University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

04 University of Plymouth Research Theses 01 Research Theses Main Collection

2013

The Processing of Accented Speech

Duffy, Hester Elizabeth Sarah

http://hdl.handle.net/10026.1/1556

http://dx.doi.org/10.24382/3232
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

THE PROCESSING OF ACCENTED SPEECH

by

HESTER DUFFY

A thesis submitted to Plymouth University in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

Faculty of Science and Technology

October 2012

Copyright Statement

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

Contents~

List of Figures

Acknowledgements

The research which went towards this thesis was funded by the EPSRC, and would not have been possible without the help of the Plymouth University participants pool, the parents and children of the Plymouth BabyLab, and the help and support of the School of Psychology's technicians.

My thanks go to my supervisors, Jeremy Goslin and Caroline Floccia for their guidance, advice and for everything I have learned from them. Also, to the members of the BabyLab and T&C (especially Sam Durrant and Emily Sullivan) who have provided sounding boards, moral support, and kindness; I couldn't have done it without you!

I would also like to thank my ever-loving, ever-patient family, and to dedicate this thesis to its newest member, Ava, with the promise that she will never have to read it.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

Work submitted for this research degree at the Plymouth University has not formed part of any other degree at Plymouth University or at another establishment.

This study was financed with the aid of the EPSRC.

Relevant scientific seminars and conferences were regularly attended at which work was often presented.

<u>Publications:</u>

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*.

<u>In preparation</u>

Duffy, H., Goslin, J., & Floccia, C. (in prep). An ERP study into repetition-suppression across accents.

Duffy, H., Floccia, C., & Goslin, J. (in prep). An ERP study into infants' segmentation within and across accents.

<u>Presentations</u>

Depaolis, R., Duffy, H., Keren-Portnoy, T. and Vihman, M. (2012). IDS differences in naturalistic speech and in stimuli affect the age at which infants succeed in Headturn Preference Procedure tasks. *Paper presented at International Child Phonology Conference, Minneapolis, June 4 - 6*

Duffy, H., Goslin, J., & Floccia, C. (2011). Infants' word segmentation across accents; an ERP study. *Paper presented at the Child Language Seminar, University of Newcastle, June 13-14.*

Duffy, H., Goslin, J., Bridges, D., & Floccia, C. (2010). Using ERP to compare the processing of regional and foreign accents. *Paper presented at the Experimental Approaches to the Perception and Production of Language Variation, University of Groningen, November 11-12.*

<u>Posters</u>

Duffy, H., Floccia, C., Goslin, J., Farag, R., Turner, J. (2012)

Infants' Word Segmentation Across Accents; an ERP Study. *Poster Presentation at XVIIIth Biennial International Conference on Infant Studies; Minneapolis, June 7-9.*

Duffy, H., Floccia, C. & Goslin, J. (2011). Electrophysiological comparison of regional and foreign accents. *Poster presented at Oxford Autumn School in Cognitive Neuroscience, University of Oxford, September 26-29.*

Duffy, H., Bridges, D., Wallington, H., Momber, S-J., Floccia, C. & Goslin, J. (2010). Electrophysiological comparison of regional and foreign accents. *Poster presentation at Psycholinguistic Approaches to Speech Recognition in Adverse Conditions workshop; University of Bristol, March 8-10.*

Word count of main body of thesis: 61,545

Signed…………………………………………………….………………………

Date……………………………………………………………………………..

Chapter 1: Main Introduction

Listening to speech in our maternal language is one of our favourite activities, and we are incredibly good at it. Whether it happens in a crowded pub or over the telephone, whether it comes from a little girl or an old man, from a Glaswegian native or a French tourist, we are generally able to understand what is said to us, with little noticeable effort or difficulty. We are even able to understand speech which has been deliberately distorted or which comes from a non-human source; we have no trouble discerning that the Daleks have a penchant for extermination, and that Polly wants a cracker. However, the ease with which we understand our interlocutor belies the complexity of the range of rapid automatic operations required, and to date, automatic speech recognition systems have failed to reproduce human-like processing (Novotney & Callison-Burch, 2010), with state-of-the-art systems, still producing an error rate of around 15%. Siri, the voice-activated virtual assistant for iPhones, has particular trouble with accented speech, spawning numerous complaints from Scottish, Southern American and non-native users. As adult listeners, and beginning in infancy, we are capable of developing robust speech recognition which is able to ignore or overcome variability in the speech signal in order to retrieve meaningful information. How then do we succeed where the most advanced technology fails?

In this thesis we examine the problems encountered by both young and mature listeners in the processing of accented speech. Accents are one specific source of variability in the natural speech signal, mainly tackling linguistic levels of processing, rather than other sources such as background noise or speaker gender, which primarily modify the speech signal at the acoustic level. However, because the field of accent processing is relatively new, our primary argument will be informed by studies looking at other sources of variation, including digital distortion and natural degradation due to noise or speech disorders.

Based on previous research (Adank, Evans, Stuart-Smith, & Scott, 2009; Chambers, 2002; Floccia, Butler, Goslin, & Ellis, 2009b; Floccia, Goslin, Girard, & Konopczynski, 2006; Girard, Floccia, & Goslin, 2008), we will examine the possibility that regional accents (used by native speakers in a particular language) and foreign accents (used by non-native speakers in this language) recruit qualitatively different mechanisms for normalisation and adaptation.

The thesis will approach these issues from two different research methods. The main stream will use neurophysiological measures with adult and infant participants, to examine the time-course of word processing in different accents (chapters 3, 4 and 8). Another stream of research will employ behavioural responses in infants to complement the use of ERPs in this population and allow comparison with previous studies (chapter 7).

Accent and variability.

Accent is here defined as a systematic pattern of variation in pronunciation. The term “dialect” is sometimes used interchangeably with “accent”. However, as discussed by Wells (1982) and Hughes and Trudgill (1996), dialect is also used in a broader sense, to refer to a linguistic variation which includes regionally-specific vocabulary (such as the word “bairn” meaning “child” in Scottish English) and syntax (for example, the typically Glaswegian construction “the woman that her dog got run over” rather than the Standard British English form “the woman whose dog got run over”). In this thesis, therefore, the term “dialect” will be avoided, and “accent” will be used to refer to regional variations in pronunciation.

Most human languages share a certain amount of systematic inter-speaker variability in the form of regionally- and socially-determined accents. Increasing geographical mobility and worldwide media such as film and TV mean that we routinely come into contact with a wide variety of both regional and non-native accents, which in turn means that we are exposed to a range of variant pronunciations of words.

Although it could be argued that all speakers have an accent and that the idea of a “non-accent” is a myth (Lippi-Green, 1997), in many languages it is possible to identify one particular accent which is viewed as being neutral or standard, and may thus be used as a baseline for describing accents. In the case of British English, this baseline is known as Received Pronunciation (RP) or Standard British English (SBE) (Hughes, et al., 1996). SBE is not geographically specific, but rather is associated with a particular demographic group; it is the accent of the educated middle classes, and until relatively recently, of the media, hence its nickname, “BBC English”. Similarly in the United States, the accent of the Mid-West and West has come to be seen as a standard “non-accent” (Bonfiglio, 2002), and a baseline to which other American accents can be compared. In this thesis, for the purposes of description, SBE will be taken as a baseline, while recognising that SBE is as much an accent as any of the regional variants described.

Within this thesis we will be looking at differences in the processing of regional and foreign (or native and non-native) accents. Regional accents are defined as being systematic phonetic and phonological variations within the speaker’s first language, which depend on (and thus reflect) the speaker’s geographical, social and cultural background (Hughes, et al., 1996; Wells, 1982), but in which the variations remain for the most part within the set of legal phonemic contrasts for the language being spoken. Thus a native English speaker who was raised in Yorkshire and who speaks English using typically Northern English pronunciations of words (such as “bath” using /æ/ rather than the typically Southern /ɑː/, and “strut” using /ʊ/ rather than the typical Southern /Λ/) may be said to have a regional accent (Wells, 1982).

A non-native or foreign accent, on the other hand, manifests in the speaker’s second language, and “arises from the interaction between the segmental and suprasegmental characteristics of a speaker’s first (L1) and second (L2) languages” (Adank, et al., 2009, p 520) and may include phonetic, phonotactic and prosodic variations from the target language. It reflects the speaker’s own native language and their degree of proficiency in the second language (in which the foreign accent manifests).

Regional accents in English tend to differ mainly in the realisation of their vowels (with some notable exceptions, such as the use of the rhotic /r/ in America, Scotland and the South-West of England, and the use of /t/ and /d/ in place of /Θ/ and /ð/ in many Irish accents), and are often identifiable by a single vowel shift (Wells, 1982). Foreign accents may also include specific phonemic shifts which allowed them to be recognised as foreign. For example, in Japanese, /l/ and /r/ are not contrastive, and thus native Japanese speakers may not differentiate between the two liquids when speaking English (Adank & Janse, 2009a). Similarly, in German no distinction is made between /ɛ/ and /æ/, and thus a native German speaker may pronounce minimal pairs such as “bat” and “bet” in exactly the same way (Bohn & Flege, 1992). However, foreign accents will often vary on a number of different characteristics, including phonemic realisations of both vowels and consonants (as in regional accents) but also prosodic and stress patterns, so that in some cases, a foreign accent would be difficult to perceive or identify from single words or phonemes, and is only noticeable or recognisable from longer utterances (Vaissière & Boula De Mareüil, 2004). Longer utterances also allow for the presence of accent-related co-articulation word-boundary phenomena. These occur when the terminal sound of one word is affected by the initial sound of the next. Co-articulation can be seen when, for example, a word ending with /n/ is followed by a word beginning with a labial consonant such as /b/; in natural speech, the place of articulation of the /n/ sound often takes on more labial elements, creating something resembling an /m/; so “*green beans*” may sound more like “*greem beans*”(Ranbom, Connine, & Yudman, 2009).

This type of intermittent variation may be particular to an accent. For example, in the accent of Bristol in the South-West of England, words ending in an /l/ may be pronounced without the /l/ sound (hence “Bristol” becomes *“*brɪstəʊ*”*), and Standard British English may introduce an intrusive /r/ between certain long vowels (including /ə/, /ɜː/, and /ɔː/) and any following vowel, so that “I saw it” becomes “aɪ sɔːr ɪt”. These types of intermittent variations are also dependent on the phonemic context; before a vowel a terminal /n/ sound is not substantially altered, as, for example, in “*green olives”*, so that in order to correctly interpret the words we hear, we must attend both to the sounds we hear and the contexts in which we hear them.

Accents, both foreign and regional, may therefore vary widely both between speakers (even speakers with similar socio-linguistic backgrounds) and within a single speaker but across utterances, making it difficult to quantify them. Rather, they consist of a set of variations on pronunciation, each of which individually may be neither necessary nor sufficient for the detection or identification of an accent, but which combine in such a way as to allow listeners to achieve quite high accuracy rates in accent detection (Flege, 1984). It may be hard for us to define the exact characteristics of a foreign or a regional accent but, like Justice Potter Stewart, we know it when we hear it, and the evidence suggests that these differences affect speech processing.

As we see, accents uncover a wide range of linguistic phenomena that span all across the whole sentence, from the local, phonemic level to cues to word boundaries and supra-segmental properties. To encompass this wide range of variation, we chose not to reduce our study of the impact of accent variation on speech processing to the word level, but to present listeners with continuous speech.

For experimental purposes, we may sometimes wish to compare accented speech to speech without an accent; since this is not possible, as seen above, we have two alternatives. We can use an accent like SBE or MWAE, which is generally held to be neutral in the country or region from which participants come, or we can use an accent as similar as possible to the participants' own accent, for example testing participants from Glasgow with a Glaswegian accent as their baseline. This latter approach is the one which will be taken in this thesis; participants from the South-West of England will be tested with accents from the South-West as a "home accent" baseline.

Another working hypothesis that we found useful and investigated further in this thesis is the possible functional difference between native and non-native accents which may cause them to be processed differently, and to require different mechanisms for both normalisation and adaptation.

Relatively little work has been done examining the ability of adult listeners to identify foreign or regional accents, but one study which examined this found accuracy levels of between 63% and 95% in detecting a foreign accent (Flege, 1984), although participants in this study were not asked to identify or classify the accent in question. Other studies have focussed on participants' ability to classify accents by geographical region. Clopper and colleagues (Clopper & Bradlow, 2008; Clopper & Pisoni, 2007) asked native speakers of American English to classify speech from speakers from four regions of America, in varying levels of masking noise. This was a free classification task, so no extra information about the accents was provided to participants. The results showed that participants were significantly more accurate at classifying accents in a no-noise condition than with even moderate levels of background noise. Overall, however, performance was poor, suggesting that classifying regional accents, even without the added difficulty created by additional noise, is a demanding task.

Two further studies (Van Bezooijen & Gooskens, 1999; Woehrling & Boula de Mareüil, 2006) tested adult listeners' ability to identify regional variations of their native language. Woehrling and Boula de Mareüil asked participants from Northern and Southern France to classify accents from six regions in France and Switzerland. Unlike Clopper's studies, this study asked participants to select from a list of possibilities for each accent, thus providing them with extra information, over and above that contained in the speech samples themselves. Participants classified the accents correctly on average around 43% of the time, which represents a much more successful response than in Clopper and colleagues' studies, as would be predicted by the use of a forced-choice procedure.

Van Bezooijen and Gooskens carried out two similar experiments, one using Dutch speakers from Belgium and the Netherlands, and one using English speakers from England, Scotland and Wales. They asked participants to classify accented speech with three hierarchical levels of accuracy; country of origin (The Netherlands or Belgium for experiment 1, England/Wales or Scotland/Northern Ireland for experiment 2), Region within the country of origin (North, East, South, West, Middle), and specific province or area (for example, Groningen or Antwerp for experiment 1, East Midlands or Belfast Area for experiment 2). Participants performed well at the country level, correctly identifying accents by their country of origin between 64% and 98% of the time, and were progressively poorer at distinguishing between more specific regional variations, scoring 60-80% at the Region level and only 40-52% correct for Province or Area. However, this study used spontaneous utterances as its stimuli, so many of the stimulus sentences included dialect-specific word choices, and these appear to have contributed to participants' success in classifying them. When standardised sentences were used, accuracy dropped for two of the accents/dialects, (although a statistical comparison was not available due to the small number of speakers), suggesting that their identification relied heavily on dialectal information rather than simply accent.

Given that the tasks in these studies were not identical, and that none of them directly compared the identification of foreign and of regional accents, it is hard to draw conclusions about the differences that might exist between accent categories in adult listeners. However, two studies have looked at foreign and regional accent detection in children (Floccia, Butler, Girard, & Goslin, 2009a; Girard, et al., 2008) and have found an apparent difference between the two. This is will be further discussed later in this chapter.

The fact that performance is generally relatively poor in accent identification tasks suggests that, while it may contribute towards understanding, identification is not essential for the accurate processing of accented speech.

Accents and Speech Processing

Accents are just one source of speech variability; other sources span both production, with intra- and inter-speaker variability affecting the acoustic and phonetic information transmitted, and perception, where the signal may be disrupted by interference, masking, or distortion from external sources (Mullennix & Pisoni, 1990). Despite this variability, as adults we are able to identify and comprehend the majority of what is said to us in our native tongue with a high level of accuracy, and to achieve this rapidly enough that we are generally unaware of the complexity of the underlying cognitive processes.

According to some accounts of word processing, in order to achieve lexical activation of the words we hear spoken, we must first go through several earlier stages including normalisation of the speech stream in an abstract, phonological code used to contact the lexicon (Gaskell & Marslen-Wilson, 1998; Goldinger, 1998; Klatt, 1979; Pallier, Colomé, & Sebastian-Gallés, 2001; Pisoni & Luce, 1987; Ranbom, et al., 2009). These models or propositions are called abstract-entry models. Other accounts, specifically exemplar based accounts (Connine, 2004; Johnson, 1997; McClelland & Elman, 1986; Sumner & Samuel, 2005), bypass the need for normalisation by suggesting that variant word-forms are lexically represented; such accounts will be addressed in detail in chapter 4. According to abstract-entry accounts, firstly, we must extract from the acoustic signal a set of phonological information which can be translated into phonemic "chunks", the building blocks of spoken language. This process will be complicated by several possible factors. For example, background noise can mask the acoustic signal. Distortion or changes in the signal itself such as those incurred when speaking over the phone or by some forms of recording device can also disrupt processing. Acoustic processing can also be affected by natural intra- and inter-speaker variability, including that caused by disordered speech or, and this is what we will be concentrating on, when the speaker has a regionally distinct accent or a foreign accent.

Secondly, the natural speech stream is not cleanly segmented into separate words; an examination of the acoustic trace of a spoken phrase demonstrates clearly that while there are pauses or breaks in the signal, these are as likely to occur in the middle of a word as between two words, making them an unreliable source of information for segmentation (McQueen, 1998). Our next task, therefore, is to identify where each word in the speech stream ends, and the next begins. To do this, we may use several different types of information, including language-specific cues such as prosody (Echols, Crowhurst, & Childers, 1997) and stress patterns (Jusczyk, Houston, & Newsome, 1999b), phonotactic rules (Mattys, Jusczyk, Luce, & Morgan, 1999), and a computational analysis of the probabilities of co-occurrence of pairs or sets of syllables (Thiessen & Saffran, 2007), as well as pattern-matching using existing lexical knowledge (Mattys, White, & Melhorn, 2005). Because accents, and particularly foreign accents, can affect the stress and prosodic cues, and the phonotactic rules, which assist in segmentation, the added variability associated with accents can disrupt this process, resulting in mis-segmentation of words in the speech stream (for example, “four candles” instead of “fork handles”, caused in part by the accent-related elision of the /h/ at the beginning of “handles”, as illustrated in the famous comedy sketch by the Two Ronnies). Because of such phenomena, it seems reasonable to consider the impact of accents on speech perception in the context of fluent speech, and not simply within single words.

Having isolated an individual word, we then have to match it to a stored representation in our lexical system in order to identify it, but this process may be hampered by variability; if a familiar word is pronounced in an unfamiliar way (for example, in a novel accent, or using some form of digital distortion) and if our speech recognition system is not able to normalise this variation pre-lexically, this should create a mismatch between the acoustic input and the existing lexical representation. This in turn should make it much more difficult to access the correct lexical entry, resulting in poor comprehension. However, despite these difficulties, we are generally capable of understanding speech even when it is spoken in an unfamiliar accent, or otherwise distorted, suggesting that the speech signal is in some way normalised in order to allow accurate lexical access, and/or that processing accented speech requires a larger involvement of top-down information to compensate for the imprecise input.

Foreign and Regional accents

If we assume, as outlined by Klatt (1979), that the speech signal undergoes a process of normalisation in order to clean out variation and allow accurate lexical access, one question which must be addressed is whether the same processes are applied to different types of accents, or whether foreign and regional accents recruit different processes. The literature contains two main hypotheses; one, the Perceptual Distance hypothesis, suggests that accents can be arrayed along a continuum of difference from the listener's own accent, and that differences in the normalisation process are purely quantitative, corresponding with the perceptual distance of a given accent from the baseline (Clarke & Garrett, 2004). A second, the Different Processes hypothesis, suggests that foreign and regional accents may recruit qualitatively different processing mechanisms which would translate, partly, in the use of different types of normalisation (Adank, et al., 2009; Floccia, et al., 2009b; Girard, et al., 2008; Goslin, Duffy, & Floccia, 2012). A further proposal (Cristià, Bradlow, Vaughn, Schmale, Seidl, & Floccia, In preparation) is that differences in the processing of foreign and regional accent depend not on perceptual distance but rather on the cost of processing, with accents that are harder to process (such as foreign over regional accents, or accents with the added cognitive load created by the addition of noise or distortion) causing greater reliance on top-down cues.

In support of the Perceptual Distance hypothesis, Clarke and Garrett (2004) found that English-speaking listeners adapted more rapidly, with less exposure, to Spanish-accented English than to Chinese-accented English. They suggest that this may be because the Chinese accent varies more from the listeners' own accent than the Spanish accent does, and thus presents a harder task. They further propose that the variation presented by accents can be seen as an extreme form of the normal inter-speaker variation encountered when listening to multiple speakers within a familiar accent and that accents which vary more perceptually from the listener's own accent would therefore be expected to incur longer delays and higher error rates in processing. This in turn implies that accents can be placed along a continuum of perceptual difference from the listener's own "home" accent, with foreign accents typically falling further away than regional accents, and thus suffering greater delays and error rates than regional accents.

There is certainly evidence that foreign accents incur greater delays and higher error rates than regional accents. Adank et al. (2009) presented listeners with statements spoken in a familiar or unfamiliar regional (Southern English and Glaswegian English) and foreign (Spanish) accent, and asked them to make a true/false judgement about each one. They found that unfamiliar regional accents incurred slower response times and more errors than a familiar accent (which matched that of the listener), and that the foreign accent increased both reaction times and error rates even further. Similarly, Floccia, Goslin, Girard and Konopczynski (2006) presented French-speaking listeners with sentences spoken in a familiar accent, an unfamiliar regional accent of French, and a foreign (English) accent, and measured their performance on a go/no-go lexical decision task. They found slower responses to the unfamiliar regional accents than to the home accent, and again this effect was stronger still for the foreign accent. In addition, they found adaptation to the regional accents with longer exposure, so that response times to long sentences were shorter than those to medium-length sentences. However, the foreign accents did not show this adaptation. These findings could be interpreted as showing that foreign accents fall further along the continuum of difference from the listener's accent, and thus the disruption caused by the accent-based variation is more severe, and allows less adaptation.

However, these results are also compatible with the second hypothesis, namely that regional and foreign accents recruit qualitatively different processes of normalisation. The fact that Floccia et al. (2006) found no evidence of adaptation to the foreign accent, even though there was rapid adaptation to the regional accents, may imply that the normalisation processes which allow adaptation in the case of regional accents are not sufficient for foreign accents. Two studies carried out with children also suggests that different processes may be at work; Girard, Floccia and Goslin (2008) found that French-speaking children were able to distinguish between foreign and native accents but did not distinguish between different regional accents. Floccia, Butler, Girard and Goslin (2009) carried out a similar study, in which British English children were asked to discriminate between their own accent, and two unfamiliar accents of similar strength, one of which was a regional accent (Irish) and one a foreign accent (French). Their results showed that at five years old, children performed poorly at this task, with results only slightly above chance, whereas at seven years old, they could reliably detect the accents. Further, the seven-year-olds were significantly better at detecting the foreign accent than the regional accent.

In addition to the experimental evidence, Chambers (2002) notes in a sociological discussion that the children of immigrant families will typically speak the language of their home without any trace of their parents' accent (and indeed may be unaware that their parents have an accent at all). On the other hand, the children of parents from a different region will often, in the early years, speak with a mixture of their parents' accent and the accent of the place in which they are growing up, or code-switch between the two (Hazen, 2002). Supporting Chambers' hypothesis, Khattab (2007) showed that Arabic/English children living in Yorkshire, England produced English-sounding vowels when speaking to a monolingual English-speaker, in spite of the fact that their Arabic-speaking parents produced those vowels with an Arabic accent. A similar study with the London-based children of Gujarati parents again showed that the children produced British-accented vowels rather than the Gujarati-accented vowels they heard from their parents (Evans, Mistry, & Moreiras, 2007). However, further work shows that the children of immigrants can and do produce foreign accented speech under certain circumstances, such as when code-switching between languages (Khattab, 2007) and when talking to their parents rather than to a native speaker (Sharma & Sankaran, 2011). This suggests that the acquisition of non-native accents in bilinguals may be a complex and context-dependent process, but for the sake of our discussion it calls into question the reality of the early foreign–accent filter posited by Chambers.

In summary, the studies looking at children's perception and production of foreign and regional accents support the idea of a qualitative rather than simply a quantitative difference in the way the two types of accents are processed. Investigating these possible differences is, however, a complex matter, and it is not yet clear what specific features of foreign accents as compared to regional accents might be driving differences in the processing of the two accent types. Clarke and Garrett's hypothesis of a continuum of difference might seem to imply that accents differ along a single dimension, and that they can be accurately and objectively ranked. However, accurately defining and classifying that accent would be a near-impossible task. Accents may differ, for example, in the production of specific vowel sounds (Schmale & Cristià, 2009), in the place of articulation of consonantal sounds (Gonet & Piétron, 2004), in their prosody (Russell, 2007), and their phonotactic rules (Mattys & Jusczyk, 2001), as well as their strength (Flege, Munro, & MacKay, 1995) and various inter-speaker differences such as rate of articulation and fluency (particularly in the case of foreign accents; L2 speakers will tend to speak more slowly, with more hesitation (Chambers, 1997), than L1 speakers). Clarke and Garrett follow Nygaard and Pisoni (1998) in assuming that foreign accents are more extreme, or more different from a baseline "home" accent, than regional (L1) accents. However, it is not clear that this is always the case.

Anecdotally, many English-speakers find the accents of some regions of the United Kingdom, such as Newcastle and Glasgow, very hard to understand, and we might predict that they would perform better in tests of comprehensibility and intelligibility with the accents of many fluent non-native speakers.

The distinction between comprehensibility and intelligibility is a further confounding factor here. Intelligibility is defined as the accessibility of the speaker's intended meaning (Floccia, et al., 2009b), and is measured via transcription or verbal repetition tasks, categorisation or word-spotting tasks, or other tasks which seek to test that the semantic content of speech has been correctly interpreted. Comprehensibility is a measure of the effort required to process or understand the stimuli, and is most often measured using response times. Thus it is possible to measure the intelligibility of stimuli without taking their comprehensibility into account. Where studies do so, they may report that complete adaptation to an accent or other form of variability or distortion has occurred, because the error rate has dropped to baseline. This ignores the possibility that while intelligibility has recovered, comprehensibility has not, and may therefore give an inaccurate impression of the difficulty involved in processing variant speech. It is also difficult to compare a study which tests only intelligibility with one which tests comprehensibility, and this may help to explain why studies such as Bradlow and Bent's (2008), which tested intelligibility using a word transcription task, and Floccia et al.'s (2009), which tested comprehensibility using a lexical decision task, seem to show mutually incompatible results. Bradlow and Bent showed successful adaptation in their transcription task, as evidenced by increased accuracy after exposure, while Floccia et al. found only partial recovery. Unlike Floccia et al., Clarke and Garrett (2004) found adaptation to non-native accents when using a measure of comprehensibility. However, where Floccia et al. used a lexical decision task, Clarke and Garrett used a cross-modal matching task; since this task involves presenting participants with a written word-form which either matches or does not match the spoken target, it may allow participants to use top-down information to achieve lexical access, while Floccia et al.'s task did not provide participants with extra information to inform their decision. This means that despite the apparent similarity of the measures used (response times as well as error rates), the two studies may not be truly comparable.

Further work is needed, therefore, in order to understand whether the processing of unfamiliar accents is determined purely by the perceptual distance from the listener's own accent (taking into consideration the many ways in which accents can differ), whether there are specific types of differences which are most important in determining how accented speech is processed, or whether there is also a categorical effect of the type of accent (whether regional or foreign). In particular, it is necessary to use a paradigm which is less reliant on task demands, in order to draw generalisable conclusions about the way in which different types of variant speech are processed.

Adaptation to accents

After having discussed the challenges raised by recognising words in accented speech, we now turn to what happens with repeated presentation of the same accented signal. When we first encounter an unfamiliar accent, anecdotal reports suggest that we will often find it hard to follow, but we quickly adapt, becoming faster and more accurate at recognising and understanding the words spoken in a novel form (Pinet, Iverson, & Evans, 2011). There is, however, some debate as to the extent of this adaptation; while some researchers have found that responses to variant pronunciations of various sorts, including time-compressed speech (Adank & Devlin, 2010), vocoded speech (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005) and accent (Clarke & Garrett, 2004; Magen, 1998) return to baseline levels within a few sentences, other studies have failed to replicate this finding with accented speech (Adank & McQueen, 2007; Bradlow & Bent, 2008; Floccia, et al., 2009b). The differing results in the accent adaptation studies may be in part due to differing methodology, such as the use of multiple speakers (Bradlow & Bent, 2008; Kraljic & Samuel, 2007) versus one single speaker (Clarke & Garrett, 2004). Bradlow and Bent (2008) actually found that adaptation to a single speaker did not generalise well to a novel speaker, whereas adaptation to multiple speakers generalised much better to novel speakers. They concluded that for speaker-independent adaptation to occur, we require exposure to greater variability than that provided by a single speaker, although we are able, perhaps through a different process, to adapt to a single speaker simply by exposure to their speech. It is not yet clear whether exposure to multiple speakers could assist with speaker-independent adaptation across accents; that is, whether exposure to the accent of one region or language would improve performance on other, perhaps related, accents.

Processing variant forms.

While there has been an increasing focus in recent years on accent and dialect as a source of naturally-occurring inter-speaker variability (see the recent review by Cristià, et al., In preparation), it is not the only source. Speech disorders can also cause a person's speech to be naturally degraded by comparison with that of healthy people. Dysarthria, or impaired articulation, can be caused by neurological damage such as that seen in Parkinson's disease, or by damage to the musculature or nerves of the face and tongue, and impaired hearing can also result in imprecise articulation and intonation. Apraxia of speech, characterised by syllable transposition and difficulties with articulation and prosody, may occur developmentally or after cerebral trauma, for example, after a stroke or a head injury (Ziegler, 2008). Relatively minor speech disorders such as stuttering or lisping are common in childhood, although some are successfully treated before adulthood. Even so, as many as one in six people may suffer from some form of communication disorder (National Institute of Health, 2007), so our exposure to naturally degraded speech is likely to be relatively common. The variability caused by such disorders shares some features with that caused by accents; it will tend to result in a set of relatively consistent changes in pronunciation, within a given speaker, affecting the same phonemes in the same way in a given context. Specific disorders may also share perceptual features (Mattys & Liss, 2008), thus providing some level of inter-speaker consistency, just as speakers sharing a geographical and social background will have similar accents.

Neuropsychologists have long studied speech production in people with such disorders but less attention has been paid to the way in which this natural degradation affects the perception and normalisation of such speech. However, if we assume that different forms of variance are processed in similar ways, differing quantitatively rather than qualitatively, the studies looking at the perception of naturally disordered speech can inform research into the effect of variability in the speech signal just as accent studies do. Mattys and Liss showed that word recognition with naturally degraded speech is significantly improved with speaker consistency; that is, a previously heard word is recalled more quickly and more accurately when repeated in the same voice than in a different voice, even when the speaker suffers from severe dysarthria and thus the word itself has poor intelligibility (Mattys & Liss, 2008). This is consistent with Bradlow and Bent's finding (2008) that we are able to adapt to the indexical features of a single speaker, and supports the idea that we are able to adapt to variant forms of our own language (for example, see Clarke & Garrett, 2004). Mattys and Liss also propose that degraded variants may encourage a greater reliance on surface acoustic characteristics of the speech signal, resulting in slower responses. This offers some support to the idea of a dual-process route, where effortful tasks rely more heavily on indexical information than easy tasks (McLennan & Luce, 2005).

As well as natural variation in speech, distortion of the speech signal can also be created artificially, and a number of studies have used artificially altered speech to study the processing of variant speech. Davis et al. (2005) asked participants to transcribe vocoded speech, and found that accuracy improved from around 0% to around 70% over the course of 30 sentences, suggesting adaptation to the distorted speech. Performance was also improved if listeners were given information about the identity of target words, suggesting that top-down processes can help with adaptation. Similarly, Dupoux and Green

(1997) Adank and Janse (2009b) found adaptation to time-compressed speech, as shown by improved recall (Dupoux & Green, 1997) or reduced reaction times in a true/false task over the course of ten blocks (Adank & Janse, 2009b). These studies are sometimes cited as demonstrating adaptation to variant speech forms, and offered as analogous to accented speech.

However, if it is the case, as proposed by the Different Processes hypothesis, that foreign and regional accents recruit different normalisation mechanisms, it is likely that disordered speech and digitally altered speech may also recruit their own unique processes. If this is the case, it may not be useful to compare them to intact but accented speech.

Summary

We have seen that the acoustic forms of spoken words vary according to a number of inter- and intra-speaker accent-related factors, which will affect the listener's ability to process what they are hearing. Lexical access of variant word-forms may be achieved via a process of normalisation to an abstract representation (as outlined by Pallier, et al., 2001, and further discussed in chapter 4) or through comparison to multiple representations, as proposed by Exemplar-based models (Johnson, 1997). We also discussed whether accents might be categorised differently by the perceptual system depending on whether they are native or not, or whether their perceived differences from the listener's own accent can be ranged along a simple perceptual continuum, with processing becoming more effortful and less successful the further along the continuum a given speaker's accent falls. We have underlined the parallel between perception of distorted speech due to aphasia or other speech disorders and the perception of accented speech (Mattys & Liss, 2008). In particular we pointed out Mattys et al.'s conclusion that this kind of speech might recruit more reliance on the acoustic signal, supporting a dual-route approach (McLennan & Luce, 2005). The evidence from artificially distorted speech, however, suggests that making top-down information available aids adaptation, once again emphasising the complex range of operations necessary for accurate speech processing.

Overall, then, distortion and variation is an intrinsic element of the language we hear on a day-to-day basis, and therefore a necessary part of the normalisation and processing of natural speech, but also a tool which can be used in order to investigate these processes in more depth.

In order to investigate the effects of natural variation, and specifically that caused by accents, two approaches will be taken. Adult studies look at the established cognitive systems which are used to process language in fluent listeners, treating language as a *fait accompli*. In order to pick apart the effects of different types of accent, we will use event-related potentials (ERPs) to compare responses to home, regional, and foreign accents at the neurophysiological level. Infant studies, on the other hand, look at language as a developing system, allowing us to look not simply at an end-point (that of fluent language use) but of the processes involved in acquiring language in the early years. In the early stages, infants' lexical and phonetic representations are relatively rigid, with only a limited ability to normalise variant forms (Johnson & Jusczyk, 2001; Schmale & Cristià, 2009; Singh, Morgan, & White, 2004). With age and experience, they gradually become more flexible, accommodating more variation in input without losing accuracy, and existing studies suggest that this process occurs earlier for some forms of variation than for others (Jusczyk & Aslin, 1995; Rost & McMurray, 2010; Singh, 2008). We will therefore pit foreign and regional accents against a home accent in a set of behavioural and ERP studies intended to cast light on the age at which infants are able to normalise speech within and across accents. While these two approaches may overlap, it is important to bear in mind the differences in the participant groups, and thus adjust our direction of enquiry and the methods we use accordingly. For this thesis, the two branches will be addressed separately, in order to allow us to focus on the relevant questions and methods for each one.

Chapter 2: Introduction to Adult work

Behavioural methods

Traditionally, psycholinguistics studies have used behavioural techniques with adults to investigate the processing of language. Simple tasks such as lexical decision tasks (in which the participant is asked to judge whether a stimulus is a real word or a non-word) allow us to look at both accuracy and response times, giving us a measure of the comprehensibility of varying types of stimuli. These tasks can also be manipulated, for example by adding a masking or a priming stimulus, allowing further insight into the underlying processes of language processing. We can also look at the intelligibility of stimuli, using transcription tasks (in which the participant is simply asked to repeat or write down the words they have heard) or categorisation tasks (in which the participant must choose which of two semantic categories a stimulus belongs to); the former give us accuracy data and also allow us to investigate the types of errors being made, while the latter give us response time data as well as accuracy/error rates (Floccia, et al., 2009b).

However, there are some disadvantages to these methods. In some cases, the distinction between intelligibility (that is, the ability to access the meaning intended by the speaker) and comprehensibility (the effort required to do so) may not be clear, which can result in apparently conflicting results from different studies, due to their use of subtly different measures. For example, Bradlow and Bent (2008) report full adaptation to foreign accents using a word transcription task to measure intelligibility, while Floccia et al. (2009b) report only partial adaptation when using a lexical decision task to measure comprehensibility. When looking for differences between two types of stimulus, these measures rely on differences of degree, in accuracy or in speed of response. This means that they can distinguish stimuli which are processed more slowly or less accurately than a baseline, but cannot necessarily distinguish different types of variation; if two stimuli both result in reduced accuracy and/or reduced response speeds, they will be seen as equivalent to each other. However, this is a somewhat one-dimensional approach, which may not necessarily be appropriate, as it may lead to misleading assumptions. An example, which will be tackled in this thesis, is the question of regional versus foreign accents. As discussed in the main introduction, some researchers have pointed to lower accuracy rates and slower responses in regional accents compared to a baseline "home" accent, and in foreign accents compared to both "home" and regional accents, as evidence that foreign and regional accents differ from a listeners' own accent along a continuum, such that foreign (L2) accents differ in the same way as, but to a greater degree than, regional (L1) accents (Clarke & Garrett, 2004; Nygaard & Pisoni, 1998). There is, however, some evidence that foreign and regional accents may not, in fact, be processed in the same way, differing only by degree of difference, but that they recruit different mechanisms for normalising and processing (Floccia, et al., 2009b; Girard, et al., 2008). It may not be possible for traditional behavioural tasks to pick apart these differences, making it necessary to use other approaches.

A growing number of studies contrasting behavioural and neurophysiological responses has shown that in some cases, a disassociation may exist between overt behaviour and neural responses to a task (Chee, 2009). Dehaene et al. (2001) used masked and clear visual stimuli, and showed that participants were unable to make behavioural responses to the masked words either at the time of presentation or later, when asked to recognise them. However, ERP data showed that there was a significant P1 component in response to masked words, over and above that towards a masked blank, demonstrating that some unconscious processing of the masked words was taking place. Unconscious processing was also shown by a study in which Chinese/English bilingual adults were asked to make a lexical relatedness decision on pairs of English words (Thierry & Wu, 2007). They were not informed that some of the word pairs shared a character when written in Chinese. The behavioural data showed no effect of the "concealed" repeated character, but ERP data revealed a significant difference between responses to pairs which did share a character and pairs which did not, regardless of the relatedness of the pair. This finding demonstrates that Chinese/English bilinguals, whose first language was Chinese, were unconsciously translating the written words, but behavioural studies have failed to show any effect of unconscious translation; it is only by examining the evidence directly from the brain that this discovery could be made. In some cases, stimuli may elicit both behavioural and neurophysiological responses, but these different types of data may show different patterns. Holcomb, Grainger and O'Rourke (2002) asked participants to make a lexical decision to words and pseudo-words with large or small orthographic neighbourhoods. The behavioural responses demonstrated that a large orthographical neighbourhood results in a faster lexical decision to words but a slower decision to pseudowords. However, the ERP data showed no such interaction; the N400 component, associated with semantic information, was larger for stimuli with larger orthographic neighbourhoods regardless of word/pseudoword status. This suggests that different strategies are being used to perform the behavioural task for words and non-words, with greater N400 activation leading in quicker positive responses to the words, but inhibiting negative responses to pseudo-words. The ERP data make it possible to examine the underlying processes which result in the behavioural responses, thus complementing the behavioural data. Neurophysiological studies, then, offer a deeper and more detailed way of approaching linguistic tasks which may provide information unavailable through purely behavioural approaches.

ERP approaches

ERP studies have been used to examine responses to a range of different forms of variation in the speech signal, including acoustic filtering (Aydelott, Dick, & Mills, 2006), and time reversal (Boulenger, Hoen, Jacquier, & Meunier, 2011), but also, of more relevance to this thesis, regional (Brunelliere, Dufour, & Nguyen, 2011; Brunelliere, Dufour, Nguyen, & Frauenfelder, 2009; Conrey, Potts, & Niedzielski, 2005) and foreign (Berman, Mandelkern, Phan, & Zaidel, 2003) accents. However, although a number of studies have looked at the disruption to processing caused by accents (Adank & McQueen, 2007; Bradlow & Bent, 2003; Clarke & Garrett, 2004; Floccia, et al., 2006; Maye, Aslin, & Tanenhaus, 2008; Weil, 2001), the majority of these studies have looked at only one type of accent (foreign, or regional, or in some cases artificial accents). Relatively few studies have used both foreign and regional accents, and even those that have (Adank, et al., 2009; Floccia, et al., 2009b), have generally compared each type of accent to a baseline without directly pitting different types of accent against each other. Because of this, we are able to say that accented speech is processed somewhat more slowly and less accurately than familiar speech (as discussed earlier, it is not accurate to refer to unaccented speech, so "familiar speech" will be used throughout, to indicate speech in an accent which is close to that of the participants, so as to serve as a baseline), but we are less able to make generalisations about the relationships between different types of unfamiliar accent. This poses a problem for Clarke and Garrett's (2004) continuum of difference, which assumes that the difficulty of processing an accent will be directly proportional to its distance on the continuum from the baseline, and that foreign accents fall further along the continuum than regional accents. In order to support this assumption, we need to compare foreign and regional accents against each other as well as against a baseline, in order to demonstrate that their processing differs in the same manner as, but to different degrees from, that of the baseline accent, and that the degree of difference in difficulty of processing is correlated with the distance from baseline (Nygaard & Pisoni, 1998). However, if different types of accents are in fact processed qualitatively differently (Floccia, et al., 2009b), we might expect to be able to pinpoint these differences more accurately using ERP studies than with behavioural studies. This thesis will take two approaches to this problem. Firstly, ERP can be used to look directly at the processing of speech in different types of accent - familiar speech, regional accents, and foreign accents – and to compare each of the classifications against the others. Secondly, we can use ERP to examine the priming effect of different types of accented speech on familiar speech, using repetition suppression. Using both of these approaches allows us an in-depth picture of the way different types of accented speech is processed and represented.

Examining accented and familiar speech directly using ERP allows us to look not only at gross differences in processing, but at the exact time-course of the processing of individual words. This is important, because it is already well-established that words are processed in different stages (Marslen-Wilson & Welsh, 1978). Raw acoustic information is processed first, followed by acoustic information, which is mapped onto phonetic features, which are then converted into phonological segments. This bottom-up process is informed by top-down constraints, which allow listeners to restore or normalise a distorted acoustic signal, often without being aware of the distortion. Finally, once this has been accomplished, we can achieve lexical access, giving way to semantic and syntactic information (Pisoni & Luce, 1987). Lexical access, then, involves an interaction between bottom-up and top-down processes, but it nonetheless appears to occur in stages, starting with acoustic processing and moving through phonetic and phonological phases. Here, we will focus on the data-driven sequential processes, while acknowledging that top-down influences are important throughout word recognition. Because these processes happen sequentially, and because EEG data gives us extremely fine temporal resolution, we are able to follow the time-course of word recognition, and look for differences at each stage. For example, a study looking at perception of non-native (Hindi) consonant contrasts (dental /d/ and retroflex /d/) in English speakers (Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000) showed that participants were unable to report a distinction between two sounds not contrasted in their own language, but that the two sounds were distinguished in the ERP data, within the first 300 ms after stimulus onset. This early effect represents auditory/acoustic processing, and precedes more language-specific phonological processing (which is more affected by perceptual narrowing), and lexical/semantic access. This suggests that the non-native contrasts were retained at the pre-attentive auditory/acoustic level, even though they were not distinguished phonetically.

Brunellière and colleagues also used ERPs to examine responses to a native vocalic contrast (/ø/ versus /y/), and a second vocalic contrast (/e/ versus /ɛ/) which was no longer native to the (Swiss French-speaking) participants (Brunelliere, et al., 2009), and found an early component, the N100, around 95ms after onset, in response to both contrasts, whereas only the familiar contrast elicited a later component at P200, indicating that the contrast could be distinguished at the purely acoustic level, but that perceptual narrowing had eliminated the distinction between the two vowels which were not contrastive in the participants' native accent at the phonetic/phonological level. A further study (Brunelliere, et al., 2011) demonstrated that the responses of Southern French speakers, for whom /e/ - /ɛ/ is not contrastive, differed from those of Standard French speakers, for whom the contrast is preserved. This set of studies demonstrates the importance of linguistic and dialectal experience in forming phonetic perceptions, but also shows us that ERPs can be used to inform our understanding of the underlying processing involved in distinguishing linguistically relevant contrasts and ignoring those which are not native. Brunellière and colleagues have also demonstrated that it is not only foreign contrasts which can be examined in this way, but also regionally-specific contrasts within a given language, making their work particularly relevant to this thesis.

Studies which examine responses to single phonemes or syllables can thus inform us about fine-grained phonological representations. However, much of the variation which goes to make up an accent (and particularly in the case of foreign accents) exists at the suprasegmental level. To examine the brain's responses to accented speech, then, we need to look at responses to whole words or sentences. In this way, we can allow for top-down as well as bottom-up processing, and we can also look not just at auditory/acoustic and phonetic/phonological processing, but also at the later stages of word processing; lexical and syntactic access (Connolly & Phillips, 1994). Since many of the behavioural studies we have discussed use intelligibility as a measure, and since intelligibility is a measure of lexical access, our studies must use whole words, ideally in the context of sentences, in order to provide comparable results and allow us to look at the complete time-course of word processing. Connolly and Phillips showed, using high and low cloze probability sentences, that phonological aspects of words are processed earlier after word onset than semantic aspects. Sentence-final words beginning with an unexpected phoneme produced a phonological mismatch negativity between 270 and 300 ms after onset, while a semantic violation produced a separate component around 400 ms after onset (the N400). The ability to separate out responses to two different types of violation demonstrates the usefulness of the fine temporal resolution of ERP data. Using ERP to look directly at sentence-final words in familiar and accented speech should allow us to look at the phonological processing, but also to examine later effects, which should serve as a measure of intelligibility; that is, whether we're able to access the intended meaning of the speech stimulus.

Studies using whole sentences rather than individual phonemes as stimuli may also allow us to side-step one potential confounding factor; that of adaptation to variant speech (including accent). If listeners adapt to accented speech during the course of exposure, this might create order effects in studies using a large number of accented stimuli, with relatively large accent effects during the early trials, but much reduced effects later, as the listeners adapt. Norris, McQueen and Cutler (2003) demonstrated that exposure to ambiguous tokens of a phoneme such as /s/ or /f/ can cause a shift in listener's phonemic boundaries, and Eisner and McQueen (2005) went on to show that this adaptation may be talker-specific rather than being generalised to novel talkers. Kraljic and Samuel (2005) exposed adult listeners to sentences containing ambiguous tokens of either /s/ or /ʃ/. Following exposure to 20 such tokens, participants' phonemic categories were shown to have shifted to allow the ambiguous tokens to be interpreted as "good" representations. However, this effect was limited to the speaker used in training, and did not generalise to a different speaker, and it could be neutralised by listening to the same speaker producing "good" tokens of the sounds which had originally been affected. These studies suggest that adaptation to specific phoneme shifts can occur under certain conditions, and that relatively few tokens are required to elicit adaptation. It has not yet been clearly established to what extent we are able to adapt to accents, which include not just simple phoneme shifts, but also variation at the supra-segmental level. Clarke and Garrett (2004) found adaptation to a novel accent within only two sentences, but their study used only a single speaker of each accent, and adaptation was measured only in terms of intelligibility and not comprehensibility. By contrast, Floccia and colleagues (Floccia, et al., 2009b) found only limited adaptation, with long-lasting perturbation of processing, when speech was delivered by multiple speakers and reaction times on a lexical decision task were used as a measure of comprehensibility. This suggests that we are able to adjust our processing of accented speech in order to extract the intended meaning, but at a cost in terms of the effort required to do so. While these two studies use different measures and find somewhat conflicting results, both lead to the same prediction; a study directly examining the brain's responses to familiar and accented speech using whole sentences should allow adaptation to take place more quickly and effectively than the use of single words (since much of the variation which comprises an accent is suprasegmental) and therefore they should not suffer from adaptation or practice effects. According to Clarke and Garrett any adaptation which takes place will do so within the first two sentences (and will thus have little or no effect on the results overall), and according to Floccia and colleagues any adaptation should in any case be minimal, and mainly affect later ERP components related to post-lexical processing.

Components of interest

Existing ERP research into the processing of variant word-forms has shown a number of electrophysiological components which respond to specific elements of speech processing. Two in particular are likely to be of interest for this thesis; the Phonological Mismatch Negativity, or PMN, and the N400.

The PMN typically peaks around 250-300 ms after stimulus onset, and occurs only in the auditory modality (although it can also be found in cross-modal studies in which visual stimuli are used to set up expectations of the up-coming speech). It is also known as the P2, and is described as the earliest effect elicited by a violation of expectations about speech sounds. It responds to a mismatch between the acoustic/phonetic input and phonological expectations, and is thus clearly demonstrated by studies which use high- and low-cloze probability sentences; that is, sentences in which the final word is very predictable (high-cloze probability), or which do not have a predictable final word (Connolly & Phillips, 1994). By manipulating the semantic and phonological expectedness of the final word of such sentences, Connolly and Phillips were able to show that the PMN occurs when the final word of the sentence was phonologically unexpected, regardless of its semantic viability. For example, a sentence such as “Don caught the ball with his glove” (where the highest probability final word is “hand” but “glove” is semantically acceptable) elicits a PMN response, but “The gambler had a bad streak of luggage” (where the initial syllable of the final word matches that of the highest probability word, “luck”, but represents a semantic violation) does not. In other words, the PMN is elicited specifically when the listener has a pre-existing expectation, generally driven by the context but also influenced by their prior knowledge, of the phonological form of a target, and when this expectation is violated by the incoming stimulus.

The N400, on the other hand, is elicited by semantically incongruous words, but not by phonological violations, as demonstrated by the fact that “Don caught the ball with his glove” does not elicit an N400, but “The gambler had a bad streak of luggage” does. This double disassociation makes the PMN and N400 particularly useful in investigating the processing of variant speech, since it allows the PMN to be used as a measure of the phonological distance between the acoustic input and the listener’s phonological expectations, while the N400 can be used as a measure of lexical access. However, ERP measures are considerably more precise than behavioural measures, and allow for much more detailed analysis.

Experiment 1

An ERP-based approach, then, offers a way to compare directly the neural processing of different types of variation in speech, such as that elicited by regional and foreign accents. By presenting participants with whole sentences of speech in a home accent, and in regional and foreign accented speech, Study 1 will make a direct comparison of the brain's response to variant word-forms. An examination of the PMN and N400 components elicited by the different types of accent should make it possible to shed some light on the way regional and foreign accents are normalised, and, in particular, to see whether the processes recruited by regional and foreign accents differ from those recruited by a home accent purely quantitatively, as a function of their perceptual distance from the home accent, or qualitatively. The Perceptual Distance hypothesis predicts that the processing of foreign accents should differ from that of a home accent in the same way as that of regional accents, but to a greater extent. However, the Different Processes hypothesis gives rise to a different prediction; that because regional accents include variability consistent with the phonetic set of the language being spoken, while foreign accents include sub- and supra-segmental variation from the speaker's L1, therefore the normalisation processes recruited will be different. These differences should affect not only the amplitude of the PMN component, but also possibly its direction (taking processing of the Home accent as a baseline). In addition, it is predicted that differences may be found in the N400, reflecting incomplete pre-lexical normalisation (Floccia, et al., 2009b) in the case of foreign accents, whereas regional accents are predicted to be indistinguishable in this component from the Home accent, since it is expected that normalisation during the pre-lexical phase would be complete, leading to successful lexical access.

Experiment 2

A second approach to accented speech using ERP is the use of repetition-suppression studies. A number of neuro-imaging studies have found robust evidence of reduced activity in some areas of the brain in response to repeated stimuli (Baldeweg, 2007; Bergerbest, Ghahremani, & Gabrieli, 2004; Garrido, 2009; Grill-Spector, Henson, & Martin, 2006), which has been interpreted as a neural correlate of priming, allowing for quicker, more accurate performance. What makes repetition-suppression a useful phenomenon in psycholinguistics studies is not simply that it is robust and ubiquitous, but that it does not require that the stimulus being repeated is identical; rather the degree to which activity is suppressed is dependent on the similarity of the two iterations (Noppeney & Penny, 2006); if two stimuli are distinguishable but share some attributes, the first will produce a suppression or priming effect on the second. The degree to which responses are suppressed, then, is a reflection of the importance to processing of the shared attributes versus the distinguishing characteristics. This allows repetition suppression to be used in a range of studies, from those examining phonological priming (Vaden, Muftuler, & Hockok, 2010) or indexical effects such as the speaker's gender (Orfanidou, Marslen-Wilson, & Davis, 2006), through to those using much higher-level attributes; Phillips and colleagues (Phillips, Klein, Mercier, & de Boysson, 2006) showed a repetition-suppression effect in bilinguals in response to words spoken in their L1 following presentation of the equivalent word in their L2, indicating that when listening to the target words in their second language, the equivalent words in their first language were being activated (although the same was not true in reverse). This shows that repetition-suppression can reveal complex underlying processes which cannot be directly observed or measured and that it can be used to capture information about the full range of features involved in the processing of speech, from the acoustic/auditory stages all the way through to post-lexical semantic, morphological and syntactic processing. Exploiting this technique to examine the processing of accented speech as compared to familiar speech should allow us not only to see whether there are coarse differences in response times or accuracy, as shown by behavioural studies, but to look for differences at the auditory/acoustic level, and the phonetic/phonological level, and the post-lexical level, giving us a much more in-depth understanding of the processing taking place. This allows different types of variation, such as that represented by regional versus foreign accents, to be compared in terms of their similarity to a familiar baseline (the listener's own accent), at a number of different stages of processing or normalisation.

Repetition-Suppression studies have a further advantage when comparing accented and familiar speech; because they enable us to separate the different stages of word processing, they can also allow us some insight into the question of whether variant word-forms, including words spoken in an unfamiliar or rarely-heard accent, are represented individually, as suggested by the Exemplar-based approach (Johnson, 1997), or whether they are normalised pre-lexically, as in the Abstract entry approach (Pallier, et al., 2001). According to the Exemplar-based approach, accented pronunciations of a given word should be stored separately, and only reconciled post-lexically. The strength of a given representation should be dependent on how commonly encountered it is (Sumner & Samuel, 2005), rather than on the relative strength of the accent or the degree of difference from the familiar speech. If this is the case, we would expect target words in accented speech to prime the same target words in familiar speech at the post-lexical phase, but not prior to that, and therefore we would expect to see suppression only in the later epochs of the ERP data. Further to this, we would expect to see the degree of suppression correlate with the familiarity of the accent, rather than its degree of difference or its comprehensibility. This would therefore predict that we would find little difference in the repetition-suppression prompted by regional and foreign accents, as long as both types of accent are similarly unfamiliar to the listeners. Both types of accent, however, should differ from familiar speech throughout the time-course of word processing.

In contrast, the Abstract entry approach would make a different set of predictions. Under this approach, accented target words should be normalised pre-lexically, and we would therefore expect to find repetition-suppression at or even before the lexical stage of processing. The ease with which words are normalised would be dependent either on the degree of difference from familiar or canonical speech (according to the Perceptual Distance hypothesis), or on whether the accent is a regional or a foreign one (according to the Different Processes hypothesis) and therefore we would expect to find more and earlier suppression in response to weak or regional accents than to strong or non-native accents. In other words, either the comprehensibility or the type of an accent will determine the ease with which it is normalised to allow access to the abstract or canonical word-form. Since these two approaches make different predictions about the way in which accented speech primes familiar speech, a repetition-suppression study which pits foreign and regional accented speech against each other, and which looks specifically at the effect on a repetition in familiar speech, should help to provide evidence to determine which of the approaches is more successful in explaining the representation of variant forms, and specifically the way in which accented speech is represented.

Experiment 2 will therefore use a repetition-suppression design, presenting participants with pairs of sentences which differ in the accent used. Rather than looking directly at the processing of accented speech, as in experiment 1, experiment 2 will look at the processing of sentences in familiar speech (in the participants' home accent) when preceded by identical sentences presented in familiar, regionally-accented, or foreign-accented speech. By comparing the first and second items in each pair, it should be possible to gain a greater understanding of the activation elicited by the different types of accent. Once again, the PMN, representing pre-lexical normalisation processes, and the N400, representing the post-lexical phase, will be examined. The predictions made about the findings come from the Abstract Entry approach and the Exemplar-based approach to lexical storage and access. According to the Exemplar-based approach, accented variants of words are represented as separate entries in the lexicon, which means that the accented sentences should result in little or no suppression in the familiar-speech forms in the pre-lexical phase. However, the Abstract entry approach indicates that any variant form, whether resulting from minor intra-speaker variation or more substantial inter-speaker differences, will be normalised to access the same abstract lexical entry. Therefore, this approach would predict that priming with the accented forms should give rise to suppression in the familiar speech when compared with unprimed speech, and that the degree of suppression may be dependent on the ease with which those variant forms are normalised, or the type of accent in which they are spoken. Under this approach, it would be predicted, then, that regional accents would elicit more suppression than foreign accents, in the PMN. Both approaches would predict that in the post-lexical N400 epoch, the familiar and accented primes would cause reduced activity in the target words. The predictions are also influenced by the Perceptual Distance and Different Processes hypotheses. According to the Perceptual Distance hypothesis, differences between the suppression elicited by the regional and foreign accents should be in direct relation to their perceived distance from the listener’s ambient accent. However, the Different processes hypothesis would predict different patterns of activation, and thus of suppression, for the regional and foreign accents. Under this hypothesis, regional accents are expected to elicit suppression in the familiar speech in both the PMN and the N400 epochs, whereas foreign accents are expected to elicit less or no suppression, or indeed to elicit increased activity in both epochs.

Chapter 3: an Adult ERP study

Introduction

As outlined in the introductory chapters of this thesis, accents are a commonly-encountered source of variation in the speech signal, which have an impact on the speed and accuracy with which speech is processed (Adank, et al., 2009; Adank & McQueen, 2007; Bradlow & Bent, 2003; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Clopper & Bradlow, 2008; Floccia, et al., 2009b; Floccia, et al., 2006; Maye, et al., 2008). It has been suggested (Clarke & Garrett, 2004) that accents can be arrayed along a continuum of perceptual difference from the accent of the listener, with non-native accents generally falling further along the continuum than native accents, and thus being harder to process accurately. According to this hypothesis, regional and foreign accents should recruit the same neural mechanisms for normalization, with both foreign and regional accents deviating from a baseline home accent in the same direction, but with greater deviations from the baseline in the foreign than the regional accent. Support for this hypothesis is found in Floccia et al.'s (2006) study, in which reaction times in a go/no-go lexical decision task to regional accents were around 30 ms slower than to a familiar accent, with responses to foreign accents being slower still (around 100 ms).

However, there is also evidence which suggests that these differences may be due to the recruitment of qualitatively different processes for regional and foreign accents (Adank, et al., 2009; Chambers, 2002; Floccia, et al., 2009b; Floccia, et al., 2006; Girard, et al., 2008). Regional accents, that is, the accents of first-language speakers in their native tongue, consist of a set of deviations in the phonetic, phonological, phonotactic and prosodic information within the speech signal (Wells, 1982), but these deviations generally remain within the unwritten rules of the language. For example, a speaker of Standard British English will produce "bath" with a medial /A:/, whereas a speaker of Northern English will produce it with a medial /{/ (Hughes & Trudgill, 1996). Both these phonemes are found in both varieties of English, albeit with differing lexical distributions, so for a native English speaker, adapting from one accent to the other is a fairly straightforward task. Foreign accents, however, arise from an interaction of the speaker's native language and the language they are speaking, so that normalisation requires processes other than the recalculation of lexical distributions; rather, a foreign accent may introduce sounds not normally heard in the language being spoken, and may also introduce changes in stress or prosody in ways which are not native to the listener. For example, a native French speaker speaking in English may alter familiar stress patterns in a way which forces the listener to recruit more complex normalisation processes and to rely more on top-down information in order to comprehend the speaker's intent (Dupoux, Pallier, Sebastian, & Mehler, 1997).

To date there is relatively little research comparing the processing of regional and foreign accents, as most previous studies compare one or the other against a baseline (the listener's own accent), so there is limited evidence to support either the Perceptual Distance hypothesis or the Different Processes hypothesis. However, the developmental literature does provide some useful evidence. Studies with five-year-old French- (Girard, et al., 2008) and seven-year-old English- (Floccia, et al., 2009b) speaking children show that while they are able to spot and classify foreign accents with relative ease, they have great difficulty spotting or classifying regional accents, even when the regional and foreign accents have been rated as being of equal strength (Floccia et al., 2009). This would seem to indicate that the typical features of foreign accents are more salient than those of regional accents, at least at this age. There is also evidence from sociological and observational studies that children moving to a different region will generally pick up the local accent very quickly (Chambers, 1992) whereas even with a great deal of exposure to their native tongue spoken by non-native speakers, children do not pick up non-native accents (Chambers, 2002).

Another area of research which gives us some insight into possible differences between the processing of foreign and of regional accents is that of adaptation to accents. Here, however, the evidence is not clear-cut. Maye and colleagues (Maye, et al., 2008) exposed participants to 20 minutes of an artificial accent in which front vowels were lowered (so that, for example, “witch” became “wetch”) and then gave them a word/non-word categorisation task which included items resembling real words with lowered vowels. They found that the exposure to the artificial accent reduced error rates (a measure of intelligibility) but did not reduce response times (a measure of comprehensibility). Adank and McQueen (2007) found a delay in response times to an unfamiliar regional accent versus a familiar one, and did not show any signs of adaptation after short-term exposure to the accents, while Clarke and Garrett (2004) found full adaptation to a foreign accent within only a few sentences, when using a cross-modal matching task (although for a discussion of problems with this interpretation, see Floccia, et al., 2009b). A study comparing foreign and regional accents with added background noise (Evans & Taylor, 2010) found more adaptation to foreign accents than to regional, but also showed that performance for foreign accents was worse overall than for regional accents, so the degree of adaptation may simply have been a function of the amount of room for improvement. Partial adaptation to both foreign and regional accents has been found using a repetition task (Pinet, et al., 2011), which measures intelligibility. So far, then, there is evidence to support both the Perceptual Distance Hypothesis and the Different Processes Hypothesis.

An Exemplar-based model does not offer a clear route to differentiating between these two hypotheses, since it states that discrete lexical entries are stored for variant forms, dependent on the frequency with which they are encountered rather than on their distance form a canonical form or the ease with which they can be processed. Thus, no normalisation is required, rendering processing differences in the pre-lexical stages of word recognition across different accent types harder to explain. Here, then, we will adopt the Abstract entry model in attempt to lend further support to one or other hypothesis. According to the Abstract entry approach to modelling word representation and recognition (Marslen-Wilson, 1987; McClelland & Elman, 1986; Pallier, et al., 2001), it can be assumed that in the pre-lexical stages of word processing, variant pronunciations are being normalised to fit the prototypical lexical entry, and that this process involves low-level phonological mechanisms (Sumner, 2011). There is also evidence that these processes may be guided by top-down information (Davis, et al., 2005; Newman & Connolly, 2009; Norris, et al., 2003), with lexical representations feeding down to inform pre-lexical normalisation. According to the Perceptual Distance hypothesis, these processes should be occurring in a similar way for both foreign and regional accents, but to a lesser degree for regional compared to foreign accents. On the other hand, the Different Processes hypothesis would predict that regional and foreign accents would recruit different types of normalisation processes, perhaps at different neural loci. We might expect to find that regional accents can be normalised quickly and easily at the phonological level, with little need for top-down or lexical information, whereas because foreign accents deviate in ways which are not specific to the listener's native language, phonological normalisation alone may not be sufficient to achieve comprehension, and there may be more reliance on lexical (top-down) information. This distinction, then, offers a way of providing more definitive support for one hypothesis over the other. However, behavioural measures such as accuracy scores or response times cannot give us the level of resolution required to interrogate different stages in the process or normalisation. Instead, we must look to ERP methodologies, which allow us to make a direct comparison between familiar speech, regional accents, and foreign accents. It is thereby possible to look at the activity recruited by each type of accent, both pre-lexically and post-lexically, rather than simply at the success or otherwise with which participants have reached the end-point of recognition or categorisation or a word.

Berman, Mandelkern, Phan and Zaidel (2003) used EEG and PET measures while listeners heard words in Australian, British, French, and Hindi accents, and found activity in the left hemisphere associated with word detection and in the right hemisphere associated with accent detection, but did not compare the accents against each other, nor did they identify specific components associated with the tasks. Conrey, Potts and Niedzielski (2005) looked at ERP responses to a phonemic contrast, /i/ vs. /e/, which is merged in some dialects and not in others, in a cross-modal matching task. Surprisingly, they found no effect in the MMN (Mismatch Negativity, a component associated with oddball or deviant stimuli) or the N400 (associated with words or word-like stimuli), both components where we might expect to see an effect of congruity. However, they did find an effect of congruity in the Late Positive Component (or LPC), from 400-750 ms after onset, indicating that participants from the merged dialect group had an attenuated response to the merged contrast when compared with those use used the unmerged dialect. In other words, where the /i/ vs. /e/ distinction was still contrastive, ERP responses reflected the distinction, whereas for participants who did not distinguish the two sounds behavioural or in normal speech (for whom the contrast represents an unfamiliar accent) the distinction was reduced at the neural level too. A similar studied carried out with French-speaking participants (Brunelliere, et al., 2009) and using a French vowel contrast (/ɛ/ vs. /e/), which is merged in Southern French but not in Standard (Northern) French, elicited differences in the MMN in Standard French-speaking listeners, while a control contrast (/ø/ vs. /y/), which is contrastive in all French accents, elicited differences in the P200, the MMN, and a later component, between 372 and 486 ms after onset. The authors suggested that exposure to merging dialects may be responsible for attenuating the difference in the merging contrast even in standard French speakers. A follow-up study (Brunelliere, et al., 2011) found that the /ɛ/ vs. /e/ contrast elicited different cortical topographies in the Standard French group and the Southern French group for the MMN component.

These studies indicate the potential usefulness of ERP approaches in examining differences in neural responses to accented speech, but so far, the results are not consistent. Some point towards asymmetrical responses (Berman, et al., 2003), and others highlight components from the early stages of acoustic processing in the MMN (Brunelliere, et al., 2009) through to much later phonological or even semantic processing (Conrey, et al., 2005). The studies discussed also use different behavioural tasks, so that differences found may be confounded by task effects.

The Current Study.

This study uses ERP to make a direct comparison of responses to three types of accent; familiar speech (that is, speech in an accent which closely matches the participants' own accent, and which can therefore be considered un-accented), regional accents (that is, accents of native English speakers from different geographical regions) and foreign accents (that is, the accents in English of speakers whose native tongue is something other than English).

Previous studies have used a range of different experimental tasks and measures, which means that it is sometimes difficult to separate out responses to the stimuli themselves from responses to the task demands of the experiment. This study will tackle this problem by limiting the task demands involved; whole sentences spoken in the three different accent types will be used, and participants will be given a simple go/no-go task, requiring a response to only a small proportion of trials (which will not then be included in the analysis), in order to ensure that they attend to the stimuli throughout the duration of the experiment. In this way, the ERP responses to the stimuli can be examined without any contamination from more complex task effects. ERPs will be time-locked to the final word of each sentence, allowing a direct comparison between responses to the baseline Home accent, the two regional accents, and the two foreign accents. While some studies have used single words or phonemes, and have controlled the accent-specific content of the stimuli, this study takes a more holistic approach to accents, using whole sentences, which have not been tightly controlled in terms of their phonology, phonotactics, or supra-segmental variation (See Floccia, et al., 2006 for a similar approach). This ensures that participants are exposed to both short-term accent-related deviations, such as vowel shifts or differences in voicing, and longer-term changes such as differences in stress timing and prosody in a way which reflects natural speech.

By looking in detail at the time-course of word processing for the three different types, it is hoped to shed light on the accuracy of the Perceptual Distance and Different Processes hypotheses. To do this, analysis will focus on two well-established ERP components which are likely to be implicated in the normalisation of accented speech; the Phonological Mismatch Negativity or PMN (Newman, Connolly, & Mcivor, 2003) and the N400 (Connolly, Phillips, Stewart, & Brake, 1992; Kutas & Hillyard, 1984).

The PMN, which typically peaks around 250-300 ms after stimulus onset, is thought to index pre-lexical phonological processing, and to be involved in lexical selection when the phonological competitors are distinguished by their acoustic/phonetic content and the expected input given their context (Connolly & Phillips, 1994; Desroches, Newman, & Joanisse, 2009; Newman & Connolly, 2009; Newman, et al., 2003). It is sensitive only to auditory stimuli, a fact which distinguishes it from the N400, which responds to both auditory and visual stimuli (Connolly & Phillips, 1994; Desroches, et al., 2009; Newman & Connolly, 2009), and seems to be involved in determining goodness-of-fit of stimuli by reconciling the acoustic/phonetic input with the expected phonetic input (Newman & Connolly, 2009). This role in normalising speech input makes the PMN a likely component for the processing of accented speech. This is because our phonetic expectations are determined by our own lexical representations of words, which will reflect our own accent, or at least the accents to which we are most commonly exposed. While we may be able to adapt to an unfamiliar accent, it seems likely that this adaptation takes the form of quicker and more accurate normalisation of the incoming speech signal, rather than a global shift in phonetic expectations (which would interfere with the processing of familiar speech). We would therefore expect to find a mismatch between the incoming signal of accented speech, and our phonetic expectations, and so we would expect accented speech to elicit a PMN. This mismatch between our expectations and the input received may help to differentiate between the Perceptual Distance hypothesis and the Different Processes hypothesis. According to the Perceptual Distance hypothesis, the deviations from the baseline Home accent in the Regional and Foreign accents should increase the amplitude of the PMN, with the foreign accents causing greater negative-going deviations than the regional accents. The Different Processes hypothesis, on the other hand, would predict that the regional accents will recruit more pre-lexical normalisation than the foreign accents (with the foreign accents recruiting more post-lexical activity). This would result in an increase in the amplitude of the PMN in response to the regional accents. However, assuming that this normalisation process allows the removal of accent-related “noise” in the signal (Luce, McLennan, & Charles-Luce, 2003) and that the nature of foreign accents leaves this noise intact, this would lead to a decrease in the PMN in response to the foreign accents, as is seen in response to low-intelligibility speech in noisy conditions (Martin, Kurtzberg, & Stapells, 1999).

The N400 is a negative-going component which typically peaks around 400 ms after stimulus onset, in response to semantic or lexical mismatches or incongruities (Desroches, et al., 2009; Kutas & Hillyard, 1984). In the current study, if normalisation takes place pre-lexically, there should be no differences between the accent conditions in the N400 response (Sumner, 2011). However, if the pre-lexical normalisation is incomplete, this should be reflected by deviation in the N400 in response to the accented speech. The Perceptual Distance hypothesis would predict incomplete pre-lexical normalisation for both foreign and regional accents, relative to their perceived distance from the home accent, which would be reflected in deviations in the N400 relative to the Home accent. The Different Processes hypothesis would predict that pre-lexical normalisation of the Regional accents should be successful, resulting in an absence of deviation in the N400 between the Home and Regional accents, whereas in the case of the Foreign accent, a failure to normalise the accent-related features of the speech should cause a significant deviation in this component from both the Home and Regional accents. The direction of the deviations in the N400 is somewhat harder to predict. If N400 increases as a function of the difficulty of the task, as is the case with some components (Goodin, Squires, & Starr, 1983; Philiastides, Ratcliff, & Sajda, 2006) then the Perceptual Distance hypothesis predicts that N400 should be more negative for regional than for home accents, and more negative still for foreign accents. However, there is evidence to suggest that in the N400, while incongruity increases the negative-going amplitude of the component, poor intelligibility due to noise or distortion reduces it (Aydelott, et al., 2006; Boulenger, et al., 2011). Boulenger et al. (2011) presented participants with sentences with either a high- or low-cloze probability final word, containing varying durations of temporally reversed signal. As expected, the low-cloze probability words elicited a more negative N400 than the high-cloze probability words, but the signal distortion was found to interact with cloze probability such that the amplitude of the negative-going N400 was more positive for low-cloze probability words in temporally reversed words than intact words, but in the high-cloze probability words, the N400 was more negative than for intact words. Similarly, Aydelott et al (2006) presented participants with sentences ending with a congruent or incongruent word, and in which the context was either intact or acoustically filtered, while the target word was always intact. Incongruent words elicited a larger N400 deviation than congruent words, but this effect was greatly reduced by filtering, despite the fact that participants were very accurate in deciding whether the sentences “made sense” or not. This being the case, if accent-related variance from the familiar baseline is treated as noise, analogous to temporal reversal or acoustic filtering, the N400 component should be attenuated by unfamiliar accents. According to the Perceptual Distance hypothesis, this should result in a smaller N400 response in the Regional accents than in the baseline Home accent, and a smaller still response to the Foreign accents. On the other hand, the Different Processes hypothesis would predict a reduced N400 in the Foreign accent condition, due to a failure to normalise the accent-related variance, but not in the Regional accent condition, where the variance will have been normalised pre-lexically.

In summary, the Perceptual Distance hypothesis would predict an increase in the negative-going PMN component for both the Regional and Foreign accents, in direct relation to their perceived distance from the Home accent, and either an increase in both Regional and Foreign conditions, or a decrease in both, in the N400. The Different Processes hypothesis would predict an increase in the PMN in the Regional condition and a decrease in the Foreign condition, followed by either an increase or a decrease in the N400 in the Foreign condition, but no difference between the Home and Regional conditions.

Method and Design

Participants

Sixty participants, all originating from the South-West of England and thus accustomed to South-West accents, were recruited from the student population at the University of Plymouth. Of these, 15 were male and 45 female. Their ages ranged from 18 to 40 (M = 20.2). Participants were offered course credits for participation. The data from five participants were discarded due to hardware or software problems (2) or poor-quality data (3 participants had more than 30% contaminated segments in a single condition, and were excluded).

Materials

A total of 12 male speakers recorded 62 sentences each. Four speakers, aged 21, 21, 23, and 26, had South-Western (Plymothian or Cornish; from now on, SW) accents, two speakers, aged 20 and 21, had Welsh accents ,two speakers, aged 21 and 26, had Yorkshire accents (specifically, both were from Leeds), two, aged 34 and 35, were native Italian-speakers from the North of Italy, and two, aged 22 and 23, were native Polish-speakers from the Central region of Poland. Participants heard either SW, Yorkshire and Polish accents (group 1; N = 26), or SW, Welsh and Italian accents (group 2; N = 34). Sentences had no embedded clause. Each sentence ended with a two-syllable trochaic noun, some of which were animals and some were not. All final words had a frequency of between 0 and 33 per million according to the COBUILD corpus, and a phonetic neighbourhood of between 0 and 27.

An example of a sentence ending with a non-animal is “There on the wall she could clearly see a buttress”, and one ending with an animal is “With all that dark make up she looked like a panda”. The speakers were asked to read the sentences in a natural way, as close to normal conversational speech as possible. Sentences were recorded and edited using Adobe Audition software (Chavez, Day, Deyell, Ellis, Fazio, & Green, 2003), in a quiet room, at 44 KHz with 16 bits per sample. Sound files were normalised to the same mean amplitude level. The sound files were trimmed to remove any sounds preceding or following the sentences.

The sentences were rated for identifiability and strength of accent. 20 participants, all of whom grew up in the South-West of England, each heard 76 of the sentences, counterbalanced so that each sentence was rated by two participants, and each participant heard sentences from each of the twelve speakers. The sentences were presented by the E-Prime Professional v2.0 software (Schneider, Eschman, & Zuccolotto, 2002), over headphones, in randomised order. After each sentence, participants were asked, via text on the screen, to indicate which accent they thought the sentence had been spoken in. They were then asked, via a second screen of text, to indicate how confident they were of their identification of the accent, by choosing a number between 0 (“Not at all confident”) and 5 (“Very confident”). Finally, they were asked how strong they felt the accent to be, using Received Pronunciation or Standard British English as a baseline, and selecting a number between 0 (“Not at all strong”) and 5 (“Very strong”). Analysis of the results showed that participants correctly identified the accent in 64% of cases (n = 971). Unsurprisingly, participants were better at identifying the South-West accent (correctly identifying it 82.87% of the time), and also the Welsh accent (75% correct), than the Polish (58.8% correct), Italian (49.2% correct), and Yorkshire accents (34.52%). For the latter, participants most often misidentified it as a South-West accent (see matrix of confusion in Table 2).

A t-test showed that participants' confidence was higher for correct identifications (m = 3.68) than for incorrect identifications (m = 2.94; p < .001), and correctly identified sentences were perceived as having a stronger accent (m = 3.20) than incorrectly identified sentences (m = 2.79; t = -9.324, p <.001). When accents were combined into three categories, "Home" (South-West), "Regional" (Welsh and Yorkshire) and "Foreign" (Italian and Polish), participants were significantly more likely to correctly identify Home accent than Regional or Foreign accents (p < .001), but were equally likely to correctly identify the Regional (m = .55) and Foreign (m = .54, p = .747) accents.

The South-West accented sentences were perceived as having the weakest of the five accents; this difference was significant in the case of the Polish (mean difference = .82, p < .001), Italian (mean difference = .82, p < .001) and Welsh accents (mean difference = .82, p < .001) but not in the case of the Yorkshire accent (mean difference = .20, p = .93). The Yorkshire accent was perceived as less strong than the Polish (mean difference = .62, P < .001), the Italian (mean difference = .62, p < .001) and the Welsh accent <mean difference = .62, p < .001).

Table 1: Confusion matrix of accent identification for experimental stimuli; number of responses (proportion of responses)

Accent	Identification					
	South-West	Polish	Yorkshire	Italian	Welsh	Don't Know
South-West	**421 (.83)**	2 (.00)	35 (.07)	1 (.00)	14 (.03)	35 (.07)
Polish	2 (.01)	**147 (.59)**	3 (.01)	92 (.37)	3 (.01)	3 (.01)
Yorkshire	135 (.54)	5 (.02)	**87 (.35)**	1 (.00)	9 (.04)	15 (.06)
Italian	3 (.01)	108 (.43)	9 (.04)	**124 (.49)**	3 (.01)	5 (.02)
Welsh	22 (.09)	9 (.04)	27 (.11)	4 (.02)	**192 (.75)**	2 (.00)

Table 2: Characteristics of stimuli across accents (SD in brackets)

Variable	Home	Yorkshire	Welsh	Italian	Polish
Sentence Duration (MS)	1846.72 (526.70)	1965.51 (542.02)	1838.48 (489.36)	2153.80 (496.77)	1944.09 (553.37)
Word Duration (MS)	526.70 (88.16)	542.02 (135.74)	489.36 (76.96)	496.77 (107.03)	553.37 (82.06)
Phonological Uniqueness	5.05 (1.10)	4.82 (0.87)	4.91 (1.00)	4.82 (0.87)	4.91 (1.00)
Phonological Neighbourhood	3.05 (4.23)	3.29 (4.64)	3.80 (5.18)	3.29 (4.64)	3.80 (5.18)
Number of Morphemes	1.43 (0.5)	1.45 (0.5)	1.45 (0.5)	1.45 (0.5)	1.45 (0.5)
Number of Phonemes	5.77 (1.02)	5.65 (1.05)	5.58 (1.02)	5.65 (1.05)	5.58 (1.02)
Lexical Frequency (Subtlex)	3.08 (1.31)	3.57 (2.62)	3.80 (2.97)	3.57 (2.62)	3.80 (2.97)
Phonological levenshtein distance (PLD20)	1.63 (3.47)	1.30 (1.73)	1.81 (3.43)	1.30 (1.73)	1.81 (3.32)

When Yorkshire and Welsh accents are combined into a single “Regional” category and Italian and Polish accents are combined into a single “Foreign” category, no significant difference in sentence duration was found between the South-West (“Home”) accent and the Regional group, but the Foreign group had significantly longer sentences than both the Home (mean difference = 191 ms, $p < .001$) and the Regional (mean difference = 151 ms, $p < .001$) accents.

Durations of the final word of each sentence were also compared. South-West (m = 527 ms), Polish (m = 553 ms) and Yorkshire (m = 542 ms) accents did not differ from each other, but the final words of Italian (m = 497 ms) and Welsh (m = 489 ms) sentences were significantly shorter, as shown in Table 2.

Table 3: differences in the duration of the sentence-final words, by accent.

	SW	Polish	Yorkshire	Italian	Welsh
SW		NS	NS	NS	mean diff = 37, p = .005
Polish			NS	mean diff = 57, p < .001	mean diff = 64, p < .001
Yorkshire				mean diff = 45, p < .001	mean diff = 53, p < .001
Italian					NS
Welsh					

There were no significant differences in word duration between Home, Regional and Foreign accent groups.

In summary, the Yorkshire accent was rated the hardest to identify, due to both speakers having a relatively weak accent, but participants did not have trouble identifying it as a regional, rather than a foreign, accent. Despite having shorter than average final words, the Italian accent had the longest sentence duration, due to the tendency in Italian speakers to add a strong vocalic offset to words terminating in a consonant.

Procedure

Sentences were delivered over headphones, and presentation was controlled by the E-Prime software version 1.1 (Psychology Software Tools, 1996). A ten-trial training block preceded the main experiment, to ensure that participants had understood the instructions. The main experiment was divided into two blocks of 185 sentences (370 in total), separated by a rest period. Trials were separated by an inter-stimulus interval of 600-800 seconds. Between trials participants were prompted to blink, in order to reduce the likelihood of blinks mid-trial. Out of these 370 sentences, 40 were sentences ending with an animal name, for which participants were required to press a button in an animal detection go/no go task. This task was purely intended to maintain attention to the task; the data from Go trials were discarded, leaving only data from trials in which no overt behavioural response was given. The remaining 330 sentences were divided in an equal sample of 110 sentences per accent condition. Therefore, each participant heard 123 sentences in each of three accents (110 “no go” trials plus 13 “go” trials which were not analysed), spoken by one of two speakers of each accent; thus they heard 61 or 62 sentences spoken by each of six speakers, in three accents, with 55 sentences from each speaker being used in the final analysis. These were delivered across two blocks, each containing 185 sentences. The order of the sentences within each block of 185 sentences was randomised. For a given participant, no sentence was ever repeated. Sentences used in Experiment 1 were also used in Experiment 2, spoken by a different speaker. Sentences spoken by a SW speaker in experiment 1 were spoken by a different SW speaker in Experiment 2. Sentences spoken by a Yorkshire speaker in Experiment 1 were spoken by an Italian speaker in Experiment 2, and sentences spoken by a Polish speaker in Experiment 1 were spoken by a Welsh speaker in Experiment 2, so across the complete study, no sentences were always spoken in a foreign accent, or always in a regional (non-SW) accent.

Accents

The different characteristics of L1 languages result in speakers from different linguistic backgrounds having different, and often identifiable, accents in a given L2. Here, a brief summary is presented of some of the most salient characteristics of the two non-native accents used in this study, but note that the following descriptions are far from exhaustive. Inter- and intra-speaker differences are not accounted for here, and foreign-accented speech generally includes many subtle but noticeable characteristics which are almost impossible to describe in writing.

Italian

Italian has only seven vowels (Adler, 1967), in contrast to the much richer vocalic system in standard British English (SBE), which includes seven short vowels, five long vowels, and eight diphthongs. L1 Italian speakers, therefore, will tend to differentiate less between certain vowels when speaking English; for example, *lip* and *leap* will be pronounced in the same way, with a medial vowel longer than that used in *lip* in SBE. In the corpus of sentences used in this study, one sentence ended with the word *bailiff*; as spoken by an Italian speaker the second syllable was pronounced with a longer vowel that in SBE or Plymothian English, so that the word could easily have been mistaken for *bay leaf*. Similarly, no distinction is made in Italian between the short /ʊ/ and the long /uː/ (exemplified in this corpus by a long /uː/ in the word *cooker* where speakers of SBE would use a short /ʊ/), while there may also be confusion between /æ/ and /e/ (in this corpus, the word *back-up* spoken by an Italian speaker, was pronounced with a /e/ rather than a /æ/ in the first syllable), and between /ʌ/ and /æ/ (so that the first vowel sound in the words *tantrum* and *puppet* in this corpus is almost identical). The schwa vowel does not exist in Italian, so words which typically use it in English are often produced with clearly articulated vowels, which can interfere with appropriate stress patterns in SBE. Thus where the second syllable in the word *dragon* in SBE is a very short schwa, in this corpus it is much more clearly enunciated by the Italian speaker, so that the second syllable carries a similar weight to the first. In English words with a medial or terminal /r/, the /r/ sound is generally not articulated in SBE (although many regional British accents do rhoticise it, especially in the South West where this study is based). In Italian accents, the /r/ sound is frequently either rhotic or trilled. This is particularly noticeable in these stimuli on words such as *clover* and *bomber*, which in SBE would end in a schwa with no articulated /r/, but which, in the Italian accent, have a pronounced trilled /r/.

Italian does not contain the two dental fricatives, / ð/ and /Ө/. /Ө/ is therefore often replaced with either /f/ or /t/, while / ð/ is generally rendered as /d/. This is particularly common in this corpus in short, common words such as *the* and *that*, where enunciation seems to be less careful and deliberate, and particularly in cases where co-articulation favours a non-canonical pronunciation, such as in the phrase "*The black and white stripes always marked the badger*", in which the / ð/ follows a /d/, and is thus co-articulated with it, rather than enunciated separately.

In some cases, Italian L1 speakers do not produce aspiration on voiceless plosive consonants, thus failing to distinguish them clearly from their voiced counterparts; thus /t/ can sound like /d/. This study's corpus for example, included the sentence "*You should take care when teeing off as you may hit a caddie*". The /t/ at the beginning of *take* has very little aspiration, and thus approximates a /d/ sound.

In SBE, stresses occur at approximately regular intervals, with some unstressed syllables being reduced or elided. Italian, on the other hand is a syllable-timed language and this often manifests in Italian-accented English; prosodic patterns carry over from Italian, resulting in more evenly-stressed syllables, with less reduction and elision than in SBE. In this corpus, the stimuli were read out, rather than spontaneously produced, so the enunciation is clearer and more deliberate than it might be with spontaneous speech.

In Italian most words (with the exception of some borrowings from other languages) end with a vowel. This results in a tendency to produce a vocalic offset to English words with a terminal consonant. In marked cases this can sound like an extra syllable, and gives rise to a characteristic "bouncing" rhythm in fluent speech. These rhythmic and phonotactic differences largely account for the finding that in this corpus, the Italian-accented sentences are longer than in any other accent.

Polish

The Polish accent in English is similarly determined by characteristics of the Polish language which are transferred across to the L2. Like Italian, Polish does not make use of the interdental fricatives, / ð/ and /Ɵ/, and tends to render them as /v/ or /d/ and /f/, /t/ or /s/ respectively (Gonet & Piétron, 2004). Thus, in this corpus, *thought* is pronounced with a /f/ sound rather than a /Ɵ/, and, as with the Italian-accented stimuli, words like *the*, *that*, and *they* are pronounced with a sound approximating /d/ rather than / ð/.

Voiced terminal plosives are often insufficiently voiced in the Polish accent, resulting in a confusion between a terminal /b/ and /p/, so that, for example, in this corpus, in the phrase "*When reaching up, the climber grabbed the piton"*, the terminal consonant of *up* is almost indistinguishable from the medial consonant in *grabbed*.

Polish contains eight vowels, and so the richer vocalic system of English can cause difficulties of distinction in both hearing and production, resulting in errors in vowel quality. Vowels may be incorrectly rhotacised, and schwa vowels may be over-articulated. There is also a notable confusion between /ɪ/ and /iː/, as in Italian, so that for example the word *sick* in this corpus was pronounced, with a Polish accent, so that it sounded very like *seek*.

Palato-alveolar consonants (/ʒ/, /ʃ/, / dʒ / and /tʃ/) are not used in Polish and may cause difficulties for some speakers, usually being articulated further forward, as alveolar or even interdental consonants. For example, the word *shadow* in pronounced in this corpus with a /tʃ/ instead of /ʃ/, while *cartridge* is pronounced with a /tʃ/ instead of / dʒ /. The English palato-alveolar /r/ may be rhotacised or trilled. In this corpus, the word *garlic* is pronounced with a strongly rhotic medial /r/.

Polish-accented prosody will also differ from SBE prosody. Typically, stress in Polish is on the penultimate syllable, and reduction or elision is not common. In SBE, stress is typically on the penultimate or antepenultimate syllable (Cutler & Carter, 1987), but because English is stress-timed rather than syllable stress-timed, unstressed syllables may be shortened or elided, so that the stressed syllables are perceived as occurring at roughly regular intervals, although as Peter Roach (1982) points out, this so-called isochrony may be “more apparent than real” (p 74). Polish-accented English will therefore be irregularly stressed, with stress sometimes occurring on the wrong syllable, and with over-articulation of syllables which would normally be reduced or elided.

Leeds, Yorkshire, UK

The Leeds accent shares many features with other Northern English dialects, including a tendency towards dropping the initial /h/ in unstressed words, replacing medial and terminal /t/ with a glottal stop, and a shortening of some vowels (British Library, 2009), such that in this sample, *casket*, pronounced /kɑːskɪt/ in Standard British English, is pronounced as /kæskɪt / in the Leeds accent. /aɪ/ may be pronounced as /æ/, and /Λ/ is often pronounced as /ʊ/. Distinctive features of the Leeds accent specifically include a reduction in the shaping of /əʊ/ such that it is pronounced like /Eː/; thus in the Leeds accent, the word *no* may sound similar to the SBE pronunciation of *ner.* However, this was not obvious in the sample used here; both Yorkshire speakers produced /əʊ/ rather than /Eː/ in words such as *coma* and *old.* /ð/ is often pronounced as /v/, particularly in the middle of words such as *without*.

In this study’s sample, the two speakers from Leeds both enunciated carefully, resulting, in some cases, in exaggerated initial /h/ sounds, rather than dropped /h/ sounds. For example, in the sentence *“He could hardly hear the carol”* the /h/ at the beginning of both *hardly* and *hear* is over-articulated. Thus the typical shortened vowels in words like *gusto* (pronounced /gʊstəʊ/ rather than /gΛstəʊ/) are the main identifiable traits of the Yorkshire accent in this sample.

South Wales, UK

For the South Wales area a description of the intonation system of this dialect is provided by Walters (2001) who analysed samples produced in the Rhondda Valley, an area of South-East Wales. The Welsh dialect of English has borrowed many prosodic features from the Welsh language, which resulted in a shortening of stressed vowels and lengthening of succeeding consonants, a pitch-rise from the stressed syllable and an increase in phonetic strength of the post-tonic syllables, and finally a shift of word stress from initial to penultimate or ultimate syllable in polysyllabic words. Intonational phrases are of two main kinds: a sequence of rising contours that can end with an ultimately rising nuclear contour, or with an ultimately falling contour. All these features contribute to the popular feeling that Welsh English is a "sing-song" dialect (Wells, 1982, p. 392). At the segmental level, according to Hughes and Trudgill (1988) Welsh English is characterised by its non-rhoticity (no post-vocalic "r"), as demonstrated in this corpus by the lack of rhoticity or trill in words such as *doorbell* and *blackbird*, the distribution of /{/ and /A:/ which follows that found in the North of England (in this study's corpus, the vowel sound in the word *class* and the first vowel sound in *bathtub* actually fall somewhere between /{/ and /A:/, being pronounced as notably shorter than the /A:/ which would be used in SBE, but longer than the short, typically Northern English /{/), and the vowel /E:/ in "bird" being rounded to approach /2:/; in this corpus the word *girdle* provides a typical example of this. In addition, the phoneme /λ/ is never dark, that is, it is not velarised after a vowel as in English Received Pronunciation. In this corpus, the word *clear* in the sentence *"It was clear that she couldn't stand the taste of curry"* would normally have a velar /l/ in Standard British English, but the Welsh-accented speaker articulates the /l/ sound further forward.

South West, UK

The South West dialect of English belongs to the family of Southern English dialects (Wells, 1982), and thus has intonation patterns that do not depart significantly from that of the Received Pronunciation English. Bolinger (1989) notices in RP English a high proportion of high initial pitches, leading to more frequent and more extended falls than in Network Standard American English (p. 29). There is also a higher proportion of terminal rises in BE than in AmE. However, in the South West short vowels tend to be longer than in other South of England accents, especially in monosyllabic words in phrase-final or prominent position (Wells, 1982, p. 345), resulting in the popular feeling that the South West dialect is slow, although within this corpus, sentences spoken in the South West accent had a shorter duration that those spoken in Yorkshire, Polish or Italian accents. At the segmental level, it is distinct from RP English in its rhoticity, the loss of the /æ/ and /a:/ distinction (Hughes &Trudgill, 1988), and by the fact that words like boat and gate have usually retained their monophthong pronunciation (Wells, 1982). This is illustrated in this corpus by the pronunciation of the name *Kate* in the sentence *"Kate was nice and warm snuggled next to her bedspread"*.

Recording System

ERP data were recorded using a 30 channel Ag/AgCl ActiCap system (actiCap, Brain Products GmbH), mounted on an elasticated cap (see figure 1) with Vision Recorder software. Two further electrodes below and lateral to the right eye were used to monitor eye movements and blinks, but were not included in the final analysis. EEG was referenced using the left mastoid electrode and re-referenced offline to the averaged activity from left and right mastoid electrodes, with the AFz electrode providing the ground. Processing of the ERP data was carried out using the Vision Analyser software. EEG epochs, beginning at 100ms before onset of target words and ending 800ms after onset, were averaged for each accent and for each participant. A bandpass filter (0.1 – 40Hz) was applied, and data were corrected by a baseline of 100 ms before target onset. A dc-detrend function which corrected relative to the first and last 100 ms was used to remove voltage drift artifacts in the data. Artifacts, including blinks and muscular artifacts, were removed, resulting in 4.7% of the no-go trials being discarded, distributed evenly across the three accent conditions ($F<1$). “Go” trials and those where the participant had erroneously responded were also discarded, so only trials on which no response was required or given were included in the analysis. Separate ERPs were calculated for each electrode site, participant, and accent type. Analyses were performed across three Anterior-Posterior columns of electrodes (see figure 1) using a separate ANOVA analysis for each column. These were; a Midline column (Fz, FCz, Cz, CPz, and Pz), an outermost column on each hemisphere, named L1 (FP1, FP2, F7, F8, T7, T8, P7, P8, O1 and O2) and a column nearer the midline in each hemisphere, named L2 (F3, F4, FC3, FC4, C3, C4, CP3, CP4, P3, and P4). The analyses included within participant factors of Accent condition (with three levels; Home, Regional and Foreign), Electrode Anterior/Posterior condition (five electrodes), and, for L1 and L2 but not the Midline, Hemisphere (left and right). There was also a between-participants factor of stimulus set (Study 1 vs. study 2). Only significant ($p < .05$) effects and interactions will be reported, adjusted using the Greenhouse-Geisser (1959) correction for violation of sphericity.

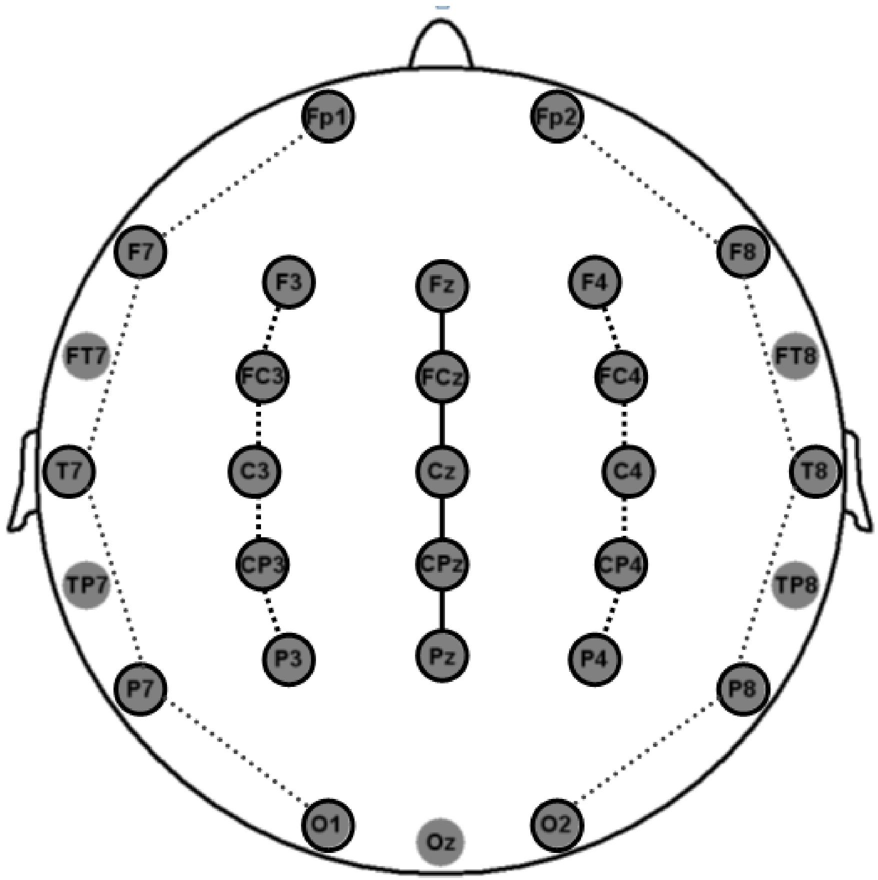


Figure 1: Schematic of electrode montage (standard 10-20 system) from which EEG activity was recorded. Highlighted electrodes used during L1 (grey dotted column), L2 (black dashed column), and midline (solid black column) electrode analyses.

Results

A visual inspection of the data allowed two time periods to be identified as being of interest; these were then used for subsequent analysis. These were 200-350 ms after onset, and 350-600 ms after onset. This earlier epoch should include the PMN or Phonological Mismatch Negativity, and captures phonological processing, while the later epoch includes a later N400, and should correspond to post-lexical processing.

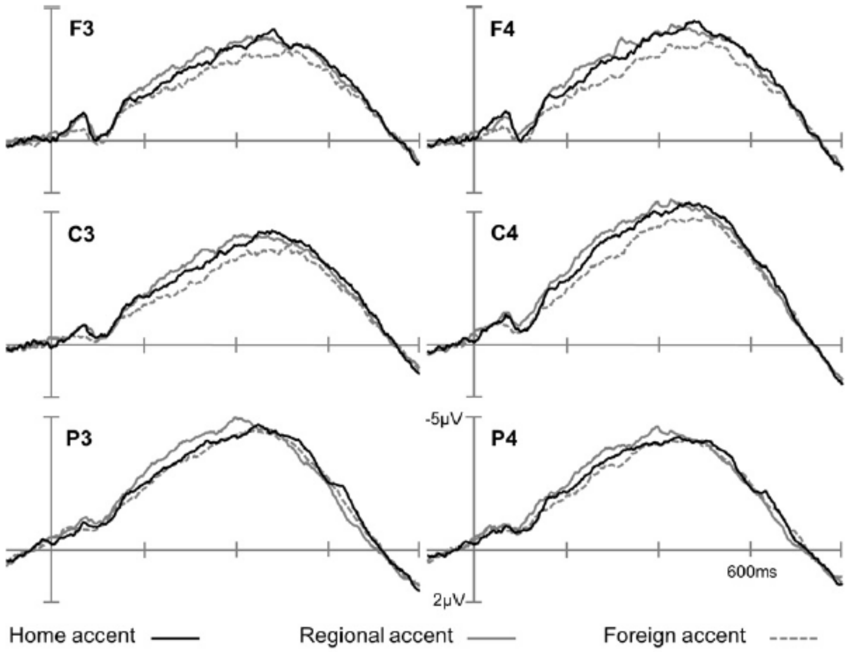


Figure 2: Grand Average ERPs to final words in three accent conditions (Home, Regional, and Foreign).

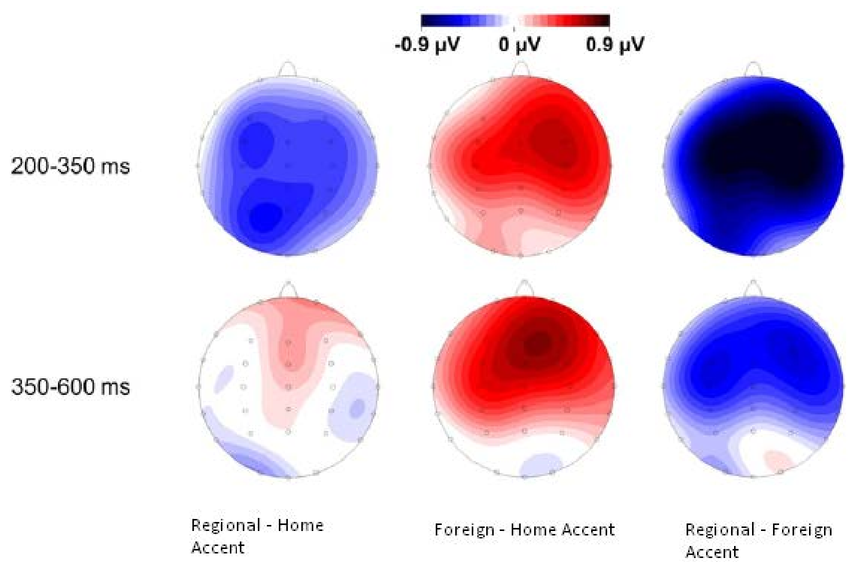


Figure 3: Headmaps showing topographic distribution of differences in voltages between conditions at 200-350 and 350-500 ms after onset of final word.

Early epoch – 200-350 ms

A significant main effect of Accent Condition was found in the early epoch in the Midline ($F(2,106) = 12.33$, $p < 0.001$, $\eta^2 = .19$), L1 ($F(2,106) = 5.62$, $p < 0.005$, $\eta^2 = .1$), and L2 ($F(2,106) = 13.11$, $p < 0.001$, $\eta^2 = .2$) columns. Planned pairwise comparisons between each pair of Accent Conditions revealed significant differences in this epoch between the Home and Regional accents in the Midline ($F(1,53) = 6.27$, $p = 0.015$, $\eta^2 = .11$) and L2 ($F(1,53) = 16.54$, $p < 0.001$, $\eta^2 = .13$) columns, and between the Home and Foreign conditions in the Midline ($F(1,53) = 6.74$, $p = 0.012$, $\eta^2 = .11$) and L2 ($F(1,53) = 9.30$, $p < 0.005$, $\eta^2 = .11$) columns. Differences were found between the Regional and Foreign accents in the Midline ($F(1,53) = 22.28$, $p < 0.001$, $\eta^2 = .3$), L1 ($F(1,53) = 15.24$, $p < 0.001$, $\eta^2 = .22$), and L2 ($F(1,53) = 31.12$, $p < 0.001$, $\eta^2 = .37$) columns. Accent Condition interacted significantly with Electrode Position and Hemisphere in L1 ($F(8,424) = 4.36$, $p < 0.005$, $\eta^2 = .08$), and planned comparisons showed that this interaction was due to significant differences between the Home and Foreign accents in three electrodes in the right hemisphere; FP2 ($F(1,53) = 4.90$, $p = .031$, $\eta^2 = .09$), F8 ($F(1,53) = 5.72$, $p = .02$, $\eta^2 = .1$), and T8 ($F(1,53) = 5.72$, $p = .005$, $\eta^2 = .14$), and between Foreign and Regional accents in four right hemisphere electrodes (FP2 ($F(1,53) = 4.42$, $p = .04$, $\eta^2 = .08$), F8 ($F(1,53) = 9.52$, $p = .003$, $\eta^2 = .15$), T8 ($F(1,53) = 14.79$, $p < .001$, $\eta^2 = .22$), and P8 ($F(1,53) = 14.95$, $p < .001$, $\eta^2 = .22$) and one left hemisphere electrode (O1 ($F(1,53) = 7.09$, $p = .01$, $\eta^2 = .12$)). No significant differences between the Home and Regional accents were found in any L1 electrodes.

Overall, the amplitude of the negative-going early component was significantly higher for Regional accents than for Home accents, while the early component for Foreign accents was significantly lower than for the Home accent.

Late epoch = 350-600 ms

A significant main effect of Accent was found in the later epoch in the Midline ($F(2,106) = 5.06$, $p < 0.001$, $\eta^2 = .09$), L1 ($F(2,106) = 3.8$, $p = 0.025$, $\eta^2 = .07$), and L2 ($F(2,106) = 6.69$, $p < 0.005$, $\eta^2 = .1$)

columns. Planned pairwise comparisons between each pair of Accent Conditions showed no differences between the Home and Regional accents in any of the three electrode columns ($F(1,53) < 1$). Responses to the Foreign accents were found to differ significantly from the Home accent in the Midline ($F(1,53) = 9.55$, $p < 0.005$, $\eta^2 = .15$), L1 ($F(1,53) = 5.38$, $p = 0.024$, $\eta^2 = .09$), and L2 ($F(1,53) = 10.81$, $p < 0.005$, $\eta^2 = .15$) columns, and from the Regional accents in the midline ($F(1,53) = 4.64$, $p =0.036$, $\eta^2 = .08$), L1 ($F(1,53) = 5.51$, $p = 0.023$, $\eta^2 = .09$), and L2 ($F(1,53) = 9.22$, $p < 0.005$, $\eta^2 = .14$) electrode columns. A significant interaction between Accent Condition and Anterior/Posterior position in the midline ($F(8,424) = 3.38$, $p = 0.033$, $\eta^2 = .06$) and L2 ($F(8,424) = 5.04$, $p = 0.003$, $\eta^2 = .09$) columns. Planned comparisons revealed that this interaction was due to significant differences ($p < .005$) between the Foreign accents and either Home or Regional accents in all Midline electrodes apart from Pz ($F < 1$) and all bi-hemispheric pairs of electrodes in L2 apart from O1 and O2 ($F < 1$). No significant differences were found between Home and Regional accents in any of the Midline ($F < 1.80$) or L2 electrodes ($F < 1$).

Overall, the amplitude of the negative-going later component was significantly smaller in the Foreign accent than in either the Regional or Home accent, with no significant differences in this component between the Home and Regional accent conditions.

Discussion

The aim of this study was to examine the electrophysiological responses to speech in regional and in foreign accents, compared to a familiar baseline accent, in order to investigate whether the two different types of unfamiliar accent require normalisation processes which are substantially the same, differing only in degree, or whether they recruit qualitatively different processes. ERP data was collected and two time epochs were examined, one (200-350 ms) corresponding with the PMN and one (350-500 ms) with the N400, in order to examine the amplitudes, directions and distributions of deviations in those components. While the peaks identified here may reflect a number of different processes, and may indeed be influenced by interactions between processes or by overlapping latent components, their latencies, directions, and apparent relationship to the stimuli used here make it convenient to refer to them as a PMN and an N400. In the earlier epoch, the regional accents were found to elicit larger negative-going responses than the baseline home accent, while the foreign accents elicited an attenuated component compared to both the regional and home accents. In the later epoch, the foreign accents elicited an attenuated N400 compared to both the home and regional accents, while no differences were found between the home and regional accents in this time period.

Both the Perceptual Distance and Different Processes hypotheses would predict a larger response to regional accents than the home accent in the PMN. However, the Perceptual Distance hypothesis would predict that the foreign accents, which are perceived as being further along the continuum of difference from the home accent than the regional accents are, would elicit an even larger response than the regional accents. Instead, the foreign accents elicit a smaller response than both the regional and the home accents; they differ from the home accents in the opposite direction from that predicted by the Perceptual Distance hypothesis, and this finding is therefore much more compatible with the predictions made by the Different Processes hypothesis. According to this hypothesis, foreign accents include elements which cannot be filtered out by a simple process of normalisation, as is the case for regional accents, and these elements may therefore act like noise on the speech signal. Noise masking, which reduces the intelligibility of speech, has been found to attenuate the PMN (Martin, et al., 1999), and listeners trying to process speech in the presence of high levels of background noise have been shown (Mattys, Brooks, & Cooke, 2009) to rely less on lexical or semantic top-down information and more on salient acoustic cues. Conversely, if foreign accents reduce the reliability of acoustic cues, due to the non-native features of the accent acting like noise on the signal, listeners might need to rely more on lexical/semantic cues, thus explaining the attenuation of the PMN component.

If this is the case, then the intelligibility of the accented speech should be the defining variable. And indeed, a transcription task using the accented stimuli used in this study found greater transcription accuracy in the home (93.83%) and regional (90.71%) accents compared to the foreign (80.60%) accents. This supports previous work which found that foreign accents reduce intelligibility (Adank, et al., 2009; Munro, 2008), and indeed a reduction in intelligibility may be considered one of the defining characteristics of a foreign accent. However, if intelligibility alone is driving the deviation in the PMN, as implied by the Perceptual Distance hypothesis, the foreign accents should elicit a greater PMN than the regional accents, which in turn elicit a large PMN than the home accents. This is not the case in the findings of this study; instead, the regional accents elicit a larger PMN than the home accent but the foreign accents elicit a smaller PMN than either of the other accent types. This might be explained by a step reduction in the PMN once an intelligibility threshold is passed, but this idea is not supported by the literature. Martin et al. (1999) presented two syllables, /b/ and /d/, in several different levels of noise masking, and found that increased noise produced progressively reduced amplitudes and increased latencies of ERP components, with no step change in evidence. The deviations in the PMN in this study would therefore seem to support the Different Processes hypothesis rather than the Perceptual Distance hypothesis.

As well as the differences in the polarity of the deviations in the PMN, there were also differences in the distribution of deviations across the scalp between the three accent conditions. In the L1 column, which contained the outermost electrodes, foreign accents elicited greater activity in the right anterior area, while regional accents recruited greater activity in the left posterior and temporal regions. Berman et al. (2003) found right hemisphere activity associated with accent detection and left hemisphere activity associated with word detection, but there has been little work seeking to identify the neural correlates of accent processing. Adank, Noordzij and Hagoort (2012) used a repetition suppression task with a change in speaker and a change in accent, and measured BOLD responses under fMRI, and found that a change in accent recruited activity in the left posterior temporal regions, and in frontal areas, while a change in speaker recruited mainly right-hemisphere activity. The accent they used was an artificially constructed accent, analogous to a regional rather than a foreign accent, so while EEG data has much poorer spatial resolution than MRI the results of this study would appear to be broadly compatible with their findings.

Overall, then, the findings of this study appear to indicate that different pre-lexical normalisation processes are being used in response to regional and foreign accents, as predicted by the Different Processes hypothesis but not the Perceptual Distance hypothesis. The lack of difference between the home and regional accents in the later epoch suggests that these processes are, in the case of the regional but not the foreign accents, sufficient to normalise speech in a regional accent by the lexical/semantic stage. In foreign accents, the persistence of deviations from the home and regional accents would indicate that the non-native phonological, phonotactic, and prosodic elements of the accent have not been normalised by the lexical stage, and are therefore having an impact on the lexical and semantic processing of speech.

In the current study, ERP responses to final words demonstrated the online normalisation involved in accented and familiar speech. The findings indicate that pre-lexical normalisation is recruited for both regional and foreign accents, but that it is only for foreign accents that post-lexical normalisation is incomplete. The nature of the differences is compatible with the Abstract Entry approach to lexical representation, which suggests that variant pronunciations must be normalised pre-lexically in order to be matched to a single representation of the target word (Pallier, et al., 2001). The Exemplar-based approach, on the other hand, suggests that variant forms are represented individually, along with indexical information about the speaker who provided the variant form, including information about their accent (Johnson, 1997). According to this theory, the ease with which variant forms are accessed depends on their similarity to existing stored representations and on their base level of activation, which will in turn depend on their frequency. It might be argued that native English speakers will have more experience with regional accents of English than with non-native accents, and therefore more exemplars linked to regionally-accented indexical information. If this is the case, the results of the current study might reflect, not a normalisation process, but rather a process of selection of the best-fit representation in the participants' lexicons, with the regional accents but not the foreign accents being successfully matched to an existing exemplar. By presenting participants with paired stimuli which differ in accent but not in semantic content, it should be possible to determine whether regional and foreign accented speech elicits the same lexical representations as familiar speech, or whether this is true only of regional and not of foreign accented speech. This in turn would lend support to either the Exemplar-based approach (if accented speech does not result in habituation to semantic content in the familiar speech) or the Abstract Entry approach (if regional accented speech does elicit semantic habituation in the familiar speech. A repetition-suppression study will therefore be carried out in order to attempt to differentiate between the two theories.

Chapter 4: Repetition-suppression to accented speech

Introduction

The previous chapter showed that the processes used to normalise foreign and regionally-accented speech differ, such that regional accents seem to be fully normalised by the post-lexical period but foreign accents do not. This study seeks to examine in more depth the way in which accented speech is processed, and by doing so, to gain a greater understanding of the way in which variant forms of familiar words are stored and accessed.

Two Approaches

There are two main theories to explain our ability to process familiar words spoken in unfamiliar variant forms. The first, referred to as the Exemplar based approach, proposes that a given word may have many discrete entries in the lexicon, one for each possible pronunciation, and all of which contribute to an overall category prototype, linked to the same semantic information (Johnson, 1997). This should allow for quick recognition, but at the cost of an extremely large lexicon, and with limited flexibility; a novel pronunciation of a familiar word would need to have a new lexical entry created in order for it to be recognised. A less extreme version of this approach is that commonly-used variant forms will have their own lexical entry, while less-frequent forms will not, and will therefore be subject to some form of normalisation process. Connine (2004) looked at words with a typical American medial flap (in which a /t/ sound in the middle of a word becomes a /d/ sound, for example turning “preTTy” into “preDy”), in a New York population, for whom the flapped variant is more common than the supposedly canonical articulated /t/ version (Connine, 2004). Listeners heard stimuli based on words which could carry either the articulated /t/ or the flapped variant, and which began with an ambiguous consonant along a voicing continuum (such as /p/ to /b/). In each case, this could result in the stimulus being heard as a word or a non-word (for example pretty vs bretty). They were asked to identify whether they heard the voiced or unvoiced consonant. Connine found that listeners were more likely to hear the ambiguous consonant as forming a real word with the flapped variant, suggesting that the flapped variants were lexically represented. This supports the idea that common variants may be explicitly encoded in the lexical system.

A priming study which used two legal variations on words with a terminal /t/ (a glottal stop and a co-articulated glottalisation) along with a mismatch condition (in which the /t/ was replaced by a different phoneme) shows that several different legal variants of a word may be equally effective as a prime, with no advantage conferred by frequency; the most common variant is not more effective than a less common but equally valid form (Sumner & Samuel, 2005). However, a further study (Sumner & Samuel, 2009) used separate participant groups, representing rhotic and /r/-less accents. Participants heard prime and target words ending with –er, and pronounced either with (as in a so-called "General American" accent) or without the terminal /r/ (as in a New York accent), and the results showed a significant interaction between participant group and accent condition; while the General American primes were effective across the board, the rhotic participants showed significantly less facilitation to the New York targets, with which they have much less experience. Taken together, this evidence suggests that common variant forms may indeed be represented lexically, and that the strength of their activation may depend on the frequency with which a listener encounters them.

According to the Exemplar based approach, speaker variance should be encoded as part of the representation of variant pronunciations. One result of this is that it should be easier to process speech by a single consistent speaker than by multiple novel speakers, and indeed a number of studies have shown this to be the case (Mullennix, Pisoni, & Martin, 1989). Intra-modal and cross-modal word naming tasks were used with either one consistent speaker or a number of randomly-ordered speakers, and the results showed significantly better performance when only one speaker was heard, lending support to the idea that indexical information (that is, characteristics specific to an individual speaker, including accent) is encoded along with lexical representations.

The second approach, known from here on as the Abstract entry approach, proposes a single lexical entry for each word (thus reducing the size of the lexicon), combined with a pre-lexical normalisation system (Pallier, et al., 2001). This system would need to include a number of rules for normalisation, for example, describing the circumstances under which a medial or final consonant such as /t/ can be flapped, elided, or glottal stopped. The system may potentially be context-sensitive, so that it "knows" which rules are likely to apply at any given time (that is, which systematic variants are common in a given accent or to a specific speaker, or which phonemes may change, and in what ways, due to co-articulation), and by using both top-down and bottom-up information, it can be flexible in establishing new rules, so that unlike the Exemplar-based approach, it can quickly adapt to novel words, or words which have never previously been heard with a specific variant pronunciation. This approach is supported by a study which compared bilingual Spanish and Catalan speakers' performance on a repetition priming task using minimal pairs of words, differing only by a Catalan-specific contrast or by a contrast common to both Spanish and Catalan (Pallier, et al., 2001). The results showed that the Spanish-dominant participants demonstrated a strong repetition priming effect for the Catalan-specific minimal pairs, indicating that they were treating the two different acoustic traces as the same word, in spite of the fact that they spoke both Spanish and Catalan fluently, and thus had valid lexical representations for both words in the minimal pairs. This finding is contrary to predictions made by the Exemplar theory.

Further support comes from a lexical decision study using ambiguous fricative sounds (McQueen, Cutler, & Norris, 2006). Dutch listeners underwent a training phase in which they heard words ending in a fricative sound. For some participants, /f/ was consistently replaced with an ambiguous [f-s] sound (/?/), while for others, /f/ was intact but /s/ was replaced with the same ambiguous sound. After training, participants heard stimuli based on minimal pairs, such that they could be completed with either /f/ or /s/ (for example knife/nice), none of which had been used in the training phase, and using the ambiguous /?/ sound in place of the fricative. These were used as primes in a cross-modal priming task, in which listeners were asked to make a lexical decision to visually-presented words, some of which matched one of the auditory candidate words. The findings showed that the ambiguous auditory prime (such as /naɪ?/) had a priming effect in the direction of the training condition; that is, for those participants trained with an ambiguous /f/, /naɪ?/ primed "knife" but not "nice". Since the target words had not been heard during the training phase, this priming effect cannot be due to the ambiguous targets being lexically represented, as would be suggested by the Exemplar-based approach. Instead, the findings indicate that the training phase had shifted listener's phonemic boundary for the ambiguous sound, and that the training was therefore allowing participants to employ the new phonemic boundary during normalisation, as would be predicted by the Abstract entry approach.

However, there is a potential time-cost involved in adding an extra stage of processing between input (the acoustic patterns of speech) and recognition (lexical access), which would not apply to the Exemplar-based approach in the case of commonly-heard variants. This time-cost may be off-set in the case of less commonly-heard variant forms, because under the Exemplar-based approach but not the Abstract entry approach, low frequency variant forms will suffer from lexical competition from other, more common forms of the same word. Human listeners are slower to identify words which have many phonological neighbours than those with few, due to competition (McClelland & Elman, 1986; Norris, 1994), and also slower to identify low-frequency than high-frequency words (Dahan, Magnuson, & Tanenhaus, 2001). Under the Exemplar approach, therefore, a word which has many different possible pronunciations will have, in effect, a larger phonological neighbourhood (since each variant will be lexically represented), and will be identified more slowly.

According to the Abstract entry approach, the context in which individual words are heard should affect the way in which they are processed. This has been supported by a study in which participants were presented with final-flap and articulated /t/ variants either with or without following context, which either did or did not provide a licensing context for the variant (Ranbom, et al., 2009). A primed lexical decision task was used, and showed that the following context significantly improved performance, especially in the case of the final-flap variants. This supports the idea of a rule-based normalisation process, in which the phonetic and linguistic context in which a word is spoken triggers the application of specific rules. Further support comes from a similar study carried out with Dutch listeners (Ernestus, Baayen, & Schreuder, 2002), who were presented with reduced or shortened variants of common words, presented either with or without the context which allows them to be shortened. A transcription task showed much better performance in the presence of following context. The Abstract theory suggests that the shortened forms will be normalised in order to allow lexical access to the canonical (unreduced) form, and that this normalisation will be assisted by licensing context. Thus these studies lend support to the Abstract approach. By contrast, the Exemplar approach would suggest that the reduced forms, which are common in Dutch colloquial speech, should be represented separately from the unreduced variants, and should thus be as easy to recognise, regardless of context, which is not supported by these findings.

With evidence supporting both the Abstract entry and the Exemplar based approach, it may be useful to consider the possibility of a compromise position, combining elements of both theories. One study which may help to guide the way in this used a series of lexical decision and word shadowing tasks, under various conditions which rendered the tasks easier or harder (McLennan & Luce, 2005). The results indicate that, when the task is relatively slow and difficult, participants relied more heavily on indexical information, whereas when the task was easy and processing was quick, indexical information had much less impact. This might suggest that lexical access is subject to a dual-route model, like those already proposed by a number of researchers for aspects of morphology (Baayen & Schreuder, 1999), auditory processing (Hanley, Kay, & Edwards, 2002), and even higher cognitive functions such as reasoning (Evans, 2003). Dual route models propose that cognitive tasks can be achieved in one of two different ways; a heuristic route, which allows quick but potentially imprecise decisions to be made on the basis of only minimal information, and a more effortful route which is slower but may be more accurate, and which draws on all the available information. McLennan and Luce's findings are compatible with such a model, with participants using the quick heuristic route for easy tasks, and resorting to the more effortful route for harder tasks, helped by drawing on indexical information which was less important for easier tasks. Some support for this position is offered by Boomershine (2006), who looked at Mexican and Puerto Rican dialects of Spanish, in Mexican and Puerto Rican groups. Participants completed a number of linguistic tasks including accent identification, lexical decision, and word naming, using words featuring phonological features which differ between the two dialects. For example, in Mexican Spanish, a syllable-final /s/ for a voiced consonant is retained, but in Puerto Rican Spanish it is frequently aspirated or deleted altogether. Boomershine found that this distinctive feature was associated with the poorest performance (as measuring by reaction times) in the word naming and lexical decision tasks, but also with the most accurate dialect identification. This was interpreted as offering support for the exemplar theory but in fact, the Exemplar theory would predict that a significant interaction would be found between the phonological variable and the listener dialect; we would expect the Puerto Rican listeners to have a strong representation of both the aspirated or deleted and the retained variants of the stimulus words (because Mexican Spanish is commonly heard via the media and has the status of a "standard" or formal form of Spanish in much of Latin America), while the Mexican listeners would not be expected to have a representation of the aspirated or deleted variant since it is uncommon in Mexico. Thus we would expect the Mexican listeners to perform well with the retained version and poorly with the aspirated/deleted version, and the Puerto Rican listeners to perform well with both variants. Since this is not the pattern of results Boomershine found, the data cannot be said to offer strong support to the Exemplar theory; instead, they suggest that hard-to-process features may be most useful in the identification of dialects or accents, and may thus be useful in determining which set of normalising rules to employ.

The findings reported in the previous chapter, demonstrating different patterns of processing for regional and foreign accents, point to pre-lexical normalisation in the case of the regional accents. This in itself lends support to the Abstract entry approach, which proposes the use of pre-lexical normalisation to access a canonical lexical entry. However, it might also be argued that the differences found are due to regional variants of the words used being lexically represented, while the foreign accented variants were not lexically represented, and could therefore not be accessed in the same way, instead relying on post-lexical top-down information in order to achieve lexical access. Priming studies such as some of those described (Ernestus, et al., 2002; Ranbom, et al., 2009) offer a way of shedding light on this question, by allowing us to look not only at the online processing of variant forms, but at the relationships between variant and canonical forms of words, and specifically the perceived similarity of variant forms to a familiar form. In order to make comparisons between both pre-lexical and post-lexical processing, however, a more sensitive measure than the behavioural methods used previously is required. Repetition-suppression, a form of priming seen in neurophysiological responses to repeated stimuli, is therefore an ideal approach to this question.

Repetition-Suppression

Repetition-suppression is a term used to describe a systematic reduction in neurophysiological response to a stimulus on repeated presentation, indicative of habituation to the stimulus at a neural level, and, like habituation in behavioural studies, it can be used as a proxy for the perceived similarity between two stimuli. If no difference is perceived, the response to the second stimulus will be suppressed whereas if a difference is perceived there will be little or no suppression, but rather the response will show signs of dishabituation, returning to levels typical of novel stimuli (Hasson, Nusbaum, & Small, 2006). This fact allows us to use repetition-suppression to examine the importance of different types of variation in a stimulus to their processing, much as habituation to a visual grid can allow us to examine visual acuity, by looking for dishabituation to a very similar grid in order to establish the degree of difference necessary for dishabituation to occur. The use of neurophysiological techniques such as EEG or fMRI also allow us to localise both activity in general and differences in activity specifically, within the brain, giving us insight into the way in which different stimuli are processed (See for example Belin & Zatorre, 2003; Orfanidou, et al., 2006). For example, since semantic content and indexical auditory information are processed in different areas in the brain and at different time-points, repetition-suppression studies in which stimuli repeat either semantic information or indexical information can isolate differences in the processing of each variable. Orfanidou, Marslen-Wilson and Davis (2006) used fMRI to examine repetition suppression to single words and pseudowords, spoken by either the same speaker or different speakers. They found evidence of repetition suppression, in the form of reduced neural activation to the second presentations of the stimuli as compared to the first presentations, to both types of stimuli, regardless of whether the voice remained the same, or changed between presentations of the stimulus, suggesting that acoustic dissimilarity does not prevent lexical priming from taking place. They were also able to show that suppression occurred in anterior and posterior regions, and that there were correlations between behavioural priming and neural suppression in the frontal region, suggesting a link in these areas between suppression and behavioural responses.

ERP studies have also revealed an effect of repetition suppression to repeated words. Phillips, Klein, Mercier and de Boysson (2006) presented repeated words (such as “bed… bed… bed… bed...”) to English/French bilingual participants, in both their first and second languages (L1 and L2). They found reduced activity in the N400 epoch to repetitions of the words, as compared to the first presentation. Activity then returned to baseline levels on presentation of a related (“sleep”) or unrelated word (“sky”) in the same language, or a translation of the same word in the other language (“lit”; that is, “bed” in French), indicating that repetition suppression occurred only when both the word-form and the meaning remained unchanged. Holcomb and Grainger (2006) found a similar reduction in the amplitude of responses to repeated presentations of written words.

Using a partial replication of Orfanidou et al.’s design would make it possible to compare the extent of repetition suppression in response to sentences spoken not only by the same speaker versus different speakers, but by different speakers with the same versus different accents, and specifically regional versus foreign accents. Further, by using EEG rather than fMRI, it is possible to look at the time-course of repetition suppression, in order to pinpoint the specific stages at which similarities or differences between stimuli are perceived or processed.

The current study

The current study therefore compares regional and foreign accents directly, by making use of a repetition-suppression paradigm and EEG data, in order to examine differences from a familiar baseline accent. Participants from the South-West heard sentences spoken twice; the second repetition was always in a baseline local accent while the first was in a local, regional or foreign accent. Both the Exemplar-based and the Abstract Entry approach would predict post-lexical suppression, since both approaches suggest that variant forms eventually activate the same semantic entry in the lexicon. It is therefore predicted that in a late epoch, between 350-600 ms after word onset, the responses to the repeated words will show repetition suppression in both the familiar speech conditions and the accented conditions; in other words, the amplitudes of responses for the second repetitions will be lower than those for the first repetitions. Exact patterns of suppression, however, are likely to differ for regional and foreign accents, since during the post-lexical period, the previous study shows that foreign accents are still reliant upon top-down information to make up for the lack of pre-lexical normalisation. However, in the pre-lexical period, the two approaches would make different predictions. The Exemplar-based approach would indicate that a first repetition in speech in a familiar (South-West) accent should elicit suppression in the second repetition, since both the repetitions will activate the same exemplar. Since accented variations of the word are, according to this approach, stored individually, this approach would predict that the two accented conditions, Regional and Foreign, should not show suppression in the PMN (250-350 ms after word onset).

The Abstract Entry approach, on the other hand, posits a single lexical entry for each word, which is activated after a process of normalisation. The PMN is associated with lexical selection, and would therefore be expected to show suppression for the accented conditions, according to this approach. Further, the degree of suppression should correspond to the ease with which the accented speech is normalised. Therefore the Abstract Entry approach would predict that in the PMN, suppression will be found in the familiar speech conditions, and also in the Accented conditions. As demonstrated in the previous chapter, different normalisation processes are used for regional and foreign accents, with regional accents being fully normalised prior to lexical access, while foreign accents were not fully normalised by this stage, and therefore relied more heavily on post-lexical top-down information. This being the case, it is predicted that the Regional accent condition should show greater repetition suppression pre-lexically than the foreign-accented condition.

Table 4: Predictions made by the two approaches for two epochs

		PMN (250-350 ms)	**Post-Lexical epoch (350-600 ms)**
Exemplar-based Approach	Familiar Speech	Suppression	Suppression
	Accented Speech	No Suppression	Suppression
Abstract Entry Approach	Familiar Speech	Suppression	Suppression
	Accented Speech	Suppression; Regional > Foreign	Suppression

The previous study pitted the Perceptual Distance hypothesis against the Different Processes hypothesis, and provided support for the latter. In the current study, the Perceptual Distance hypothesis would once again predict that the degree of repetition suppression should vary along a continuum, with the Familiar accent conditions eliciting the greatest suppression, and the Foreign accent condition eliciting the least, with the Regional accent condition falling in between the two. The Different Processes hypothesis would predict different patterns of suppression for the Regional and Foreign conditions. In particular, this approach would predict that the Regional accent condition should elicit suppression pre-lexically (in the PMN) and post-lexically (in the later epoch), while the Foreign accent should not elicit suppression in the PMN, since the previous study shows that foreign accents are not normalised pre-lexically in the way that regional accents are. In the later epoch, suppression is expected in the Familiar and Regional accent conditions (since lexical access should already have been achieved by the end of the earlier epoch. Given the findings of the previous experiment, which showed differences in the post-lexical period between the Foreign accent conditions and the Home and Regional conditions, it is predicted that suppression will be reduced or absent in the Foreign condition, reflecting a greater reliance on post-lexical top-down processing when dealing with foreign-accented speech.

Method

Participants

37 right-handed adults (22 female) with a mean age of 31 years (range 19 years and 10 months – 65 years and 11 months) were recruited via Plymouth University's paid participants pool; a further two were excluded due to only providing a small number of usable segments, and two were excluded due to technical issues or equipment failure. All the participants had grown up in English-speaking monolingual households in the South-West of England, and were therefore familiar with the baseline accent.

Materials

A total of 13 female speakers recorded 420 sentences each. Three speakers (ages 21, 27 & 31) were from Plymouth or the surrounding area and had South-Western (from here on, SW) accents, two speakers (ages 24 & 28) had Southern Welsh accents, two speakers (ages 24 & 25) had Yorkshire accents (from Hull and Leeds respectively), two (aged 21 & 29) were from Hong Kong and were native Cantonese speakers and two (aged 30 & 42) were native French-speakers from Southern France. Sentences were all English sentences with no embedded clause, and they all ended with a bi-syllabic trochaic noun, some of which were animals names and some were not. Since all speakers recorded all the sentences, there were no differences in stimulus characteristics across conditions. The mean Phonological Uniqueness point for the target words was 4.83 (std dev = 1.17), the mean phonological neighbourhood size was 2.13 (std dev = 2.82). The mean number of phonemes was 5.64 (std dev = 1.06). The mean frequency, taken from Subtlex (Baayen, Piepenbrock, & van Rijn, 1993) was 4.35 (std dev = 7.91) and the mean phonological Levenschtein distance was 2.15 (std dev = .57).

An example of a sentence ending with a non-animal is "Down at the end of the field there was a small paddock". The SW speakers recorded two versions of these sentences. Sentences ending with an animal also had a non-animal version; for example “Hunting would be easier with the skills of a kestrel” and “Hunting would be easier with the skills of a trapper”. All speakers were asked to read the sentences in a natural, conversational tone. Recordings were made in a quiet room using Adobe Audition, which was also used to remove noise where necessary, and to normalise sound levels. The sound files were trimmed to remove any noise preceding or following the sentences.

Procedure

Stimuli were presented over headphones, using E-Prime software, version 1.1 (Schneider, et al., 2002). A 22-trial training block preceded the main experiment in order to ensure that participants understood the task. The main experiment then consisted of three 160-trial blocks (420 sentences in all), separated by rest periods. Paired sentences were never split across blocks. Participants were asked to listen for sentences ending in the name of an animal, and to press a response button only for those sentences, which constituted around 5% of the total; these trials and the sentences paired with them, which did not end in an animal name but which were included to prevent participants from predicting “Go” trials, were not included in the analysis.

Participants were given a “blink” command between trials, and were encouraged to try to avoid blinking at other times. Between blocks they were able to take a short rest period and re-commence when they were ready to do so.

Accents

SW

The baseline accent used was a South-West accent; specifically, the three SW speakers were brought up in Plymouth, Devon. The West Country accent is a Southern English accent the most salient feature of which is a rhotic /r/ in words such as “graveyard” and “clutter”. In intonation it does not differ significantly from Received Pronunciation or Standard British English (SBE), although word-final vowels are sometimes elongated, giving the impression of a slower pace of speech. In our corpus, however, the South-West speakers were among the fastest, with a mean sentence duration of 2265 ms, and a final word duration of 589 ms.

French

Non-native accents arise as a result of the interaction between the phonology and prosody of their native tongue and the L2 in which they are speaking. French is often described as being a syllable-timed language, whereas English is considered a stress-timed language, and French-accented English speech is therefore often perceived as having a different rhythm from native English, with fewer elisions or reduced vowels, and less elongation of word-initial and word-final syllables; instead, syllables are closer to each other in duration, spaced more regularly, and are stressed more equally than in native speech.

The English interdental fricatives / ð/ and /Ө/ are not used in French, so French speakers will often render them as /v/, /z/ or /d/, and /f/, /t/ and /s/ respectively. This is particularly noticeable in our corpus following fricatives such as /s/ and /z/, as in the sentence "Damion enjoyed going running with his trainer". French has an uvular /r/ rather than the approximant /r/ used in English, so a French speaker make render /r/ as closer to /w/ than in naïve English speech. French also does not use an aspirated /h/, so in English /h/ is frequently either deleted or exaggerated; in our corpus this is clearly illustrated in the sentence "Harry had never seen such a large concourse", which the /h/ at the beginning of "Harry" is enunciated much more than by any of the native English speakers, but the /h/ at the beginning of "had" is missing entirely. Terminal consonants are often elided in French, and may therefore be deleted or reduced in English. This is exemplified in our corpus in the sentence "For her prickly nature she was nicknamed cactus", in which the /d/ at the end of "nicknamed" is deleted by the French speakers but enunciated by the native English speakers.

The use of vowels in French differs from that in English. French does not use /iː/ and so “seen” approaches “sin” when spoken with a French accent. French also does not use the English diphthong /əʊ/, which may be realised in French accented English as a monophthong /o/, or as a somewhat exaggerated diphthong (both exemplified in our corpus in the sentence “Both were pretty fed up with all the packing”).

Cantonese

Cantonese is thought of as a syllable-timed language, like French, and in contrast to English. Cantonese-accented English can therefore differ in its prosody from native English. This, combined with the lack of double consonantal endings, can give it a clipped sound and a “machine-gun” rhythm. Cantonese does not include double consonants at the end of words, so words ending with a double consonant may be reduced, or enunciated slowly and in an exaggerated manner, as exemplified in our corpus by the sentence “Benjamin’s voice screeched like an enraged preacher”, in which one of our Cantonese speakers reduced “screeched” to “screech” and “enraged” to “enrage” while the other enunciated the /d/ sound without the co-articulation used by our native English speakers, thus lengthening the duration of the sentence and changing its rhythm. Some other double consonants may also be reduced or changed; in our corpus the /ks/ in “except” was reduced to /s/.

In Cantonese, there is no contrastive distance between /s/ and /sh/, and so in Cantonese accented English, both tend toward a central point; in our corpus this is particularly well illustrated I the sentence “As she was Scottish, Sue had never heard a cockney”, which, despite a lack of co-articulation between “Scottish” and “Sue”, “Sue” is pronounced as “shoe”.

The English interdental fricative /Ө/ is frequently produced in a Cantonese accent as /v/, /f/ or /sh/, as illustrated in “Dave impressed everyone with his display of daring”, in which “with” is pronounced with a

/v/ sound at the end. The voiced fricative / ð/ is often produced as /d/, as in “Deep down in the cellar there was plenty of cognac”, where both “the” and “there” were produced with an initial /d/ sound.

Final consonants in Cantonese are generally devoiced, so terminal /d/ and /t/ are often confused, as in the sentence “It was hard not to laugh at his downfall”, in which our speakers did not voice the /d/ are the end of “hard”.

Cantonese does not distinguish between /æ/ and /e/, so in the sentence “The performer was happy to find his backer”, the /æ/ sound in both “happy” and “backer” tended towards /e/.

Welsh and Yorkshire

The Welsh and Yorkshire accents in English are described in the previous chapter.

Stimulus ratings

The sentences were used in an intelligibility and rating task, by 11 participants, all of whom originated in the South-West of England. They each heard 8 sentences from each of the 11 speakers, presented in random order over headphones, by the EPrime software. After each sentence they were asked to enter the final word of the sentence. They were then asked to identify the accent in which the sentence had been spoken by pressing a key on the keyboard; 1 for South-West, 2 for Cantonese, 3 for Yorkshire, 4 for French, and 5 for Welsh. There was also the option of pressing 0 for “No Idea”, but participants were encouraged to guess if possible. Next they were asked to rate their confidence in their identification on a scale of 0 (“Complete guess”) to 5 (“Absolutely certain”), then to rate the strength of the accent, also on a scale of 0 (“Not at all strong”) to 5 (“Very strong”), and finally to say how easy it was to understand the sentence on a scale of 0 (“Impossible to understand”) to 5 (“Very easy to understand”).

Overall, participants correctly identified the accents 67% of the time (n = 968). South-West and Cantonese accents were correctly identified on 75% of trials, with Yorkshire accents correctly identified

70% of the time, French accents identified 60% of the time, and Welsh accents (surprisingly, given the higher identification rate in the previous study) just 50% of the time. Welsh accents were most commonly mistaken for South-West accents (on 31% of trials) (see confusion matrix in Table 5).

The South-West accent was rated as being the weakest of the five accents (m = 2.19), and differed significantly from all four other accents ($p < .001$ in all cases). The Welsh accent was the next weakest (m = 2.89), differing significantly from the French accent (m = 3.51, $p < .001$) and the Cantonese (m = 3.81, $p < .001$) but not from the Yorkshire accent (m = 3.13, p = .098). The Yorkshire accent also differed from the Cantonese ($p < .001$) and French accents (p = .009). The French and Cantonese accents also differed from each other (p = .033).

Table 5: Confusion matrix of accent identification for experimental stimuli; number of responses (proportion of responses)

Accent	Identification					
	Cantonese	French	South-West	Welsh	Yorkshire	Don't Know
Cantonese	**132 (.75)**	16 (.09)	2 (.01)	5 (.03)	3 (.02)	18 (.1)
French	40 (.23)	**105 (.60)**	4 (.03)	5 (.03)	2 (.01)	20 (.11)
South-West	2 (.01)	2 (.01)	**199 (.75)**	6 (.02)	16 (.06)	39 (.15)
Welsh	0 (0)	0 (0)	54 (.31)	**88 (.50)**	14 (.08)	20 (.11)
Yorkshire	0 (0)	0 (0)	29 (.16)	9 (.05)	**123 (.70)**	15 (.09)

A t-test showed that participants' confidence was higher for correct identifications (m = 3.17) than for incorrect identifications (m = 2.71, p = .001), and incorrectly identified sentences were rated as having a stronger accent ($m = 3.40$) than correctly identified sentences ($m = 2.93$, $p < .001$). Correctly identified sentences were rated as being easier to understand ($m = 3.77$) than incorrectly identified sentences ($m = 3.02$, $p < .001$). Ease of understanding and Strength correlated negatively ($r^2 = -.35$, $p < .001$).

Target words were judged to have been correctly identified if the participant transcribed them accurately, or replaced them with a homophone or phonetic spelling which accurately represented the word in their own dialect; for example, since “bridle” and “bridal” are homophonic in the South-West dialect, participants were judged to have correctly identified the word “bridle” if they entered “bridal”. Overall, 80% of target words were correctly identified. Target words were identified correctly most often in the South-West (mean = .86), Welsh (mean = .84) and Yorkshire (mean = .82) accents than in the French (mean = .78) and Cantonese (mean = .68) accents. Word identification scores did not differ for the South-West and Welsh ($p = .52$) or Yorkshire ($p = .18$) accents, and the scores for the Welsh and Yorkshire accents also did not differ significantly ($p = .69$). South-west scores were higher than those for the French ($p = .049$) and Cantonese ($p < .001$) accents. The Welsh scores were higher than the Cantonese ($p< .001$) but not the French ($p = .23$) scores, and the Yorkshire scores were also higher than the Cantonese ($p < .001$) but not the French ($p = .42$) scores. The French and Cantonese scores also differed significantly ($p = .015$). When the accents were grouped into three categories, “Home” (South-West), “Regional” (Welsh and Yorkshire) and Foreign (Cantonese and French), participants were more likely to identify the accent correctly in the Home condition ($m = .75$) than the Regional condition ($m = .60$, $p < .001$) or the Foreign condition ($m = .67$, $p = .035$). They were more likely to correctly identify the Foreign than the Regional sentences ($p = .036$). They were also more likely to correctly identify the target word in the Home condition ($m = .86$) and the Regional condition ($m = .83$) than in the Foreign condition ($m = .73$, $p = .002$). There was no difference between the Home and Regional conditions in the number of words identified correctly.

Table 6: Durations of Sentences and Target Words across accents (SD in brackets)

Variable	Cantonese	French	South-West	Welsh	Yorkshire
Sentence Duration (MS)	2703 (378.6)	2559.08 (269.41)	2264.69 (277.58)	2465.59 (323.25)	2204.99 (347.31)
Word Duration (MS)	611.83 (111.20)	637.62 (110.47)	588.78 (123.50)	628.72 (139.75)	564.26 (103.56)

Differences were found in both word and sentence durations across all three conditions, with the Familiar sentences having a mean duration of 2264.69 ms, significantly less than the Regional sentences ($m = 2335.66$ ms, $p < .001$) and the Foreign sentences ($m = 2631.04$ ms, $p < .001$). The Foreign and regional sentences also differed significantly ($p < .001$). Similarly, the Familiar Target word durations ($m = 588.78$ ms) were shorter than those in the Regional condition ($m = 596.47$ ms, $p = .045$) and the Foreign condition ($m = 624.73$ ms, $p < .001$), and the Foreign and Regional word durations also differed significantly ($p < .001$).

The stimuli were sorted into pairs of sentences, falling into four conditions, with the second sentence in each pair always being spoken by one of the three SW speakers. In Conditions 1 (Plymouth – same speaker), the first sentence in the pair was spoken by the same speaker as the second sentence. In Condition 2 (Plymouth – different speaker), the first sentence was spoken by a different SW speaker. In condition 3 (Regional), the first sentence in the pair was spoken by one of the four Regional speakers, and in condition 4 (Foreign), the first sentence of the pair was spoken by one of the four Foreign speakers. The two sentences in each pair were identical in content in 90% of trials (although the two sentences were always different tokens, even when spoken by the same speaker); in the other 10% of cases, one sentence of the two ended with the name of an animal, while the other did not. The order of the animal and non-animal endings within pairs was counter-balanced to ensure that animal endings were not predictable, and the animal sentences were used in a go-no-go task to ensure that participants were attending to the stimuli. This task required participants to press a button when they heard a sentence which ended with the name of an animal. Neither of the sentences in the go-trial pairs was used in the analysis.

The pairs of sentences were arranged into pseudo-randomised lists of 160 sentences (80 pairs), using the Mix software (van Casteren & Davis, 2006), such that the members of a pair appeared no fewer than 8 and no more than 16 sentences apart (Orfanidou, et al., 2006), and so that repeating patterns were avoided. Each participant heard three such blocks, and no pairs were repeated within or across blocks for any participant. Sentences were counterbalanced across accents, so that a sentence heard in one condition or spoken by one speaker by one participant was spoken by a different speaker and appeared in a different condition for another participant. There was also a practice block, which was identical for all participants, consisting of ten pairs of sentences. None of the sentences used in the practice block were repeated in the trial blocks.

Recording System

ERP data were recorded using a 62 channel Ag/AgCl ActiCap system (actiCap, Brain Products GmbH), mounted on an elasticated cap (see figure 4) with Vision Recorder software. Two further electrodes below and lateral to the right eye were used to monitor eye movements and blinks, but were not included in the final analysis. EEG was referenced using the left mastoid electrode and re-referenced offline to the averaged activity from left and right mastoid electrodes, with the AFz electrode providing the ground. Processing of the ERP data was carried out using the Vision Analyser software. EEG epochs, beginning at 100ms before onset of target words and ending 800ms after onset, were averaged for each accent and for each participant.

Comparisons were made between the first and second presentations of sentences, in order to give a measure of repetition-suppression, with the Plymouth – same speaker condition being taken as a baseline, since it was predicted that suppression was most likely to be found in this condition. A bandpass filter (0.1 – 40Hz) was applied, and data were corrected by a baseline of 100ms before target onset. Artefacts, including blinks and muscular artefacts, were removed, resulting in 4.7% of the no-go trials being discarded, with no difference found in the number of discarded trials between the three accent conditions ($F<1$). “Go” trials and those where the participant had erroneously responded were also discarded, so only trials on which no response was required or given were included in the analysis. Separate ERPs were calculated for each electrode site, participant, and accent type. Analyses were conducted across four anterior-posterior columns of electrodes, each consisting of five electrodes in each hemisphere. There were L1 (Fp1, F7, T7, P7, O1, Fp2, F8, T8, P8, and O2), L2 (AF3, F5, C5, P5, PO3, AF4, F6, C6, P6, and PO4), L3 (F3, FC3, C3, CP3, P3, F4, FC4, C4, CP4, and P4), and L4 (F1, FC1, C1, CP1, P1, F2, FC2, C2, CP2, and P2).

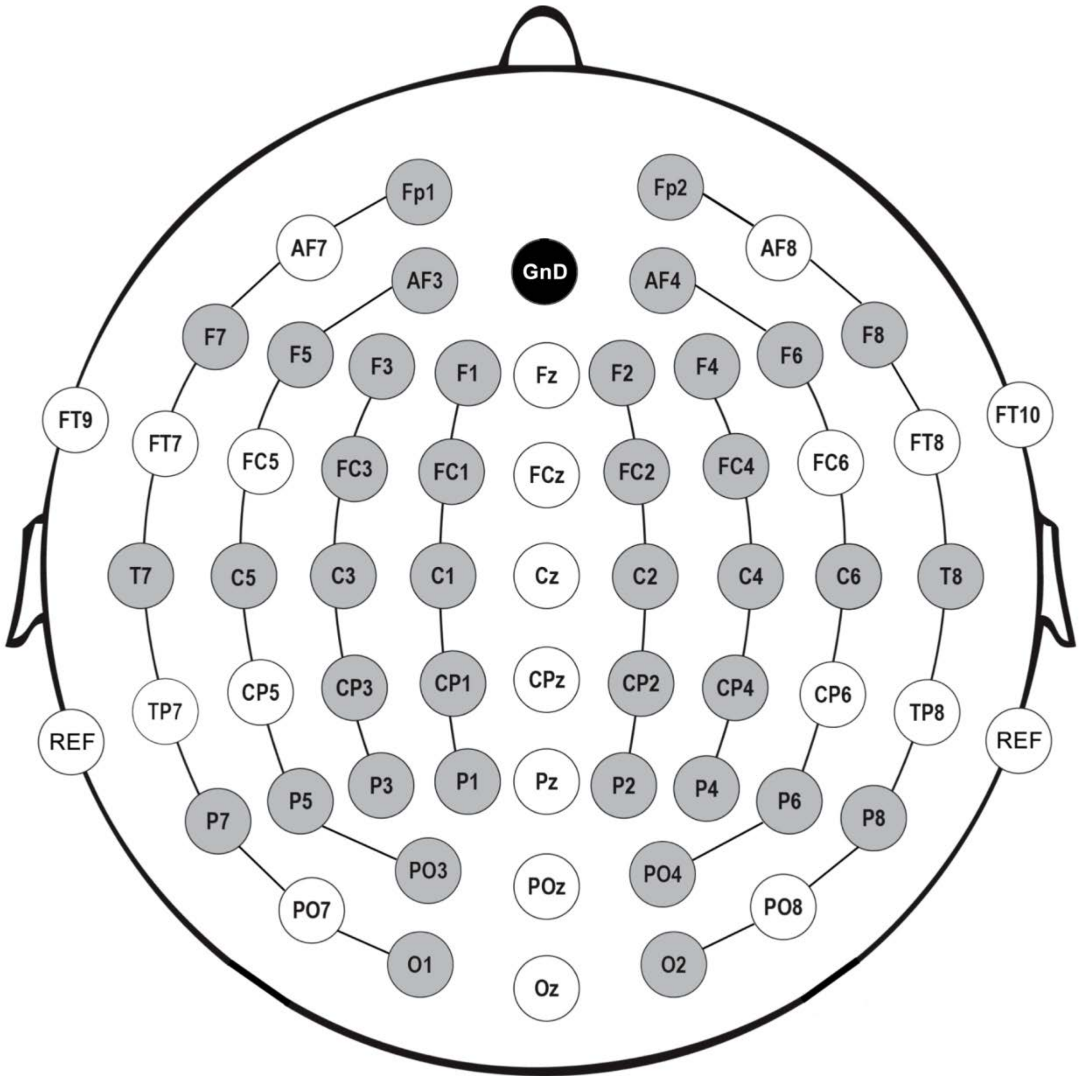


Figure 4: Schematic of electrode montage from which EEG activity was recorded. Highlighted electrodes used during electrode analyses.

Analyses therefore included within-participants factors of Accent Condition (with four levels; Plymouth – same speaker, Plymouth – different speaker, Regional, and Foreign; in each case the accent named is that of the first presentation of the sentence, as the second presentation was always in a Plymouth accent), Electrode line (with four levels), Electrode Anterior/Posterior position (with 5 levels), Hemisphere (left and right), and Repetition (first vs second). Only significant main effects and interactions ($p < .05$) of Condition and Repetition will be reported, adjusted using the Greenhouse-Geisser (1959) correction.

Results

A visual inspection revealed a positive component peaking at 200 ms after word onset, followed by a negative component peaking at 300 ms after word onset. This was followed by a broad negative component which continued until around 600 ms after word onset. Two epochs were therefore investigated; an early epoch between 250-350 ms, coinciding with the PMN and therefore indexing phonological processing and lexical selection, and a later epoch, between 350-600 ms after word onset, which incorporates the N400 and is thought to index post-lexical processing. These two epochs closely correspond to the two epochs investigated in the previous chapter, and will again be referred to as the PMN and N400, while recognising that they may include influences from other components.

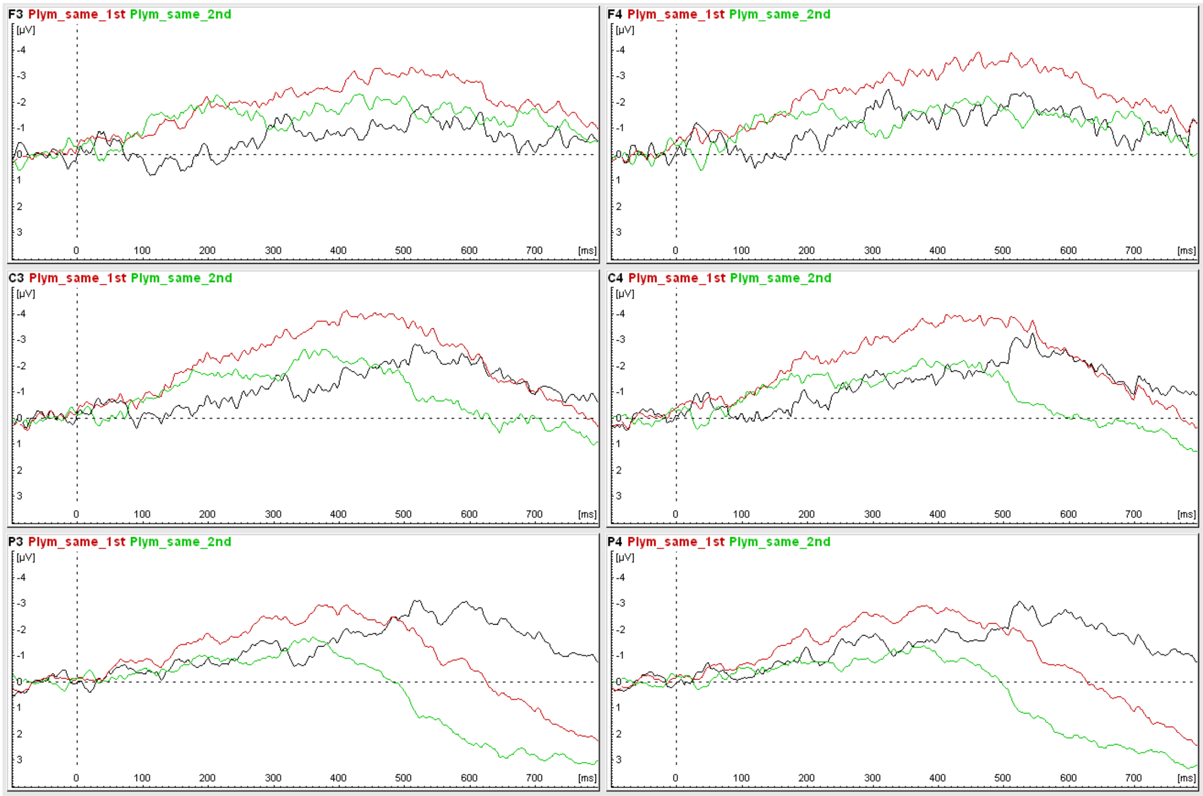


Figure 5: Averaged waveforms and difference wave (black) for first (red) and second (green) presentations of sentences in Plymouth – same speaker condition, across six electrodes (F3, F4, C3, C4, P3 and P4).

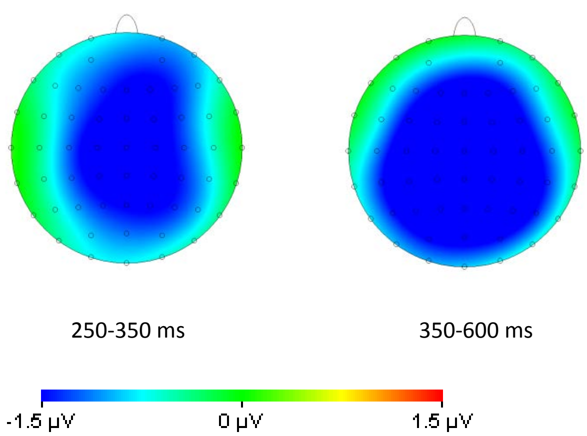


Figure 6: ERP Headmaps showing differences between first and second repetitions in the Plymouth – same speaker condition, between 250-350 ms (left) and 350-600 ms (right).

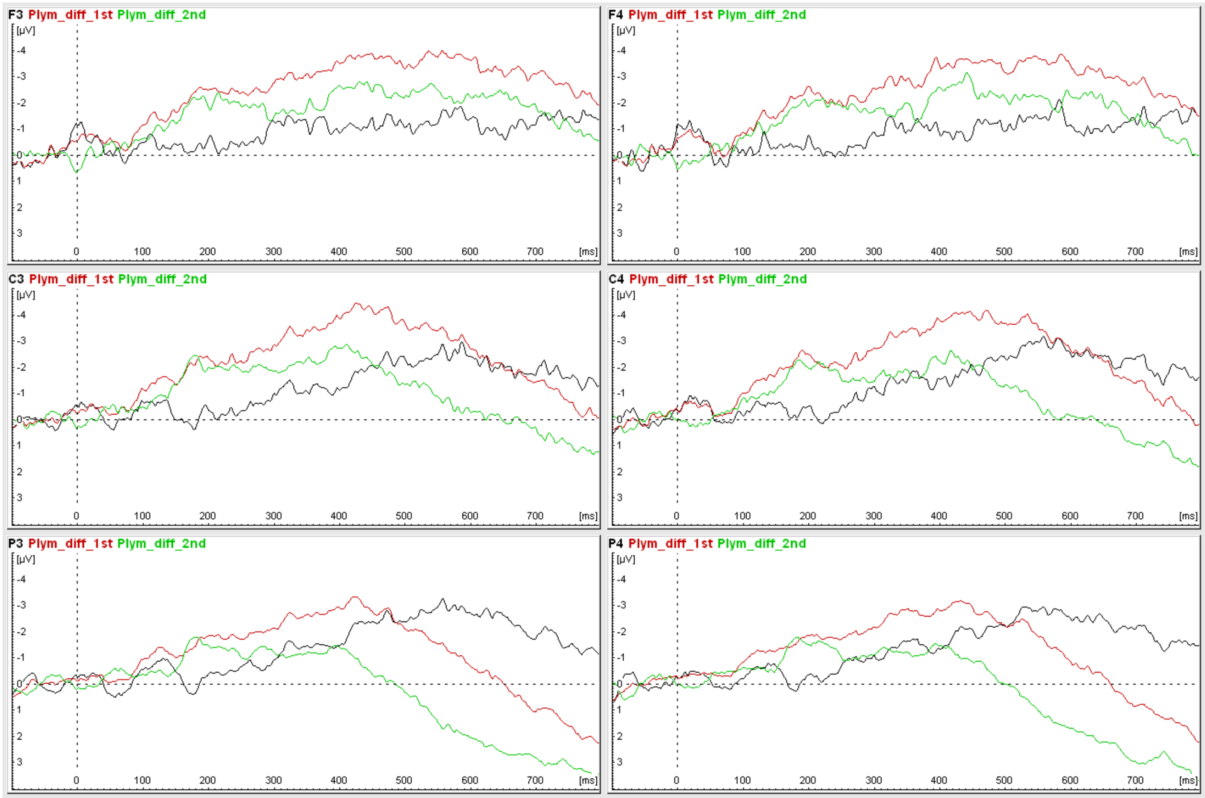


Figure 7: Averaged waveforms and difference wave (black) for first (red) and second (green) presentations of sentences in Plymouth – different speaker condition, across six electrodes (F3, F4, C3, C4, P3 and P4).

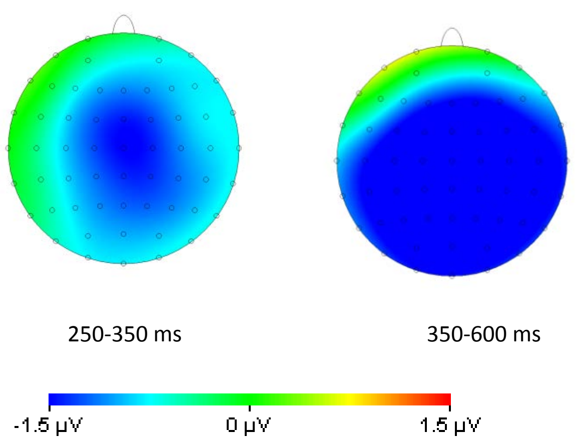


Figure 8: ERP Headmaps showing differences between first and second repetitions in the Plymouth – different speaker condition, between 250-350 ms (left) and 350-600 ms (right).

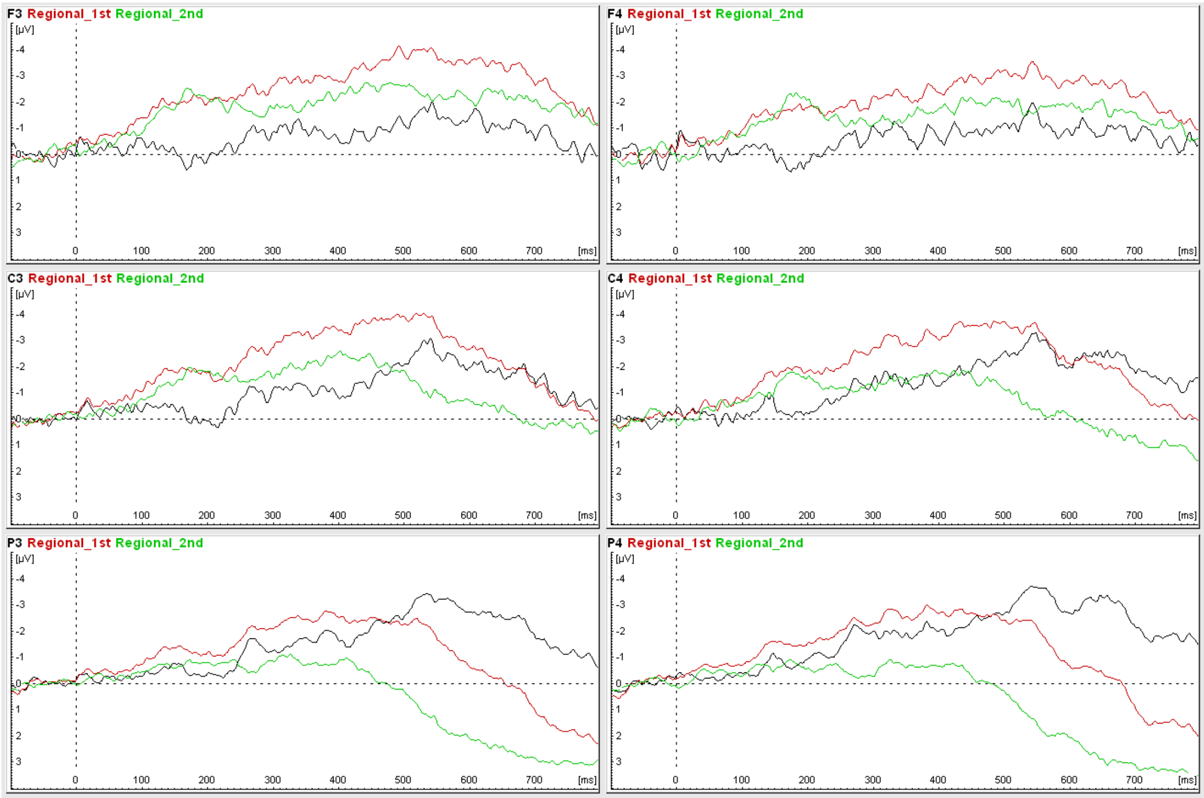


Figure 9: Averaged waveforms and difference wave (black) for first (red) and second (green) presentations of sentences in the Plymouth-Regional accent condition, across six electrodes (F3, F4, C3, C4, P3 and P4).

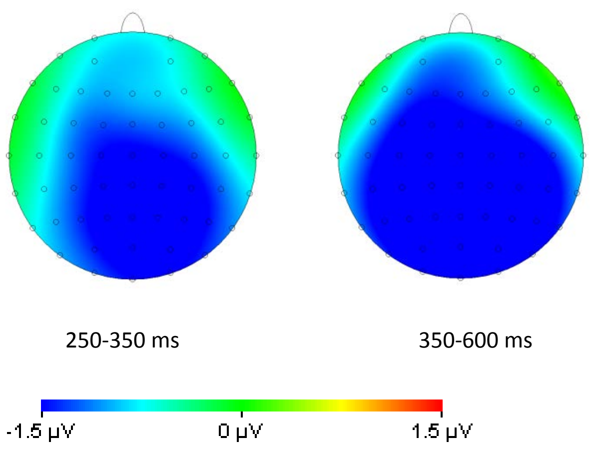


Figure 10: ERP Headmaps showing differences between first and second repetitions in the Plymouth-Regional accent condition, between 150-250 ms (left), 250-350 ms (centre) and 350-600 ms (right).

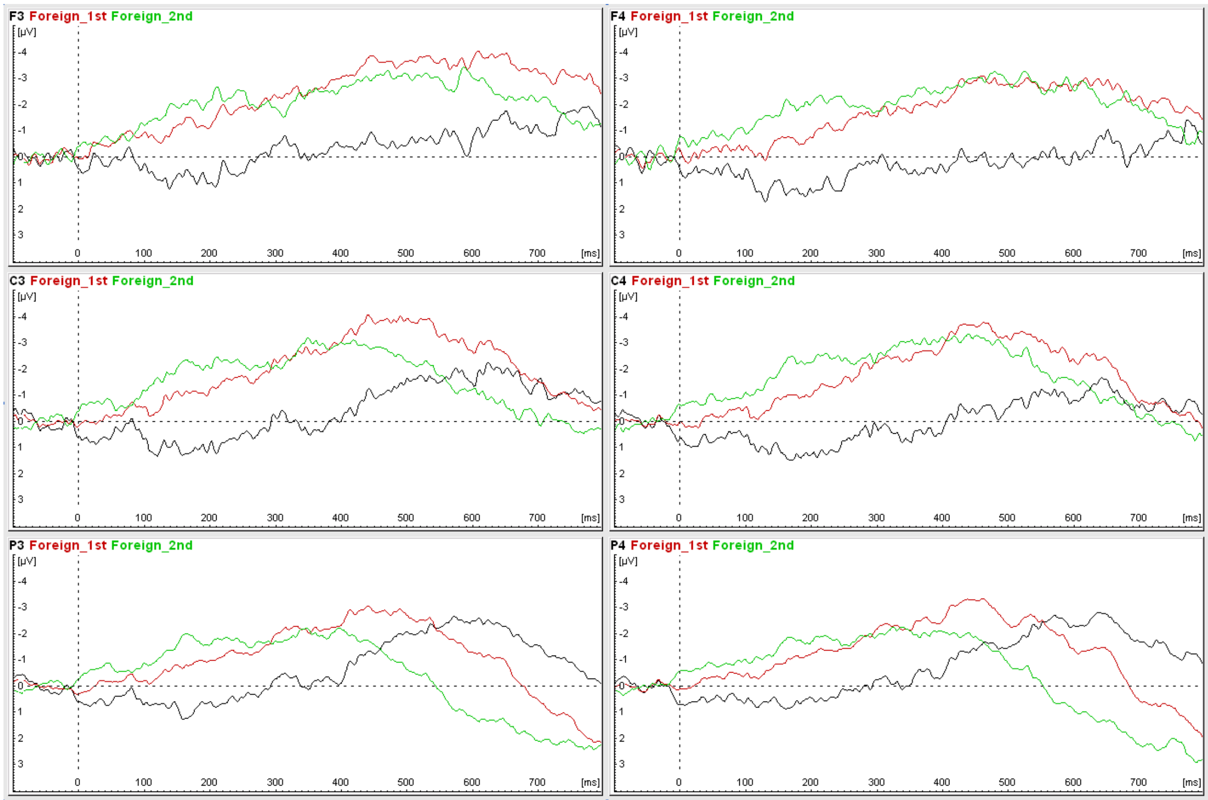


Figure 11: Averaged waveforms and difference wave (black) for first (red) and second (green) presentations of sentences in the Plymouth-Foreign accent condition, across six electrodes (F3, F4, C3, C4, P3 and P4).

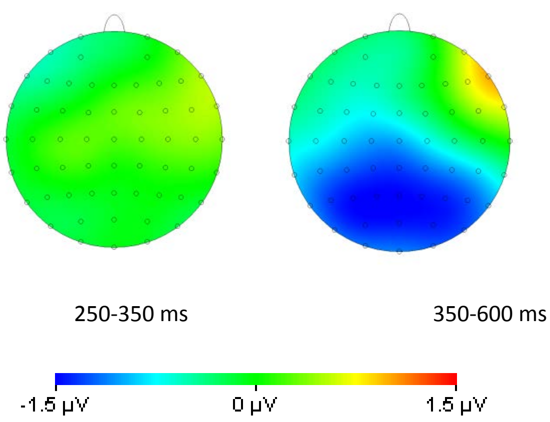


Figure 12: ERP Headmaps showing differences between first and second repetitions in the Plymouth-Foreign accent condition, between 250-350 ms (left) and 350-600 ms (right).

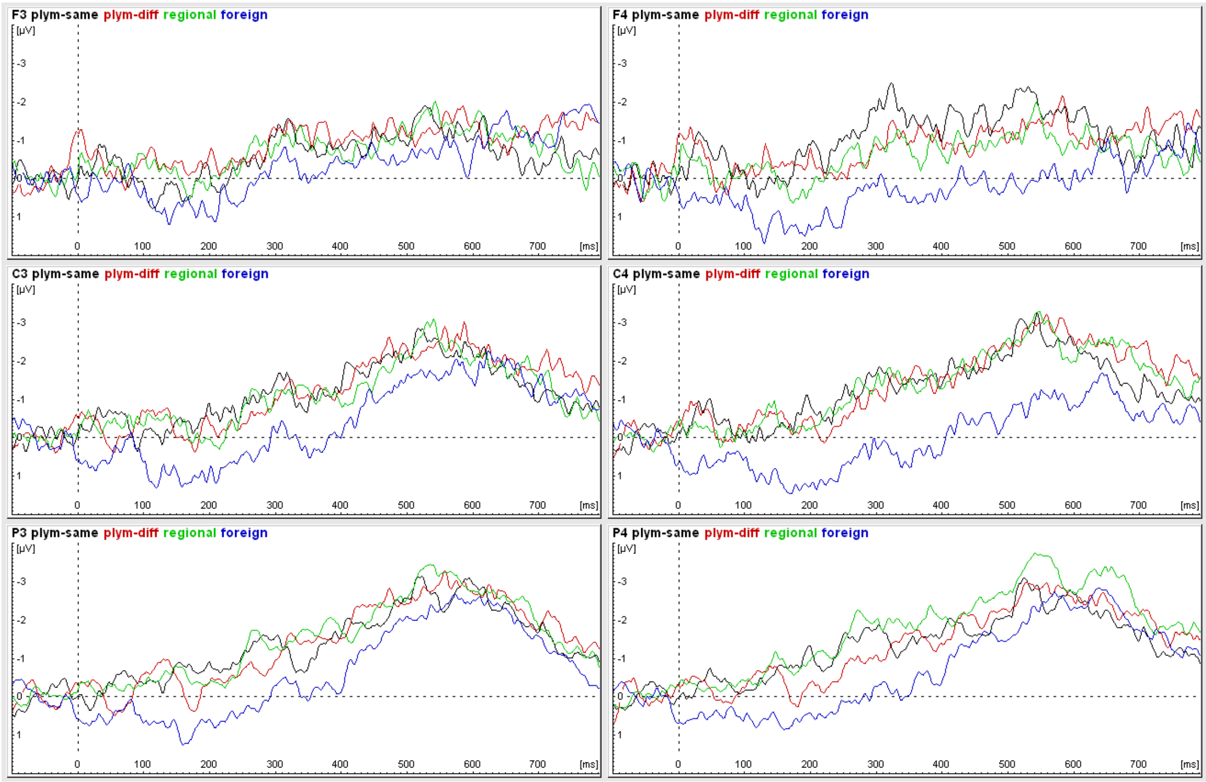


Figure 13: Difference waves (first repetition – second repetition) for four conditions, across six electrodes (F3, F4, C3, C4, P3, P4).

PMN – 250-350 ms

In this time period, there was no significant main effect of Condition, but there was a significant effect of Repetition (F (1, 36) = 14.24, p = .001, η^2 = .28), and an interaction between Condition and Repetition (F (3, 108) = 3.22, p = .029, η^2 = .08). There was also a three-way interaction between Condition, Repetition, and Line (F (9, 324) = 6.97, p < .001, η^2 = .16), as well as interactions between Repetition and Line (F (3, 108) = 17.1, p < .001, η^2 = .32) and between Repetition and posterior/anterior position (F (4, 144) = 3.43, p = .045, η^2 = .09), and Repetition, Line, and Position (F (12, 432) = 3.24, p = .008, η^2 = .08).

Pairwise comparisons were carried out between each pair of conditions.

When comparing the Plymouth –same speaker and Plymouth – different speaker conditions, no effects of or interactions with Condition were found, but a main effect of Repetition was found ($F (1, 36) = 15.8$, $p < .001$, $\eta^2 = .31$), with less negative responses in the second repetition than in the first. As in the previous epoch, the Plymouth – different speaker condition was used in pairwise comparisons with the two other conditions.

No significant effects of or interactions with Condition were found when comparing the Plymouth – different speaker and Regional conditions, but again, a Repetition main effect was found ($F (1, 36) = 16.17$, $p = .001$, $\eta^2 = .28$); again, responses to the second repetition less negative than those to the first. There was also an interaction between Repetition and Line ($F (1, 36) = 16.27$, $p < .001$, $\eta^2 = .31$). A main effect of Repetition was found in all four lines (L1 $F (1, 36) = 8.13$, $p = .007$, $\eta^2 = .18$; L2 $F (1, 36) = 10.55$, $p = .003$, $\eta^2 = .23$; L3 $F (1, 36) = 15.8$, $p < .001$, $\eta^2 = .31$; L4 $F (1, 36) = 17.74$, $p < .001$, $\eta^2 = .33$), and a significant effect of Line was found in both the first ($F (3, 108) = 47.8$, $p < .001$, $\eta^2 = .57$) and second ($F (3, 108) = 13.31$, $p < .001$, $\eta^2 = .27$) repetitions, so this interaction is likely to be due to differences in the amplitudes of first and second repetition responses in the four lines. The scalp maps in figures 8 and 10 above suggest that repetition effects are concentrated around the midline, and the effect sizes confirm that the differences are greater in L3 and L4, nearest the midline, and smaller in L1 and L2. In all lines, the second repetition elicited less negative responses than the first.

When comparing the Foreign accent condition to the Plymouth – different speaker condition, a Condition by Repetition interaction was found ($F (1, 36) = 5.22$, $p = .028$, $\eta^2 = .13$); this was due to a significant main effect of Repetition in the Plymouth – different speaker condition ($F (1, 36) = 6.6$, $p = .014$, $\eta^2 = .16$) but not in the Foreign accent condition ($F < 1$). There was also a Condition by Repetition by Line interaction ($F (3, 108) = 8.48$, $p = .002$, $\eta^2 = .19$). This is due to a Repetition by Line interaction in the Plymouth – different speaker condition ($F (1, 36) = 9.09$, $p = .002$, $\eta^2 = .20$), which was absent in the

Foreign condition ($F < 1$). A line by line analysis revealed that in the Regional condition, there was a main effect of Repetition in L2 ($F(1, 36) = 4.34$, $p = .044$, $\eta^2 = .11$), L3 ($F(1, 36) = 8.35$, $p = .007$, $\eta^2 = .19$) and L4 ($F(1, 36) = 9.13$, $p = .005$, $\eta^2 = .20$), but not in L1 ($F(1, 36) = 2.73$, $p = .107$, $\eta^2 = .07$).

When comparing the Regional and Foreign accent conditions, there was also a Condition by Repetition interaction ($F(1, 36) = 5.48$, $p = .025$, $\eta^2 = .13$), due to a main effect of Repetition in the Regional accent condition ($F(1, 36) = 10.76$, $p = .002$, $\eta^2 = .23$) but not in the Foreign accent condition ($F < 1$). A Condition by Repetition by Line interaction was also found ($F(3, 108) = 10.1$, $p = .001$, $\eta^2 = .22$), and was due to an interaction between Repetition and Line in the Regional condition ($F(3, 108) = 13.04$, $p < .001$, $\eta^2 = .27$) but not in the Foreign condition ($F < 1$). In the Regional condition, there was a significant main effect of Repetition in all four lines (L1 $F(1, 36) = 5.79$, $p = .021$, $\eta^2 = .14$; L2 $F(1, 36) = 7.7$, $p = .009$, $\eta^2 = .18$; L3 $F(1, 36) = 12.95$, $p = .001$, $\eta^2 = .27$; L4 $F(1, 36) = 14.03$, $p = .001$, $\eta^2 = .28$). The scalp maps in figures 10 and 12 indicate that this effect may be somewhat reduced in L1 and L2, furthest away from the midline, and where the smallest effect sizes are found, compared to L3 and L4, nearest the midline. In all cases, the second repetition elicited less negative responses than the first.

Late epoch – 350-600 ms

In the late epoch, there was a significant effect of Condition ($F(1, 36) = 2.86$, $p = .048$, $\eta^2 = .07$) and a significant effect of Repetition ($F(1, 36) = 32.74$, $p < .001$, $\eta^2 = .48$). There was also an interaction between Condition, Repetition and Line ($F(9, 324) = 5.16$, $p = .001$, $\eta^2 = .13$), and between Condition, Line and anterior/posterior position ($F(36, 1296) = 4.97$, $p < .001$, $\eta^2 = .05$). There was also a four-way interaction between Condition, Repetition, anterior/posterior position and Hemisphere ($F(12, 432) = 2.38$, $p = .027$, $\eta^2 = .06$).

Pairwise comparisons were carried out between each pair of conditions.

When comparing the Plymouth – same speaker and Plymouth – different speaker conditions, there was no main effect of, nor any interactions with, Condition, but there was a main effect of Repetition ($F (1, 36) = 35.41$, $p < .001$, $\eta^2 = .5$), with second repetitions eliciting less negative responses than the first repetitions.

When comparing the Plymouth – different speaker and Regional accent conditions, no effects of or interactions with Condition were found. There was a significant main effect of Repetition ($F (1, 36) = 33.19$, $p < .001$, $\eta^2 = .48$), with responses to the second repetitions being less negative than those to the first repetitions.

When comparing the Plymouth – different speaker and Foreign conditions, there was no main effect of Condition, but there was a main effect of Repetition ($F (1, 36) = 13.6$, $p = .001$, $\eta^2 = .27$), with responses to the second repetition being less negative than those to the first repetition. There was also a Condition by Repetition interaction $F (1, 36) = 4.89$, $p = 033$, $\eta^2 = .12$), which was due to a main effect of Repetition in the Plymouth – different condition ($F (1, 36) = 21.07$, $p < .001$, $\eta^2 = .37$), with the second repetitions eliciting less negative responses than the first, but not in the Foreign condition ($F (1, 36) = 3.49$, $p = .07$, $\eta^2 = .09$). There was also a three-way interaction between Condition, Repetition, and Line ($F (3, 108) = 6.38$, $p = .008$, $\eta^2 = .15$). This was due to a significant interaction between Repetition and Line in the Plymouth – different speaker condition ($F (3, 108) = 18.2$, $p < .001$, $\eta^2 = .37$) but not in the Foreign condition ($F (3, 108) = 1.08$, $p = .32$, $\eta^2 = .03$). In the Plymouth condition, there was a significant main effect of Repetition in all four lines (L1 $F (1, 36) = 6.39$, $p = .016$, $\eta^2 = .15$; L2 $F (1, 36) = 15.22$, $p < .001$, $\eta^2 = .3$; L3 $F (1, 36) =28.38$, $p < .001$, $\eta^2 = .44$; L4 $F (1, 36) = 24.46$, $p < .001$, $\eta^2 = .41$), and no further effects were found, so the Repetition by Line interaction in this condition is most likely due to differences in the amplitudes between the four lines. As indicated by the scalp maps in figures 8 and 10 above and by the effect sizes, the repetition effect is smaller in L1 than in the other lines, although in all cases, the second repetition elicited less negative responses than the first repetition.

When comparing the Regional and Foreign conditions, there was a significant effect of Condition (F (1, 36) = 7.59, p = .009, η^2 = .17), with the Foreign condition eliciting more negative responses than the Regional condition. There was also a Condition by Repetition by Line interaction (F (3, 108) = 7.67, p = .002, η^2 = .09). This was due to a significant interaction between Repetition and Line in the Regional condition (F (3, 108) = 23.6, p < .001, η^2 = .4) but not in the Foreign condition (F (3, 108) = 1.08, p = .32, η^2 = .03). In the Regional condition, there was a significant main effect of Repetition in all four lines (L1 F (1, 36) = 7.53, p = .009, η^2 = .17; L2 F (1, 36) = 19.4, p < .001, η^2 = .35; L3 F (1, 36) = 29.9, p < .001, η^2 = .45; L4 F (1, 36) = 34.69, p < .001, η^2 = .49), but no other effects were found, so the interaction in this condition is likely to be due to differences in the amplitudes of the responses in the four lines; as indicated by the effect sizes and seen on the scalp maps in figures 10 and 12 above, the repetition effect seems to be less pronounced in L1, furthest away from the midline, than in the other lines, but in all lines, the second repetition elicited less negative responses than the first.

Discussion

This study aimed to differentiate between the Exemplar-based approach to lexical storage and the Abstract entry approach, by eliciting repetition suppression to familiar speech and accented speech. Both approaches would predict post-lexical repetition suppression, but only the Abstract entry would predict repetition suppression in the PMN when the first iteration of a stimulus is spoken in an unfamiliar accent and the second iteration is spoken in a familiar accent.

In the early epoch (250-350 ms after word onset, which is likely to include the PMN), repetition-suppression was found in both the Plymouth-accented conditions and the Regional accent condition, but not in the Foreign accent condition. Since this stage of processing is associated with phonological processing, this finding suggests that the phonological processes recruited when listening to and normalising foreign-accented speech are different from those used with both familiar and regional accents. In addition, this stage of processing has been associated with lexical selection (Hagoort & Brown, 2000); that is, during this epoch, a match is being made between the acoustic input and the most appropriate lexical candidate. According to the Exemplar-based approach, the lexical entry accessed in the regional condition should be different from that accessed in the familiar-speech conditions, as variant forms should be stored separately. If this were the case, we would expect to find no repetition suppression in the regional accent condition, as the lexical candidate selected in the first repetition, in a regional accent, should be different from that chosen during the second repetition, in the Plymouth accent. The fact that there is suppression in this period in the regional condition, and that, in fact, the regional condition is indistinguishable from the two Plymouth-accented conditions in this epoch, is not compatible with this approach. However, the Abstract entry approach suggests that prior to lexical selection, variant forms are normalised to allow lexical access of a single canonical form, so that the same candidate is selected regardless of whether the stimulus is heard in a familiar accent or a regional one. In other words, according to the Abstract entry approach, we would expect to find repetition suppression in this epoch in the regional condition. The findings of repetition suppression in the early epoch in both the Plymouth-accented conditions and the regional accent condition, but not in the Foreign accent condition, therefore support the Abstract entry approach, as well as the Different Processes hypothesis.

The late epoch (between 350-600 ms after word onset) is likely to index lexical integration and post-lexical processing. In this stage, it was predicted that repetition suppression would be found in all conditions, since by this stage processing is dealing with semantic context rather than acoustic or phonological variation. The previous chapter indicates that in this stage, there are still persistent differences between the processing of foreign accents when compared to either familiar or regional accents, but the findings from the ratings carried out on the stimuli used in this experiment suggest that even in the foreign accent, target words are accurately identified 73% of the time. Thus the foreign-accented words in the first repetition should act as a prime for the Plymouth-accented words in the second repetition, although this priming effect (which should result in repetition suppression) should be less strong in the foreign condition than in the regional and familiar conditions. In fact, repetition suppression was found in this epoch in the two Plymouth conditions and the Regional condition, but did not reach significance in the Foreign accent condition. This finding is compatible with the idea, presented in the previous chapter, that in this stage, foreign accents require greater reliance on post-lexical top-down information than regional accents, because regional variations have been successfully normalised pre-lexically, leading to accurate lexical selection, whereas foreign accents recruit different normalisation processes, leading to incomplete normalisation.

Chapter 5: Discussion of Adult work

The experiments reported in the two previous chapters set out to investigate the processing of speech in familiar, regional, and foreign accents, using ERP measures. The primary aim was to find evidence which would support one of two hypotheses regarding the processing of regional and foreign accented speech; either the Perceptual Distance hypothesis, which states that accents can be arrayed along a continuum of perceptual difference from the listener's own accent, and that their processing depends on the degree of difference (Clarke & Garrett, 2004), or the Different Processes hypothesis, which suggests that foreign accents recruit different mechanisms for normalisation, prior to lexical access (Chambers, 2002; Floccia, et al., 2009b; Floccia, et al., 2006). A secondary aim, addressed in chapter 4, was to find support for one of two approaches to the storage and lexical access of variant word-forms (including accent-based variants). The Exemplar-based approach states that previously-encountered variant forms are individually lexically represented, and thus do not require a normalisation process in order to achieve lexical access (Johnson, 1997), whereas the Abstract Entry approach suggests that only a single canonical or abstract form is lexically represented, and variant forms undergo a process of normalisation prior to lexical access (Pallier, et al., 2001).

Experiment 1 directly compared the online processing of accented speech, by recording ERP data from participants listening to stimuli spoken in a familiar or home accent, one of two regional accents, and one of two foreign accents. Analysis of their neurophysiological responses provided strong support for the Different Processes hypothesis, by showing that responses to foreign-accented speech differed from those to familiar speech and to regional-accented speech in the PMN epoch, which indexes phonological processing and lexical selection (Connolly & Phillips, 1994; Desroches, et al., 2009; Newman & Connolly, 2009), and that foreign and regional accents elicited activity that differed from familiar accents not only in amplitude but in polarity. In a later, post-lexical period, responses to home-accented and regional-accented speech did not differ from each other, but the responses to the foreign accent still differed from both home and regional accents. One interpretation is that regional accents elicit a normalisation process which allows for early, successful lexical access, while foreign accents elicit a different set of processes for normalisation of the acoustic input, but also rely more heavily on top-down information in the post-lexical period. These findings are not compatible with the Perceptual Distance hypothesis, which would predict that regional and foreign accents recruit the same normalisation mechanisms, and that the degree to which these processes are required depends on the degree to which the speech perceptually differs from the home accent (Clarke & Garrett, 2004).

A third possible explanation is that of processing cost (Cristià, et al., In preparation). Cristia et al. propose that, just as listening to many speakers incurs an added processing cost over listening to just one consistent speaker (Nusbaum & Morin, 1992), so listening to an unfamiliar accent may incur an added processing cost over listening to a familiar accent. Indeed, Nygaard and Pisoni (1998) suggest that the variability constituting an accent is, in effect, simply a more extreme version of normal inter-talker variability. Differences in processing, then, might reflect the cost incurred by a given exemplar, with foreign accents typically incurring a higher cost than regional accents, and accents which are more similar to the listener's own accent incurring a lower cost an those which are more perceptually distant, even if they are within the same accent category. This effect would be mediated by familiarity (Adank & McQueen, 2007; Floccia, et al., 2006; Munro & Derwing, 1995) and intelligibility (Bradlow & Bent, 2008). This hypothesis incorporates the flexibility of the Perceptual Distance hypothesis, in that accents which differ more from the listener's own accent incur a greater cost to process, but it also allows for the recruitment of different processes for lexical access. In favour of this hypothesis, Burki-Cohen, Miller and Eimas (2001) presented listeners with a phoneme-detection task and a lexical task, with familiar speech and foreign-accented speech. They found that the phoneme detection task was performed for both accents using pre-lexical (acoustic) information when it was presented alone, but when presented with the lexical task, post-lexical information was used. However, when the stimuli were obscured by the addition of noise, increasing the difficulty of the task, listeners used post-lexical information to perform the phoneme detection task in the foreign accent even when it was presented without the lexical task. This suggests that when a high processing cost is incurred, such as when listening to accented speech, it may cause listeners to rely more heavily on post-lexical information, as is indicated by the findings in chapter 3. The processing cost hypothesis, then, may be a way of bringing together aspects of the Different Processes and the Perceptual Distance hypotheses.

Experiment 2 used paired sentences, the second of which was always spoken in a home accent while the first could be spoken in a home, regional, or foreign accent, to examine the way in which accented speech primes subsequent familiar speech, in a repetition suppression design similar to that used by Orfanidou, Marslen-Wilson and Davis (2006). As in experiment 1, foreign accents were found to elicit a different pattern of activation from either home-accented or regional-accented speech, lending further support to the Different Processes hypothesis. In addition, the fact that foreign accents failed to elicit repetition suppression during the early epoch is compatible with the Abstract entry approach (Pallier, et al., 2001), and is more difficult to account for by the Exemplar-based approach (Johnson, 1997).

Taken to the extreme, the Abstract entry approach and the Exemplar-based approach may seem entirely incompatible. An extreme version of the Abstract entry approach would suggest that all speaker-specific information should be normalised out of the speech signal pre-lexically, and then discarded as irrelevant, while an extreme version of the Exemplar-based approach suggests that speaker-specific information is stored at the lexical level, and is essential for lexical access. However, as discussed by Cutler, Eisner, McQueen and Norris (2010), neither of these extreme accounts is fully supported by the existing evidence. A number of studies have shown that speaker-specific information can facilitate word processing (Assmann, Nearey, & Hogan, 1982; Mullennix & Pisoni, 1990; Mullennix, et al., 1989;

Nusbaum & Morin, 1992; Nygaard & Pisoni, 1998), which is not compatible with the idea that indexical detail is discarded after normalisation. On the other hand, there is also evidence that phonemic boundaries can be moved through training, and that these phonemic shifts can then be generalised to novel words (Eisner & McQueen, 2005; Kraljic & Samuel, 2005; Norris, et al., 2003). Purely episodic or Exemplar-based models cannot explain or replicate this generalisation (Cutler, et al., 2010). Cutler et al. therefore propose that indexical information is indeed stored, but not at the lexical level. Rather, it is represented in long-term memory, and is used to facilitate word processing, including pre-lexical normalisation, when appropriate.

Overall, then, these two studies indicate that variation in the acoustic signal resulting from foreign and regional accents recruits different types of normalisation processes, dependant not just on the degree of perceptual difference from the home accent but on the type of accent, in order to achieve lexical access. This normalisation process begins during the auditory/acoustic stage of processing (as shown by the findings within the early 150-250 msec window in experiment 2), and, in the case of regional accents, is completed within the phonological stage (as shown by the results in the PMN window in both experiment 1 and 2). In foreign accents, it is not completed in this phonological phase, but continues into the post-lexical phase, resulting, in experiment 1, in differences between the processing of foreign accents from that of familiar and regional accents between 350-600 msec and in experiment 2, in reduced suppression in the foreign-accent condition compared to the other conditions.

While these studies do provide strong support for the use of different normalisation processes for regional and foreign accents, they cannot entirely rule out the possibility that, within accent categories, some accents may incur a higher processing cost than others, as suggested by Cristia et al. (In preparation).

Although Experiment 2 found no differences in the suppression caused by home and regional accents in the PMN epoch, Experiment 1 did show a difference in the online processing of the home and regional conditions at this stage. This indicates that while the regional accents resulted in successful lexical access, the processing required to do so was not identical to that used for more familiar accents. Cristia et al. suggest that unfamiliar regional accents incur higher processing costs than familiar speech, due to their unfamiliarity or to their acoustic distance from the listener's own accent, and the findings of Experiment 1 could reflect this processing cost. According to this hypothesis, we would expect accents which are more similar to the listener's own accent to incur lower costs than those which, while falling within the same category of accents (regional or familiar), differ more. This is likely to be reflected in ERP data, with higher-cost accents eliciting greater deviations from the home accent those with a lower processing cost. Thus, we might expect these deviations to be more marked in response to a strong Glaswegian accent than in response to a South-East English accent, using a South-West accent as the baseline. Similarly, experiment 1 found a reduced PMN in foreign accents as compared to the home accent, and we might expect this component to be reduced further for an accent judged to be very different from the home accent (such as the Cantonese accent used in experiment 2) than for an accent judged to be less different (such as a Dutch accent).

One possible future direction for research, then, would be to focus on a single category of accents (either regional or foreign) and use stimuli spoken in a number of accents which vary on their perceptual distance from the baseline accent. It might be predicted that accents judged as more different would elicit more extreme deviations from the baseline in the PMN epoch (but not in the later, post-lexical period) than those judged more similar. Another issue, touched on in the introductory chapter of this thesis, is that of adaptation to accented speech. While some studies (Bradlow & Bent, 2008; Clarke & Garrett, 2004) have found rapid adaptation to accented speech within only a few sentences, other studies suggest that the perturbations in performance caused by accents, and particularly foreign accents, are much more persistent (Floccia, et al., 2009b). One possible explanation for these contradictory findings is the use of different measures. Clarke and Garrett used a cross-modal word recognition paradigm, providing listeners with extra top-down cues, while Bradlow and Bent used a transcription task, which by its nature can only provide a measure of the degree to which a listener has understood the intended meaning of the sentences they heard, and not of the effort required to do so. Floccia et al. (2009b) used a lexical decision task, which enabled them to measure not simply the intelligibility of the stimuli but also their comprehensibility, without adding additional top-down cues which might simplify the task. ERP studies would clearly provide a way of examining adaptation to variant speech in the absence of task-related demands.

Of particular relevance for this study, a previous fMRI study of activation to time-compressed speech has found signs of neural adaptation over the course of four blocks of stimuli (Adank & Devlin, 2010), suggesting that we might expect to find adaptation to accented speech using ERP. In addition, the fine temporal resolution of ERPs would allow us to investigate adaptation at different stages of processing, in a way which neither purely behavioural nor MRI studies can.

Behavioural studies which measure intelligibility have tended to show a high level of adaptation (Bradlow & Bent, 2008; Clarke & Garrett, 2004), suggesting that after adaptation, listeners are able to achieve lexical access with accented stimuli. This should be reflected in the post-lexical period when looking at ERP data, leading to a prediction that in the late epoch, incorporating the N400, neurophysiological responses should show rapid adaptation to unfamiliar accents. However, comprehensibility measures such as those used by Floccia et al. (2009b) show that response times to foreign-accented stimuli remain slower than those to regional or home accents, suggesting a higher processing cost or the use of different normalisation processes. We would expect to see this reflected in the PMN epoch in ERP data, so that deviations in this time-window would be expected to adapt more slowly and less completely. Thus, a finding of slow and/or incomplete adaptation between 250-350 msec after word onset, with a quicker, more complete return to baseline after 350 msec, in response to foreign-accented stimuli as compared to familiar speech, would go some way to reconciling the apparently contradictory findings in the area of adaptation to accents.

We would also expect to find more rapid and more complete adaptation to regional than to foreign accents, although this effect might also be affected by the degree of perceptual similarity to the baseline accent. Thus we might expect quicker and more complete adaptation to a Dutch accent than to a Cantonese accent, and to a South-East English accent than to a Glaswegian accent, when using a South-West English accent.

The findings might also be affected by the design used. The two experiments described in chapters 2 and 3 were not intended to address the question of adaptation, and therefore the sentences, and the accents in which they were presented, appeared in a pseudo-randomised order. We might expect much greater or more rapid neural adaptation when using a block design, in which the same unfamiliar accent is heard consistently across a number of sentences, than in a pseudo-randomised design, in which the accent heard changes frequently. This would allow adaptation to be studied at two levels. Firstly, by comparing the responses to sentences at the beginning of each block with responses to sentences at the end of each block, rapid online adaptation could be tested; Clarke and Garrett (2004) reported that adaptation sometimes required as few as two sentences to take place, although Floccia et al. (2009b) did not find any reduction in response times across a four-sentence block in their regional and foreign accent conditions. Secondly, by comparing early blocks with later blocks within an accent, it would be possible to examine slower and/or longer-lasting adaptation. Clarke and Garrett found a reduction in response times across three blocks, suggesting that adaptation was on-going. However, Floccia et al. found only marginal adaptation across blocks, and also found that a change in accent elicited an increase in response times. Thus we might expect reduced deviations from the home-accented baseline between the first and last sentences of each block, with a more gradual pattern of longer-term adaptation between early and late blocks.

A number of studies have found that exposure to variable input leads to richer representations than a less variable input; in particular, exposure to multiple speakers with an unfamiliar accent allows listeners to adapt to the accent better than exposure to only a single accented speaker (Bradlow & Bent, 2008; Kraljic & Samuel, 2007), and generalise that adaptation to a novel speaker of the same accent. Bradlow and Bent also looked at whether adaptation to one foreign accent generalises to another foreign accent. They exposed listeners to either one or several Chinese-accented or American-accented talkers in two training sessions, over two days. They then tested their ability to transcribe sentences spoken by a Slovakian-accented speaker, and found that while there was an effect of training, this effect did not differ between those trained with a familiar accent and those trained with a foreign accent. This suggests that the training effect was less to do with the accent itself, and more to do with task demands. However, only one accent was used for training. Just as exposure to multiple speakers allows listeners to form richer representations than those exposed to only one speaker, we might also expect exposure to multiple foreign accents to generalise better than just a single accent.

If different processes are indeed recruited for the normalisation of regional and foreign accents, we would not expect adaptation to regional accents to generalise to foreign accents, since different mechanisms are at work. However, within an accent category, if the same processes are used (as suggested by the lack of within-category differences in chapter 3), then we might expect that listeners exposed to speech in a number of different regional accents would perform better with a new regional accent, and that listeners exposed to multiple foreign accents would fare better with a novel foreign accent. The two accents used by Bradlow and Bent, Chinese and Slovakian, are likely to be quite dissimilar, since they come from entirely unrelated languages. Some degree of generalisation might therefore be expected if accents from within a single language family (such as French and Italian, or Slovakian and Polish) were used. Similarly, we might expect to find generalised adaptation when using regional accents which are similar to each other, such as Irish and Scottish accents in English, but not when using very distinct regional accents, such as Welsh and South African.

These two studies together have demonstrated for the first time that in adult listeners, regional and foreign accents recruit different mechanisms in the early, pre-lexical stages of word processing, and that these differences are not merely quantitative, but, since the polarity of the differences from the familiar baseline differs, that the processing of these two classes of accents is qualitatively different. This finding lends substantial support to the Different Processes hypothesis of accent processing, and is not consistent with Clarke and Garrett's proposal of a continuum of difference. In addition, Experiment 2 has provided strong support for the Abstract Entry theory of lexical representation over the Exemplar-based theory.

Chapter 6: Introduction to Infant work

Contrary to adults who are using a stable, robust and established speech perception and word recognition device, infants are actively engaged in the development of such a system from the incoming linguistic input. In the first year of life, children are going through the process of perceptual narrowing, a process which, when completed, allows them to attend preferentially to cues which are relevant to their native language, and to ignore variation which is not useful (Scott, Pascalis, & Nelson, 2007). In addition, infants in the first few months do not possess lexical, syntactic or morphological knowledge, which is used by adults in a top-down manner to inform word processing. Because infants lack this body of knowledge, they must rely almost entirely on bottom-up information, which can be extracted directly from the acoustic signal. This is particularly interesting when looking at the processing of variant speech, since, in adults, top-down information may be used to help compensate for unfamiliar forms of variation. Infants, on the other hand, cannot draw on this top-down processing, which means that infant studies allow us to look at low-level, auditory/acoustic processes of normalisation or compensation for variation, rather than the later stages (although , naturally, it is not clear that the processes used in childhood to learn language are the same than those used for processing in adulthood) .

While many psycholinguistics studies with adults focus on lexical access (that is, the ability to access the meanings of words from acoustic input), research on the earliest stages of language acquisition is concerned not with the meaning of language but with acoustic processing in the absence of lexical knowledge. Before an infant can associate words with meanings, they must first establish that human speech is unique among the sounds to which they are exposed (Vouloumanos & Werker, 2004), that there is some consistency both within and between speakers despite the natural variation in the speech signal (Jusczyk, Pisoni, & Mullennix, 1992), and that a speech stream can be broken up into individual linguistic units (such as phonemes, syllables or morae, and words) and arranged in language-specific patterns (Jusczyk, et al., 1999b). These early stages of learning, then, require attention to a number of intersecting cues, and a process of establishing a distinction between contrastive and non-contrastive variability, in order to build a consistent and yet flexible model of the native language which then allows associations between sound and meaning to be made. For this reason, the following studies will compare the processing of accented speech against the processing of familiar speech in infants, in order to further our understanding of the ways in which a still-developing system normalises variant speech without the benefit of top-down information.

Infants' processing of variability

As with adults, infants' speech perception can be disrupted by variability in the speech signal, such as that caused by an unfamiliar accent. Unlike adults, though, infants are not yet expert users of language. In their early years, they are still acquiring not just vocabulary or grammatical rules but also learning to master phonology; the inter-relationship between phonetics, phonemics and meaning. In this early stage, variability is not necessarily a disadvantage; in fact, a certain amount of variation in their input may be essential to their learning (Fennell & Werker, 2003; Rost & McMurray, 2009, 2010). Indeed, if a language system is extremely inflexible having experienced only perfectly homogenous input, any variant form would be treated as an entirely novel word. Clearly this would be an uneconomical system, and it does not match what we know of the human speech processing system; some variation in the speech signal is not only unavoidable, but perhaps actually desirable and even necessary. Rost and McMurray (2009) showed that infants at 14 months of age were better able to discriminate two newly-learned phonological neighbours when they heard the names presented by multiple speakers than when presented by a single speaker, indicating that inter-speaker variability allowed the infants to form more robust and thus discriminable representations of the words. Singh (2008) found similar results in 7.5 month old infants in a classic Head-turn Preference Procedure (HPP) study. Infants were familiarised with two target words, and then heard passages containing those target words, and passages containing matched distractor words. One of the target words was spoken with a consistent affective tone, while the other target word was presented in a range of different affective tones (happy, scared, angry, excited). The infants showed longer looking times to passages containing words which had been familiarised with variable affect than to passages containing words familiarised with a consistent affect, suggesting that they were better able to segment the passages and attend to the trained word when it had been encoded with a certain amount of variability. Again, this suggests that the variability in the input has allowed, even at this early age, more robust representations to be stored and accessed than when the input is less variable.

However, a system which is trained with very variable output is likely to form poorly-specified representations of words, which may therefore overlap with each other. This is likely to lead to a high proportion of false positives which, once again, will lead to slow vocabulary acquisition (as novel words may be mistaken for existing items rather than allowed their own identity). Understanding the effects of variability on the developing system therefore adds a separate layer of issues; we must question not only whether variability prevents us from matching input to an existing representation but how it affects the way in which we acquire new information. The use of regional and foreign accents gives us a useful source of relatively consistent inter-speaker variability with which to attempt to answer some of these questions.

Too much variation in the speech signal can disrupt processing and retention (and therefore learning), as shown by Jusczyk, Pisoni and Mullinex (1992). They habituated two-month-old infants to several exemplars of a syllable (for example, /bʌg/) spoken by either a single speaker, or multiple speakers. They then exposed them to a syllable with a single phonetic change (/dʌg/) , either immediately or after a two minute break. They found that if the phonetic shift occurred immediately, infants dishabituated whether they had heard a single speaker or multiple speakers, but when there was a two-minute break between habituation and the phonetic shift, only the single-speaker group dishabituated to the shift.

Houston and Jusczyk (2000) showed that at 7.5 months, infants were able to generalise trained word forms within speaker gender but not across; when trained with a female speaker, they attended longer to target word forms spoken by a different female speaker but not when the targets were produced by a male speaker, whereas when familiarised with a male speaker, they attended longer to a novel male speaker but not to a female speaker. A similar study showed that at the same age, infants attended longer to targets which matched the vocal affect they had heard in the training phase, but not to targets spoken with a different vocal affect (Singh, et al., 2004). These studies suggest that at this early age, variability in the speech signal can prevent robust phonetic representations from being stored and retained.

However, as the system gains experience and develops more robust, more flexible representations, it becomes able to compensate for a certain amount of variability and to normalise deviant signals (Schmale & Cristià, 2009; Singh, 2008; Singh, et al., 2004). Singh et al. (2004) used a head-turn preference procedure with infants at seven and a half months and ten months of age, familiarising them with words in either a happy or a neutral tone, and then testing them with passages containing the trained target words and untrained distractor words, in either the same affect or an unmatched affect. They found that the younger age group were able to demonstrate recognition only when the affect of the test passages matched that of the familiarisation phase, whereas by ten months, the infants were able to generalise across affect conditions. Building on this finding, Singh (2008) presented seven and a half month old infants with a familiarisation phase featuring variable affect; that is, they heard words spoken in a range of different affective tones, such as happy, sad, scared, and so on. When they were then tested with passages spoken in a single affect, they were able to recognise the trained target words and attend to them for longer than they attended to untrained distractor words. Singh concludes, as did Houston and Jusczyk (2000), that by this age, variable tokens of a stimulus enable infants to form more abstract and thus more robust representations of words, as compared to the less flexible representations formed by more uniform exemplars. This suggests that in the second half of the first year of life, infants become more able to form and store abstract representations of word forms, and to generalise them across variable inputs. This process seems to occur, or at least reach completion, at different ages for different sources of variation. This in itself suggests that not all variation is equal, but rather that different types of variation are treated differently at this early stage. This is likely to be driven in part by the degree of difference represented by different types of variability. White and Morgan (2008) demonstrated that 19-month-olds' looking times to a familiar object varied according to the pronunciation of the object's name, with longest looking times to correctly pronounced objects, and a reduction in looking times proportional with the number of feature changes in mispronunciations. Thus when "shoe" was mispronounced with a single feature change (as /fuː/), it elicited shorter looking times than the correct pronunciation, but longer than a mispronunciation with a two-feature change (/vuː/), while a three-feature change (/guː/) elicited even shorter looking times. This demonstrates that infants are sensitive to the degree of difference in pronunciation, and this might be expected to have an impact on their perception of accented speech, with accents which are more different from their ambient accent causing more difficulty in processing than accents similar to their own.

Infants' processing of regional- and foreign-accented speech

An individual accent may encompass a number of different types of variation, but it is also important to consider that different types of accents may themselves represent different types of variation. As discussed earlier with regard to adult studies, it is possible that all accents can be thought of as falling along a simple perceptual continuum of difference (Clarke & Garrett, 2004) from the familiar Note that in pre-verbal infants, we cannot consider the listener's own accent as the baseline, so must instead take the accent or accents to which the child is most commonly exposed in the home as the standard one. The studies already presented in this thesis, however, point to qualitative differences in the way regional and foreign accents are processed in the adult brain (Goslin, et al., 2012, and see also chapters 3 and 4 of this thesis). It may also be that in the infant system, different types of accents (such as regional versus foreign accents) are treated as qualitatively different.

However, infants are engaged in the process of creating phonetic and phonemic categories through perceptual narrowing, and so what holds true for adults may not hold true in infants. Perceptual narrowing occurs at different ages for different linguistic features (Scott, et al., 2007). In the earliest months, perceptions seem to be arrayed naturalistically, on a sliding scale so that infants are able to make fine distinctions between percepts which are not contrastive in their native tongue. As they gain experience, they learn to divide these sliding scales into language specific categories, which then inform their perception. Existing literature demonstrates that infants are able to differentiate some prosodic qualities of speech from birth. For example, English babies show a preference for stress-timed languages (Nazzi, Bertoncini, & Mehler, 1998) over syllable- or mora-timed languages. They quickly refine their discrimination so that they are able to show a preference for their own language over other languages (even those which share timing patterns and other prosodic qualities) at four months (Bosch & Sebastián-Gallés, 1997). Continuing this trend should allow them to discriminate speech in a familiar accent from an unfamiliar one even before they are extracting meaning from what they hear, and indeed there is some evidence to suggest that such discrimination arises after three and before six months (Butler, Floccia, Goslin, & Panneton, 2011; Kitamura, Panneton, Deihl, & Notley, 2006a). Butler et al. showed that five-month-old British infants from the South-West of England were able to distinguish between a South-West accent and a Welsh accent, but not between two unfamiliar regional accents, while Kitamura et al. showed a preference for familiar Australian speech over unfamiliar

American speech in three-month-old infants from Australia. This preference for familiar speech persists through infancy and into childhood (Kinzler, Shutts, DeJesus, & Spelke, 2009), and may be driven by the cognitive difficulty of processing accented speech; infants may choose to attend to speech which is easier to comprehend, and not to speech which is difficult to process. However, it is also possible that a preference for familiar speech causes infants to attend more, and therefore allows them to process the stimuli more successfully, while a relative dislike of unfamiliar accented speech reduces their attention, causing them to be less successful in processing it.

In summary, the literature shows that infants are able to successfully normalise some forms of variation, such as speaker gender (Houston & Jusczyk, 2000), and affect (Singh, et al., 2004), but since variant accents may act as carriers for multiple types of variation, it may take the infant system longer to get to grips with unfamiliar accents. For example, a foreign accent such as that of a native French speaker, speaking in English, may introduce sounds not used in English, (such as the uvular or alveolar /r/ used in French but almost entirely absent in English), and eliminate sounds which are used in English (for example, since /h/ is not used in initial position in French, it may often be dropped by French speakers when speaking English). It may also alter stress patterns, since for example French is generally held to be syllable-timed rather than stress-timed, and typically uses iambic stress patterns more than trochaic stress patterns. Therefore the stress system used in English by a native French speaker is often different from that used by a native English speaker. A foreign accent might also affect the duration of some sounds (since final-syllable lengthening is more characteristic of French than of English; see Levitt & Aydelott Utman, 1992). In short, we might expect foreign accents (which may include sounds which are not used in the infants' own tongue, phonotactic violations of the rules of the infants' own tongue, and unfamiliar prosodic patterns, as well as the simpler phonetic shifts which typically comprise regional accents) to be normalised with more difficulty, and therefore later, than regional accents.

So far there has been little work directly comparing infants' processing of foreign versus regional accents, but several studies have looked at the impact of accent on word learning and recognition. Best and colleagues (Best, Tyler, Gooding, Orlando, & Quann, 2009; Mulak, Best, Tyler, Kitamura, & Bundgaard-Nielsen, 2010) examined infants' ability to recognise known words in familiar versus unfamiliar accents, using a head-turn preference procedure. Fifteen- and nineteen-month-olds were exposed to lists of familiar and unfamiliar words, spoken by an American speaker (whose accent was familiar to the infants) and a Jamaican speaker (whose Jamaican English Mesolect was unfamiliar). At 19 months, the infants showed a preference for the familiar words regardless of the accent in which they were spoken, but at 15 months, this preference appeared only when listening to the American speaker, and not the Jamaican speaker. A word-learning study carried out with 24- and 30-month-olds used a preferential looking procedure to test infants' ability to generalise novel word-learning across accents (Schmale, Hollich, & Seidl, 2011). Infants were trained with two novel objects, which were assigned novel names. These names were presented in either the infants' ambient (American) accent or a South American Spanish accent, and paired with the matching object in several training trials. In the test phase, the infant was then presented with two objects, a trained target and a distractor, and heard the word matching the target. The results showed that at 30 months, infants were able to show a preference for the named target despite a change in accent between the training and test phases. However, at 24 months, the infants trained in their ambient accent and tested in a non-native accent spent as much time looking at the distractor as they did at the target, suggesting that they were not able to generalise their learning across accents. When trained in the Spanish accent and tested in their ambient accent, they did show a target preference.

A set of studies published after the current research had begun looked at segmentation within and across accents (Schmale & Cristià, 2009; Schmale, Cristià, Seidl, & Johnson, 2010), and used both foreign and regional accents. Infants were habituated to single-word utterances of target words, and then tested with passages containing the trained target word forms versus passages containing counterbalanced distractor word forms. As would be predicted by a number of segmentation studies (Johnson, Jusczyk, Cutler, & Norris, 2003; Jusczyk & Aslin, 1995; Jusczyk, Hohne, & Bauman, 1999a), at nine months Schmale and Cristià (2009) found that infants showed a preference for sentences containing disyllabic trained target word forms over sentences containing distracter words, indicating successful segmentation, when both the training phase (consisting of single-word utterances) and the test phase (consisting of sentences) were presented in a familiar accent, but by two different speakers with dissimilar voices. They also showed that at the same age, when both the training and test phases were presented in an unfamiliar (Spanish) accent, by a single speaker, the infants were able to show a preference for the sentences containing trained target word forms. However, this preference disappeared when two different Spanish-accented speakers were used, one for the training phase and the other for the test phase, suggesting that infants were only able to accommodate accent variability at a very local level, and could not easily generalise across speakers. By 13 months, infants preferred the trained target sentences across speakers in the unfamiliar accent, and across accents. A follow-up study (Schmale, et al., 2010) showed that at 12 months, but not at nine, infants were able to show a preference for sentences containing trained target words when either the training phase was presented in a familiar accent and the test phase was in an unfamiliar regional accent, or when the training phase was presented in an unfamiliar regional accent and the test phase in a familiar accent. Unlike the word-learning study described above (Schmale, et al., 2011), which found that infants could recognise a target named in a familiar accent when trained in a foreign accent, but could not recognise the target named in a foreign accent after training in a familiar accent, these studies did not show the same type of asymmetry; infants were able to segment equally successfully when trained in a familiar accent and tested in an unfamiliar one, or when trained in an unfamiliar accent and tested in a familiar one.

Schmale et al.'s studies support Houston and Jusczyk (2000) and Singh's (2008) suggestion that infants towards the end of the first year of life are beginning to develop the ability to use varied input to form more robust representations and to generalise their learning across variant forms of speech. However, it appears that this process may not be completed. A study comparing mono-dialectal and bi-dialectal infants (that is, infants exposed to only one accent or dialect at home, versus those exposed to more than one) from the South-West of England, examined the ability of 20-month-old infants to recognise familiar object names when spoken with a rhotic accent (native to the South-West of England) or a non-rhotic accent (Floccia, Delle Luche, Durrant, Butler, & Goslin, 2012). For example, the word "horse" is pronounced /hɔːs/ in standard British English, but as /hɔːrs/ in the South-West of England. The results showed that both groups looked longer to the target only when pronounced in a rhotic accent, suggesting that even at this age, they were not able to generalise existing lexical representations across accents, and that the more variable input experienced by the bi-dialectal infants did not, contrary to expectations, give them an advantage in recognising the non-rhotic forms. Rather, they seemed to have formed abstract representations based on the ambient accent of the local community, rather than drawing on the non-rhotic speech of their parent/s. Taken along with Schmale et al.'s (2011) word-learning study with 24- and 30-month old infants, and Best et al.'s (2009) familiar word recognition study, this suggests that by the end of the first year, infants are able to compensate for variability in the online processing of novel stimuli, but that their lexical representations of known or trained words are still somewhat more rigid, and cannot as easily be generalised across accents.

Methodological considerations

The majority of an infant's linguistic input comes, not in the form of single-word utterances, but in the form of multi-word phrases or sentences (Brent & Siskind, 2001; van de Weijer, 1998). Before the infant can attach meaning to individual words, therefore, he or she must establish where the barriers between meaningful units (words) are; they must segment the speech stream into linguistic units. Since fluent speech, unlike written language, does not typically include noticeable gaps at the end of each individual word, and may include pauses in the middle of words, this task requires other cues, often language-specific. These may include prosodic patterns (Echols, et al., 1997), phonotactic cues (Mattys & Jusczyk, 2001), statistical probabilities (Thiessen & Saffran, 2007) and, once the infant has learned some words, lexical identification (McClelland & Elman, 1986). As mentioned above, Jusczyk and Aslin's seminal study (1995) demonstrated that at seven and a half months, infants listened longer to passages of fluent speech which contained words to which they had been habituated to than to passages which did not contain habituated words, suggesting that they were able to recognise and segment the familiar word forms from the speech stream. However, in the early stages, this ability is not robust, and can easily be disrupted by variability in the speech signal. A change in affect (Singh, et al., 2004) or gender (Houston & Jusczyk, 2000) between the training and test phases disrupts recognition of a learned word when heard as part of a sentence, in infants of seven and a half months, although at ten and a half months, this is no longer the case, and infants can successfully recognise trained words even with a change of affect or gender. This suggests that in the early stages of language acquisition, infants are forming over-specific representations of the words they hear, and that as the infant develops and gains more experience, this over-specificity is then relaxed, creating more flexible representations which allow for normal inter- and intra-speaker variability, without preventing infants from distinguishing contrastive differences. These studies show that the task of segmentation allows us to track infants' ability not just to discriminate between different stimuli, but to generalise across individual stimuli within a single category and to accommodate variation, both within and between speakers, including that caused by co-articulation in fluent speech.

The use of fluent speech rather than just single words or phonemes has a further advantage. As with the adult studies, using whole sentences rather than individual phonemes or words should allow adaptation to take place rapidly, so that it can be discounted as a confounding factor in our studies. It also provides more acoustic information, which the infant system may be able to use to inform the normalisation process.

In adult psycholinguistic studies, a range of behavioural techniques is available to test word recognition and lexical access under different experimental conditions. However, the majority of these rely on an explicit response to a task, such as categorising stimuli by semantic category or acoustic features, or transcribing speech. Infant studies usually rely instead on paradigms which make use of infants' spontaneous responses to stimuli. The majority of studies described here therefore use infants' direction of gaze as a measure of attention or recognition. In older infants, preferential looking is used as a measure of infants' word-learning, but this requires word forms to be paired with referents. The head-turn preference procedure, though, allows infants' processing or word forms without referents to be examined. This technique, first used by Fernald (1985), pairs word forms with a visual fixation stimulus (a flashing light or a non-lexical visual pattern such as a grid pattern), and uses infants' attention to the paired stimuli as a measure of their preference for the auditory stimulus. Since the same visual stimuli are used across conditions, any preference shown must be for the auditory stimulus rather than the visual fixation stimulus. This technique has been used to measure infants' preference for infant-directed speech over adult directed speech (Fernald, 1985), and for their ambient accent over a non-native accent (Best, et al., 2009). It was then adapted to allow for a separate training or familiarisation phase (see for example Nazzi et al., 2000), a technique which is used in chapter 7. This is particularly useful when looking at segmentation, as it allows researchers to measure infants' ability to extract newly-learned word forms from fluent speech, or to learn new word forms from fluent speech. Under this adaptation, infants are first familiarised to stimuli, such as single-word utterances of target word forms, and are then tested with passages of fluent speech containing those target words (Jusczyk & Aslin, 1995; Schmale & Cristià, 2009; Singh, 2008).

While the head-turn preference procedure is widely and successfully used, it nonetheless has its practical flaws. Although the task demands involved in HPP are minimal, this paradigm does require the attention of the infants. If they are distracted from the visual fixation stimuli, it may be impossible to collect useable data from them. In addition, because HPP experiments require repeated presentations of the stimuli, they can be repetitive and therefore uninteresting for infants, which results in a high drop-out rate. In some cases, it may also be difficult to establish whether poor performance is due to task difficulty, leading to inattention, or whether a simple preference is leading to inattention, making a task seem difficult. One solution to these problems is to use a task which does not require any behavioural response, and which therefore does not need infants' attention, and which allows large amounts of data to be collecting even when the infants are distracted. Neurophysiological studies, which measure directly the brain's responses to stimuli can be used to by-pass these practical issues, and thereby to complement the behavioural data, giving a more in-depth look at processing.

Summary

Just as accents affect adult processing of fluent speech (Adank, et al., 2009; Clarke & Garrett, 2004), they also affect infants' processing (Best, et al., 2009; Cristià, et al., In preparation; Schmale & Cristià, 2009; Schmale, et al., 2010), but the extent of this disruption is not yet clear. When approaching the question of children's ability to segment speech across accents, it must be borne in mind that we cannot necessarily treat all unfamiliar accents as equal. Whether the differences fall simply along a continuum of perceptual distance from the home accent, or whether regional and foreign accents employ different cognitive processes, as appears to be the case in adults (Goslin, et al., 2012), we would expect foreign accents to prove harder to segment than regional accents. Both regional and foreign accents are expected to disrupt infants' ability to segment fluent speech in behavioural tests (Schmale & Cristià, 2009; Schmale, et al., 2010), while electrophysiological methods offer a way to examine the time-course of accented speech processing.

The following chapters aim to investigate infants' processing of accented speech, using two different experimental approaches, and taking segmentation of novel word forms from fluent speech as a proxy measure. Chapter six will take a behavioural approach, using the Head-turn Preference Procedure, and comparing home, regional and foreign accent processing, with a view to measuring the effect of accents on infants' ability to segment speech. Chapter seven will use ERP as a more online measure of segmentation at the neural level, comparing the infant brain's ability to segment within a familiar accent, within an unfamiliar, non-native accent, and across accents.

Chapter 7: Infants' ability to segment across accents.

As outlined in the previous chapter, variability in the speech signal is inevitable and necessary for robust learning, but it can also be disruptive in a number of ways. When adults hear speech which is degraded, masked, or which varies from familiar speech, they are able to use top-down processing to help reconstruct the speaker's intended meaning, through the use of semantic and syntactic cues. This is likely to be helped by their wider experience of variation, which will result in adults having more robust lexical representations. According to an exemplar-based approach of word recognition, variant pronunciations of words are lexically represented in the adult system, even if they are only rarely encountered (Johnson, 1997; Sumner & Samuel, 2005), while an abstract-entry approach suggests that adults are capable of normalising variants pre-lexically, using a context-specific algorithm in order to access the intended word (Pallier, et al., 2001).

However, in their early years, infants lack the experience or linguistic exposure which might allow them either to hold multiple representations of the words they encounter (many of which may not have been encountered at all, let alone in different accents or forms), or to have the systems in place to normalise variant forms in order to reach the abstract, "canonical" representation. By the age of 30 months, on average infants have a productive vocabulary of around 600 words (Beckage, Smith, & Hills, 2010; Mayor & Plunkett, 2011), and much of the speech to which they attend during their first two to three years will be infant-directed speech (IDS), which typically has a simplified, repetitive vocabulary (Spencer, Blumberg, McMurray, Robinson, Samuelson, & Tomblin, 2009).

Despite this limited input, infants quickly start to discriminate between familiar and unfamiliar forms of speech. At five and seven months, infants familiarised to speech in either their own ambient accent (South-West English) or an unfamiliar regional accent (Welsh) showed a preference for the accent to which they were familiarised over the other accent, in a head-turn discrimination procedure (Butler, et al., 2011), although they did not appear to distinguish between two unfamiliar accents (Welsh and Scottish English). At six months, American infants showed a preference for a novel native accent (Australian) over an American accent, as measured both by looking behaviour and heart-rate deceleration (Diehl, Varga, Panneton, Burnham, & Kitamura, 2006). While the direction of their preference was somewhat unexpected, it does provide evidence that they are able to distinguish between their own accent and an unfamiliar native accent at this early age; indeed, this ability can be found at three months of age (Kitamura, Panneton, Diehl, & Notley, 2006b).

As infants grow older and gain experience, it appears that they discriminate less between regional accents and their own ambient accent. Phan and Houston (2006) found that seven-month olds, but not eleven- or eighteen-month olds, discriminated between single words pronounced in their ambient accent and an unfamiliar non-native accent, suggesting that as they develop, infants learn to ignore phonological distinctions which are not contrastive in their ambient accent, while Polka and Werker (1994) have shown a gradual reduction in discrimination of non-native contrasts across the first year of life. Perhaps as a result of this reduced sensitivity, later in childhood, children may still have difficulty using accent-based variation as a basis on which to discriminate between or classify speakers. Girard, Floccia, and Goslin (2008) carried out a set of experiments in which French-speaking five- and six-year old children were asked to classify speakers into groups, either within or across accents, so that accent could be used as a cue to distinguish one speaker from another in some conditions but not others. Girard et al. found that regional accents did not serve as an effective cue for classification, although foreign accents did. However in this study the perceived strength of accents was not controlled so that the foreign accent speakers had a stronger accent than the regional accent ones. A replication of this study with more controlled material and with five- and seven-year old English-Speaking children found similar results (Floccia, et al., 2009a), with regional accents proving difficult to classify in both age groups, while the older children were able to classify the foreign accents. This would seem to suggest that older infants and young children are able to normalise, and thus ignore, at least some of the variation due to regional accents, while the difficulties associated with foreign accents persist until a later age even when strength of accent is controlled, suggesting differences between the way regional and foreign accents are processed.

As well as a developmental progression in the way in which infants and young children recognise or discriminate variant forms, there also seems to be a developmental change in the way accents impact the learning and processing of words (Best, et al., 2009; Schmale & Cristià, 2009; Schmale, et al., 2011). This has been shown both in well-established and newly-learned words. A word-recognition study showed that at 15 months, infants show a preference for highly familiar words over rarely-encountered words in a familiar accent, but fail to do so when the words are heard in an unfamiliar accent until 19 months (Best, et al., 2009; Mulak, et al., 2010). In a word-learning study using a preferential looking paradigm 24-month-olds failed to show a preference for a newly-learned word-object pair when it was presented at test in an unfamiliar accent, although they could do so at 30 months (Schmale, et al., 2011). Interestingly, at 24 months, they were able to identify the newly-taught word-object pair when it was taught in an unfamiliar accent and tested in a familiar accent. This asymmetry suggests that by this age, they are beginning to develop strategies to deal with variant speech, but that these are not yet complete.

In what follows, we will examine the impact of accent-related variability on early normalisation abilities in continuous speech. To do so, we will focus on infants' capacity to segment words out of the speech signal when an unfamiliar accent is presented. During the first year of life, infants' vocabulary is still extremely limited, and one of their main linguistic tasks (and one which has been extensively explored in the literature) is that of segmentation, namely the ability to extract individual word-forms from the speech stream. Infants use a variety of cues in order to achieve this, including phonotactics (Brent &

Cartwright, 1996; Mattys & Jusczyk, 2001), prosodic cues (Echols, et al., 1997; Mattys, et al., 1999; Polka, Rvachew, & Molnar, 2008) and distributional regularity (Brent & Cartwright, 1996; Thiessen & Saffran, 2007), and several studies have found that infants can segment fluent speech effectively from as early as seven and a half months (Jusczyk & Aslin, 1995; Singh, et al., 2004). However, variation in the speech signal can disrupt segmentation (Houston & Jusczyk, 2000; Singh, et al., 2004), and this includes the variation caused by an unfamiliar accent. A set of segmentation studies with infants aged nine, twelve and thirteen months demonstrated that while the nine-month-olds were able to segment successfully within their own accent, they failed to do so across accents (that is, when habituated to individual words in one accent, and then presented with sentences containing those trained words or untrained distractors in a different accent) until twelve months in the case of a regional accent and thirteen months in the case of a foreign accent (Schmale & Cristià, 2009; Schmale, et al., 2010). However, these studies did not directly pit regional and foreign accents against each other, and indeed there is a lacuna in the literature when it comes to direct comparisons of infant's processing of regional and foreign accents. The current study attempts to address this question, with the expectation that both regional and foreign accents will have an impact on infants' ability to segment across accents, and that this impact will be more severe in the case of foreign accents.

The current study

The first experiment described here set out to examine infants' ability to segment words from fluent speech within and across familiar, regional and foreign accents. In this, it covers much the same ground as Schmale and colleagues (Schmale & Cristià, 2009; Schmale, et al., 2010), their studies not having been published at the time this work began, but with the additional remit of allowing a direct comparison between performance with regional and foreign accents. The existing literature shows that while infants are able to segment familiar speech from as early as 7.5 months (Jusczyk & Aslin, 1995), segmentation, like word recognition and word learning, can be disrupted by intra- or inter-speaker variability in the speech signal, due to the speaker's gender (Houston & Jusczyk, 2000), affect (Singh, et al., 2004), or (most relevantly for this thesis) accent. To summarise, Schmale and colleagues found that at nine months infants could segment within their ambient accent, as predicted by previous literature (Johnson & Jusczyk, 2001; Jusczyk & Aslin, 1995; Jusczyk, et al., 1999a), but that they were unable to segment across accents until 12 months for a regional accent (Schmale, et al., 2010), or 13 months for a foreign accent (Schmale & Cristià, 2009). The current study will use a similar technique to investigate infants' ability to segment both within and across accents, using a foreign and a regional accent. It will, however, differ from Schmale and colleagues' work in a number of ways, described below.

A number of accent conditions will be included, using both the regional and the foreign accent, and both within- and across-accent changes. Infants will be tested at the age of 10 months, by which age the existing literature predicts that they will be able to segment words from fluent speech successfully within their own accent and in the absence of other variation (Jusczyk & Aslin, 1995; Singh, 2008).

At this age, infants are also able to accommodate at least some variation in the speech signal, as shown by Houston and Jusczyk (2000). They showed that at 10.5 months but not at 7.5, infants displayed a preference for passages containing trained target word forms even when a male speaker was used for familiarisation and a female talker for test, or vice versa. In addition, Singh (2008) shows that infants may be better able to accommodate variation when trained with variable input. At 7.5 months, infants were better able to segment trained words from passages spoken with a happy affect when they had been trained with words spoken in a variety of affects than when trained with words spoken in a neutral affect.

Many of the existing segmentation studies discussed so far habituated infants to single words, and then tested them using passages of fluent speech which contained either those trained words, or untrained but matched distractor words. However, Brent and Siskind (2001) carried out an analysis on a large corpus of natural infant-directed speech by mothers to their infants before the age of 12 months, and found that only around 9% of infant-directed utterances consisted of isolated words. This suggests that training infants with fluent sentences containing target words, and then testing them with individual words is a more ecologically valid approach to segmentation than the reverse order. It may also represent a more rigorous test, since it requires segmentation to have taken place in order for infants to learn the target word forms, whereas training infants with individual words and then testing with fluent speech allows infants to succeed by treating the task as a word-spotting task rather than an accurate test of segmentation. However, Jusczyk and Aslin (1995) found successful segmentation of monosyllabic words in seven and a half month old infants using both orders, leading them to conclude that the two orders represent similar levels of difficulty.

For the first experiment, then, passages of fluent speech will be used in the familiarisation phase, followed by a test phase using individual words (the reverse of the order used by Schmale and colleagues (Schmale & Cristià, 2009; Schmale, et al., 2010). A number of different accent conditions were originally planned, divided into Within-Accent (where the accent remains the same for the familiarisation phase and the test phase) and Across-accent (where one accent is used for the familiarisation phase, and a different accent is used for the test phase). The Within-accent conditions are divided into Familiar (where the accent used throughout is the infants' ambient accent, from Plymouth, in the South-West of England) and Unfamiliar (where the accent used throughout is an unfamiliar regional or foreign accent). The Across-accent conditions are further divided into Regional (in which a Plymouth accent is used for one phase and a Scottish accent is used for the other phase) and Foreign (in which a Plymouth accent is used for one phase and a German accent is for the other). As well as reversing the order used by Schmale and colleagues, the first experiment in the current study always uses two different speakers, one for the training phase and another for the test phase, even when the accent remains. Schmale et al. (2010) also did not include a within-accent condition using a regional accent.

In Schmale et al.'s regional accent study (2010), a Canadian accent was used as the unfamiliar regional variant, with a Mid-American accent as the familiar baseline. In their foreign accent study (2009), a South American Spanish accent was used. The Canadian accent was chosen because it differs only minimally from the Mid-American accent (Schmale, et al., 2010), while the Spanish accent differs much more (Schmale & Cristià, 2009), with variations at both the subsegmental and suprasegmental levels. This raises the possibility that the differences found in the ages at which infants are able to segment across accents are due not to the type of accent (foreign versus regional) but rather to their distance from the baseline accent (White & Morgan, 2008). Chapters 3 and 4 of this thesis have demonstrated that in adults, ERP responses to foreign and regional accents differ qualitatively and not just quantitatively, but it has not been established that this is true for infants, who may be responding according to perceptual distance rather than qualitative differences. For the current study, therefore, it was decided to use a regional and a foreign accent which both differ from the baseline accent to a similar degree, so that a direct comparison could be made between the two types of accent.

In the current study, a Scottish accent was used as the regional variant and a German accent was used as the foreign variant. The two accents used are discussed in more detail in the Stimuli section of this chapter.

Table 7: A summary of the conditions used in Experiment 3.

Training Accent	**Test Accent**	**Condition**	
Plymouth	Plymouth	Baseline; Within - familiar	
Plymouth	German	Across-accent - Foreign	

German	Plymouth		
Plymouth	Scottish	Across-accent - Regional	
Scottish	Plymouth		
German	German	Within-accent - Foreign	Within-accent – Unfamiliar (These conditions were discontinued, and not included in the final analysis)
Scottish	Scottish	Within-accent - Regional	

The existing literature (Jusczyk, et al., 1999a; Schmale, et al., 2010; Singh, 2008) predicts that at ten months, infants will be successful in segmenting disyllabic word forms from fluent speech within a familiar accent (that is, one which matches their ambient accent). It is therefore expected that in this experiment, infants will show a preference for trained target words over matched distractors in the test phase when both phases are presented in the infants' ambient accent. The increased task difficulty associated with processing an unfamiliar accent, whether regional or foreign, is expected to have an impact on their ability to segment trained word forms from fluent speech within an unfamiliar accent, particularly when the accent is a non-native one.

<u>Within-accent conditions</u>

It was initially intended in this study to include two within-accent conditions using unfamiliar accents, one regional and one foreign, in order to investigate infants' ability to segment within an accent other than their ambient accent. However, when preliminary analysis failed to show the expected results in the familiar accent condition, these two conditions were discontinued, and will be reported here only as footnotes.

The remaining within-accent condition, which uses a familiar Plymothian accent for both the habituation and test phases, is intended as a baseline, since it is predicted, based on previous literature (Jusczyk & Aslin, 1995; Schmale & Cristià, 2009; Schmale, et al., 2010; Singh, 2008) that infant will show clear evidence of segmentation in this condition[1].

Across-accent conditions

The Across-accent conditions represent a much more difficult task than the Plymouth-Plymouth condition, as infants must compensate for accent-based variation, and generalise from training in one accent to the test phase in a different accent. At 24 months, infants show an asymmetry in their responses when generalising across accents (Schmale, et al., 2011); when trained in an unfamiliar accent, they were able to generalise to their own accent, but when trained in their own accent, they were unable to generalise to an unfamiliar accent. However, at nine and thirteen months, they do not show this asymmetry (Schmale & Cristià, 2009). Therefore at ten months, it is predicted that responses will not show an effect of the order in which the accents are presented.

Since the across-accent conditions are more difficult than the within-accent conditions, it is predicted that at this age, they will spend less time looking towards the target words in the test phase in the across-accent conditions than in the within-accent conditions. Based on White and Morgan's findings (2008) showing a gradation of responses depending on the degree of phonetic variation from the target words, it is also predicted that they will spend less time looking towards the targets in the Plymouth-German condition than in the Across-accent – Regional condition (Plymouth-Scottish).

[1] It was further predicted that infants would show a gradation in their performance with the two other within-accent conditions, as measured by looking times to the target words in the test phase. It was expected that we would find the longest looking times in the Plymouth-Plymouth condition, with somewhat shorter looking times to target in the Scottish-Scottish condition, and the shortest looking times to target in the German-German condition.

Since Across-accent conditions are more challenging than within-accent conditions, this study does not include an Across-accent condition using two unfamiliar accents.

Experiment 3

Method

Participants

Fifty-eight healthy monolingual 10-month-old infants were recruited via Plymouth Babylab's participant database (M age = 9.91 months, range = 8.92 – 11.31 months). Twenty-three were male and thirty-three were female, and all lived in or near Plymouth, and had parents from the area. Parental reports confirmed that the infants' ambient accent was that of the South-West of England. Infants were randomly assigned to each of five conditions, as listed in Table 8, where the first accent named is that used for the familiarisation phase, and the second refers to the accent in which the target and distractor word forms were presented. The two cross-accent conditions each include both Familiar-Unfamiliar and Unfamiliar-Familiar orders.

Only five infants took part in each of the Within-accent Unfamiliar conditions (German-German and Scottish-Scottish); originally it was intended to test more infants in these conditions, but the unexpected results found in the baseline (Plymouth-Plymouth) condition meant that any findings from the Within-Accent Unfamiliar conditions would be impossible to interpret, and it was therefore decided to focus on the remaining conditions.

Thirty additional infants were excluded due to fussiness (N = 27), or specific exclusion criteria (i.e. they were exposed to other languages or accents on a regular basis; N = 3).

Table 8: A list of Conditions used in Part 1 of Experiment 3, with number of participants.

	Condition	Number of Participants
Within-Accent	Plymouth-Plymouth	16
	Scottish-Scottish	5 (This condition was discontinued and not included in overall analysis)
	German-German	5 (This condition was discontinued and not included in overall analysis)
Across-accent	Plymouth-Scottish Scottish-Plymouth	8 8
	Plymouth-German German-Plymouth	8 8

Stimuli

Stimuli consisted of two types: passages containing the target words, which were used in the training phase of the experiment, and word lists, made up of several instances of one of the target or distractor words. Four words were selected (*carriage*, *pasture*, *dialect*, and *tourist*), all of which were trochaic English words (that is, disyllabic words in which the stress falls on the first syllable) which were easily distinguishable from each other, but which were unlikely to be familiar to the infants. They also all started by a plosive sound to facilitate segmentation.

For each target word, a passage was recorded containing four unrelated sentences, each of which featured the target word. The target word always appeared embedded in the sentence, never in first or last position; for example, “He gave her a **carriage** clock for Christmas”. These sentences were presented during the familiarisation phase.

The word lists in the test phase consisted of 5 unique utterances of the target word, repeated three times to give 15 instances of each word.

Passages and words were spoken in child-directed speech, and the words were recorded with a variety of different intonations. Six female speakers were used; two from Plymouth (aged 31 and 40), two from Scotland (one, aged 22, from Edinburgh and one, aged 21 from Glasgow, both of whom had lived in Scotland until the age of 20) and two from Germany who had lived in the South-West of England for at least 2 years (aged 19 and 34). They each recorded all of the sentences and words several times, with the clearest and best-matched example of each sentence and the five clearest and most varied examples of each word being used in the final design. The sentences were concatenated with a pause of around 250 ms between sentences, and the words were concatenated with a pause of around 460 ms between words. This resulted in passages with an average duration of 27.06 seconds, and sets of words with an average duration of 18.08 seconds. A one-way ANOVA showed no differences in passage ($F(2, 21) = 2.55$, $p = .1$; mean for Plymouth sentence sets = 28.09 seconds, Scottish = 27.84 seconds, German = 25.25 seconds) or word ($F(2, 21) = 3.41$, $p = .052$; mean for Plymouth word sets = 19.32 seconds, German = 16.62 seconds, Scottish = 18.3 seconds) duration between the three accents. As with the stimuli used in the adult studies already reported, the stimuli were intended to represent a range of normal variation, so fine-grained acoustic analysis was not carried out. Stimuli were normalised for amplitude using Adobe Audition (Downs, 2010).

Accents

Scottish

The Scottish accent differs from Standard British English and from the Plymothian accent used here as a baseline in a number of features (Stuart-Smith, 2004). Vowel length and realisation differs on several vowels, including, in the stimuli used here, the first vowel in “various”, which is realised as /eə/ in the Plymouth accent but as /eɪ/ in the Scottish accent, and the vowel in “May”, realised as /eɪ/ in the Plymouth accent but as /iː/ in the Scottish accent. In “clock”, the vowel is realised as /ɒ/ in the Plymothian samples but as /ɔː/ in the Scottish stimuli. /r/ is rhotic in both the Plymothian and Scottish accents, but has a noticeably different quality in the Scottish samples, for example between vowels in the words “tourist” and “carriage”, and before a consonant in “horses”, where it is strongly rhotic. Stops are generally less aspirated in the Scottish dialect than in SBE (Stuart-Smith, 2004), and some stops are glottalised or reduced; in the stimuli used here for example, in the sentence “the carriage was pulled by two big white horses” the /d/ at the end of “pulled” is present in the Plymothian sentences but absent in the Scottish sentences. There is less divergence in prosody, but several of the Scottish sentences exhibited a rising intonation at the end of the sentence, whereas Plymothian sentences typically had a falling intonation.

German

The German accent in English differs from SBE as a result of the interaction between German and English phonology and prosody. German, like English, is considered to be a stress-timed language, and the two have very similar patterns of intonation (Grabe, 1997), so the prosodic structure of German-accented English diverges very little, in general or in the corpus used here, from native English. However, where English has eleven vowels and eight diphthongs, German uses only eight vowels and three diphthongs, resulting in an altered vowel-space in German-accented English. For example, the first vowel in “carriage” and “pasture” is realised as /æ/ in the Plymothian accent but tends towards /e/ in the German-accented sentences , with the medial vowel in “primroses”, realised as /əʊ/ in the

Plymothian sentences but tending towards /ɔː/ in the German stimuli. Some consonants also differ, particularly in their voicing. In the corpus used here, /d/ was often unvoiced, as for example “the carriage was pulled by two big white horses”, resulting in the word “pulled” being realised as /pɅlt/, rather than /pɅld/ in the Plymothian accent. /Ө/ is not used in German, and this often results in it being pronounced as /s/ or /d/; in the stimuli used here, this is particularly noticeable in the word “cathedral”, realised as /kəsiːdrəl/ in the German-accented sentences.

Apparatus

The experiment used the head-turn procedure. It was conducted in a purpose-built booth at the Plymouth Babylab, using proprietary software. Presentation of visual and auditory stimuli was controlled by the experimenter via a purpose-built response box, and the infants’ behaviour during the experiment was observed via a video link-up.

Procedure

Parents were seated in the booth with their infant on their lap, wearing headphones through which they heard music in order to mask the auditory stimuli. At the beginning of each trial, a red light flashed directly in front of the infant, to get their attention. In addition, the experimenter could trigger a doorbell sound directly in front of the child, or orient their attention forward. When the experimenter, who was seated outside the booth but could observe the infant via video, judged that the infant was looking at the central light, she started the trial, causing the green light to stop and a red light to start flashing either to the left or the right side (randomised and counterbalanced) of the infant. When the infant turned towards the correct side, a sound file, consisting of either a passage containing one of the target words or a word list containing one of the target or distractor words, was played via a speaker on the same side as the flashing light. This sound file played until the end of the file (around 27 seconds), or until the child looked away for more than 2 seconds. During the test phase, any trial on which the infant looked away within 1.5 seconds was aborted and the word list was repeated. This procedure conforms closely to that used by Nazzi et al. (2000).

During the training/familiarisation phase, infants heard the recorded passages featuring two of the target words (either carriage and pasture, or dialect and tourist). Passages were presented to infants accompanied by a flashing light, and continued until the end of the file, or until the infant looked away for more than two seconds. This was then repeated, with the two passages occurring in random order, until the infant had accumulated a total of 45 seconds looking time towards each of the two passages. Once they had achieved this, the test phase began. In this phase, the infant was presented with all four word lists, in random order, with two of the words serving as targets (those that matched the passages they had heard) and two serving as unfamiliar distractors. The four word lists were each presented three times. As in the familiarisation phase, trials ended either at the end of the sound file, or when the child looked away from the target side for more than two seconds. If three consecutive trials were aborted, the experiment ended, and was considered incomplete due to fussiness, and not included in the final analysis.

Depending on the condition to which the infant had been randomly assigned, they heard the sentences in one of the three accents, Plymouth, German, or Scottish. The words could then be in the same accent, or in a different accent; German or Scottish if they had heard the sentences in a Plymouth accent, or Plymouth if they had heard the sentences in either a German or Scottish accent. Thus there were seven possible conditions; P-P, S-S, G-G, P-S, S-P, P-G and G-P. Infants were not tested with a German accent in one phase and a Scottish accent in the other, as this represents a harder task which infants at this age are unlikely to be able to achieve. When the same accent was used for both the sentences and the words, two different speakers were used for the familiarisation and test phases (as in experiments 1 and

5, Schmale & Cristià, 2009, in which two different English or Spanish speakers were used for the two phases of the experiment).

Results

Analyses were carried out on the three following conditions: Plymouth-Plymouth, Plymouth-Scottish/Scottish-Plymouth, and Plymouth-German/German-Plymouth, with 16 infants in each of the three conditions (eight each in the Plymouth-Other and Other-Plymouth orders).

For each child, looking times during test trials were extracted for the two target words and the two distractor words, not including any abandoned trials. A mean value was taken across the three trials for each word (target and distractor), and a grand mean was calculated for the target words (Target) and for the distractor words (Distractor). A positive significant difference between Target and Distractor is thus taken as a measure of segmentation, with larger looking times towards the targets' sound source than towards the distractors' being taken as evidence of recognition, and therefore, of segmentation from the previously presented passages.

A one-way ANOVA with one within-participants factor of stimulus type (target vs. distractor) and one between-participants factor of condition (Plymouth-Plymouth, Plymouth-Scottish, Plymouth-German) revealed no main effect of either stimulus type ($F(1, 45) < 1$) or condition ($F(2, 45) < 1$), and no interaction between these two factors ($F(2, 45) = 1.16$, $p = .32$). These findings indicate that segmentation was not taking place in any of the conditions, and that looking times to targets and to distractors did not differ between the conditions.

Some other infant studies (See for example Johnson & Jusczyk, 2001) have used only 16 infants for a single experiment; based on this precedent, and in order to allow the findings of this experiment to be compared to previous studies, planned comparisons were carried out within each condition. In the Plymouth-Plymouth condition (in which both training and test phase were presented in a Plymouth accent), infants looked at the targets for an average of 8.63 seconds, and at the distractors for an average of 8.76 seconds, with no difference found between the two scores ($t(15) = -.17$, $p = .87$). In the Plymouth-Scottish condition, the mean looking time for targets was 8.77 seconds and the mean distractor looking time was 8.02 seconds, again with no difference between the two scores ($t(15) = 1.35$, $p = .20$). No effect of order (Plymouth-Scottish vs. Scottish-Plymouth) was found for either the targets ($F(1, 14) = .14$, $p = .72$) or the distractors ($F(1, 14) = 1.28$, $p = .28$). In the Plymouth-German condition, the mean target looking time was 7.5 seconds, and the mean distractor time was 8.33 seconds, with no difference between the groups ($t(15) = -.94$, $p = .36$). No effect of order (Plymouth-German vs. German-Plymouth) was found for either targets ($F(1, 14) = .88$, $p = .36$) or distractors ($F(1, 14) = .35$, $p = .57$) [2].

[2] In the Scottish-Scottish condition in which only 5 children were tested, the mean target looking time was 9.19 seconds and the mean distractor looking time was 7.65 seconds, with no difference between the two scores ($t(4) = 1.4$, $p = .23$). In the Within-accent – Foreign (German-German) condition with 5 children, the mean target looking time was 8.7 seconds and the mean distractor looking time was 8.42 seconds, with no difference between the two scores ($t(4) = .29$, $p = .79$).

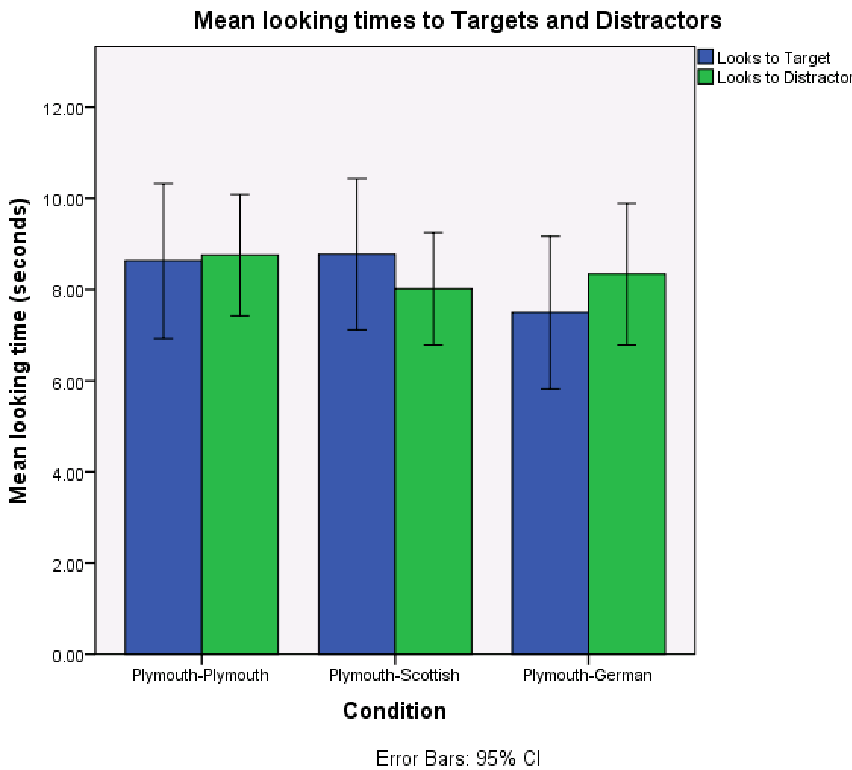


Figure 14: Mean looking times to Target and Distractor words in seconds, for three conditions in Experiment 3

Some previous studies have found that with repeated trials, familiarisation results in progressively less systematic looking behaviour, and have therefore used “first look” data instead of mean looking times across numerous trials (Schafer & Plunkett, 1998; Tincoff & Jusczyk, 1999). Data from the first test trial for each of the two target words and the two distractor words only was therefore analysed in the same way as the mean looking times. An ANOVA with one within-participants factor of stimulus type (target vs. distractor) and one between-participants factor of condition (Plymouth-Plymouth, Plymouth-Scottish, Plymouth-German) showed no effect of stimulus type ($F(1, 45) < 1$) or of condition ($F(2, 45) < 1$) and no interaction between the two factors ($F(2, 45) < 1$). No order effects were found in the two Across-accent conditions ($F(1, 14) < 1$ in both cases).

Individual condition analyses revealed no difference between targets and distractors for the Plymouth-Plymouth condition (target mean = 9.52, distractor mean = 8.44; $t(15) = 1.26$, $p = .23$), the Plymouth-

Scottish condition (target mean = 8.96, distractor mean = 9.78; t (15) < 1,), the Plymouth-German condition (target mean = 8.08, distractor mean = 8.90; t (15) < 1).[3] These results once again indicate that segmentation was not taking place in any condition.

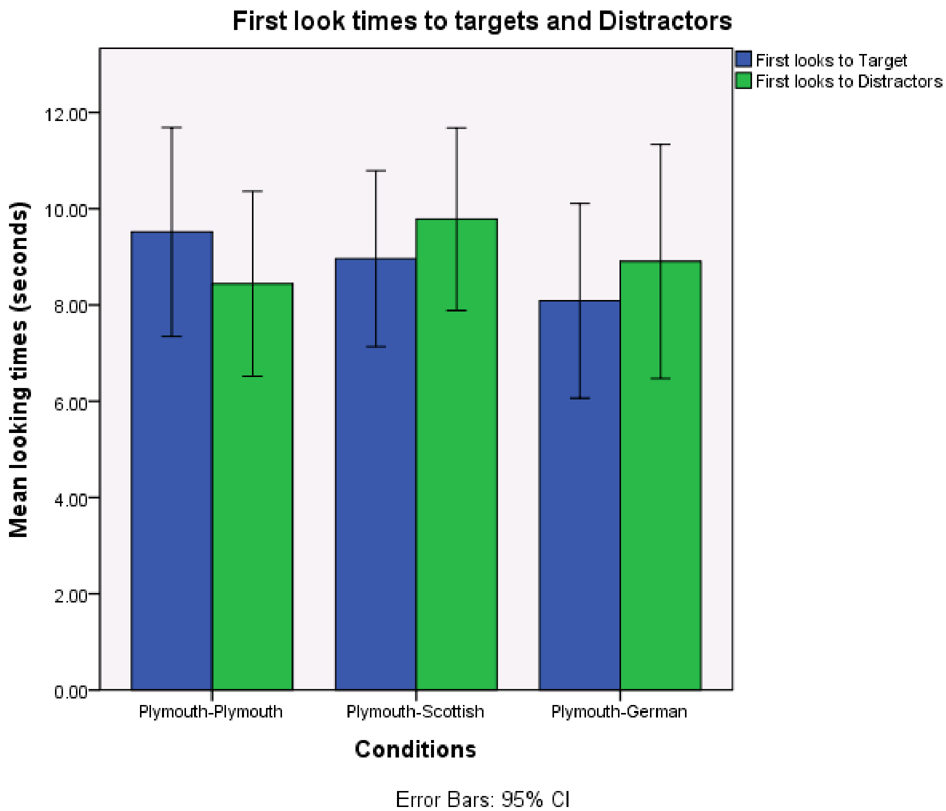


Figure 15: First look durations (in seconds) to Targets and Distractors for three conditions in Experiment 3.

Previous literature has found a consistent target preference in infants as young as 7.5 months for similar tasks, when the training and test stimuli are presented in familiar speech. The lack of a consistent preference in this experiment in the Plymouth-Plymouth condition was therefore surprising, and prompted a closer examination of the data for that condition. Of the 16 infants in this condition, seven showed an overall preference for the Target word, while nine showed a preference for the distractor. Taking the first trials only, ten showed a Target preference while six showed a Distractor preference. A

[3] No systematic differences were found in the Scottish-Scottish condition (target mean = 10.72, distractor mean = 9.68; t (4) = .63, p = .56), or the German-German condition (target mean = 9.59, distractor mean = 10.59; t (4) = -.61, p = .57).

binomial test showed that these figures represent a normal distribution (overall p = .69, first looks p = .17), so it is not possible to say that significantly more infants showed a target preference than a distractor preference.

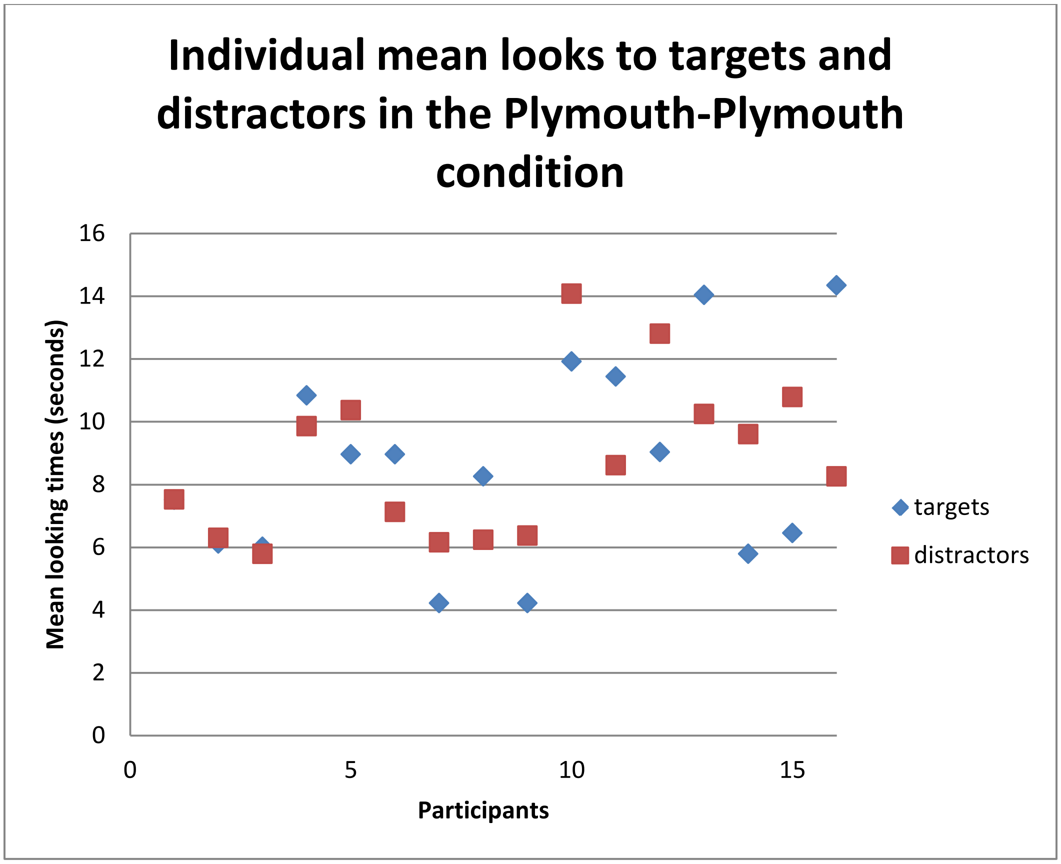


Figure 16: Individual mean looking times for Targets and Distractors, Plymouth-Plymouth condition only, ordered as a function of the magnitude of difference between target and distractor looking.

Discussion of Experiment 3

The failure of this experiment to find segmentation in the Within Familiar accent condition, contrary to previous findings (Jusczyk & Aslin, 1995; Schmale & Cristià, 2009; Singh, et al., 2004), was surprising given that the method used here closely mirrored that used by several other researchers, and that successful segmentation has been found in a younger age group than the participants in this experiment when a familiar accent has been used. The Plymouth-Plymouth accent condition was intended here to provide a baseline for Target preference, to which the other accent conditions could be compared, since the previous literature predicted that infants would show a reliable Target preference in this condition. In the absence of evidence of segmentation in the baseline condition, little can be said about infants' performance in the Across-Accent and Within Unfamiliar accent conditions, so it was considered important to investigate the Plymouth-Plymouth accent condition more thoroughly, in order to understand why the results of this experiment do not reflect those of the previous studies on which it was based.

As compared to the previous literature, two changes had been made which might have affected the results. Firstly, many of the previous studies had used only one speaker for both the familiarisation phase and the test phase, whereas in this experiment, the two phases were always presented by different speakers, even when the accent remained consistent. Schmale et al's (2009) findings show that a change of speaker is enough to disrupt segmentation within a foreign accent, suggesting that it makes the task more difficult for infants; if the stimuli used in this study were, for some reason, more difficult to segment than in Schmale et al's work, the change of speaker might have a cumulative effect, causing the infants in this study to fail to segment.

Secondly, the majority of the cited studies (Houston & Jusczyk, 2000; Jusczyk & Aslin, 1995; Schmale & Cristià, 2009; Schmale, et al., 2010; Singh, 2008) used single words for the familiarisation phase, and sentences containing target and distractor words for the test phase, whereas in this experiment, that order was reversed. Jusczyk and Aslin (1995) reported equally successful segmentation in 7.5 month olds using either a word/passage or a passage/word trial order, but other researchers (Nazzi, Mersad, Sundara, Iakimova, & Polka, In Revision) have found that the stimuli used for familiarisation versus test are important. In their case, they found that while Canadian infants from French-speaking families could segment words using a word/passage presentation order, French infants could only segment the same stimuli using a passage/word order. Although the infants in the current study have failed to show segmentation under the conditions in which Nazzi et al's infants succeeded, Nazzi et al's findings suggest that presentation order may have an impact on the task. They suggest that the passage-word order may be more difficult, as the passages do not contain any information which would allow the infants to discern the identity of the targets other than the fact that the target words occur several times in each passage. The word-passage order should therefore be easier, as in this condition the target words are heard in isolation prior to the test phase, and need only be recognised in the context of fluent speech. In trying to establish why the infants in the current study have not shown results in line with previous work, then, presentation order must be taken into account.

Experiment 4

In order to establish whether the change in speaker was responsible for the lack of segmentation found in Experiment 3, a new condition (referred to as Same Speaker) was run in which infants heard both the familiarisation and test phases spoken by a single Plymouth-accented speaker. In all other details, the procedure was identical to that used in Experiment 3. A second condition referred to as the Reverse condition, presented infants with single-words utterances during the familiarisation phase, and sentences containing either a trained target word or an untrained distractor word during the test phase, again using the same Plymouth-accented speaker throughout.

Participants

Thirty-two healthy monolingual 10-month-old infants were recruited via Plymouth Babylab's participant database (M age = 10.3 months, range = 9.57– 11.02 months). Twenty-three were male and nine were female, and all lived in or near Plymouth, and had parents from the area. Parental reports confirmed that the infants' ambient accent was that of the South-West of England. Infants were randomly assigned to the Same Speaker and Reverse conditions, with sixteen infants in each of the two conditions. A further ten infants were excluded due to fussiness.

Procedure

The procedure for the Same Speaker condition was identical to that for the original Within-Accents Plymouth-Plymouth condition, except that the same speaker was used throughout both phases of the experiment.

In the Reverse condition, the familiarisation phase consisted of the single-word utterances which had been used in the test phases of the other conditions, while the test phase used the passages previously used for familiarisation. Each child heard two target words during the familiarisation phase, in random order, until they had accumulated 45 seconds of looking time towards each one. They then heard passages featuring those two target words and two distractor words, with three test trials for each of the four passages.

Results

In the Same Speaker condition, infants looked towards the target words for an average of 9.86 seconds, and towards the distractor words for an average of 9.65 seconds. In the Reverse condition, they looked towards the target sentences for an average of 5.32 seconds, and towards the distractor sentences for an average of 5.03 seconds. An ANOVA with one within-participants factor of stimulus type (target vs. distractor) and one between-participants factor of condition (Same Speaker versus Reverse) found no significant effect of stimulus type ($F(1, 30) < 1$) and no interaction ($F(1, 30) < 1$). There was a significant effect of condition ($F(1, 30) = 21.57$, $p < .001$), due to shorter looking times towards both targets and distractors in the Reverse condition than in the Same Speaker condition.

As in the previous experiment, planned comparisons were carried out within each condition (Johnson & Jusczyk, 2001). Paired-samples t-tests revealed no evidence of segmentation in the Plymouth-Plymouth Same Speaker ($t(15) < 1$) or Plymouth-Plymouth Reverse ($t(15) < 1$) conditions. When looking just at the first test trial for each target and distractor, an ANOVA using a within-participants factor of stimulus type and a between-participants factor of condition found no effect of stimulus type ($F(1, 30) = 2.24$, $p = .14$) and no interaction ($F(1, 30) < 1$). Again, shorter looking times to both targets and distractors in the Reverse condition led to a significant effect of condition ($F(1, 30) = 27.16$, $p < .001$).

In the Same Speaker condition the mean target looking time was 12.54 seconds and the mean distractor looking time was 11.99 seconds, with no significant difference between the two scores ($t(15) < 1$), and in the Reverse condition, the mean target time was 6.89 seconds and the mean distractor time was 5.47 seconds, with no difference between the two scores ($t(15) = 1.67$, $p = .11$).

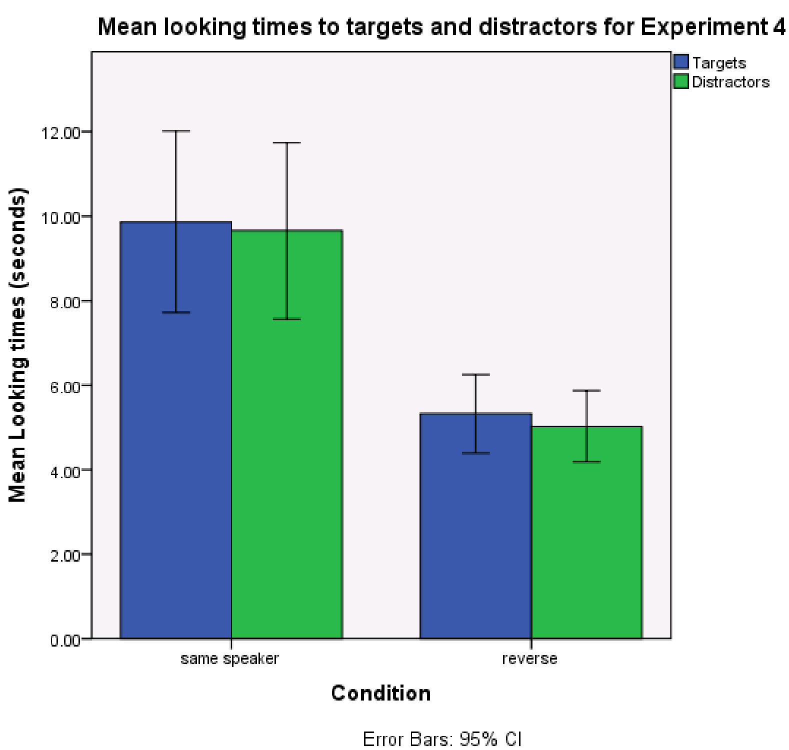


Figure 17: Average looking times to target and distractor words in two conditions in Experiment 4.

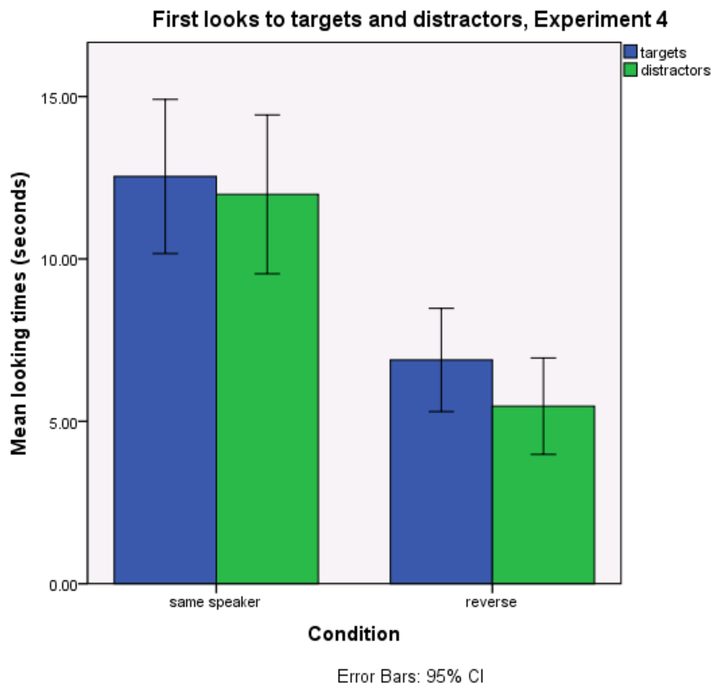


Figure 18: First Look durations to targets and distractors in two conditions in Experiment 4.

Discussion of Experiment 4

This experiment was intended as an extension of Experiment 3, in the hope of explaining the failure of Experiment 3 to replicate the results of previous segmentation studies (Jusczyk & Aslin, 1995; Schmale & Cristià, 2009; Schmale, et al., 2010), by using the same speaker through both the familiarisation and test phases, and by reversing the passage-word order, in line with Schmale and colleagues' work.

Once again, these results failed to support previous literature, as no effect of segmentation was found, despite using the same speaker throughout, and despite reversing the presentation order of the passages and words. The finding that infants looked significantly less long to the targets and distractors in the test phase in the Reverse condition was in line with previous findings: Jusczyk and Aslin (1995) who directly compared word-passage and passage-word orders reported that seven and a half month old infants looked to target passages for a mean of 8.29 seconds and to distractor passages for a mean of 7.04 seconds, while at the same age, infants looked towards target words for 10.43 seconds and to distractor words for 8.32 seconds. The passage files used in the test phase in this condition are longer than the word files, so these shorter looking times are not experimental artefacts; rather they seem to reflect reduced attention to the passages, compared to the single words, when used in the test phase, on the part of the infants. This does not necessarily represent a simple preference for the single-word utterances over fluent speech, although this is one possible explanation. It is also possible that it is specifically the switch from passages to words which results in longer looking times to the words than the switch from words to passages, perhaps because the single-word utterances used in this study encompassed a range of tones of voice, making the single-word files more varied in prosody than the passages. The single-word utterances also have a much slower speech rate and longer pauses between words than the fluent speech of the passages, perhaps making the words particularly salient to infants, as these features are all typical of infant-directed speech (Kitamura & Burnham, 1998).

Discussion – Experiments 3 and 4

These first two experiments with infants aimed to replicate previous literature showing that from the age of seven and a half months, infants are able to segment fluent speech when it is sufficiently consistent and familiar (Jusczyk & Aslin, 1995; Schmale & Cristià, 2009; Singh, et al., 2004), and to examine the effect of unfamiliar accents on segmentation. The main expectation in Experiment 3 was that at ten months, infants would show a significant target preference within their own ambient accent, but that when presented with unfamiliar accented speech, their performance would worsen. The results failed to support these predictions, in that segmentation was not found even within the ambient accent. In Experiment 4, using a single speaker and reversing the passage/word order did not redress this failure, and thus the results from Experiment 3 and 4 did not support the existing literature.

The published literature contains numerous examples of infants younger than ten months successfully segmenting both mono-syllabic (Jusczyk & Aslin, 1995; Singh, 2008) and disyllabic (Schmale & Cristià, 2009; Schmale, et al., 2010) words from fluent speech. However, there is also a small number of studies which have failed to find segmentation (DePaolis, Duffy, Keren-Portnoy, & Vihman, 2012; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006; Nazzi, et al., In Revision), and notably, the studies which have failed to find segmentation all used British or European infants, whereas the majority of the studies which have found successful segmentation used American or Canadian infants. In all cases, the papers describing the studies describe the stimuli used as being delivered in Infant-Directed Speech, but an informal comparison of the stimuli used by some of the successful studies (Johnson, et al., 2003; Schmale, et al., 2010) with those used by the current study and one other (DePaolis, et al., 2012) reveals some obvious differences in the style of speech. This will take us momentarily away from our main topic of accent perception, in a necessary digression to understand the role of Infant-Directed Speech in early segmentation abilities.

Infant-Directed Speech

When talking to babies and young children, we tend to adopt a particular style of speech, known as Babytalk, Motherese, or Infant-Directed Speech (IDS). This is typified by a slower speech rate, a higher pitch, longer vowels and longer pauses between utterances, and broader pitch contours than in adult-directed speech (Kitamura & Burnham, 2003; Kitamura, Thanavishuth, Burnham, & Luksaneeyanawin, 2001; Pegg, Werker, & McLeod, 1992); notably these are many of the same characteristics which seem to be used as cues for segmentation. IDS has been found across an enormously wide range of languages including tonal languages such as Thai (Kitamura, et al., 2001) and Mandarin (Grieser & Kuhl, 1988), and non-tonal languages generally classified as stress-timed, such as English (Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies, & Fukui, 1989) and German (Fernald & Simon, 1984), as well as those considered to be syllable-timed, such as Italian and French (Fernald, et al., 1989). While there may be some differences in IDS across different languages (Ratner & Pye, 1984), it appears that some form of IDS is more or less universal. Although it has sometimes been referred to as “motherese”, in fact it is used by fathers as well as by mothers (McLaughlin, White, McDevitt, & Raskin, 1983), and indeed it has even been observed in the older siblings of babies (Tomasello & Mannle, 1985), making it a very robust linguistic phenomenon. It appears that it serves a purpose, in capturing the attention of infants from an early age (Pegg, et al., 1992) and perhaps in making the task of processing the speech easier (Thiessen, Hill, & Saffran, 2005), which may help to explain its ubiquity. As well as occurring naturally when talking to babies, IDS is commonly used in laboratory environments for studies with infants. This ensures that the studies are ecologically valid, since the stimuli used mirror the speech input heard at home by the participants, and it also reduces the demands of the task, since IDS provides extra cues and is considered easier to process.

However, just as regional and foreign accents may vary in their strength, so too IDS may range from moderate to more extreme. While some differences in IDS have been noted in different languages (Fernald, et al., 1989; Ratner & Pye, 1984), little work has focused on the possibility that IDS also varies between cultures but within a single language, over and above differences caused by accent. Infant-Directed speech is generally described as having a slower speech rate, higher overall pitch, and greater pitch variation than adult-directed speech, although few studies give a detailed account of the exact acoustic properties of the speech used. In both sets of successful stimuli, recorded by American speakers (Johnson, et al., 2003; Schmale & Cristià, 2009), these features appeared to be particularly salient; the speech rate was slower, the overall pitch higher, and the pitch variation greater, than in the unsuccessful stimuli, which were recorded by British English speakers. This is supported both by a cross-linguistic study of natural infant-directed speech (Fernald, et al., 1989), which found more extreme prosodic modifications in IDS by American parents than in French, Italian, German, Japanese, or British-

English parents. Further support comes from a smaller study (DePaolis, et al., 2012), in which American and British-English parents were asked to “read” their child a picture book at home, in order to elicit natural IDS. An analysis of the resulting corpus of IDS showed that American IDS typically featured greater prosodic differences between consecutive syllables, and longer pauses following target words (in this case, the names of objects pictured in the book), and also featured the target word as a single-word utterance, or in clause-final or prosody-final positions more often than the English IDS. These features, which are typical of IDS in general and of American IDS in particular, are thought to facilitate segmentation (Fernald & Mazzie, 1991; Thiessen, et al., 2005). This gives rise to the possibility that the discrepancy in the results of the previous two studies as compared to Schmale and colleagues’ with American-English infants (Schmale & Cristià, 2009; Schmale, et al., 2010) is due to the style of Infant-Directed Speech, with the IDS used in the current stimuli lacking the features which allowed infants to segment the successful stimuli.

Experiment 5 therefore investigates this possibility by presenting infants with two different styles of fluent speech, one typical of the unsuccessful UK-based studies, and one more typical of the successful North American studies, in order to establish whether the type of IDS alone can be held responsible for the differing results. It was therefore predicted that infants would be more successful in segmenting NA-style speech than the UK-style passages.

Moreover, the UK-style IDS is closer in style and acoustic qualities to adult-directed speech than the NA-style IDS is, prompting a further prediction. Previous studies have shown a consistent preference, demonstrated by longer looking times, for IDS over ADS in infants from only a few weeks after birth (Pegg, et al., 1992) and throughout the first year of life (Werker & McLeod, 1989). Other studies have shown a preference for speech characterised by a positive vocal affect (Papousek, Bornstein, Nuzzo, Papousek, & Symmes, 1990; Singh, et al., 2004). It was therefore predicted that infants might find the more exaggerated and upbeat NA-style speech more interesting or attractive, and that they would therefore attend to it more in the familiarisation phase, and therefore that those in the NA-style condition would habituate more quickly than those in the UK-style condition.

This prediction however is limited by the fact that, in the typical segmentation paradigm, the familiarisation time is fixed to 45 sec for all infants before switching to the test phase. What could potentially vary between the conditions though is the time spent on each trial before reaching these 45 sec, or, in other words, the number of trials necessary to reach this familiarisation criterion. Based on the aforementioned studies on IDS preference, we would predict that the children presented with the more arousing stimuli would turn away less frequently than those exposed to the less interesting ones. That would translate in a smaller number of trials during the familiarisation phase for the NA-style condition than the UK-style condition, or to longer familiarisation trials in the NA-style than the UK-style.

Experiment 5

Participants

Thirty-two infants (20 male and 12 female) with a mean age of 11.09 months (range = 10.52 months to 11.64 months) took part in the experiment. All infants lived in the Plymouth area, and none had any reported exposure to North American accents, according to parental report. They were randomly assigned to one of the two conditions (UK-style N = 16, NA-style N = 16).

A further 19 infants were excluded from the analysis of the Test phase due to inattention or fussiness.

Stimuli

Two sets of eight target words were chosen, all of which were trochaic English words with plosive initial consonant (carriage, dialect, pasture, and tourist in set 1, and clover, dwelling, pension and trigger in set

2). Each infant heard two target words and two distractors from a single set. For each word, four sentences were recorded, each featuring the target word embedded within the sentence; for example, “The carriage was pulled by two big white horses”, “The vowels in your dialect determine how you speak”. Set 1 is identical to that used in Experiments 3 and 4; the inclusion of a second set allowed infants to take part in both Experiment 3 or 4 and this experiment, on separate occasions, without being exposed to the same stimuli. Jusczyk and Hohne (1997) showed that eight-month-old infants can remember newly-learned words for two weeks, so a second list of words avoided contamination of the results.

The stimuli were recorded by a single female speaker (aged 26) from Plymouth. She was asked to read the sentences twice “as though reading a bedtime story to a small child”, and then twice “as though at a children’s birthday party, getting the kids excited for the cake”. These two styles of reading were intended to elicit, in the first case, moderate IDS, as used in the unsuccessful UK-based studies, and in the second case, a more exaggerated style of IDS, like that used in the successful NA-based studies. For each sentence and each condition, the recording which was judged to best represent the intended style of IDS was then chosen. This resulted in sentences with a mean duration of 2.77 seconds in the UK-style speech and 2.80 seconds in the NA-style speech, with no significant difference in duration (t (18) = -.19, p = .85). The mean pitch for the UK-style passages was 226.73 Hz, while the mean pitch for the NA-style passages was 267.31 Hz, and this difference was found to be significant (t (14) = -7.29, p < .001). The minimum pitch did not differ between the UK-style passages (m = 86.23 Hz) and the NA-style passages (m = 99.16 Hz, t (14) = -.94, p = .36), but the maximum pitch did (UK-style mean = 448.94 Hz, NA-style mean = 519.09 Hz, t (14) = -6.34, p < .001), indicating that the NA-style speech had both a higher overall pitch and a greater pitch range than the UK-style speech. The amplitudes of the two styles showed a similar pattern; the minimum amplitudes were similar for both styles (UK-style mean = 20.96 dB, NA-style mean = 20.32 dB, t (14) = .68, p = .51), but significant differences were found in the mean amplitudes (UK-style mean = 69.38 dB, NA-style mean = 71.76 dB, $t(14) = -5.57$, $p < .001$) and in the maximum amplitudes (UK-style mean = 78.61 dB, NA-style mean = 82.84 dB, $t(14) = -5.57$, $p < .001$), indicating that the NA-style speech had a higher overall intensity, and greater variation in intensity than the UK-style speech.

The target and distractor words were recorded in a range of different tones of voice (happy, questioning, assertive), and four clear examples of each were used for the test phase of the study. Thus infants heard the familiarisation phase, consisted of passages of fluent speech in either UK or US style IDS, and then both groups of infants heard the same single-word utterances in the test phase of the experiment.

Procedure

The procedure in Experiment 5 was identical to that used in the Same Speaker condition in Experiment 4.

Results

For each infant, the average time spent looking to the Targets and to the Distractors in the Test phase was calculated. Overall, infants spent on average 6.02 seconds looking towards the trained Target word forms, and 5.89 seconds looking towards the untrained Distractor word forms.

An ANOVA with one within-participants factor of stimulus type (target vs. distractor) and one between-participants factor of condition (NA versus UK) revealed no main effect of stimulus type ($F(1, 30) < 1$) or of condition ($F(1, 30) < 1$), and no interaction between the two factors ($F(1, 30) < 1$).

As in the two previous experiments, planned comparisons were carried out within each group (Johnson & Jusczyk, 2001). Paired-samples t-tests revealed no difference between targets and distractors for either the UK-style (target mean = 6.25, distractor mean = 6.05; $t(15) < 1$) or the North American-style (target mean = 5.8, distractor mean = 5.73; $t(15) < 1$) conditions.

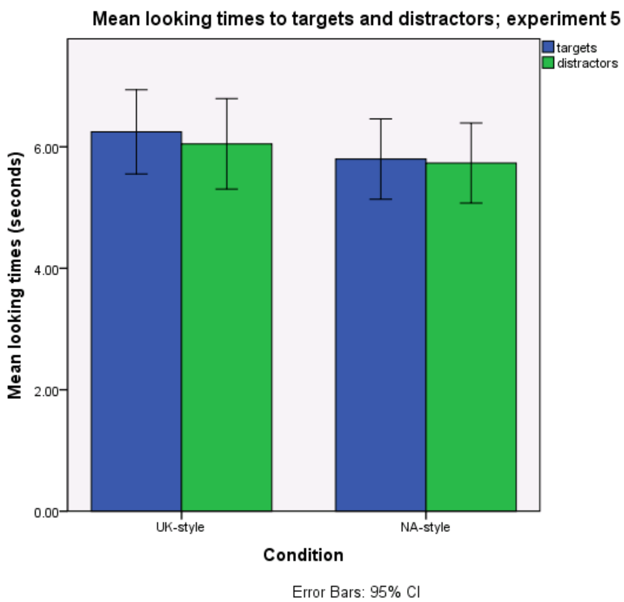


Figure 19: Mean looking times to Targets and Distractors using UK-style and North American-style speech.

Of the 16 participants in the UK-style condition, nine were found to show a target preference in overall looking times, while seven showed a distractor preference. In the NA-style condition, ten infants showed a target preference and six showed a distractor preference. A binomial test shows that this represents a normal distribution (UK condition; $p = .32$ NA condition; $p = .-17$), so it is not possible to say that significantly more infants showed a target preference.

As in Experiments 3 and 4, a similar analysis was carried out on the first test trial only, in order to establish whether infants' first looks showed a systematic difference.

An ANOVA using one within-participants factor of stimulus type and one between-participants factor of condition showed no main effect of stimulus type ($F(1, 30) < 1$) or of condition ($F(1, 30) < 1$), and no interaction between the two factors ($F(1, 30) < 1$).

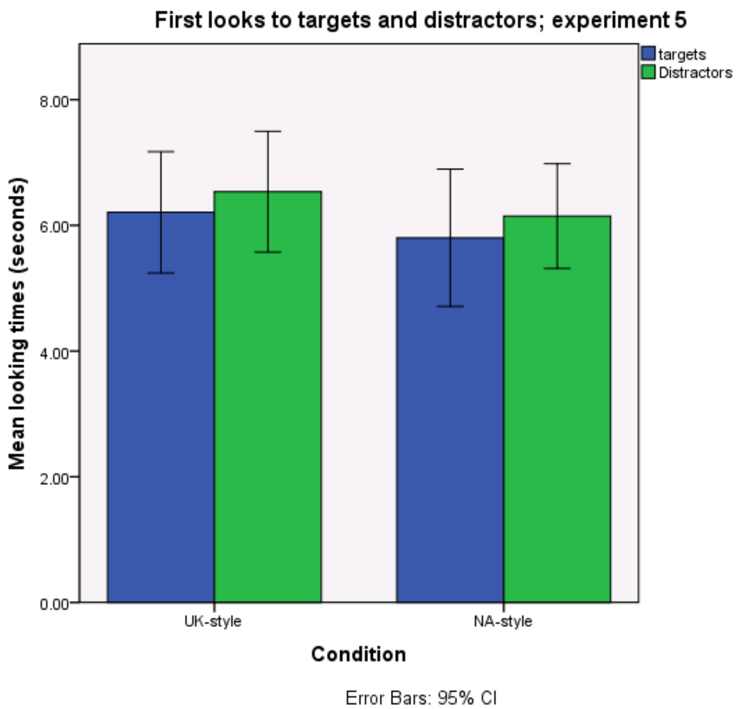


Figure 20: First look durations for targets and distractors using UK-style and NA-style speech.

In the UK-style condition, seven infants showed a target preference in their first looks, while nine showed a distractor preference. In the NA-style condition, nine infants showed a target preference and seven showed a distractor preference. A binomial test showed that these figures represent a normal distribution (UK-style $p = .32$, NA-style $p = .69$), so it is not possible to say that either group showed a consistent preference.

The Familiarisation phase required infants to spend a cumulative 45 seconds looking towards each of the two Target word forms. The number of trials taken to achieve this varies widely from one child to another, as a function of their interest in the stimuli. A child who attends more to the familiarisation trials, and thus who takes fewer trials to habituate, will have a longer average time per trial than a child who took more trials. The average looking time during the Familiarisation phase can thus be taken as a measure of the infants' interest in the training stimuli; trials in which the infant looked for less than two seconds were considered null, and removed from the analysis (as in Singh, et al., 2004). In the UK-style condition infants spent a mean time of 8.88 seconds looking towards the stimuli (std dev = 2.5 seconds), and in the NA-style condition they spent on average 9.96 seconds looking towards the stimuli (std dev = 3.8 seconds). A one-way ANOVA comparing the mean familiarisation times for the two conditions revealed no difference between the two groups ($F (1, 30) < 1$).

Discussion of Experiment 5

Experiment 5 aimed to compare how two different styles of infant-directed speech would affect word segmentation in British infants, with the idea that a more exaggerated IDS style would promote word segmentation. However, the results failed to show a difference in infants' responses to the two styles. This followed on from Experiments 3 and 4, which failed to replicate previous literature indicating that infants are able to segment words from fluent speech in their ambient accent from around seven and a half months of age (Johnson & Jusczyk, 2001; Jusczyk & Aslin, 1995; Schmale & Cristià, 2009; Schmale, et al., 2010; Singh, 2008; Thiessen, et al., 2005).

When this study began, it was expected to find successful segmentation in ten-month old infants in their ambient accent, and the subsequent publication of Schmale et al's similar studies (Schmale & Cristià, 2009; Schmale, et al., 2010) strengthened this prediction. It was also predicted that the use of unfamiliar accents, in both within-accent and across-accent conditions, would disrupt segmentation, resulting in shorter target looking times in unfamiliar accents than in the ambient accent. The results show that infants did not distinguish the targets from the distractors in the Plymouth-Plymouth conditions, even when the same speaker was used throughout (Experiment 4, Same Speaker condition), and even when the Passage/Word order was reversed to match more closely the procedures used in studies which did find successful segmentation (Experiment 4, Reverse Order). They also did not show signs of segmentation in any of the accented conditions. Experiment 5 attempted to examine one possible reason for this unexpected null result, by comparing two different styles of IDS, one of which was intended to mirror that used by American speakers. Once again, no significant difference was found between looking times towards targets and distractors, which would have indicated successful segmentation. Neither was there any difference in the familiarisation patterns between the two types of IDS.

However, this cannot rule out the possibility that the style of IDS used may have an impact on segmentation. The NA-style stimuli used in this study, while exaggerated in comparison to the UK-style stimuli, were nonetheless not as extreme as the stimuli used in Schmale et al.'s studies, and may simply not have featured clear enough cues for segmentation. Replicating the study with more extreme IDS might be sufficient to find successful segmentation, given that a study pitting American IDS against American Adult-directed speech (ADS) has shown that the exaggerated features associated with IDS allow infants to segment while the less exaggerated ADS does not (Thiessen, et al., 2005).

DePaolis et al. (DePaolis, et al., 2012) have used a very similar technique to that described here with eight- and nine-month old infants from Northern England and from America. In both cases they used IDS produced in the infants' ambient accent, but in a moderate, "UK-style" of speech, with only moderate pitch changes. They failed to find segmentation in either group, despite earlier findings from American studies in which infants at this age were able to show successful segmentation. This suggests that the style of IDS used in the task had an impact on the American infants' ability to segment fluent speech, and that the differences between the current study and the successful North American studies are not simply due to a developmental advantage for American infants.

It must be pointed out that there is a well-established (but as-yet unexplained) gap in vocabulary scores between British English and American English-speaking infants (Hamilton, Plunkett, & Schafer, 2000). Whether measured by parental report or by more objective measures, American infants' vocabulary is typically found to be slightly larger than that of their British peers throughout their first three years. This raises the possibility that there are cross-cultural differences in early linguistic abilities, which might help to explain why the English infants who took part in this study were unable to segment at ten and eleven months while their American peers have been found to segment successfully (Jusczyk & Aslin, 1995; Schmale, et al., 2010) at younger ages. However, DePaolis et al.'s finding that American infants were unable to segment the English-style but American-accented stimuli suggests that a broad difference in developmental ability is not the full story.

Nazzi et al. (In Revision) noted that French-speaking infants from Canada had shown successful segmentation at eight months (Polka & Sundara, 2012) while French-speaking infants from Europe had failed to do so at eight or twelve months, only succeeding at 16 months (Nazzi, et al., 2006). Interestingly, Polka and Sundara used two dialects of French, one produced by a Canadian speaker of French and one by a European French-speaker, and found that Canadian infants were able to segment both accents at eight months. However, when Nazzi et al. (In Revision) tested European French infants with Polka and Sundara's European French and Canadian French stimuli under identical conditions to those used with the Canadian infants, the European French still infants failed to show segmentation at eight and twelve months. They were able to elicit successful segmentation in the European French infants by extending the duration of the familiarisation phase, but only when presenting the passages of fluent speech in the familiarisation phase and the single-utterance word forms in the test phase. In contrast, the Canadian infants were able to segment with a shorter familiarisation period and when the words were used in the familiarisation phase and the passages in the test phase. This asymmetry suggests that the European French infants are using different processes to segment speech, and incurring a greater cost in segmentation than the Canadian infants.

Discussion of Infant experiments

Taken together, these studies point to an interaction between the style of IDS used in segmentation experiments and the ambient accent experienced by the infants. Infant-directed speech is marked by a number of features, including higher overall pitch, increased pitch variability, slower speech rate, and longer pauses after target words, all of which are thought to assist the segmentation of fluent speech (Fernald & Mazzie, 1991; Thiessen, et al., 2005), and these features are more exaggerated in American IDS than in British English (Fernald, et al., 1989) and some other European languages. Canadian French and European French appear to differ in much the same way, with Polka and Sundara's Canadian French stimuli being slower than the European French stimuli. Notably, their European French stimuli also differed from Nazzi et al.'s European French stimuli, having a slower speech rate, higher F0 frequencies, and more variable pitch contours. This suggests that both the ambient style of speech experienced throughout an infant's lifetime and the style of speech used in experimental stimuli may influence infants' ability to segment under experimental conditions, and may even interact to some degree.

In order to understand the effects of both experimental stimuli and ambient dialect on infants' performance, there are a number of possibilities for future research. Nazzi et al. (In Revision) used European and Canadian French stimuli with European French babies, mirroring Polka and Sundara's use of both accents with Canadian French infants. A similar set of cross-cultural studies using American and British English stimuli with American and British English infants, making use of carefully controlled experimental manipulations like those used by Nazzi et al, might establish the conditions under which

British English infants are able to segment. If, as Nazzi suggests, European French infants are recruiting different mechanisms to process fluent speech from those used by Canadian French infants, due to their lifelong exposure to a less-easily segmentable dialect, the same may be true of British English infants, as compared to American English-speaking infants. This being the case, it would be predicted that English infants can in fact segment at an earlier age than has currently been shown, but only under specific conditions, and with easily-segmentable stimuli. Tightly controlling the acoustic features of the stimuli would allow a fair comparison of the infants' abilities, separate from the impact of the style of IDS. Once the conditions under which English and American infants segment have been established, a set of studies using graded levels of IDS could then be used, in order to establish how much exaggeration in the typical features of IDS is required by American and British infants in order to be successful.

Behavioural studies may not be able to pinpoint the exact mechanisms being used to segment, and more generally, to process accented speech, since behavioural measures such as looking times are relatively imprecise. Another direction for future research is therefore to use more direct measures to approach this question. By using ERP (Event-related potentials), it should be possible to look more directly at segmentation at the neurophysiological level, and thus shed light on the kind of processes that may be used to process familiar and unfamiliar speech.

Chapter 8: Infants’ segmentation within and across accents; an ERP study

Introduction

As discussed in the previous chapter, behavioural studies show that variability in the speech signal, including variation due to regional and foreign accents, can disrupt infants’ processing of speech. This is seen particularly clearly in segmentation studies, which show that infants are able to demonstrate segmentation within a familiar accent at nine months, but not until 12 or 13 months across accents (Schmale & Cristià, 2009; Schmale, et al., 2010). However, the previous experiments failed to find successful segmentation even within a familiar accent in 10- and 11-month old British infants, supporting findings from other British (DePaolis, et al., 2012) and European (Nazzi, et al., In Revision) studies. Taken together, these results point to the importance of both infants’ linguistic experience and the stimuli used in a laboratory setting. They also highlight the importance of using sufficiently sensitive measures of infants’ processing, and the inherent difficulties in using experimental paradigms which rely on infants’ attention to the stimuli used.

Kooijman, Hagoort and Cutler (2005) were among the first to use ERPs to look at segmentation of words from fluent speech in infants. They used a passive listening task, in which 10-month-old infants heard a familiarisation phase consisting of isolated target words, followed by a test phase in which they heard passages of sentences containing the target words or counterbalanced distracter words. ERPs arising from the target and distracter words were compared in order to see whether the infants were able to distinguish between them. They found that in the test phase, the responses to familiar words differed from those to unfamiliar words between 350-500 ms after word onset, with familiar words eliciting less positive waveforms than unfamiliar words. This epoch coincides with the late PMN (phonological mapping negativity), and extends into the N400.

The PMN is typically found in response to phonological violations; Kujala, Alho, Ilmoniemi and Connolly (2004) give the example of a task in which participants are presented with a sentence in which the final word violates phonological expectations but is semantically and syntactically viable, such as "*During the powercut the house became quiet*" (where the expected final word is "dark"). This task elicits a PMN response, indicating a phonological violation. In infants, the PMN can be elicited when purely phonemic expectations are violated, in the absence of semantic or syntactic information. An ERP study with 14- and 20- month old infants (Mills, Prat, Zangl, Stager, Neville, & Werker, 2004) showed that 14-month olds exhibited a response similar to a PMN (although not identified as such by the authors) to nonsense words which did not sound like known words (such as "kobe"), but not to either known words or non-words which differed by only one feature from known words (such as "bear" and "gare"). However, the more experienced 20-month olds showed a similar response to both similar and dissimilar non-words.

The very well documented N400 is typically found in adults in response to semantic violations; for example, in the sentence "*the gambler had a streak of bad luggage*", the word "luggage" would elicit an N400 response (Connolly, Stewart, & Phillips, 1990) as well as a PMN response.

Kooijman et al.'s findings (2005) are particularly interesting when we look at the time-course of the responses. The target and distractor words they used had a mean duration of 710ms, and infants were found to differentiate between them between 350-500 ms after word onset; that is, before the end of the spoken word. Additionally, the target and distractors were words which were unlikely to be familiar to the 10-month-old infants, and they were presented only as auditory stimuli, with no visual referents. Kooijman et al. point out that this means that they are unlikely to have been processed as lexically meaningful words, but rather as word-forms, with no lexical or semantic information available. The authors do not explicitly identify the components likely to be responsible, but since the N400 is found in adults only in response to lexical and semantic violations, it is unlikely that the infants' response can be attributed to an N400 response; rather the reduced activity in response to the target word-forms must represent a reduced PMN, indicating that infants were able to recognise the phonological structure of the trained word-forms but not that of the entirely novel distractors. In adults the PMN peaks at or just before 350 ms, with the N400 peaking around or just after 400 ms, meaning that Kooijman et al.'s findings, between 350-500 ms, correspond more closely in timing to the adult N400 than the adult PMN. However, a number of studies have found that infant components occur slightly later than their adult counterparts (Haan, Pascalis, & Johnson, 2002), making the PMN a more likely candidate to explain Kooijman et al.'s findings.

In addition, because the infants are unlikely to be processing the semantic content either of the target and distractor words nor of the carrier sentences, and will therefore not be predicting the target or distractor words, we would not expect a phonological mismatch response to occur immediately after word onset, as it would in adults in response a sentence such as "*During the powercut the house became quiet"*. Instead, the initial phoneme of the target and distractor words serves as context, and, in the case of the target words, creates a phonological expectation about the phonemes that follow. The fulfilment of those expectations leads to the negative deviation in the PMN, and this also helps to explain why, in Kooijman et al.'s findings, it occurred slightly later than would be expected in adults. In other words, in adults who are drawing on semantic context to make predictions about the coming words, the PMN is elicited immediately after word onset, whereas in infants who do not have access to top-down semantic information, it is expected to be slightly delayed.

Goyet, de Schonen and Nazzi (2010) used a similar technique to that used by Kooijman et al. (2005) to look at whole-word segmentation from fluent speech in 12-month-old French infants. Their participants heard several tokens of four target words, all of which were iambic (weak-strong stress patterned) bisyllabic French words unlikely to be represented in the infants' lexicon. They then heard blocks of sentences containing the familiarised target words, and counterbalanced distractor words, allowing a direct comparison of the infants' neural responses to familiar and unfamiliar word-forms. This mirrors the behavioural segmentation experiments reported in the previous chapter, but Goyet et al. recorded ERP responses to the stimuli, time-locked to the onset of the target and distractor words, rather than relying on behavioural responses such as looking times. They also found a familiarity effect at 350-500 ms after word onset in the test phase, again indicating that segmentation was occurring at the phonological level. Since the mean duration of the target and distractor words was 464 ms (representing a slightly faster speech rate than that of the stimuli in Kooijman et al.), this suggests that segmentation was beginning during the presentation of the target words. Going one step further in their analysis, they also looked for differences between targets and distractors with the ERPs time-locked to the onset of the second syllable of the bi-syllabic words. This time, they found no effect of familiarity, even though the second syllable of the target and distracter words was stressed, and even though the second syllable should have been as familiar as the first. They interpreted this as indicating that segmentation was occurring at the word level and not simply at the syllable level, and that infants were therefore treating the word-forms as being word-like, even though they had no semantic information about them. This study indicates that infants at 12 months are capable of segmenting familiarised bisyllabic words from fluent speech, and that this segmentation is based on whole word-forms, not merely on the presence of a single familiar syllable.

To date the literature does not include any ERP studies with infants which look at the effect of accents on speech processing, but Rivera-Gaxiola and colleagues have used a similar approach to look at infants' discrimination of native and non-native contrasts (Rivera-Gaxiola, Silva-Pereyra, Klarman, Garcia-Sierra, Lara-Ayala, Cadena-Salazar, & Kuhl, 2007; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005) using an oddball paradigm. Adults, who can be considered expert users of their native language, are typically insensitive to deviant sounds which do not cross a phonemic boundary in their L1 (Dehaene-Lambertz, 1997), but infants who have not yet undergone the process of perceptual narrowing have previously been found to distinguish contrasts which are not used in their ambient language (Werker & Tees, 1984). In Rivera-Gaxiola et al.'s studies, seven-month-old and 11-month-old English-speaking infants heard a string of a repeated consonant-vowel (CV) syllable, which is typically perceived as /ta/ in Spanish and as /da/ in English. Interspersed with this, and making up around 10% of presentations each, were two deviant syllables, differing only in their voicing; one is phonemic in Spanish (known as Spanish /da/), but not in English (and thus is unlikely to be distinguished from the baseline syllable, English /da/, by adults), and the other is phonemic in English (known as English /ta/) but not in Spanish. They found that at 7 months, the infants discriminated both the native (English /ta/) and non-native (Spanish /da/) deviant syllables, as demonstrated by a difference in responses to the native and non-native syllables, compared to the baseline syllable, in two epochs; 150-250 ms from onset, and 250-550 ms after onset. This early effect is consistent with a secondary finding from Kooijman et al. (2005), from the familiarisation phase of their word segmentation study, although the time epochs studied in the two papers are not identical. They compared the responses to the first two and last two familiarisation trials in order to establish whether there was evidence of familiarisation during this phase, and if so, when that familiarity effect arose during the time-course of the word presentation. They found that in 16 electrodes, mostly in the anterior region, their familiarity criterion was reached around 160-190 ms after word onset; that is, within the earlier auditory epoch identified by Rivera-Gaxiola and colleagues. Similarly, Goyet et al. found a familiarity effect within the familiarisation phase of their experiment in one electrode (FC4) from 130 ms after word onset. This very early epoch is likely to reflect purely auditory processing; it corresponds roughly to a late N100, which has been found to index acoustic properties of speech such as voicing onset times (Steinschneider, Volkov, Noh, Garell, & Howard III, 1999).

By 11 months, in Goyet et al. (2010), the responses to the native contrasts had become stronger, while the responses to the non-native contrasts had disappeared. This indicates that both age groups were distinguishing the native contrast at the auditory and phonological level, and that perceptual narrowing was occurring between 7 and 11 months, so that the older children were able to ignore a difference which was not relevant to their linguistic experience. This suggests that by the age of 11 months, infants are able to disregard at least some non-contrastive variability, while still discriminating variations which are contrastive in their native tongue. The implication of this is that within-language variation, such as that resulting from a regional accent, is likely to have less of a disruptive effect on infants' processing by this age than the variation caused by foreign accents, which may include unfamiliar sounds imported from the speaker's own language into their L2 (which is the listener's native tongue). Thus, at 11 months a foreign accent and a regional accent would be predicted to result in different responses at the neural level, particularly in the later 250-500 ms epoch (which incorporates the 350-500 ms epoch used by Kooijman et al. and Goyet et al.), which reflects responses to familiar versus unfamiliar phonological forms. The earlier epoch identified by Rivera-Gaxiola and colleagues (Rivera-Gaxiola, et al., 2007; Rivera-Gaxiola, et al., 2005), being linked to purely auditory processing, is likely to show less distinction between different types of accent, as it is not associated with the learned phonemic or phonological characteristics of a specific language.

In summary, then, electrophysiological studies have been successful in pinpointing the segmentation of whole words from fluent speech, and in distinguishing between different types of variability in the speech signal, complementing the behavioural studies already discussed. The previous chapter described a set of behavioural experiments which failed to show behavioural evidence of segmentation in 10- and 11-month old British infants, despite a large body of evidence showing that American infants are able to segment words successfully at this age. Similarly, the experiments described failed to show the effects on segmentation or processing more generally of unfamiliar accents, although both British and American studies have previously shown that regional and non-native accents can affect preferences and processing within the first year of life. Given this failure to replicate previous findings using behavioural evidence, the studies described here, which show fine differences in the way familiar and unfamiliar speech stimuli are processed by infants, ERP approaches offer a way of investigating more closely the ability of UK-based infants to segment newly-learned word forms from fluent speech. While behavioural studies rely on infants' attention to both visual and auditory stimuli, ERP measures can be taken in the absence of overt attention, and also offer a much more sensitive measure of processing than is available through infants' behaviour. In addition, where a head-turn preference paradigm offers only a very coarse measure of segmentation, in which longer looking times to targets than to distractors is taken as evidence of segmentation and equal looking times fail to demonstrate segmentation, ERPs can provide information not only on whether segmentation is occurring, but on the time-course of the processing, so that we can tell whether the segmentation is occurring at a purely acoustic level, or whether phonological levels of processing are involved. The current study, then, uses ERP data to investigate segmentation in British infants in more detail than those described in the previous chapter.

The current study

The current study therefore uses event-related potentials as a measure of the developing ability to segment words from fluent speech, both within and across accents in eleven-month-old infants, with a view to examining differences in the auditory MMN period (150-250 ms after onset) as well as a later epoch representing a late PMN. Differences between targets and distracters in the MMN would indicate that infants are able to extract new word-forms from fluent speech and respond to similar-sounding isolated word-forms at the auditory level, while differences in the later epoch would indicate that they are differentiating between familiarised targets and unfamiliar distracters at a more sophisticated phonological level, and treating the targets as word-like. Infants were exposed to passages of fluent speech, each containing several repetitions of a target word, in either a Plymothian or a German accent.

They then heard single-word utterances of the target word and a previously unheard distracter word, in an accent that either matched that of the preceding passage, in the two within-accent conditions, or did not match, in the across-accent condition.

Table 9: Accent conditions used in Experiment 6

Passage accent	Word accent	Condition
Plymouth	Plymouth	Within-accent - Familiar
German	German	Within-accent - foreign
German	Plymouth	Across-accent

Only one cross-accent condition was used, with the passages presented in a German accent and the words in a Plymothian accent, in order to limit the number of trials each child heard. Most importantly, Schmale and colleagues found no effect of Familiarisation order (that is, whether infants were familiarised with a foreign accent and tested with a familiar accent or vice versa) at either 9 or 13 months, so it was considered unnecessary to include a condition in which infants heard the passages in a Plymothian accent and the words in a German accent. ERPs time-locked to the onset of the targets and distracters were examined in order to determine whether responses differed systematically, in three conditions based on the accent in which the passages and words were spoken.

Since behavioural studies (Schmale & Cristià, 2009; Schmale, et al., 2010) have found that infants are able to segment words from fluent speech within an unfamiliar accent, it was predicted that systematic differences between the target words and distracter words would be found in the two within-accent conditions (in which both the passages and words were heard in a German accent, or both in a Plymothian accent), mirroring the behavioural findings. Schmale and colleagues found segmentation across accents at 13 months when using a foreign accent (Schmale & Cristià, 2009), but not at nine months.

The experiments in the previous chapter failed to find successful segmentation in behavioural paradigms at 10 or 11 months, but Nazzi et al.'s work (In Revision) suggests that even those infants who have failed to show segmentation under the same conditions as those used in Schmale and colleagues' (Schmale & Cristià, 2009; Schmale, et al., 2010) and Polka and Sundara's (2012) work may in fact be able to segment under ideal conditions. Since neurophysiological paradigms are more sensitive than behavioural tests, and should therefore identify differentiation between stimuli at an earlier age than behavioural studies, it was also predicted that successful segmentation might be found in the cross-accent condition (in which the sentences were heard in a German accent and the words in a Plymothian accent), which mirrors that used in Schmale and Christia (2009). This segmentation effect should be shown by differences between responses to familiarised Target words and unfamiliar Distracter words across all three conditions, and is likely to be found at around 150-250 ms after word onset, as predicted by Rivera-Gaxiola and colleagues (2007) and Kooijman et al. (2005), since it is likely to occur at the auditory level. In the Familiar (Plymouth-Plymouth) accent condition, which broadly matches that used by Goyet and colleagues (2010), it is predicted that a Target/Distracter difference will be found in a later epoch, between 400-500 ms, matching the late PMN component in adults, and indicating that familiarised stimuli in a familiar accent are being treated as word-like, although as infants were provided with no referents for the word-forms, we would not expect to find components associated with semantic processing, such as the N400. This Target-Distractor difference is not predicted in the German-German condition, in which the accent of both the familiarisation and test phase is unfamiliar, or in the German-Plymouth condition, in which the accent switches between the two phases.

Method

Participants

53 infants (26 female and 27 male) were recruited via the Plymouth Babylab participant database. They had a mean age of 11 months (std dev = .55 months, range 9.57 – 12.29 months). All were living in monolingual English-speaking households in South Devon, and self-report questionnaires determined that all babies had at least one resident parent whose accent originated in the South-West of England. A further five infants were excluded due to technical issues resulting in missing data, and four due to reported exposure to non-UK accents or lack of exposure to South-Western accents. All parents of infants included in the analysis reported no exposure to German or other non-native accents. None of the participants had any known cognitive delays or hearing impairments, and none had been born more than six weeks premature.

Stimuli

Forty-two English words were chosen, to act as both target and distractor words (counter-balanced across participants). All the words were di-syllabic and trochaic, and began with a plosive consonant to facilitate segmentation; for example, tourist, carriage, and dwelling. They had a mean Celex frequency score of 4.17 (with a range from 0 to 26.92) and a mean age of acquisition score, using the Bristol/Gilhooly-Logie norms (Stadthagen-Gonzalez & Davis, 2006) of 449.16, corresponding to an age of approximately 8 years (Baayen, et al., 1993), meaning that our participants were very unlikely to be familiar with any of the words we used. For each target word, four sentences were created. In each case the word appears embedded within the sentences rather than in initial or terminal position (for example, “The **carriage** was pulled by two big white horses”, or “A bungalow is a **dwelling** but so is a mansion”). The sentences were recorded by two female native English speakers who had grown up in Plymouth (aged 31 and 40), and two female native German speakers (resident in South Devon) who had lived in Germany throughout their childhood and adolescence, and thus had distinct German accents, although both were judged to speak fluent English (aged 34 and 19). All recordings used infant-directed speech.

The average duration of the individual words was 809 ms, and the average duration of the sentences was 3,126 ms. This represents a slower speech rate than that used in either Goyet et al. (2010) or Kooijman et al. (2005). An independent samples t-test showed that the German words had a longer duration, at 866ms (std dev = 325 ms) than the Plymouth words, with a mean of 753 ms (std dev = 231 ms; $t = 5.8$, $p < .01$), but that the German sentences had a shorter duration (m = 2753 ms, std dev = 1110 ms) than the Plymouth sentences (m = 3488 ms, std dev = 1373 ms; $t = 5.39$, $p < .01$).

Design and Procedure

Infants were seated in a comfortable high chair, wearing the infant ActiCap, and provided with toys, and a silent video on a TV screen. Stimuli were played through speakers using E-Prime software, and parents or carers, who were present throughout the study, were asked to avoid speaking or otherwise making any noise. The task was a passive listening task, and no behavioural measures were taken. Participants heard 21 blocks, with seven blocks in each of three conditions; Plymouth-Plymouth, German-German, and German-Plymouth (where in each case the first accent is that of the sentences used in the Training phase of the block and the second is that of the single-word utterances used in the Test phase). The blocks were presented in random order by E-Prime (Psychology Software Tools, 1996). In line with Kooijman et al. (2005) and Goyet et al. (2010), no separate training phase was included so as to maximise the number of blocks each child heard before they became bored or fussy.

Each block began with a training phase, consisting of a set of four sentences, each containing the target word for that block. The four sentences were presented twice so that the target word for each block was heard eight times in connected speech. This was followed by a test phase, during which the participant heard four presentations each of the target word for that block, and a counter-balanced distracter word (such that words heard as targets by half the infants were heard as distracters by the other half), recorded as four separate single-word utterances. The test phase was always spoken by a different speaker from that heard in the training phase, even when the accent remained the same. The order of the target and distracter words within the test phase of the block was randomised, without constraints. The experimenter then cued the next block, allowing for a brief pause if the infant was fussy or noisy.

Each participant therefore heard 21 target words, and 21 distracter words. The words heard by one child as targets were heard by another as distracters, in order to eliminate item effects. The stimuli assigned to each condition were also counter-balanced, so that stimuli heard in a Plymouth accent by one child were heard in a German accent by another.

There was an inter-stimulus interval of 1000 ms between the sentences. After the final sentence in the block, there was an inter-stimulus interval of 2000 ms before the presentation of the first single-word utterance. Between single-word utterances there was an interval of 2500 ms. The whole experiment took around 25 minutes to complete. All completed blocks were included in the analysis, and some infants did not complete all 21 blocks.

ERP recording and Analysis

Scalp voltage data were collected using 30 actively amplified Ag/AgCl electrodes (actiCap, Brain Products GmbH) fitted to an elastic cap (See fig 21). The left mastoid electrode was used as a reference, and data were re-referenced offline to left and right mastoid activity. ERPs were calculated for a 900 ms period time-locked to the onset of the target words in the training sentences, and to the onset of target and distractor words in the test phase; this period included a baseline period of 100 ms before onset and 800 ms after onset. Baseline correction was applied using the first 100 ms (up to onset) as a reference. An offline bandpass filter of 0.1 – 30 Hz was applied. Semi-automated artefact rejection was carried out to remove segments contaminated with excessive ocular or muscular movements. Since the infants were free to move their heads, this resulted in a much larger loss of data than would be experienced in an adult study, but all children contributed at least 30 uncontaminated segments.

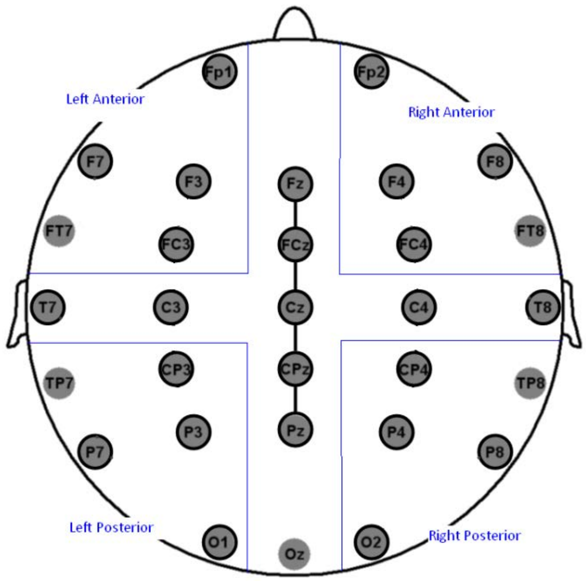


Figure 21: Schematic of Electrode montage, showing four quadrants used in electrode analysis.

Separate ERPs were calculated for each electrode site, each participant, and for the target and distracter in each of the three accent conditions. For analysis, 20 electrodes were divided into quadrants (as per Goyet et al., 2010). The Left Anterior quadrant was comprised of Fp1, F7, F3, FT7 and FC3, the Right Anterior quadrant was comprised of Fp2, F4, F8, FC4, and FT8, the Left Posterior quadrant was comprised of TP7, CP3, P7, P3, and O1, while the Right Posterior quadrant was comprised of CP4, TP8, P4, P8, and O2. Thus analyses will include within-participant factors of Condition (with three levels; Plymouth-Plymouth, German-German, and German-Plymouth), Anterior/Posterior position, Hemisphere, Electrode (with five electrodes in each quadrant), and Target vs. Distracter. Significant main effects of, and interactions with, the factors of Condition and Segmentation (Target vs. Distractor) will be reported, corrected using the Greenhouse-Geisser correction for violation of sphericity.

Results

On average, infants completed 19.74 blocks out of the total 21 (std dev = 2.88). After artefact rejection, infants contributed an average of 88.45 segments (std dev = 33.08) spread across the three conditions. A one-way Anova showed that there was no difference between the number of segments in each condition ($F(2) = .175$, $p = .84$)

Grand averages for target and distracter words in each of the three conditions are shown in figures 22, 24 and 26. A visual inspection reveals a broadly positive component peaking around 200 ms, during which the response to the distracters is more negative than to the targets, especially in the German-German and German-Plymouth conditions. This is followed by a negative component peaking around 450 ms, during which the targets are more negative in the Plymouth-Plymouth and German-German conditions, and less negative in the German-Plymouth condition. The earlier component seems to correspond to the MMN component in adults, which has also been found in infants (Rivera-Gaxiola, et al., 2007; Rivera-Gaxiola, et al., 2005), and which typifies responses to an unfamiliar auditory stimulus. The later component, between 400 and 480 ms, may correspond to that found by Goyet et al. (2010) in response to bisyllabic words in a whole-word segmentation task with 11-month-old infants; it also coincides with a late PMN in adults, which is found in response to phonological violations. Mean amplitude values were therefore calculated for two epochs, from 150-250 ms and from 400-480 ms from onset.

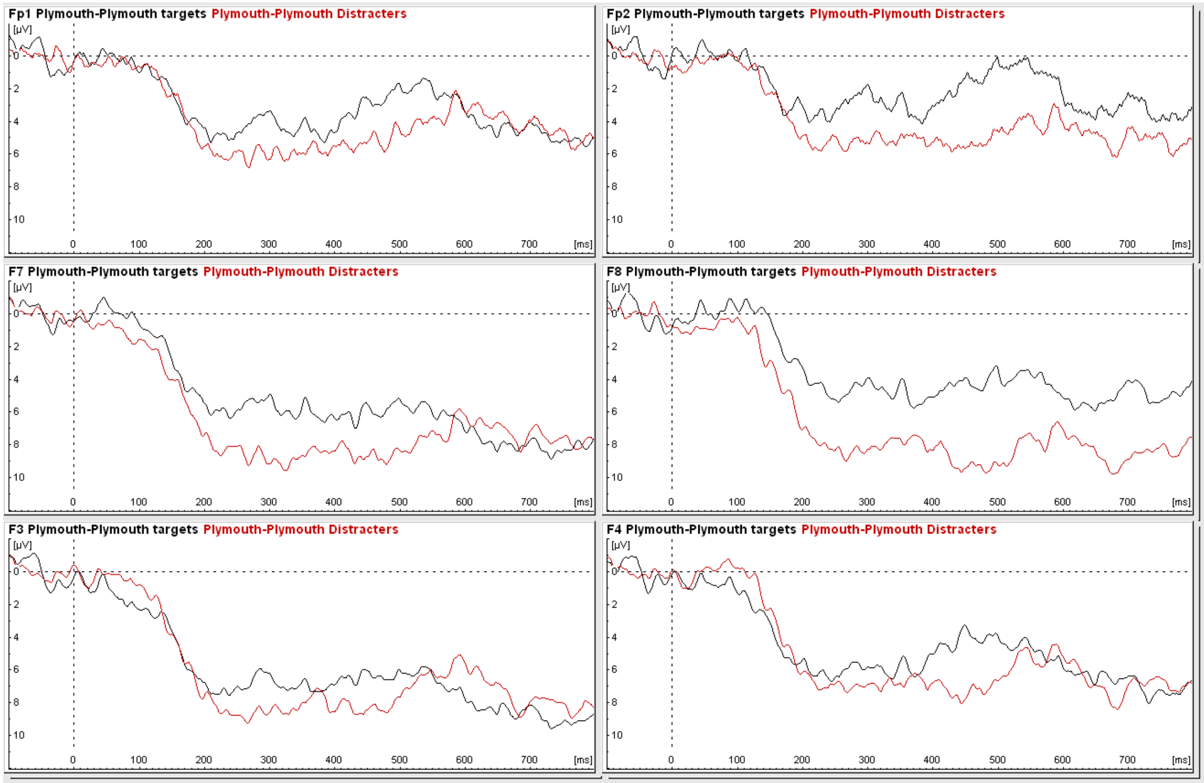
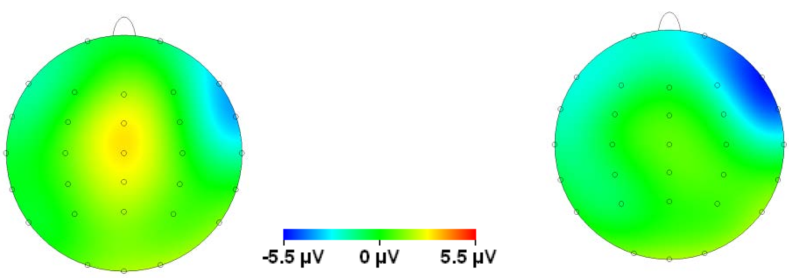


Figure 22: Averaged waveforms for targets and distracters across six anterior electrodes (FP1, FP2, F7, F8, F3, F4) for the Plymouth-Plymouth condition.



150-250 ms

400-480 ms

Figure 23: ERP head maps showing the difference between Target Responses and Distracter responses in the Plymouth-Plymouth Accent Condition between 150-250 ms (left) and between 400-480 ms (right).

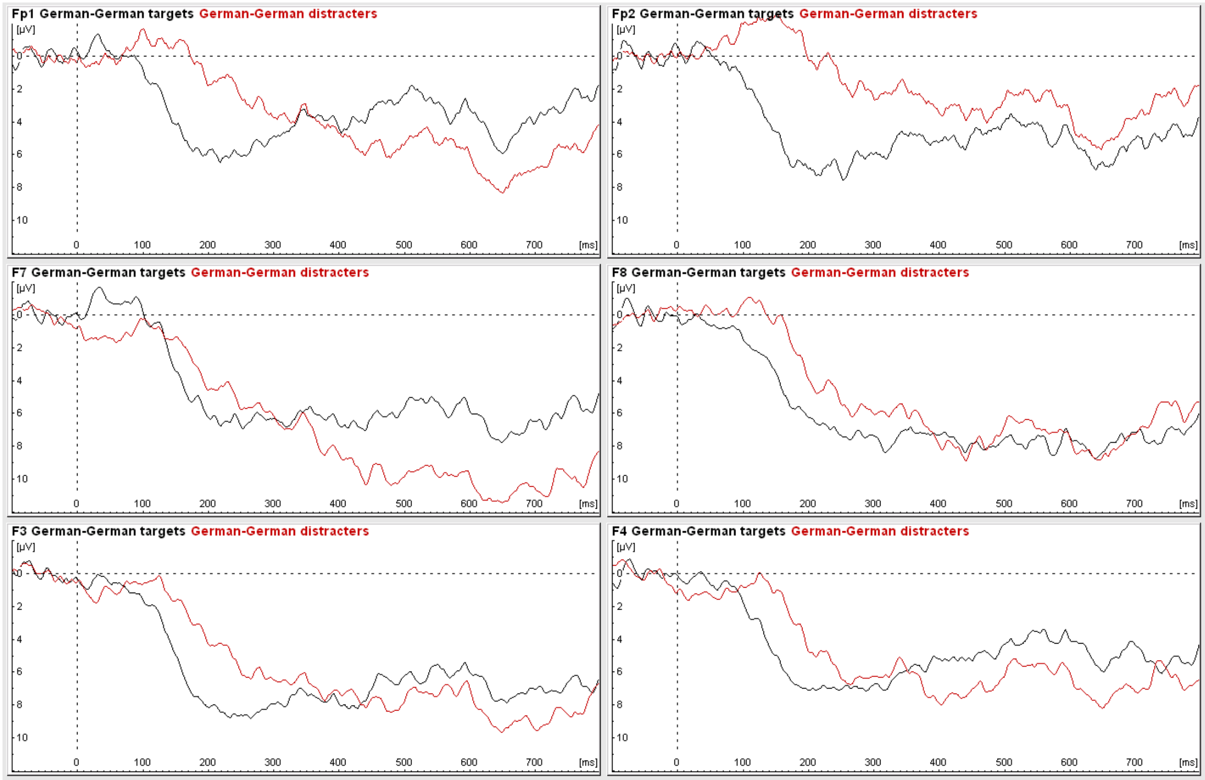
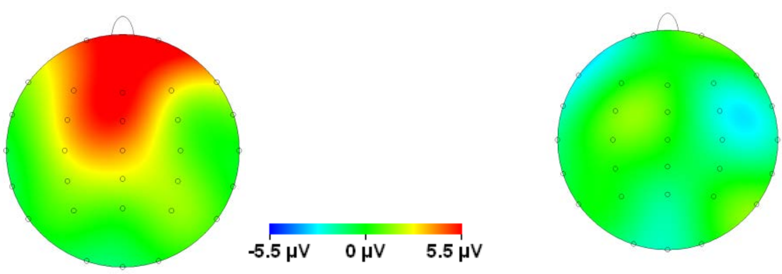


Figure 24: Averaged waveforms for targets and distracters across six anterior electrodes (FP1, FP2, F7, F8, F3, F4) for the German-German condition



150-250 ms

400-480 ms

Figure 25: ERP head maps showing the difference between Target Responses and Distracter responses in the German-German Accent Condition between 150-250 ms (left) and between 400-480 ms (right).

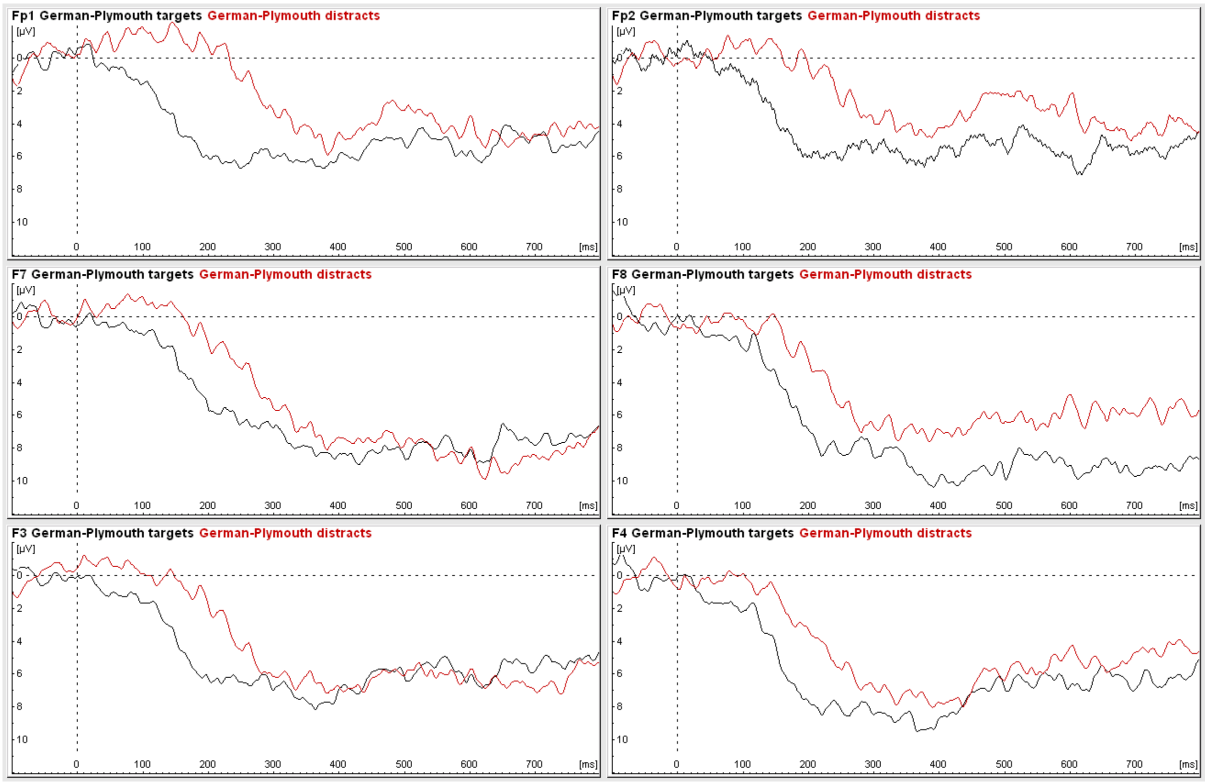


Figure 26: Averaged waveforms for targets and distracters across six anterior electrodes (FP1, FP2, F7, F8, F3, F4) for the German-Plymouth condition

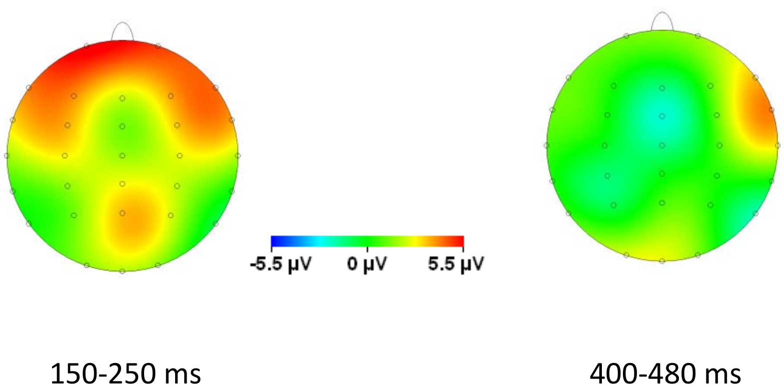


Figure 27: ERP head maps showing the difference between Target Responses and Distracter responses in the German-Plymouth Accent Condition between 150-250 ms (left) and between 400-480 ms (right).

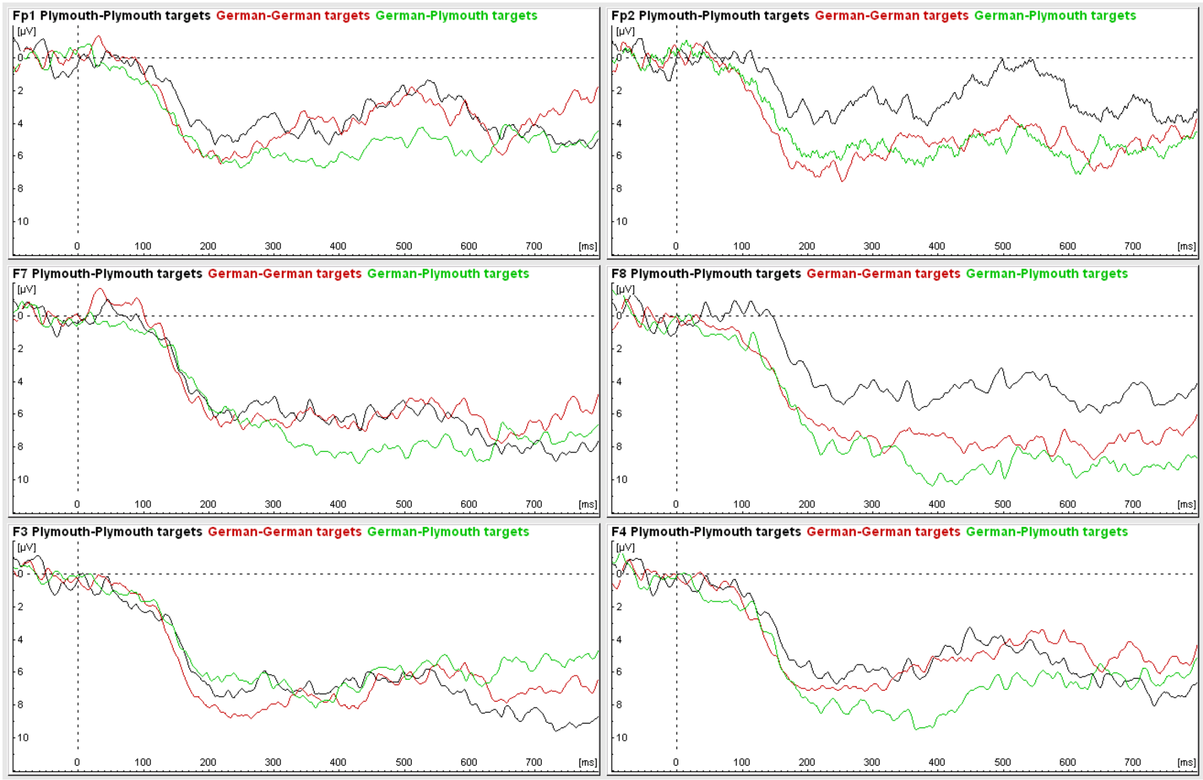


Figure 28: Averaged waveforms for targets across six anterior electrodes (FP1, FP2, F7, F8, F3, F4) for all three conditions

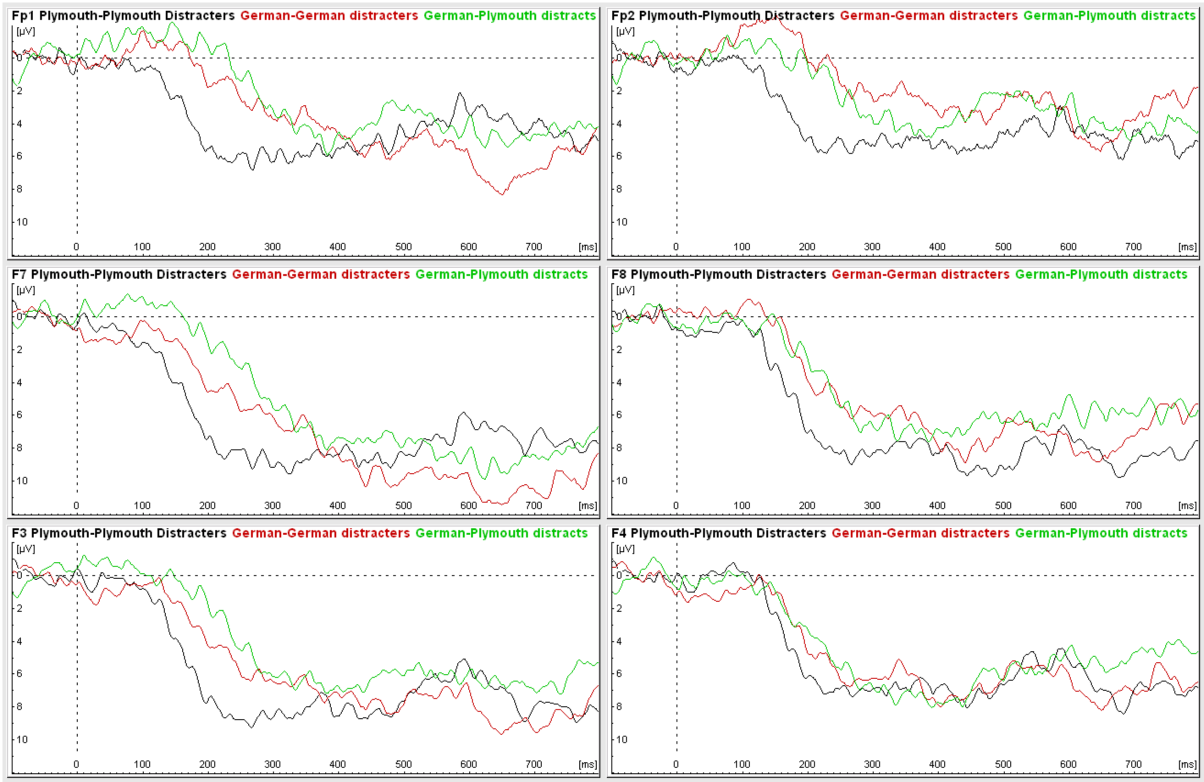


Figure 29: Averaged waveforms for distractors across six anterior electrodes (FP1, FP2, F7, F8, F3, F4) for all three conditions

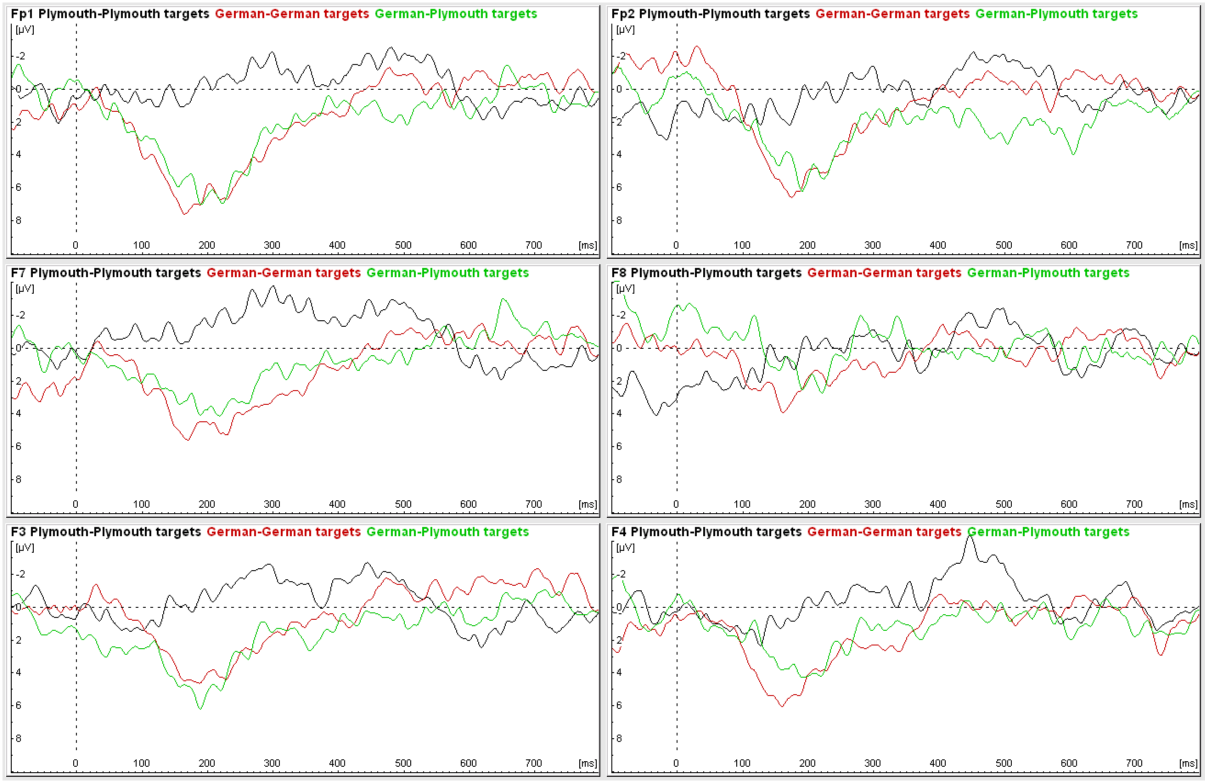


Figure 30: Difference waveform (Targets – Distractors) across six anterior electrodes (FP1, FP2, F7, F8, F3, F4) for all three conditions

Early period; 150-250 ms.

Analysis revealed main effects of Segmentation (F (1, 52) = 6.49, p = .014, η^2 = .11), with responses to targets being more positive than those to distractors. A significant interaction between Condition, Position and Segmentation (F (2, 104) = 6.50, p = .003, η^2 = .11) was also found, suggesting that localised patterns of responses to the targets and distractors differed across conditions; this could indicate that infants' ability to segment the stimuli depended on the accent condition.

A three-way interaction was also found between Position, Electrode, and Segmentation (F (4, 208) = 4.98, p = .004, η^2 = .09).

Planned pairwise comparisons were also carried out between each pair of Accent Conditions in order to explore the interaction between Condition, Position and Segmentation and establish how the three accent conditions affected segmentation.

When comparing the Plymouth-Plymouth and German-German conditions, a significant interaction was found between Condition, anterior/posterior Position, and Segmentation ($F (1, 52) = 11.75$, $p = .001$, $\eta^2 = .18$), again suggesting localised differences in the responses to targets and distractors between the two conditions. Post-hoc analyses were carried out in order to explore this interaction, and showed a significant interaction between Position and Segmentation in both the Plymouth-Plymouth ($F (1, 52) = 4.54$, $p = .038$, $\eta^2 = .08$) and German-German ($F (1, 52) = 7.29$, $p = .009$, $\eta^2 = .12$) condition. Separate analyses were therefore carried out on the anterior and posterior areas. These revealed a significant interaction between Condition and Segmentation in the anterior region ($F (1, 52) = .028$, $\eta^2 = .09$) but not in the posterior region ($F < 1$). Single condition analyses in the anterior region found a marginal effect of Segmentation in the German-German condition ($F (1, 52) = 3.91$, $p = .053$, $\eta^2 = .07$), with the targets eliciting more positive responses than the distractors, but not in the Plymouth-Plymouth condition ($F (1, 52) = 1.07$, $p = .307$, $\eta^2 = .02$). However, a comparison of the target responses within the anterior region in the two conditions showed no difference between responses to targets in the Plymouth-Plymouth and German-German conditions ($F (1, 52) = 1.66$, $p = .203$, $\eta^2 = .03$), while a comparison of the responses to distractors in the anterior region did reveal a marginal effect of Condition ($F (1, 52) = 3.98$, $p = .051$, $\eta^2 = .07$), with distractors in the Plymouth-Plymouth condition eliciting more positive responses than distractors in the German-German condition. The three-way interaction between Condition, Position, and Segmentation therefore appears to be due to a difference between the distractor responses (and not the target responses) for Plymouth-Plymouth and German-German conditions in the anterior region.

When comparing the Plymouth-Plymouth and German-Plymouth conditions, a three-way interaction was found between Condition, anterior/posterior position and Segmentation ($F (1, 52) = 11.65$, $p = .001$, $\eta^2 = .18$), once again suggesting that the accent condition was eliciting different localised patterns of response to targets and distractors. Post-hoc analyses revealed a significant interaction between Position and Segmentation in both the Plymouth-Plymouth condition ($F (1, 52) = 4.54$, $p = .038$, $\eta^2 =.08$) and the German-Plymouth condition ($F (1, 52) = 6.78$, $p = .012$, $\eta^2 =.12$). Analyses were therefore carried out in the anterior region, revealing a significant interaction between Condition and Segmentation ($F (1, 52) = 6.94$, $p = .011$, $\eta^2 = .12$), which was not found in the posterior region ($F < 1$). Single-condition analyses limited to the anterior region revealed a significant effect of Segmentation in the German-Plymouth condition ($F (1, 52) = 7.61$, $p = .008$, $\eta^2 = .013$), with the targets eliciting more positive responses than the distractors, but not in the Plymouth-Plymouth condition ($F (1, 52) = 1.07$, $p = .301$, $\eta^2 = .02$). However, separate analyses of the targets and distractors revealed a significant main effect of Condition in the responses to the distractors ($F (1, 52) = 8.1$, $p = .006$, $\eta^2 = .14$) but not in the responses to the targets ($F < 1$). This three-way interaction was therefore due to a difference in the anterior region between the responses to the distractors (and not to the targets) in the two conditions.

A four-way interaction between Condition, Hemisphere, Electrode, and Segmentation ($F (4, 208) = 2.76$, $p =. 037$, $\eta^2 = .05$) was also found. Post-hoc analyses to explore this interaction found a significant interaction between Condition, Hemisphere and Electrode in the responses to the distractors ($F (4, 208) = 4.02$, $p = .006$, $\eta^2 = .07$) but not in the responses to the targets ($F (4, 208) = 1.75$, $p = .157$, $\eta^2 = .03$). Further analysis revealed a significant Hemisphere by Electrode interaction in responses to the distractors in the German-Plymouth condition ($F (4, 208) = 3.18$, $p = .024$, $\eta^2 =.06$) but not in the Plymouth-Plymouth condition. This was due to a significant effect of Electrode in the right hemisphere ($F (4, 208) = 2.89$, $p = .031$, $\eta^2 = .05$) but not in the left hemisphere $F (4, 208) = 1.07$, $p = .365$, $\eta^2 = .02$). Paired-samples t-tests showed a significant effect of segmentation in four right-hemisphere electrodes in the German-Plymouth condition; FP2 (t (52) = 2.49, p = .016), F8 (t (52) = 2.28, p = .027), F4 (t (52) = 2.15, p = .037) and FT (t (52) = 2.49, p = .016). In all cases, the responses to the targets were more positive than those to distractors.

When comparing the German-Plymouth and German-German conditions, no significant interactions with Condition were found across the four quadrants, but a main effect of Segmentation was found (F (1, 52) = 8.893, p = .004, η^2 =.146), with targets eliciting more positive responses than distractors, as well as an interaction between Position and Segmentation (F (1, 52) = 20.26, $p < .001$, η^2 = .28). This was due to a significant main effect of Segmentation in the anterior region (F (1, 52) = 16.26, $p < .001$, η^2 =.24), with the targets eliciting more positive responses than the distractors, but not in the posterior region (F <1). This indicates that the anterior region of the brain was differentiating between the familiarised targets and the unfamiliar distractors in these two conditions.

Late period; 400-480 ms.

A whole-head analysis in the later epoch, between 400-480 ms from onset, found a marginal interaction between Condition, Position, and Segmentation (F (2, 104) = 2.994, p = .056, η^2 = .054), suggesting that in this later epoch there might also be differences, as in the earlier epoch, in the local patterns of activation in response to targets and distractors across the three conditions, although these differences are smaller and less reliable.

Planned paired comparisons were carried out for this epoch.

When comparing the Plymouth-Plymouth and German-German conditions, there were no effects of, or interactions with, condition, but there was a significant interaction between position and segmentation (F (1, 52) = 4.18, p = .046, η^2 = .07) in the four quadrants suggesting that patterns of activation to the targets and distractors differed across the brain. Post-hoc analyses were carried out to investigate this interaction, but no main effect of segmentation was found in either the anterior region (F (1, 52) = 2.56, P = .115, η^2 =.05) or the posterior region (F (1, 52) = .01, p = .92, η^2 =0). Separate analyses of the targets and distractors revealed that there was a main effect of Position in both the targets (F (1, 52) = 58.05, p < .001, η^2 = .53) and the distractors (F (1, 52) = 86.9, p < .001, η^2 = .63), with more positive responses in the anterior than the posterior region. The interaction is most likely due to differing amplitudes in the anterior and posterior regions. As indicated by the scalp maps in figures 23 and 27, the difference between targets and distractors is greater in the anterior region than the posterior.

When comparing the Plymouth-Plymouth and German-Plymouth conditions, there was a significant interaction between condition, position and segmentation (F (1, 52) = 5.25, p = .026, η^2 = .09), and an interaction between condition, position, hemisphere and segmentation (F (1, 52) = 6.81, p = .012, η^2 = .12) in the four quadrants. Once again this suggests a different pattern of local responses to targets and distractors across these two conditions. Post-hoc analyses to investigate the three-way interaction revealed that there was a significant interaction between position and segmentation in the Plymouth-Plymouth condition (F (1, 52) = 6.29, p = .015, η^2 = .11) but not in the German-Plymouth condition (F <1). In the Plymouth-Plymouth condition, this interaction was due to a marginal effect of segmentation in the anterior region (F (1, 52) = 4, p = .051, η^2 =.07) but not the posterior region (F < 1). In investigating the four-way interaction, a three-way interaction between position, hemisphere, and segmentation was revealed in the Plymouth-Plymouth condition (F (1, 52) = 4.38, p = .041, η^2 = .08) but not in the German-Plymouth condition (F < 1). This was due to a significant interaction between position and segmentation in the right hemisphere (F (1, 52) = 8.68, p = .005, η^2 = .014) but not the left hemisphere (F (1, 52) = 1.37,

$p = .248$, $\eta^2 = .03$). Further analysis revealed a significant effect of segmentation in the Plymouth-Plymouth condition in the right anterior quadrant ($F(1, 52) = 5.63$, $p = .021$, $\eta^2 = .1$) but not the right posterior quadrant ($F < 1$). In this quadrant, responses to the distractors were more positive than to the targets.

When comparing the German-German and German-Plymouth conditions, no significant effects of or interactions with condition or segmentation were found.

Discussion

The aim of this study was to examine the electrophysiological responses to trained target word-forms versus untrained distracter words within and across accent conditions in infants, in order to establish the extent to which the natural variation provided by an accent disrupts the normalisation of the speech signal, and thus infants' ability to segment fluent speech. Following work by Rivera-Gaxiola et al. (2007) and Goyet et al. (2010), it was predicted that infants would differentiate between the targets and distracters across all three conditions between 150-250 ms from word onset, indicating that at an acoustic level, they are differentiating between the familiarised targets and the unfamiliar distractors. In addition, it was predicted that they would differentiate between targets and distracters in the Plymouth-Plymouth condition from 400-480 ms, indicating that, in the Familiar Speech condition, they were treating the stimuli as word-like.

In line with the adult literature and with Rivera-Gaxiola et al.'s findings, a significant auditory mismatch negativity was found between 150-250 ms after onset, in response to the untrained distracter words, compared to the trained target words, across the three accent conditions. As in both Rivera-Gaxiola et al.'s work and that of Goyet et al., this effect was found to be confined to the anterior electrodes. This difference between responses to targets and distracters indicates that learning of the sounds comprising the target words has taken place at the auditory level, and that this is being generalised, in the German-Plymouth condition, from the accented speech to the familiar speech.

Most notably, separate analyses comparing the targets and the distractors across all three conditions found that the condition by segmentation interactions were due, not, as might be predicted, to differences in the responses to the targets (which would indicate differing levels of learning across the accent conditions) but to differences in responses to the distractors, with the distractors in the Plymouth-Plymouth condition eliciting responses very similar to those of the targets in all conditions. In other words, the finding that there was no main effect of segmentation in the Plymouth-Plymouth condition was not due to infants treating the targets in this condition as unfamiliar word-forms, as we would expect if they were failing to segment the target words from the sentences in the training phase; rather, the infants were responding to the distractors as if they were familiar words, as shown by the fact that the distractor words in the Plymouth condition elicited activity similar to that elicited by the target words in all three conditions. This was not the case in the German-German and German-Plymouth conditions.

These results are difficult to interpret, since they suggest at first glance that infants at 11 months are able to segment words spoken in an unfamiliar accent and across accents, but that they are unable to segment words within a familiar accent, but this does not seem consistent with existing research (for example, Schmale et al., 2010). Given the failure, in the previous chapter, to find behavioural evidence of segmentation within or between accents at 10 or 11 months, it is plausible that infants might also fail to show electrophysiological evidence of segmentation at this age within a familiar accent, but in this case we would also expect them to show no segmentation within an unfamiliar accent or across accents. Given the finding in this experiment that the infant brain does in fact differentiate between trained targets and untrained distractors both within an unfamiliar accent (in the German-German condition) and across accents (in the German-Plymouth accent), it must surely follow that they are also capable of doing so within a familiar accent. And indeed, this assumption is supported by the finding that responses to the trained target words do not differ across the three conditions; differences are found only in the distractors. This may suggest that in the German-German and German-Plymouth conditions, the distracters are being treated as unfamiliar stimuli because they are novel words presented in an unfamiliar accent in the German-German condition, while in the German-Plymouth condition, the training phase is presented in a German accent while the test phase, containing the target and distractor words, is presented in a Plymouth accent. This switch from one accent to another may be enough to cause the untrained distractor words to be treated as entirely unfamiliar. However, in the Plymouth-Plymouth condition, although the distracters are untrained, they are less unfamiliar or less unexpected than in the other two conditions. The fact that the responses to the targets do not differ across the conditions suggests a ceiling effect; in all conditions, infants may be treating the targets as familiar stimuli, showing that they have successfully segmented them from fluent speech, even in the German-Plymouth condition, which represents a switch from an unfamiliar to a familiar accent.

Kooijman et al. (2005) found reduced positivity in familiarised stimuli versus unfamiliar stimuli in this time epoch; this is the reverse of the pattern found in both the German-German and German-Plymouth conditions, in which the targets elicited more positive responses than the distractors. Rivera-Gaxiola et al. (2005) found increased positivity in response to discriminable deviant stimuli compared to non-deviant stimuli or contrasts which cannot be discriminated, in the same epoch. In the current study, greater amplitudes were found in the responses to distracters in the German-German and German-Plymouth accent conditions than to those in the Plymouth-Plymouth condition. Usually, the more complex a task and the larger the corresponding neurophysiological activation will be (Goodin, et al., 1983; Philiastides, et al., 2006). If we were to rate a priori the complexity of the task of having to segment words from connected speech, we would have expected the German-German and German-Plymouth conditions to represent a more difficult task than the Plymouth-Plymouth condition. In the case of the German-German condition infants are processing speech in an unfamiliar accent, and in the case of the distractors, they are also dealing with entirely unfamiliar words, making the processing of the distractors a complex task. In the German-Plymouth condition, they are having to switch from an unfamiliar accent to a familiar one, which in itself represents a challenging task. Thus there are two levels of familiarity to be considered; the overall familiarity of the accent in which they hear a given target or distractor word (that is, whether it is spoken in the infant's ambient accent or an unfamiliar one), and the within-block consistency (that is, whether the targets and distractors are spoken in the same accent as the training phase of the block, or a different accent). In the Plymouth-Plymouth condition, the distractors, although they are unfamiliar word-forms, are spoken in a consistent and familiar accent. In the German-German condition, they are spoken in a consistent but unfamiliar accent, and in the German-Plymouth condition, they are spoken in a familiar but inconsistent accent. The results of the current study suggest that in the Plymouth-Plymouth condition, the consistency and familiarity are enough to allow infants to treat distractor word-forms as familiar, whereas in the German-German and German-Plymouth conditions, the unfamiliarity or inconsistency of the distractors makes this impossible; instead, the added difficulty makes it possible to classify the distractors as unfamiliar easily. However, in all three conditions, they are able to treat the trained targets as familiar, due to their exposure to them in the training phase. This is also consistent with the existing behavioural literature (Schmale & Cristià, 2009; Schmale, et al., 2010), which finds segmentation within an unfamiliar accent at a later age than segmentation within a familiar accent, and that segmentation across accents only emerges later again; Schmale and colleagues found successful across-accent segmentation at 13 months when using a foreign accent.

Goyet et al. (2010) found that French infants showed a significant effect of familiarity between 350-500 ms after word onset, and they interpreted this as meaning that at twelve months, infants are capable not only of segmenting bisyllabic words from fluent speech, but that they are treating the trained word-forms as word-like. It was therefore predicted that the current experiment would find a similar result in the Plymouth-Plymouth condition, which closely corresponds to Goyet et al.'s task. The current results showed that in this condition, infants showed a significant effect of familiarity in the right anterior quadrant between 400-480 ms after word onset. Goyet et al. (2010) also found segmentation effects in the right anterior quadrant, however, they also found a significant target/distractor difference in two posterior electrodes (C6 and CP4); no effects were found in posterior regions in the Plymouth-Plymouth condition of this experiment.

The German-German and German-Plymouth conditions did not differ significantly from the Plymouth-Plymouth condition, but despite this, a localised significant effect of segmentation was only found in the Plymouth-Plymouth condition and not in the other two conditions. This finding must be treated with caution, but may indicate that at this age, while infants are able to treat trained word-forms as word-like in optimal conditions (that is, in a familiar accent), they struggle to do so if there is accent-based variability in the speech signal, even though they are able to distinguish between familiar and unfamiliar word-forms at the auditory level.

Overall, then, the results of the current study offer only limited support for Schmale and colleagues' behavioural findings (Schmale & Cristià, 2009; Schmale, et al., 2010). These two studies, which used a head-turn preference procedure, found that infants showed a preference for trained target words over untrained distractors within a familiar accent at nine months, and across accents as twelve (for regional accents) or thirteen (for foreign accents) months. The current study suggests that the infant brain is able to discriminate trained from untrained word-forms at the auditory level (as demonstrated by the findings in the early epoch) both within and between accents, but the responses suggest only a marginal tendency to discriminate at the phonological level within a familiar accent, but not within an unfamiliar accent or across accents. As discussed in the previous chapter, infant-directed speech may be an aid to the segmentation of new word-forms from fluent speech (Thiessen, et al., 2005). Informal comparisons between the stimuli used by Schmale and colleagues and those used in the current study suggest that the former are spoken in a more exaggerated infant-directed register than the latter, making the segmentation task in the current study harder than that in Schmale et al.'s studies. It is possible that this increased difficulty is, in part, responsible for the lack of discrimination found in the later epoch in this study.

In order to explore this possibility, one potential course of action would be to replicate the current study but using stimuli which are rated by expert listeners as highly infant-directed. Thiessen et al.'s (2005) work would predict that this would aid whole-word segmentation, allowing infants to discriminate trained from untrained words at the phonological as well as the auditory level, so that responses to targets in the later epoch would be significantly reduced in amplitude, compared to the responses to the distractors. The findings of the current study and those of Schmale and colleagues would further predict that this effect would be greater within a familiar accent (the Plymouth-Plymouth condition) than across accents (the German-Plymouth condition). A further line of enquiry would involve the use of stimuli ranged along a continuum of infant-directedness, so that some are spoken in adult-directed speech, some in extremely infant-directed speech (that is, with a much slower speech rate, a much higher overall pitch, and much greater pitch variation within sentences), and some falling in between the two extremes.

If infant-directed speech does indeed aid segmentation, it would be predicted that target words trained in a more infant-directed register would elicit reduced responses as compared to distractor words, while those trained in an adult-directed register would show responses with a greater amplitude, with little or no distinction in the phonological epoch between targets and distractors. Given the results of the current study, we would expect to see discrimination in the auditory epoch even in the adult-directed condition. Indeed, it might be predicted that discrimination would be greater in the adult-directed condition, which represents a greater level of difficulty, than in the more infant-directed condition. In this study, there was a main effect of segmentation in the more difficult German-Plymouth condition but not in the easier Plymouth-Plymouth condition, and the lack of discrimination in the easier condition appears to have been due to infants responding to distractors as though they were familiar, rather than because they treated targets as unfamiliar. A more difficult condition, either a cross-accent condition or a condition using adult-directed speech, may make it easier for infants to classify distractors as unfamiliar, resulting in a larger effect of segmentation in the early epoch.

Chapter 9: Discussion of infant work

The previous two chapters have described two sets of experiments looking at the processing of accented speech in infants towards the end of their first year of life. Both used segmentation of disyllabic word forms from fluent speech as a measure of infants' ability to process speech within and across familiar and unfamiliar accents, with varying levels of success.

The first set of experiments, described in chapter 7, used a behavioural approach, in the form of the Head-Turn Preference procedure, which has been widely used to look at infants' ability to segment speech in the presence of variability (Johnson, et al., 2003; Nazzi, Jusczyk, & Johnson, 2000; Schmale & Cristià, 2009; Singh, 2008). However, where previous work has consistently found that infants are able to segment successfully from seven and a half months when dealing with familiar speech, and to overcome variability from nine months on, the experiments described here failed to find segmentation at ten and eleven months, even using a single speaker with a familiar accent, and regardless of the order of presentation of fluent passages and individual words. An informal comparison of the stimuli used in a number of successful experiments (Johnson & Jusczyk, 2001; Schmale & Cristià, 2009; Schmale, et al., 2010) as well as those in which segmentation was not found (DePaolis, et al., 2012, and the studies described in the two previous chapters), suggested the possibility that the style of Infant-Direct Speech used might have an impact on infants' performance. This seems consistent with the findings of other studies which failed to find segmentation (DePaolis, et al., 2012; Nazzi, et al., In Revision), as well as with a cross-cultural study of IDS, which found that American IDS featured more extreme prosodic variation than that of several other cultures (Fernald, et al., 1989). Experiment 5 in chapter 7 therefore pitted two different styles of IDS against each other in a further head-turn preference study, but again failed to find successful segmentation.

The second study, described in chapter 8, used ERP to analyse the time-course of the infant brain's response to a passive segmentation task within and across accents. Two epochs were investigated, the first of which is thought to index auditory/acoustic processing, and the second of which represents phonological processing, and specifically is thought to provide a measure of infants' ability to treat sounds as word-like (Goyet, et al., 2010). Successful segmentation was found across the three accent conditions (Plymouth-Plymouth, German-German, German-Plymouth) in the early, auditory epoch, but, unexpectedly, infants showed greater distinction between targets and distractors in the German-German and German-Plymouth conditions than in the Plymouth-Plymouth condition. However, a deeper analysis revealed that this was not due to infants failing to recognise the trained targets in the familiar speech; rather, the findings indicate that they were equally successful in recognising the targets in all three conditions (suggesting a ceiling effect), and that the differences between conditions were due to their treating the distractors as more familiar in the Plymouth-Plymouth condition than in the other two conditions. In the later, phonological epoch, the findings are only marginal, and must be treated with caution, but seem to indicate that infants are more able to treat the newly-trained target words as being word-like in the familiar speech condition (as found by Goyet, et al., 2010) than when listening to a foreign accent, or when switching between accents.

In terms of the failure of these experiments to exactly replicate the findings of previous studies such as those carried out by Schmale and colleagues (Schmale & Cristià, 2009; Schmale, et al., 2010), there are two obvious lines of enquiry to follow. The first of these addresses the style of speech used in segmentation studies in general, and the second addresses the question of how different accents, and different types of accents, might be processed in infants.

Nazzi et al. 's (In Revision) study with French-speaking children indicates both that stimuli recorded in a more exaggerated IDS style (typical of Canadian French but not European French speakers) are easier to segment, and that European French children are at a disadvantage compared to their Canadian French peers; they are able to segment at eight months, but only with a simplified task. The differences found between Canadian French and European French IDS are mirrored in American English and UK English (DePaolis, et al., 2012), and results of the behavioural study reported in chapter 7 are also in line with earlier work which failed to find segmentation in 8-month old European French infants (Nazzi, et al., 2006). A cross-cultural study replicating Nazzi et al.'s work (In Revision), using English-speaking British and American infants, and both British and American style stimuli, might therefore allow the impact of both the lab-based IDS used and the ambient speech heard by the infants to be assessed. Careful control of the acoustic qualities of the stimuli would allow for a gradation of infant-directed speech, in order to establish the degree of infant-directedness necessary for segmentation to occur. It might also provide a scale against which to compare stimuli, so that rather than simply describing stimuli as being spoken in "infant-directed speech", researchers could give a more objective measure of the degree of infant-directedness of their stimuli.

If, as suggested by Polka and Sundara (2003, 2012) and Nazzi et al. (In Revision) it is correct that regional variations of French elicit different segmentation strategies, then it is fair to assume that the same may be true of other languages, including English; indeed, given the huge variety of regional accents of the English language, both within the United Kingdom and in other English-speaking countries, English may be particularly susceptible to this phenomenon. In Nazzi et al.'s work (In Revision), it has been pointed out that the prosodic quality of Canadian French varies from that of European French, and that this may in part be responsible for differences in segmentation studies. In attempting to replicate these results in English, it would therefore be advisable to use infants exposed to varieties of English with variant prosody, such as Welsh-accented English (Peppe, 2011). In order to be truly comparable, the stimuli used for the different populations would need to be carefully controlled, in order to avoid confounding effects from accent or style of Infant-directed speech. In the absence of a truly neutral accent of English (particularly when looking at international variations), two possible solutions present themselves. Either, stimuli could be recorded by a native speaker on each region from which infants are tested, but with acoustic characteristics such as stress, speech rate, overall pitch and pitch variation carefully controlled, or an artificial language or accent could be designed in such a way that these characteristics are controlled. This would allow the various cues which have been shown to be effective for segmentation in infants, such as prosody (Echols, et al., 1997), phonotactics (Mattys, et al., 1999), stress (Jusczyk, et al., 1999b), and distributional probabilities (Thiessen & Saffran, 2007), to be examined individually, across several different regional varieties of English. If infants exposed to different varieties of the same language do indeed develop different strategies for segmentation, we would expect to see this reflected in the age at which they succeed in using different cues to segment newly-learned word forms from fluent speech.

A second line of enquiry is more specific to the processing of accented speech. In chapter 8, a German accent was used as the unfamiliar accent. A native German speaker speaking in English will typically produce a number of phonological shifts, compared to a native English speaker, but German, like English, is generally thought of as a stress-timed language, so German accents may only present limited supra-segmental variation compared to an English accent. Since there is evidence that stress cues are used in order to assist in the task of segmentation (Jusczyk, Cutler, & Redanz, 1993; Thiessen & Saffran, 2007), it might be expected that German accents are easier to segment than an accent produced by a native speaker of a syllable-timed language such as French. If the ERP study with infants reported in chapter 8 were replicated using a French accent rather than a German accent, we might expect larger differences between the conditions, with less successful discrimination between targets and distractors when the French accent is used in either the training phase alone, or both training and test phases. Additionally, Schmale and colleagues' work (Schmale & Cristià, 2009; Schmale, et al., 2010) suggests that regional accents may be easier to segment than foreign accents, so if a regional accent were used instead of the German accent, this might also affect the findings. A number of studies have highlighted infants' ability to discriminate small phonological shifts at the auditory level (Goyet, et al., 2010; Rivera-Gaxiola, et al., 2007; Rivera-Gaxiola, et al., 2005), leading to a prediction that infants would differentiate between familiar speech and regionally-accented speech in the early epoch (150-250 ms after word onset) identified in chapter 8. However, perceptual narrowing within the first year of life allows infants to ignore differences in the acoustic input which are not contrastive in their native language. Rivera-Gaxiola et al. found that by 11 months, infants did not distinguish non-native contrasts in a later epoch (250-550 ms after onset), although at 7 months they did discriminate during this epoch. We might therefore expect to find differences between the responses to familiar speech and a regional accent in the early epoch, but not in the later epoch.

Testing infants across a range of ages might also allow a clearer understanding of the development of their ability to discriminate, and to generalise across, regional and foreign accents towards the end of their first year.

These two sets of studies add a new dimension to the existing literature on infants' processing of accented speech, since previous findings have suggested that infants (largely in North America) have no difficulty in segmenting speech within their own accent and even in unfamiliar accents or across accents from the age of around eight months. The current studies suggest that, at least in British infants, the task is more complex than previously thought, thus challenging previous findings. The wider implications of this are that we may need to be cautious about generalising findings across cultures; infants' linguistic context, beyond simply their maternal tongue may have a significant impact on their language development.

In addition, these studies have shown, for the first time, electrophysiological evidence of segmentation both within an unfamiliar foreign accent and, perhaps more interestingly, across accents, in infants within their first year of life. These electrophysiological findings compliment and extend the behavioural findings made by Schmale and colleagues, and the ERP findings of Nazzi and colleagues.

Chapter 10: General Discussion

This thesis has addressed the question of the processing of accented speech, in adults and in infants, through the use of two sets of studies. The first set used ERP measures of the adult brain's electrophysiological activation while listening to stimuli in familiar, regional and foreign accents, while the second looked at infants' ability to segment familiar and accented speech, using a combination of behavioural and electrophysiological methods. These two sets of studies can be seen as addressing two endpoints in a developmental process; in early infancy, infants struggle to recognise newly-learned (Schmale & Cristià, 2009; Schmale, et al., 2010) or familiar (Best, et al., 2009) words across accents, as evidenced by the fact that they do not show a preference for accented sentences containing trained target words at nine months (Schmale & Cristià, 2009), or for lists of familiar accented words at fifteen months (Best, et al., 2009). However, by adulthood, while unfamiliar accents may cause initial disruption in processing (Bradlow & Bent, 2003; Floccia, et al., 2009b), we are generally able to accommodate the accent-based variation and understand accented speech. This accommodation, however, incurs a processing cost, which may manifest in slower reaction times (Adank & McQueen, 2007; Munro & Derwing, 1995) and in reduced accuracy (Gass & Varonis, 1984).

The findings of the experiments described in this thesis show further evidence for this developmental change. The ERP findings reported in chapter 8 indicate that in the phonological phase of processing, infants did not distinguish between trained target words and untrained distractors in a foreign accent, while there were some indications that they were able to do so in a familiar accent. According to Goyet et al. (2010) discrimination during this period indicates an ability to treat the newly-trained word-forms as being word-like, so the finding of a localised difference between targets and distractors in the Plymouth-Plymouth condition suggests that by this age, infants are starting to segment familiar speech into word-like units. No such difference was found in the foreign accent. However, by adulthood we generally have little difficulty in understanding foreign-accented speech, and both previous literature and the findings reported in chapters 3 and 4 support the idea that with experience, we are able to develop mechanisms to allow us to normalise variant speech. The behavioural studies discussed in previous chapters (Schmale & Cristià, 2009; Schmale, et al., 2010; Singh, et al., 2004) suggest that these mechanisms start to emerge around the end of the first year of life, but a study with four and seven year olds suggests that between these ages, children are still developing the ability to normalise phonetic variability, thus "converting" accented speech into familiar speech (Nathan, Wells, & Donlan, 1998). Nathan et al. asked London-based children to identify words spoken in a Glaswegian accent. They found that at four years old, children most often gave phonetic responses, that is, they repeated the Glaswegian-accented words as they heard them and often incorrectly defined the target word. In contrast, by seven years old, children most often gave phonological repetitions, that is, they repeated the target word in their own accent and were able to accurately define it. Further work is required in order to understand the time-course of this development, and to track the changes in the processing of accented speech, from inability to recognise accented word-forms in infancy through to adult-like normalisation of accent-based variation in childhood.

Historically it has often proved difficult to compare infant studies with adult work, since with infants, we are constrained to undemanding paradigms which measure spontaneous behaviour such as direction of gaze. The development of ERP studies allows the use of passive-listening tasks without the need for overt instructions or deliberate responses, making it possible to compare findings across different age groups. Replicating the adult ERP experiment described in chapter 3, which used a passive listening task in home, regional and foreign accents, with participants ranging in age from late infancy through to adolescence would provide insight into the lifetime progression in the processing of accented speech. The work of Nathan et al. might lead to a prediction that adult-like processing would be achieved by the age of seven, but Kerswill and Williams (2000) suggest that children may be in a critical period for the acquisition of regional accents up until around puberty; if we assume that acquisition and perception are closely related, then this might suggest that children's processing of accented speech might not stabilise in an adult-like state until the end of this critical period.

Both Nathan et al.'s study and Kerswill and Williams' work examined comprehension or production of regional variations rather than foreign accents. Given the evidence presented in this thesis that in adults, foreign accents seem to be processed differently from regional accents, there are two possible developmental paths. Either infants process foreign accents differently from regional accents from the earliest days, or in the early stages of linguistic learning they do not differentiate, and this difference emerges later. Schmale et al. (Schmale & Cristià, 2009; Schmale, et al., 2010) found that infants are able to segment across regional accents at twelve months and foreign accents at thirteen months, suggesting that at the end of the first year foreign accents are being treated differently. However, these studies used only one example of a foreign accent (a South American Spanish accent) which diverges on a number of levels from the infants' ambient accent, and one regional accent (Southern Canadian), which was chosen specifically because it diverges only minimally from the ambient accent. Any differences in their processing of the foreign and regional accent could therefore be due to the degree of difference from the home accent rather than to the type of accent used. Not only this, but infants were tested with the regional accent only at nine and twelve months, and with the foreign accent only at nine and thirteen months. This set of studies, then, cannot be taken as definitive evidence of a categorical difference between infants' processing of regional and foreign accents at this age, since they were not tested with accents matched for strength, at the same age. The current dearth of studies which directly compares infants' processing of foreign and regional accents makes it difficult to pinpoint the earliest existence of this difference. Girard et al. (2008) and Floccia et al. (2009) did make a direct comparison, by asking children at five and six years (Girard, et al., 2008) or five and seven years (Floccia, et al., 2009a) to categorise accented speech. Floccia et al., who controlled for strength of accent, found that at seven, children categorised the foreign accent much more accurately than the regional accent, while at five years, their tendency to do so did not reach significance. Although these studies once again used only a single exemplar of a regional or a foreign accent, and thus may not generalise across a whole category, they may indicate at between the ages of five and seven, children are developing the ability to distinguish between different types of accents. This may suggest that by this age, different normalisation processes are recruited for foreign and for regional accents. It is as yet unclear at what age these different processes emerge, but if they are reliant on the ability to classify an accent as regional or foreign, we might expect to find them emerging between the ages of five and seven years (Floccia, et al., 2009a). Prior to this, we might expect to find that differences in processing are dependent on the degree of difference between the accent being processed and the infants' ambient accent (White & Morgan, 2008).

This raises a further question, discussed in chapter 1 of this thesis; that of the role of the identification of an unfamiliar accent in its processing. Some automated speech recognition systems, including Apple's Siri, have attempted to use an algorithm allowing them to identify the accent being used, in order to allow the system to use an accent-specific set of normalisation routines (Angkititrakul & Hansen, 2006; Hansen & Arslan, 1995), but with limited success. Indeed, even a system attempting to correct for a given known accent, rather than trying to identify the accent used, will still result in a high error rate (Ahmed & Tan, 2011). As yet, there has been little work looking at whether human listeners' ability to identify an accent impacts on its intelligibility and comprehensibility, in either children or adults. Mullenix and Pisoni (1990) have shown that talker-specific information can help or hinder phonetic processing; they asked participants to classify either the gender of speakers or the word being presented (either "bad" or "pad"), thus encouraging them to attend to either talker-specific characteristics or to acoustic information, and to disregard the other dimension. They found that word classification was slower when the speaker's gender varied randomly, and faster when "bad" was always spoken by one gender and “pad” by the other. This suggests that indexical information is processed alongside acoustic-to-phonetic processing, and informs normalisation procedures. This is supported by the finding that listeners are slower to identify words (Nusbaum & Morin, 1992) or vowels (Assmann, et al., 1982) when the stimuli are spoken by multiple speakers rather than just one. Interestingly, it is not only the raw acoustic input which influences this effect, but also listeners’ expectations; Magnuson and Nusbaum (Magnuson & Nusbaum, 2007) exposed listeners to vowels produced by a single speaker or by two different speakers, and manipulated their expectations about how many different speakers they would hear. Those who believed they were hearing two speakers in the single-speaker condition identified the vowels more slowly than those who were expecting only one speaker, and in fact performed as if there were in fact two different speakers. This suggests that the integration of speaker-dependent information into speech processing relies in part on top-down processes, including a listener’s expectations about the speech they hear. Further evidence for the importance of expectations in processing speech, and specifically accented speech, is provided by two studies which paired images of people of different races with speech stimuli (Rubin, 1992; Staum Casasanto, 2008). In the first, speech in a standard American accent was understood less well when paired with images of Asian faces than when paired with Caucasian faces (Rubin, 1992), and in the second, a word pronounced /mæs/ was interpreted as “mass” when paired with a Caucasian face, but as a reduced form of “mast” when paired with an African American face (Staum Casasanto, 2008). Explicit information about accents may also inform the processing of accented speech in a similar way, and it might be possible to examine this possibility, by presenting listeners with sentences spoken in a home, regional, or foreign accent, and priming them with either the identity of the accent they’re about to hear, or some other unrelated information, such as the age of the speaker. If listeners draw on existing accent-specific normalisation processes using the identity of the accent, we would expect more accurate responses to accents which have been named than to those which have not been named. If this is the case, we would also expect this effect to be stronger in more experienced listeners, who have had more opportunity to determine the characteristics of accents other than their own, so that adults should show more of an identification effect than children. This would support Mullenix and Pisoni's argument, that indexical information and phonetic information are processed in an inter-dependent manner, as suggested by the literature showing that adaptation to the accent of a single speaker does not generalise easily to other speakers, even to those who share an accent (Bradlow & Bent, 2003; Kraljic & Samuel, 2007). However, Cutler, Eisner, McQueen and Norris (2010) argue that talker-specific information is not always necessary for successful normalisation, as shown by studies using whispered speech or vocoded speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), in which talker-dependent information is absent, but speech recognition is unimpaired.

In the experiments reported in this thesis, the participants were all natives of the South-West of England, and did not have any substantial reported exposure to any of the native or foreign accents used, making them essentially mono-dialectal. The experiments reported here, as well as the majority of the studies cited, focused on short-term, lab-based exposure to unfamiliar accents, rather than to long-term natural exposure. These studies can provide insight into online processing, including normalisation and adaptation, but they do not tell us about longer-term effects, or about how long-term exposure to accent-based variation affects processing. They can therefore tell us about the way in which established normalisation "algorithms" operate, but not about how the algorithms develop. Only a few studies look directly at listeners' life-long experience as a factor in their ability to normalise variant speech. Adank et al. (2009) used listeners from the South-East of England and from Glasgow, and played them speech in a South-East English accent and a Glaswegian accent (as well as a Spanish accent). They found that Glaswegian listeners responded as quickly to the South-East stimuli as they did to the Glaswegian stimuli, while the South-East listeners responded significantly more slowly to the Glaswegian stimuli than to those in their own accent. Adank et al. attribute this finding to the fact that the South-East accent is the socially dominant variety of English in the United Kingdom, and is widely heard in the media, so that Glaswegian listeners are likely to have considerable long-term exposure to South-East English. South-East listeners are likely to have had much less exposure to Glaswegian English, since it is much less common in the British media. Thus the Glaswegian listeners were essentially bidialectal, while the South-East listeners were monodialectal. Similarly, Sumner and Samuel (2009) compared the performance of listeners from New York, who had an r-less accent, with listeners who had a General American accent, which is rhotic, in a priming experiment using words ending in "-er" such as "baker". They found that the r-less primes were effective only for the listeners from New York, while the rhotic primes were effective for both New Yorkers and the General American listeners, suggesting that life-long exposure to a General American accent through the media had given the New Yorkers an advantage in dealing with an accent which was not their own, but was highly familiar. A further study with French-speaking listeners finds similar results (Floccia, et al., 2006); listeners from the Eastern region of France heard speech in their own (home) accent, a familiar regional accent (that of Paris, which is considered the "standard" French accent and is heavily represented in the French-speaking media), and an unfamiliar regional accent (from the South of France). Reaction times in a lexical decision time did not differ for the home and familiar accents, but were considerably slower for the unfamiliar accent, suggesting that Floccia et al.'s participants, like those used by Adank et al. (2009) and Sumner and Samuel (2009), were effectively bidialectal. This finding suggests a further direction for future research. The studies in this thesis have examined the processing of regional and foreign accents in monodialectal listeners, whose linguistic input has been for the most part limited in its accents-based variation. If, as suggested by Bradlow and Bent (2003), greater variation in our input produces more robust representations, which are better able to generalise to novel speakers, we might expect bidialectal speakers to adapt better to novel accents than their monodialectal peers. Not only that, but we might expect infants growing up in a bidialectal environment (such as those whose parents' accent does not match their ambient accent, or who have two resident parents with different accents) to differ in the development of normalisation processes.

There is currently very little research looking at bidialectalism in infants, but one of the few studies which does address this issue suggests that the community accent, rather than that of an infant's parents, determines infants' phonetic representations (Floccia, et al., 2012). Twenty-month old infants growing up in the South-West of England, and therefore exposed to rhotic accents, but with at least one resident parent with a non-rhotic accent, were exposed to familiar words, pronounced in either a rhotic or non-rhotic form. Infants consistently identified the rhotic versions in an Intermodal Preference Looking task, but failed to identify the non-rhotic versions; their looking times did not differ significantly from those of infants exposed only to rhotic accents at home. However, there is also evidence that exposure to an accent other than the community accent may have an effect in early infancy; in a visual fixation preference task, American infants were found to attend for longer to an Australian accent than to an American accent, indicating that the Australian accent was being treated as novel, while Australian infants did not distinguish between the two accents at six months (Kitamura, et al., 2006b). At three months, the Australian infants did differentiate between the two, which suggests that by six months, they have received enough exposure to American accents via popular media to affect their preference. A forthcoming study with mono- and bidialectal infants (Durrant, Delle Luche, & Floccia, In preparation) showed that at 20 months, monodialectal infants are sensitive to mispronunciations of known words, looking longer to target pictures only when they were correctly named. However, bidialectal infants (who had at least one resident parent with an accent other than the community accent) looked longer to target pictures even when their referents were mispronounced. This suggests either that the variation in their day to day exposure had resulted in reduced sensitivity to mispronunciations, or that they had broader, more robust word representations than the monodialectal infants. In either case, we might therefore expect bidialectal infants to show less disruption when dealing with accented speech than is seen in monodialectal infants.

This thesis set out to explore the processing of accented speech in infants and in adults. The infant head-turn study described in chapter 7 has highlighted some methodological considerations impacting on behavioural segmentation studies and perhaps on the stimuli used in lab-based infant studies more generally, while the infant ERP study in chapter 8 has begun to explore the time-course of word-processing, and to reveal differences in the way home and foreign accents are segmented towards the end of the first year. The two adult studies (chapters 3 and 4) shed new light on the time-course of adults' processing, and particularly on the differences between the processing of regional and foreign accents.

Support was found for the hypothesis that foreign and regional accents recruit different normalisation processes (Adank, et al., 2009; Floccia, et al., 2006; Girard, et al., 2008), and that it is not the case that accents are processed simply as a function of their perceptual distance from the listener's own accent (Clarke & Garrett, 2004). Support was also found for the Abstract entry approach to the lexical representation of variant word-forms (Pallier, et al., 2001). The findings of the passive-listening ERP study reported in chapter 3 indicate that pre-lexical normalisation is taking place in both the foreign and regional accents, as compared to the home accent, and the repetition-suppression study reported in chapter four indicates that the lexical selection process is indistinguishable for home and regional accents.

Further research is now required to clarify the way in which infants process accent-based variation, and to plot the development of normalisation processes for regional and for foreign accents between infancy and adulthood.

References

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage, 49*.

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of Familiar and Unfamiliar Native Accents Under Adverse Listening Conditions. *Journal of Experimental Psychology - Human Perception and Performance, 35*(2), 520 - 529.

Adank, P., & Janse, E. (2009a). COMPREHENSION OF A NOVEL ACCENT BY YOUNG AND ELDERLY LISTENERS.

Adank, P., & Janse, E. (2009b). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America, 126*, 2649.

Adank, P., & McQueen, J. M. (2007). The Effect of an Unfamiliar Regional Accent on Spoken Word Comprehension. *ICPHS*.

Adank, P., Noordzij, M. L., & Hagoort, P. (2012). The role of planum temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping, 33*(2), 360-372.

Adler, K. (1967). *Phonetics and Diction in Singing: Italian, French, Spanish, German*. Minnesota: University of Minnestoa Press.

Ahmed, B. H. A., & Tan, T. P. (2011). *Non-native Accent Pronunciation Modeling in Automatic Speech Recognition*.

Angkititrakul, P., & Hansen, J. H. L. (2006). Advances in phone-based modeling for automatic accent classification. *Audio, Speech, and Language Processing, IEEE Transactions on, 14*(2), 634-646.

Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *The Journal of the Acoustical Society of America, 71*, 975.

Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology, 43*(5), 454-464.

Baayen, H., & Schreuder, R. (1999). War and Peace: Morphemes and Full Forms in a Non-interactive Activation Parallel Dual-Route Model. *Brain and Language, 68*.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The {CELEX} lexical data base on {CD-ROM}.

Baldeweg, T. (2007). ERP Repetition Effects and Mismatch Negativity Generation. *Journal of Psychophysiology, 21*(3-4).

Beckage, N. M., Smith, L. B., & Hills, T. (2010). *Semantic network connectivity is related to vocabulary growth rate in children*.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*(16), 2105.

Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. E. (2004). Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience, 16*(6), 966-977.

Berman, S. M., Mandelkern, M. A., Phan, H., & Zaidel, E. (2003). Complementary hemispheric specialization for word and accent detection. *NeuroImage, 19*(2), 319-331.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of Phonological Constancy Toddlers' Perception of Native-and Jamaican-Accented Words. *Psychological Science, 20*(5), 539-542.

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14*(02), 131-158.

Bonfiglio, T. P. (2002). *Race and the rise of standard American* (Vol. 7): Walter de Gruyter.

Boomershine, A. (2006). *Perceiving and Processing Dialectal Variation in Spanish: An Exemplar Theory Approach.* Paper presented at the 8th Hispanic Linguistics Symposium.

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*(1), 33-69.

Boulenger, V., Hoen, M., Jacquier, C., & Meunier, F. (2011). Interplay between acoustic/phonetic and semantic processes during spoken sentence comprehension: An ERP study. *Brain and Language, 116*(2), 51-63.

Bradlow, A. R., & Bent, T. (2003). *Listener adaptation to foreign-accented English*.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*.

Brent, M. R., & Cartwright, T. A. (1996). Distributional Regularity and phonotactic constraints are useful for segmentation. *Cognition*, 93-123.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), 33-44.

British Library. (2009). Leeds. *Archival Sound Recordings* Retrieved 06/09/2010, from http://sounds.bl.uk/View.aspx?item=021M-C0900X08603X-0200V0.xml

Brunelliere, A., Dufour, S., & Nguyen, N. (2011). Regional differences in the listener's phonemic inventory affect semantic processing: A mismatch negativity (MMN) study. *Brain and Language, 117*(1), 45-51.

Brunelliere, A., Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2009). Behavioral and Electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition, 1*, 390-396.

Bürki-Cohen, J., Miller, J. L., & Eimas, P. D. (2001). Perceiving non-native speech. *Language and Speech, 44*(2), 149-169.

Butler, J., Floccia, C., Goslin, J., & Panneton, R. (2011). Infants’ discrimination of familiar and unfamiliar accents in speech. *Infancy, 16*(4), 392-417.

Chambers, F. (1997). WHAT DO WE MEAN BY FLUENCY? *System, 25*(4), 535-544.

Chambers, J. K. (1992). Dialect acquisition. *Language*, 673-705.

Chambers, J. K. (2002). Dynamics of dialect convergence. *Journal of Sociolinguistics, 6*(1), 117-130.

Chavez, M., Day, R., Deyell, S., Ellis, P., Fazio, S., & Green, P. (2003). Adobe Audition (versión 1.0).

Chee, M. W. L. (2009). fMR-adaptation and the bilingual brain. *Brain and Language, 109*(2), 75-79.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116*(6).

Clopper, C. G., & Bradlow, A. R. (2008). Perception of Dialect Variation in Noise: Intelligibility and Classification. *Lang Speech, 51*(3), 175-198.

Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics, 35*(3), 421-438.
Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review, 11*(6).
Connolly, J. F., Phillips, N., Stewart, S., & Brake, W. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43*(1), 1-18.
Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*(3), 256-266.
Connolly, J. F., Stewart, S., & Phillips, N. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language, 39*(2), 302-318.
Conrey, B., Potts, G. F., & Niedzielski, N. A. (2005). Effects of dialect on merger perception: ERP and behavioral correlates. *Brain and Language, 95*(3), 435-449.
Cristià, A., Bradlow, A., Vaughn, C., Schmale, R., Seidl, A., & Floccia, C. (In preparation). Developmental Changes in Accent Perception from Infancy to Adulthood.
Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.
Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory Phonology, 10*, 91-111.
Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*(4), 317-367.
Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General, 134*(2).
Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport, 8*(4), 919.
Dehaene, S., Naccache, L., Cohen, L., Bihan, D., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature neuroscience, 4*(7), 752-758.
DePaolis, R., Duffy, H., Keren-Portnoy, T., & Vihman, M. M. (2012). *IDS differences in naturalistic speech and in stimuli affect the age at which infants succeed in Headturn Preference Procedure tasks*. Paper presented at the International Conference on Child Phonology.
Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience, 21*(10), 1893-1906.
Diehl, M., Varga, K., Panneton, R., Burnham, D., & Kitamura, C. (2006). *Six-month-old infants' perception of native speech accent*. Paper presented at the the XVth Biennial International Conference on Infant Studies.
Downs, M. (2010). Journey Education, 2-Adobe Audition 3.0 $299.96. *Approved by CIO on November, 8*.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human perception and performance, 23*(3), 914.
Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing" deafness" in French? *Journal of Memory and Language, 36*, 406-421.
Durrant, S., Delle Luche, C., & Floccia, C. (In preparation). TBA.
Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language, 36*, 202-225.
Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics, 67*(2), 224-238.
Ernestus, M., Baayen, H., & Schreuder, R. (2002). The Recognition of Reduced Word Forms. *Brain and Language, 81*, 162-173.
Evans, B., Mistry, A., & Moreiras, C. (2007). *An acoustic study of first-and second-generation Gujarati immigrants in Wembley: Evidence for accent convergence*.
Evans, B., & Taylor, E. (2010). Investigating the time course of perceptual adaptation to unfamiliar accented speech.
Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Science, 7*(10).
Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech, 46*(2-3), 245-264.
Fernald, A. (1985). Four-month-old infants prefer to listen to motherese*. *Infant behavior and development, 8*(2), 181-195.
Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental psychology, 27*(2), 209.
Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology, 20*(1), 104.
Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of child language, 16*(03), 477-501.
Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America, 76*(3), 692-707.
Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*(5), 3125-3134.
Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009a). 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., Butler, J., Goslin, J., & Ellis, L. (2009b). Regional and Foreign Accent Processing in English: Can Listeners Adapt? *Journal of Psycholinguistic Research, 38*.
Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a Regional Accent Perturb Speech Processing? *Journal of Experimental Psychology - Human Perception and Performance, 32*(5).

Flocciasecond, C., Butler, J., Girard, F., & 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.

Garrido, M. I. (2009). Repetition Suppression and plasticity in the human brain. *NeuroImage, 48*.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of Phonological Inference in Speech Perception. *Journal of Experimental Psychology, 24*(2), 380-396.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning, 34*(1), 65-87.

Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology, 26*, 409-433.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review, 105*(2), 251.

Gonet, W., & Piétron, G. (2004). The Polish Tongue in the English Ear. *Zeszyty Naukowe Państwowej Wyższej Szkoły w Koninie, 1*(4), 56-65.

Goodin, D. S., Squires, K. C., & Starr, A. (1983). Variations in early and late event-related components of the auditory evoked potential with task difficulty. *Electroencephalography and Clinical Neurophysiology, 55*(6), 680-686.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*.

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research, 1332*, 75-89.

Grabe, E. (1997). *Comparative intonational phonology: English and German*.

Grieser, D. A. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental psychology, 24*(1), 14.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*(1), 14-23.

Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience, 14*(2), 199-209.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia, 38*(11), 1518-1530.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of child language, 27*(3), 689-705.

Hanley, J. R., Kay, J., & Edwards, M. (2002). Imageability Effects, Phonological Errors, and the relationship between auditory repetition and picture naming: implications for models of auditory repetition. *Cognitive Neuropsychology, 19*(3).

Hansen, J. H. L., & Arslan, L. M. (1995). *Foreign accent classification using source generator based prosodic features*.

Hasson, U., Nusbaum, H. C., & Small, S. L. (2006). Repetition Suppression for Spoken Sentences and the Effect of Task Demands. *Journal of Cognitive Neuroscience, 18*(12), 2013-2029.

Hazen, K. (2002). The family. In J. K. Chambers (Ed.), *Handbook of Language Variation and Change* (pp. 500-525). Oxford: Blackwell.

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience, 18*(10), 1631-1643.

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience, 14*(6), 938-950.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants *Journal of Experimental Psychology - Human Perception and Performance, 26*(5), 1570-1582.

Hughes, A., & Trudgill, P. (1996). English accents and dialects: an introduction to social and regional varieties of English in the British Isles.

Hughes, A., Trudgill, P., & Watt, D. (1996). *English accents and dialects: an introduction to social and regional varieties of English in the British Isles* (Vol. 3): Arnold (London and New York and New York).

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*(4), 548-567.

Johnson, E. K., Jusczyk, P. W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology, 46*.

Johnson, K. (1997). Speech Perception without Speaker Normalization. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145 - 165). San Diego: Academic Press.

Jusczyk, P., & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23

Jusczyk, P., Hohne, E. A., & Bauman, A. (1999a). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics, 61*(8), 1465-1476.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675-687.

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science, 277*(5334), 1984-1986.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999b). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology, 39*(3–4), 159-207.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*.

Kerswill, P., & Williams, A. (2000). Creating a new town koine: children and language change in Milton Keynes. *Language in society, 29*(1), 65-115.

Khattab, G. (2007). Variation in vowel production by English–Arabic bilinguals. *Laboratory Phonology, 9*, 383-410.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children’s social preferences. *Social Cognition, 27*(4), 623.

Kitamura, C., & Burnham, D. (1998). The infant's response to maternal vocal affect. *Advances in infancy research, 12*, 221-236.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy, 4*(1), 85-110.

Kitamura, C., Panneton, R., Deihl, M., & Notley, A. (2006a). *Attuning to the native dialect: When more means less*.

Kitamura, C., Panneton, R., Diehl, M., & Notley, A. (2006b). *Attuning to the native dialect: When more means less.* Paper presented at the The Eleventh Australian International Conference on Speech Science & Technology, Aukland, New Zealand.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant behavior and development, 24*(4), 372-392.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7*(312), 1-26.

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysical evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research, 24*, 109-116.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*, 141-178.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56*, 1-15.

Kujala, A., Alho, K., Ilmoniemi, R. J., & Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. *Cognitive Brain Research, 21*(1), 106-113.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Levitt, A. G., & Aydelott Utman, J. G. (1992). From babbling towards the sound systems of English and French: A longitudinal two-case study. *Journal of child language, 19*(01), 19-49.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*: Psychology Press.

Luce, P. A., McLennan, C., & Charles-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming.

Magen, H. S. (1998). The perception of foreign-accented speech. . *Journal of phonetics, 26*.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human perception and performance, 33*(2), 391.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63.

Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds/ba/and/da. *Journal of Speech, Language, and Hearing Research, 42*(2), 271.

Mattys, S., & Jusczyk, P. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91-121.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*(4), 465-494.

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception and Psychophysics, 70*(7).

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology - General, 134*(4).

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science, 32*.

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *DEVELOPMENTAL SCIENCE, 14*(4), 769-785.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86.

McLaughlin, B., White, D., McDevitt, T., & Raskin, R. (1983). Mothers' and fathers' speech to their young children: Similar or different? *Journal of child language, 10*(01), 245-252.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 306.

McQueen, J. M. (1998). Segmentation of Continuous Speech Using Phonotactics* 1,* 2,* 3. *Journal of Memory and Language, 39*(1), 21-46.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*(6), 1113-1126.

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14-and 20-month-olds. *Journal of Cognitive Neuroscience, 16*(8), 1452-1464.

Mulak, K., Best, C. T., Tyler, M. D., Kitamura, C., & Bundgaard-Nielsen, R. (2010). *Vocabulary size predicts the development of phonological constancy: An eyetracking study of word identification in a non-native dialect by 15- and 19-month-olds.* Paper presented at the 20th International Congress on Acoustics, ICA 2010, Sydney, Australia.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Attention, Perception, & Psychophysics, 47*(4), 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*(1).

Munro, M. (2008). Foreign accent and speech intelligibility. *Hansen Edward, JG & Zampini, ML [eds]. Phonology and Second Language Acquisition. Amsterdam, John Benjamins*.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning, 45*(1), 73-97.

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of child language, 25*(02), 343-365.

National Institute of Health. (2007). National Institute on Deafness and other Communication Disorders. Retrieved 28/01/2010, 2010, from http://www.nidcd.nih.gov/about/learn/mission.asp
Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance, 24*(3), 756.
Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*(3), 283-299.
Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*(1), 1-19.
Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (In Revision). Early Word Segmentation in French: Effects of dialect?
Newman, R., & Connolly, J. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom-up and top-down effects on spoken word processing. *Biological Psychology, 80*(1), 114-121.
Newman, R., Connolly, J., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology, 40*(4), 640-647.
Noppeney, U., & Penny, W. D. (2006). Two Approaches to Repetition Suppression. *Human Brain Mapping, 27*, 411-416.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*.
Novotney, S., & Callison-Burch, C. (2010). *Cheap, fast and good enough: Automatic speech recognition with non-expert transcription*.
Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. *Speech perception, production and linguistic structure*, 113-134.
Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Attention, Perception, & Psychophysics, 60*(3), 355-376.
Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural Response Suppression Predicts Repetition Priming of Spoken Words and Pseudowords. *Journal of Cognitive Neuroscience, 18*(8), 1237-1525.
Pallier, C., Colomé, A., & Sebastian-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based Versus Abstract Lexical Entries. *Psychological Science, 12*(6).
Papousek, M., Bornstein, M. H., Nuzzo, C., Papousek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant behavior and development, 13*(4), 539-545.
Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant behavior and development, 15*(3), 325-345.

Peppe, S. (2011). 6 Prosody in the World's Languages. *Multilingual Aspects of Speech Sound Disorders in Children, 6*, 42.
Phan, J., & Houston, D. M. (2006). *Infant Dialect Discrimination*. Paper presented at the The XVth Biennial International Conference on Infant Studies.
Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *The Journal of Neuroscience, 26*(35), 8965-8975.
Phillips, N. A., Klein, D., Mercier, J., & de Boysson, C. (2006). ERP measures of auditory word repetition and translation priming in bilinguals. *Brain Research, 1125*(1), 116-131.
Pinet, M., Iverson, P., & Evans, B. G. (2011). Perceptual Adaptation for L1 and L2 Accents in Noise by Monolingual British English Listeners.
Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25*(1-2), 21-52.
Polka, L., Rvachew, S., & Molnar, M. (2008). Speech perception by 6-to 8-month-olds in the presence of distracting sounds. *Infancy, 13*(5), 421-439.
Polka, L., & Sundara, M. (2003). *Word segmentation in monolingual and bilingual infant learners of English and French*.
Polka, L., & Sundara, M. (2012). Word Segmentation in Monolingual Infants Acquiring Canadian English and Canadian French: Native Language, Cross-Dialect, and Cross-Language Comparisons. *Infancy*.
Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human perception and performance, 20*(2), 421.
Psychology Software Tools. (1996). E-Prime (Version 1.1).
Ranbom, L. J., Connine, C. M., & Yudman, E. M. (2009). Is Phonological Context Always Used to Recognize Variant Forms in Spoken Word Recognition? The Role of Variant Frequency and Context Distribution. *Journal of Experimental Psychology - Human Perception and Performance, 35*(4), 1205 -1220.
Ratner, N. B., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of child language, 11*(03), 515-522.
Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research, 111*(1-2), 13-23.
Rivera-Gaxiola, M., Silva-Pereyra, J., Klarman, L., Garcia-Sierra, A., Lara-Ayala, L., Cadena-Salazar, C., et al. (2007). Principal component analyses and scalp distribution of the auditory P150–250 and N250–550 to speech contrasts in Mexican and American infants. *Developmental neuropsychology, 31*(3), 363-378.
Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science, 8*(2).
Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73-79). London: Edward Arnold.
Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *DEVELOPMENTAL SCIENCE, 12*(2), 339-349.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy, 15*(6), 608-635.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511-531.

Russell, M. (2007). Analysis of Italian children's English pronunciation. from http://www.eee.bham.ac.uk/russellm/ItalianEnglishReport/ItalianEnglish_report_v2.htm

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development, 69*(2), 309-320.

Schmale, R., & Cristià, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations *DEVELOPMENTAL SCIENCE, 12*(4), 583-601.

Schmale, R., Cristià, A., Seidl, A., & Johnson, E. K. (2010). Developmental Changes in Infants' Ability to Cope with Dialect Variation in Word Recognition. *Infancy, 15*(6), 650-662

Schmale, R., Hollich, G., & Seidl, A. (2011). Contending with foreign accent in early word learning*. *Journal of child language, 1*(1), 1-13.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime (version 1.1).

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science, 16*(4), 197-201.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science, 270*(5234), 303-304.

Sharma, D., & Sankaran, L. (2011). Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language variation and Change, 23*(03), 399-428.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition, 106*, 833-870.

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. . *Journal of Memory and Language, 51*(2), 173-189.

Spencer, J. P., Blumberg, M. S., McMurray, B., Robinson, S. R., Samuelson, L. K., & Tomblin, J. B. (2009). Short arms and talking eggs: Why we should no longer abide the nativist–empiricist debate. *Child Development Perspectives, 3*(2), 79-87.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior research methods, 38*(4), 598-605.

Staum Casasanto, L. (2008). *Does Social Information Influence Sentence Processing?* Paper presented at the The 30th Annual Meeting of the Cognitive Science Society, Washington D.C.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard III, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of neurophysiology, 82*(5), 2346-2357.

Stuart-Smith, J. (2004). Scottish English: phonology. *A Handbook of Varieties of English, 1*, 47-67.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition, 119*(1), 131-136.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*.

Sumner, M., & 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.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*(30), 12530.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*(1), 53-71.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development, 3*(1), 75-102.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172.

Tomasello, M., & Mannle, S. (1985). Pragmatics of sibling speech to one-year-olds. *Child Development*, 911-917.

Vaden, K. I., Jr, Muftuler, L. T., & Hockok, G. (2010). Phonological repetition-suppression in bilateral superior temporal sulci. *NeuroImage, 49*.

Vaissière, J., & Boula De Mareüil, P. (2004). Identifying a language or an accent: from segments to prosody.

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties. *Journal of language and social psychology, 18*(1), 31-48.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior research methods, 38*(4), 584-589.

van de Weijer, J. C. (1998). *Language input for word discovery.* University of Nijmegen, Nijmegen.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *DEVELOPMENTAL SCIENCE, 7*(3), 270-276.

Weil, S. A. (2001). *Foreign accented speech: adaptation and generalization.* The Ohio State University, Ohio.

Wells, J. C. (1982). Accents of English. 3 vols: Cambridge: Cambridge University Press.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue canadienne de psychologie, 43*(2), 230.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life*. *Infant behavior and development, 7*(1), 49-63.

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*(1), 114-132.

Woehrling, C., & Boula de Mareüil, P. (2006). Identification of regional accents in French: perception and categorization.

Ziegler, W. (2008). Apraxia of speech. *Handbook of clinical neurology, 88*, 269-285.