drag the picture of the speaker into a box on the screen (see screen shot). The children will be asked a few questions at the end of the game (e.g. Did any of the teachers sound like anyone you know?) and



Screen shot of the game

their responses will be audio recorded. The game will be played with the researcher (Ella) in a quiet corner of the classroom with a member of staff present in the room. The children will be asked by Ella if they are happy to take part in the game just before they begin and they will be told they are free to leave and stop playing at any time. Therefore, if you give consent, we would like to encourage you to explain to your child that it's fine to play the game if they want to.

**Benefits**

Although there will be no direct benefit for the children, the game has proved to be a stimulating, fun activity which other children have enjoyed playing. This research does not test the children's language ability in any way but results of the research may be beneficial in revealing new information about the abilities of primary school age children to group speakers by their accent.

**Does my child have to take part?**

Every effort will be made to ensure that the children enjoy taking part, but if they become bored or don't seem to like playing the game, they are free to stop playing at any time. All information is strictly confidential and no real names will be used in any presentations, publications or theses. All data and information will be stored securely in the Department of Language and Linguistic Science at the University of York. If you decide to withdraw from the study, your child's data information will be permanently deleted and not used in the research project.

We would like to run the game in the next few weeks so if you give permission for your child to take part, **please fill in the consent form and the background information form (attached) and hand these forms back to your child's teacher as soon as possible.**

*This study has been reviewed and approved by the Departmental Ethics Committee of the Department of Language and Linguistic Science at the University of York. If you have any questions regarding this, you can contact the chair of the L&LS Ethics Committee, Dr Traci Walker* ***(email:*** *traci.walker@york.ac.uk;* ***Tel:*** *(01904) 322671).*

If you have further questions regarding this study, please feel free to contact:

**Researcher: Ella Jeffries**
Department of Language and Linguistic Science
University of York, Heslington, York, YO10 5DD
**tel:** 07912690196
**email:** erwj500@york.ac.uk

Supervisors: Professor Paul Foulkes
**tel:** (0)1904 322669
**email:** paul.foulkes@york.ac.uk

Dr Carmen Llamas
**tel:** (0)1904 322618
**email:** carmen.llamas@york.ac.uk

# Grouping Speakers Game
# Consent Form

Please fill in to indicate that you consent for your child to take part in the Grouping Speakers Game. If there is anything you do not understand, or if you want more information, please ask the researcher (contact details on the information sheet).

| | |
|---|---|
| I understand what the research is about and what is involved | Yes ☐ No ☐ |
| I have had the chance to ask questions about the research | Yes ☐ No ☐ |
| I agree for my child to take part in the study | Yes ☐ No ☐ |
| I agree for my child to be recorded on audio | Yes ☐ No ☐ |
| I agree to excerpts from my child's audio recordings to be used in presentations or in teaching by the researcher, without disclosing their real name<br>*(You may take part in the study without agreeing to this).* | Yes ☐ No ☐ |
| I consent to the recordings being kept after the duration of the current project, to be used in future research on language | Yes ☐ No ☐ |
| I understand that the information my child provides will be held in confidence by the research team and that no identifying information will be mentioned in any publication | Yes ☐ No ☐ |
| I understand that my child may withdraw from the study at any time | Yes ☐ No ☐ |
| I understand that I can withdraw any information that I/my child provides up to six months after the information has been collected | Yes ☐ No ☐ |
| I agree to the information that I/ my child provides being kept after the duration of the current project, to be used in future research on language | Yes ☐ No ☐ |
| I agree to the researcher keeping my contact details after the end of the current project, in order that s/he may contact me in the future about possible participation in other studies<br>*(You may take part in the study without agreeing to this).* | Yes ☐ No ☐ |

Your child's name _____Your name _____

Contact email/address for summary of results *(Only fill in if you would like a summary)*

_____

Your signature _____Date_____

## Background Information

**(1)** Child's name and Date of Birth

_____

**(2)** Was your child born in York? Yes❒ No❒

If **no,** where were they born and how long have they lived in York?

_____

**(3)** Does your child have any brothers or sisters? Yes❒ No❒

If **yes,** please indicate how many and whether they are younger or older

Brother(s)_____

Sister(s)_____

**(4)** Are both your child's parents/carers originally from York? Yes❒ No❒

If **no,** please indicate where each parent/carer is from and how long they have lived in York

_____

**(5)** Please indicate the occupations of both parents/carers

_____

_____

**(6)** What is the highest level of education of all parents/carers?

High school (up to 14) ❒ High school (up to 16) ❒
High school/college (up to 18) ❒ Undergraduate degree ❒
Postgraduate degree ❒

Other_____

*Research shows that children's exposure to different languages and different accents can affect their ability to tell the difference between them. These questions ask about your child's exposure to other languages/accents (both in real-life and through television) in order for this claim to be investigated.*

**(7)** Is your child exposed to a language other than English at home? Yes❒ No❒

If **yes**, please indicate what language(s) and how much exposure they have

Language(s)_____

Main language spoken at home ❒
Daily exposure at home but not main language ❒
Occasional exposure at home ❒

**(8)** Are there any other people from a different city/town/region of England, or from a different country, who your child frequently spends time with (for example, a grandparent, a childminder or a close family friend)? Yes❒ No❒

If **yes**, please indicate their relationship to your child, the country/city/town they are from and their native language (if not English)

_____

_____

_____

**(9)** What are your child's favourite things to watch? (on TV, DVDs, cartoons, films, programmes available on the internet)

Are there any particular characters they are fond of? (e.g. they talk about them frequently/have merchandise of them)

_____

_____

_____

**Thank you!**

## Appendix 2: Gruffalo story passage and nursery phrases recorded for the Grouping experiment

**The Gruffalo**

1) A mouse took a stroll through the deep dark wood.

   A fox saw the mouse and the mouse looked good.

2) "Where are you going to little brown mouse?

   Come and have lunch in my underground house."

3) "It's terribly kind of you Fox, but no, - I'm going to have lunch with a gruffalo."

   "A gruffalo? What's a gruffalo?"

4) "A gruffalo! Why, didn't you know?

   He has terrible tusks, and terrible claws,

   and terrible teeth in his terrible jaws."

5) "Where are you meeting him?"

"Here, by these rocks, and his favourite food is roasted fox."

6) "Roasted fox!! I'm off!!" Fox said

"Goodbye little mouse" and away he sped.

7) "Silly old fox! Doesn't he know,

there's no such thing as a gruffalo?"

Hello, Nursery?

Line up to wash your hands

Who's ready for snack?

Come on It's story time soon

Get your coat

Look! Here comes mummy

Daddy's here to collect you

There's a good girl

There's a good boy

Well done!

Where's teddy?

Nevermind, why don't you try again

We need to tidy  up these toys now

Bye!

No!

## Appendix 3: Chicken Little Story passage and sentences recorded for the Second Grouping experiment

One day Chicken Little was pecking at the grass under the acorn tree, looking for tasty worms, when he felt a sharp WHACK on his fluffy head.
"Ouch! Oh dear! Disaster!" cried Chicken Little, who was a bit of a drama queen. "The sky is falling!" His head hurt, and he could feel a big bump on it.
"I'd better warn everyone!" he squawked. And off he raced, in a panicked cloud of fluff.
He found Plucky Ducky doing backstroke in the pond.
"Get out of the bath! Alert the authorities!" gasped Chicken Little. "The sky is falling, let's go!"
"Ahoy there!" quacked Plucky Ducky, who was covered in soap. "Stay calm. I'll rescue you. I'm not scared of anything. Remember that time in class–"
"No time to waste!" shrieked Chicken Little. "This is no joke. A chunk of sky just bruised my head. It's starting to throb – look at this bump!"
Chicken Little wasn't having a laugh: Plucky Ducky could see the swelling under his feathers.
"Wait there!" quacked Plucky Ducky, "We'd better grab the others."
Hurrying down the path, the pair met Weepy Sheepy, chewing gloomily on some clover.
"Run fast, before it's too late!" screeched Chicken Little. "The sky is falling. Look at my poor head!"
Weepy Sheepy burst into tears. "Please don't raise your voice at me," he said, sobbing. "I'm feeling a bit delicate."
"Have courage," said Plucky Ducky sympathetically, "I'll protect you. Join us, you don't want to be the last!"
"Oh help, I'm scared," wailed Weepy Sheepy, trotting after them down the road. "Has anyone got a tissue?" he asked.

- This is the final *task*
- I should not be *last*
- There's a new member of *staff*
- It's icy on the *path*
- I had to walk on the *grass*

- I'm already *late*
- We should be in the *shade*
- I really want to *play*
- She said that I should *wait*
- It has been a good *day*

- I have a very sore *toe*
- Let's walk along this *road*
- Tell me when I should *go*
- The reply to the question is *no*
- There was a sheep and a *goat*

- My shoes are full of *mud*
- She was following the *duck*
- I have milk in my *cup*
- It had chocolate *chunks*
- He cycled across the *bump*

- I would like to do *better*
- The sweets have too much *sugar*
- I need you to *remember*
- He would like some extra *paper*
- You should try and go *under*

- His bruise really *hurt*
- It's behind the *church*
- I think the balloon will *burst*
- She is on high *alert*
- We need to have a *word*

# References


Atagi, E., and Bent, T. (2011). *Perceptual dimensions of nonnative speech.* Paper presented at the Proceedings of the XVIIth International Congress of Phonetic Sciences. Hong Kong, China.

Atagi, E., and Bent, T. (2013). Auditory free classification of nonnative speech. *Journal of Phonetics, 41*(6), 509-519.

Atagi, E., and Bent, T. (2014). Relationship between listeners' nonnative speech recognition and categorization abilities. *The Journal of the Acoustical Society of America, 137*(1), EL44-EL50.

Audacity Team (2012). Audacity®. Version 2.0.3. Audio editor and recorder. Available from: http://audacityteam.org/.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barbu, S., Nardy, A., Chevrot, J.-P., and Juhel, J. (2013). Language evaluation and use during early childhood: Adhesion to social norms or integration of environmental regularities? *Linguistics, 51*(2), 381-411.

Baron, A. S., and Banaji, M. R. (2006). The development of implicit attitudes evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological science, 17*(1), 53-58.

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology, 27*(4), 464-472.

Beck, E. L. (2014). *The Role of Socio-indexical Information in Regional Accent Perception by Five to Seven Year Old Children.* Unpublished: The University of Michigan. PhD Thesis.

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47*(2), 383-411.

Belin, P., Bestelmeyer, P. E. G., Latinus, M., and Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*(4), 711-725.

Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in cognitive sciences, 8*(3), 129-135.

Best, C. and Kitamura, C. (2012). Accent on language development: using dialects to trace how children come to recognise spoken words. In L.Burton (ed). *Psychology*. Australia: John Wiley & Sons, 332-333.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J.F. Kroll and A.M.B.De Groot (eds), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press, 109-127.

Birdsong, D. and Vanhove, J. (2016). Age of second language acquisition: Critical periods and social concerns. In E. Nicoladis and S. Montanari (eds), *Bilingualism across the lifespan: Factors moderating language proficiency*. APA and De Gruyter, 163-182.
Blewitt, P. (1994). Understanding categorical hierarchies: The earliest levels of skill. *Child development, 65*(5), 1279-1298.

Bornstein, M. H., Hahn, C.-S., and Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24*(3), 267-304.

Britain, D. (2004). Dialect and accent. *Ulrich Ammon et al.(Hg.), Sociolinguistics: An International Handbook of the Science of Language and Society, 2*, 267-273.

Bruce, V., and Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327.

Butler, J., Floccia, C., Goslin, J., and Panneton, R. (2011). Infants' Discrimination of Familiar and Unfamiliar Accents in Speech. *Infancy, 16*(4), 392-417.

Carey, S., Diamond, R., and Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology, 16*(4), 257-269.

Carr, P. (1999). *English Phonetics and Phonology: An Introduction.* Oxford: Blackwell.

Chambers, J. K. (1992). Dialect Acquisition. *Language, 68*(4), 673-705.

Chambers, J. K. (2002). Dynamics of dialect convergence. *Journal of Sociolinguistics, 6*(1), 117-130.

Chiu, S. (2001). Exploring kindergartners' social and cognitive competence: An application of ECLS-K. Dissertation Abstracts International, 62(12), 4051.

Chomsky, N. (1967). The formal nature of language. *Biological foundations of language*, 397-442.

Clopper, C. G., and Pisoni, D. B. (2004a). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change, 16*(01), 31-48.

Clopper, C. G., and Pisoni, D. B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32*(1), 111-140.

Clopper, C. G., and Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics, 35*(3), 421-438.

Compton, A. J. (1963). Effects of Filtering and Vocal Duration upon the Identification of Speakers, Aurally. *The Journal of the Acoustical Society of America, 35*(11), 1748-1752.

Cragg, L., and Nation, K. (2009). Shifting development in mid-childhood: The influence of between-task interference. *Developmental Psychology, 45*(5), 1465.

Creel, S. C., and Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology, 113*(4), 487-509.

Cutler, A., Eisner, F., McQueen, J. M., and Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory phonology*, *10*, 91-111.

Cvencek, D., Greenwald, A. G., and Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The preschool implicit association test. *Journal of Experimental Child Psychology, 109*(2), 187-200.

Davidson, M. C., Amso, D., Anderson, L. C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078.

DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*, 1174-1176.

DeKeyser, R. and Larson-Hall, J. (2005). What does the critical period really mean? In J.F. Kroll and A.M.B.De Groot (eds), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press, 88-108.

DePaolis, R. A., Vihman, M. M and Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants?. *Infant Behavior and Development, 34*(4), 590-601.

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*(4), 642-649.

Diana, R. A., Reder, L. M., Arndt, J., and Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13*(1), 1-21.

Docherty, G. J., and Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. *Urban voices: Accent studies in the British Isles*, 47-71.

Docherty, G. J., and Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua, 142*, 42-56.

Donaldson, J., and Scheffler, A. (1999). *The Gruffalo*. London: Macmillan.

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass, 4*(7), 473-480.

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech, 54*(1), 99-121.

Drummey, A. B., and Newcombe, N. (1995). Remembering versus Knowing the Past: Children' s Explicit and Implicit Memories for Pictures. *Journal of Experimental Child Psychology, 59*(3), 549-565.

Ebert, I., and Steffens, M. (2008). *Generality and correlates of the implicit preference for women.* Paper presented at the 11th Jena workshop on intergroup processes: Positive behaviour between social groups—processes, problems, and promises.

Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Wiley-Blackwell.

Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association, 8*(1-2), 18-23.

Esling, J.H. (1998). Everyone has an accent except me. In L. Bauerand P. Trudgill (eds), *Language Myths.* London: Penguin, 169-175.

Evans, B. G., and Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America, 121*(6), 3814-3826.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., and Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, *242*(59:5), i-185.

Fernald, A., Swingley, D. and Pinto, J.P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*, 1003-1015.

Field, A., Miles, J. and Field, Z. (2012). *Discovering statistics using R*. Los Angeles, Calif.; London: Sage.

Floccia, C., Butler, J., Girard, F., and Goslin, J. (2009). Categorization of regional and foreign accent in 5-to 7-year-old British children. *International Journal of Behavioral Development, 33*(4), 366-375.

Floccia, C., Delle Luche, C., Durrant, S., Butler, J., and Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition, 124*(1), 95-100.

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology, 1*(1), 5-39.

Foulkes, P., and Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *Forensic linguistics, 7*, 180-198.

Foulkes, P., and Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*(4), 409-438.

Foulkes, P., Docherty, G. J., and Watt, D. (1999). *Tracking the emergence of sociophonetic variation.* Paper presented at the Proceedings of The 14th International Congress of Phonetic Sciences, 1–7 August 1999, San Francisco.

Foulkes, P., Docherty, G., and Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 177-206.

Foulkes, P., and Hay, J. B. (2015). The Emergence of Sociophonetic Structure. *The handbook of language emergence, 87*, 292-313.

Fry, D. B. (1979). *The physics of speech*: Cambridge University Press.

Gerfen, C., and Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics, 33*(3), 311-334.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological linguistics*, 87-105.

Giles, H., Harrison, C., Creber, C., Smith, P. M., and Freeman, N. H. (1983). Developmental and contextual aspects of children's language attitudes. *Language & Communication, 3*(2), 141-146.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition 22*(5): 1166-1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review 105*(2): 251-279.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th international congress of phonetic sciences*, 49-54.

Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.

Greenwald, A. G., McGhee, D. E., Schwartz, J. L.K. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology 74*(6): 1464–1480.

Haddican, B., Foulkes, P., Hughes, V., and Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change, 25*(03), 371-403.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist, 52*(10), 1091.

Hart, B., and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hart, B., and Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American educator, 27*(1), 4-9.

Hazen, K. (2002). The Family. In J. K. Chambers, P. Trudgill and N. Schilling-Estes (eds), *Handbook of Language Variation and Change.* Oxford: Blackwell, 500-525.

Heselwood, B., and Plug, L. (2011). *The role of F2 and F3 in the perception of rhoticity: Evidence from listening experiments.* Paper presented at the Proceedings of the 17th International Congress of Phonetic Sciences, August.

Hintzman, D.L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

Hintzman, D.L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528-551.

Howard, L. H., Carrazza, C., and Woodward, A. L. (2014). Neighborhood linguistic diversity predicts infants' social learning. *Cognition, 133*(2), 474-479.

Hughes, A., Trudgill, P., and Watt, D. (2012). *English accents and dialects: An introduction to social and regional varieties of English in the British Isles*. London: Routledge.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236-248.

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., and Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology, 43*(5), 1062-1083.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, *30.5,* 513-541.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, 145-165.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*(4), 485-499.

Johnson, K. E., Scott, P., and Mervis, C. B. (1997). Development of children's understanding of basic-subordinate inclusion relations. *Developmental Psychology, 33*(5), 745-763.

Kemler Nelson, D. G. K., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant behavior and development, 18*(1), 111-116.

Kerswill, P. (1996). Children, adolescents, and language change. *Language Variation and Change, 8*(02), 177-202.

Kerswill, P. (2006). Standard English, RP and the standard–non-standard relationship: Language in the British Isles. Cambridge: Cambridge University Press.

Khattab, G. (2007). Variation in vowel production by English–Arabic bilinguals. *Laboratory Phonology, 9*, 383-410.

Khattab, G. (2013). Phonetic convergence and divergence strategies in English-Arabic bilingual children. *Linguistics, 51*(2), 439-472.

Kinzler, K. D., and DeJesus, J. M. (2013). Northern= smart and Southern= nice: The development of accent attitudes in the United States. *The Quarterly Journal of Experimental Psychology, 66*(6), 1146-1158.

Kinzler, K. D., Dupoux, E., and Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, 104*(30), 12577-12580.

Klatt, D.H., and Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2): 820-857.

Kreiman, J., and Papcun, G. (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech communication, 10*(3), 265-275.

Kreiman, J., and Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception.* Chichester: John Wiley & Sons.

Labov, W. (1964). Stages in the acquisition of standard English. *Social dialects and language learning*, 77-104.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns.* Oxford: Blackwell

Ladefoged, P., and Ladefoged, J. (1980). The Ability of Listeners to Identify Voices. *University of California Working Papers in Phonetics, 49*(Apr), 43-51.

Lambert, W. E. (1967). A social psychology of bilingualism. *Journal of Social Issues, 23*(2), 91-109.

Laver, J. D. (1968). Voice quality and indexical information. *International Journal of Language & Communication Disorders, 3*(1), 43-54.

Laver, J.D. (1980). The phonetic description of voice quality. *Cambridge Studies in Linguistics London, 31*, 1-186.

Lawrence, D. (2014). Phonetic variation as a cue to regional identity: an experimental approach. *Proceedings of the second Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2014) Conference, Issue 2:* 42-63.

Lawson, E., Scobbie, J. M., and Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic/r/in Scottish English. *Journal of Sociolinguistics, 15*(2), 256-268.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Lev-Ari, S., and Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093-1096.

Lippi-Green, R. (2012). *English with an accent : language, ideology and discrimination in the United States,* 2nd edn. London: Routledge.

Llamas, C. (2010). Convergence and divergence across a national border. *Language and identities*, 227-236.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-886.

Lundberg, I., Larsman, P., and Strid, A. (2012). Development of phonological awareness during the preschool year: The influence of gender and socio-economic status. *Reading and writing, 25*(2), 305-320.

MacNeilage, P. F., and Davis, B.L. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (ed.), *Attention and Performance XIII: Motor representation and control.* Hillsdale, NJ: Lawrence Erlbaum, 453-476

Mann, V. A., Diamond, R., and Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27*(1), 153-165.

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*(2), 179-204.

McAllister Byun, T., Inkelas, S., and Rose, Y. (2016). The A-map model: Articulatory reliability in child-specific phonology. *Language,* to appear.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419-457.

McGlothlin, H., and Killen, M. (2010). How social experience is related to children's intergroup attitudes. *European Journal of Social Psychology, 40*(4), 625-634.

McQueen, J. M., Cutler, A. and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science 30(6),* 1113-1126.

Mielke, J., Baker, A., and Archangeli, D. (2010). Variability and homogeneity in American English/r/allophony and/s/retraction. *Laboratory Phonology, 10*, 699-730.

Milroy, J. (1992). *Linguistic variation and change: on the historical sociolinguistics of English*: B. Blackwell.

Montgomery, C. 2007. *Northern English dialects: A perceptual approach*. Unpublished: University of Sheffield PhD thesis.

Murphy, G. L. (2004). *The big book of concepts*: MIT press.

Nardy, A., Chevrot, J.-P., and Barbu, S. (2013). The acquisition of sociolinguistic variation: looking back and thinking ahead. *Linguistics, 51*(2), 255-284.

Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition, 58*(2), 221-251.

Nathan, L. I. Z., Wells, B., and Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language, 25*(02), 343-365.

Newcombe, N. S., Drummey, A. B., Fox, N. A., Lie, E., and Ottinger-Alberts, W. (2000). Remembering Early Childhood How Much, How, and Why (or Why Not). *Current directions in psychological science, 9*(2), 55-58.

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition.* Cambridge: Cambridge University Press.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological review*, *110*(4), 611-646.

Norris, D., McQueen, J.M. and Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology 47(2),* 204-238.

Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Attention, Perception, & Psychophysics, 60*(3), 355-376.

Öhman, L. (2013). All ears: Adults' and children's earwitness testimony. Unpublished: University of Gothenburg PhD Thesis.

Orton, H. (1962). *Introduction to the Survey of English Dialects.* Leeds: EJ Arnold.

Papcun, G., Kreiman, J., and Davis, A. (1989). Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America, 85*(2), 913-925.

Payne, A. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In W. Labov (ed), *Locating Language in Time and Space*. New York: Academic, 143-178.

Peirce, JW (2007) PsychoPy - Psychophysics software in Python. J Neurosci Methods, 162(1-2):8-13.

Phillips, D. A., and Shonkoff, J. P. (2000). *From Neurons to Neighborhoods:: The Science of Early Childhood Development*. National Academies Press.

Pierrehumbert, J. B. (2002). Word-specific phonetics. *Laboratory Phonology, 7*, 101-139.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*(2-3), 115-154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics, 34*(4), 516-530.

Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. *Talker variability in speech processing*, 9-32.

Prince, A.S. and Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar.* Unpublished ms., Rutgers University, New Brunswick and University of Colorado, Boulder.

Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., and Lehtonen, J. (2005). Four-month-old infants process own mother's voice faster than unfamiliar voices—Electrical signs of sensitization in infant brain. *Cognitive Brain Research, 24*(3), 627-633.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ready, D. D., LoGerfo, L. F., Burkam, D. T., and Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference? *The Elementary School Journal, 106*(1), 21-38.

Roberts, J. (1997). Acquisition of variable rules: a study of (-t, d) deletion in preschool children. *Journal of Child Language, 24*(02), 351-372.

Roberts, J., and Labov, W. (1995). Learning to talk Philadelphian: Acquisition of short a by preschool children. *Language Variation and Change, 7*(01), 101-112.

Rose, P. (2002). *Forensic speaker identification.* London: Taylor and Francis.

Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*(01), 185-205.

Rutland, A., Cameron, L., Milne, A., and McGeorge, P. (2005). Social Norms and Self-Presentation: Children's Implicit and Explicit Intergroup Attitudes. *Child development, 76*(2), 451-466.

Sander, E.K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37,* 55-63.

Sankoff, G. (2004). Adolescents, young adults and the critical period: Two case studies from 'Seven Up'. *Sociolinguistic variation: Critical reflections*, 121-139.

Schmale, R., Cristia, A., and Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental science, 15*(6), 732-738.

Schmale, R., Cristià, A., Seidl, A., and Johnson, E. K. (2010). Developmental Changes in Infants' Ability to Cope with Dialect Variation in Word Recognition. *Infancy, 15*(6), 650-662.

Schweinberger, S. R., Herholz, A., and Sommer, W. (1997). Recognizing Famous Voices: Influence of Stimulus Duration and Different Types of Retrieval Cues. *Journal of Speech, Language, and Hearing Research, 40*(2), 453-463.

Sebastián-Gallés, N. (2005). Cross Language Speech Perception. In D. Pisoni and R. Remez(eds), *The Handbook of Speech Perception.* Malden, MA; Oxford : Blackwell Publishers, 546-566.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic inquiry, 11*(3), 563-605.

Shneidman, L. A., and Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: how important is directed speech? *Developmental science, 15*(5), 659-673.

Sjöström, M., Eriksson, E. J., Zetterholm, E., and Sullivan, K. P. (2009). A switch of dialect as disguise. *Lund Working Papers in Linguistics, 52*, 113–116.

Smith, E., and Zárate, M. (1992). Exemplar-Based Model of Social Judgment. *Psychological Review, 99*, 3- 21

Smith, J., Durham, M., and Fortune, L. (2007). “Mam, my trousers is fa'in doon!”: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change, 19*(01), 63-99.

Smith, J., Durham, M., and Richards, H. (2013). The social and linguistic in the acquisition of sociolinguistic norms: Caregivers, children, and variation. *Linguistics, 51*(2), 285-324.

Sosa, A. V., and Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language, 33*(01), 31-50.

Spence, M. J., Rollins, P. R., and Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech, Language and Hearing Research, 45*(1), 214-222.

Sternberg, R.J. (2004) ‘Individual Differences’ in U. Goswami (ed.) *Blackwell Handbook of Childhood Cognitive Development*. Oxford: Blackwell, 600-619.

Stevenage, S. V., Clarke, G., and McNeill, A. (2012). The “other-accent” effect in voice recognition. *Journal of Cognitive Psychology, 24*(6), 647-653.

Strand, E. A. (2000). *Gender Stereotype Effects in Speech Processing*. Colombus, Ohio: Ohio State University PhD dissertation.

Stuart-Smith, J. (1999). Glasgow: accent and voice quality. *Urban voices: Accent studies in the British Isles*, 203-222.

Stuart-Smith, J. (2003). The phonology of modern urban Scots. In J. Corbett, J.D. McClureand J. Stuart-Smith(eds), *The Edinburgh Companion to Scots.* Edinburgh University Press: Edinburgh, UK, 110-137.

Stuart-Smith, J. (2007). A sociophonetic investigation of postvocalic/r/in Glaswegian adolescents. *Proceedings ICPhS, Saarbrücken*.

Stuart-Smith, J., Pryce, G., Timmins, C., and Gunter, B. (2013). Television can also be a factor in language change: Evidence from an urban dialect. *Language, 89*(3), 501-536.

Sumner, M., Kim, S. K., and King, K. B. M. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in psychology, 4*, 1-13.

Sumner, M., and Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487-501.

Szreder-Ptasinska, M. (2012). *Child Phonology as a Dynamic System.* Unpublished: University of York PhD thesis.

Tagliamonte, S. A., and Molfenter, S. (2007). How'd you get that accent?: Acquiring a second dialect of the same language. *Language in Society, 36*(05), 649-675.

Tagliamonte, S. A., and Roeder, R. V. (2009). Variation in the English definite article: Socio-historical linguistics in t'speech community. *Journal of Sociolinguistics, 13*(4), 435-471.

Tajfel, H., and Turner, J. (1985). The social identity theory of intergroup behavior. In S. Worchel and W G. Austin (eds), *Psychology of intergroup relations.* Chicago: Nelson-Hall, 7-24.

The Yorkshire Ridings Society (2013). *Association of British Counties Map of Yorkshire.* [Online] The Yorkshire Ridings Society. Available at: http://www.yorkshireridings.org/news/yorkshire-map.html [Accessed 12 February 2016].

Thelen, E., and Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambrige, MA: MIT Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Boston, MA: Harvard University Press.

Trudgill, P. (1974). *The social differentiation of English in Norwich.* Cambridge: Cambridge University Press.

Trudgill, P. (1981). Linguistic accommodation: sociolinguistic observations on a sociopsychological theory. *Papers from the Parasession on Language and Behavior*, 218-237.

Trudgill, P. (1986). *Dialects in contact*: Blackwell Oxford.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., and Wetherell, M. S. (1987). *Rediscovering the Social Group: A self-categorization theory.* Oxford, England: Basil Blackwell.

van der Feest, S. V., and Johnson, E. K. (2015). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition*, 1-23.

van Heugten, M., and Johnson, E. K. (2015). Toddlers' word recognition in an unfamiliar regional accent: The role of local sentence context and prior accent exposure. *Language and Speech*, 1-11.

van Lancker, D., and Kreiman, J. (1985). Unfamiliar Voice Discrimination and Familiar Voice Recognition Are Independent and Unordered Abilities. *University of California Working Papers in Phonetics, 62*(Dec), 50-60.

van Lancker, D., and Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25*(5), 829-834.

Velleman, S.L. and Vihman, M. M. (2003). The emergence of the marked unfaithful. In A. Carpenter, A. Coetzee and P. De Lacy (eds), *University of Massachusetts Occasional Papers in Linguistics 26: Papers in Optimality Theory II.* Amherst, MA: GLSA, 397-419.

Vihman, M. M. (2014). *Phonological Development:The First Two Years*. Second edition. Oxford: Blackwell.

Vihman, M. M., and Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics, 45*(4), 683-725.

Vihman, M.M., Nakai, S., DePaolis, R.A., Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, *50,* 336-354.

Wagner, L., Clopper, C. G., and Pate, J. K. (2014). Children's perception of dialect variation. *Journal of Child Language, 41*(05), 1062-1084.

Watt, D. (2002). 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics, 6*(1), 44-63.

Watt, D., and Tillotson, J. (2001). A spectrographic analysis of vowel fronting in Bradford English. *English World-Wide, 22*(2), 269-303.

Wells, J. C. (1982a). *Accents of English 1: An Introduction* (Vol. 1). Cambridge: Cambridge University Press.

Wells, J. C. (1982b). *Accents of English 2: The British Isles* (Vol. 2). Cambridge: Cambridge University Press.

Williams, A., Garrett, P., and Coupland, N. (1999). Dialect recognition. In D. R. Preston (ed). *Handbook of Perceptual Dialectology*. Philadelphia: John Benjamins, 345–358.

Williams, A., and Kerswill, P. (1999). Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. *Urban voices: Accent studies in the British Isles*, 141-162.

Wright, J. (1898-1905). *The English dialect dictionary; being the complete vocabulary of all dialect words still in use or known to have been in use during the last two hundred years, etc.* Six Volumes. London: Oxford University Press.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., and Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English/r. *The Journal of the Acoustical Society of America, 123*(6), 4466-4481.