The role of vowel hyperarticulation in clear speech to foreigners and infants

A thesis submitted for the degree of
Doctor of Philosophy
By

Jayanthiny Kangatharan

Division of Psychology

Department of Life Sciences
College of Health and Life Sciences

Brunel University London

August 2014

Abstract

Research on clear speech has shown that the type of clear speech produced can vary depending on the speaker, the listener and the medium. Although prior research has suggested that clear speech is more intelligible than conversational speech for normal-hearing listeners in noisy environments, it is not known which acoustic features of clear speech are the most responsible for enhanced intelligibility and comprehension. This thesis focused on investigating the acoustic characteristics that are produced in clear speech to foreigners and infants. Its aim was to assess the utility of these features in enhancing speech intelligibility and comprehension. The results of Experiment 1 showed that native speakers produced exaggerated vowel space in natural interactions with foreign-accented listeners compared to native-accented listeners. Results of Experiment 2 indicated that native speakers exaggerated vowel space and pitch to infants compared to clear read speech.

Experiments 3 and 4 focused on speech perception and used transcription and clarity rating tasks. Experiment 3 contained speech directed at foreigners and showed that speech to foreign-accented speakers was rated clearer than speech to native-accented speakers. Experiment 4 contained speech directed at infants and showed that native speakers rated infant-directed speech as clearer than clear read speech. In the fifth and final experiment, naturally elicited clear speech towards foreign-accented interlocutors was used in speech comprehension tasks for native and non-native listeners with varying proficiency of English. It was revealed that speech with expanded vowel space improved listeners' comprehension of speech in quiet and noise conditions. Results are discussed in terms of the Lindblom's (1990) theory of Hyper and Hypoarticulation, an influential framework of speech production and perception.

Declaration

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Jayanthiny Kangatharan

Acknowledgements

I would like to thank my supervisors Dr Maria Uther and Professor Fernand Gobet for their magnificent support and their exceptional assistance they have provided throughout my time as their student. Their mentoring was crucial in completing my research and writing up this thesis.

Their patient guidance was significant in helping me make the right decisions about how to proceed with my research. Their advice and suggestions on how to effectively conduct experiments, analyze different types of data, and write up research have been absolutely invaluable during this intellectually inspiring journey. I have been extraordinarily fortunate to have supervisors who were so supportive of my research, and who replied to my queries so swiftly. I also would like to thank Dr Adrian Williams for kindly taking on the supervisory role at Brunel University during my supervisors' absence, and for actively supporting my academic progress.

I must express my gratitude to my family who helped me keep things in perspective. I would like to thank them for their continued patience and encouragement. Also, completing this research would have been all the more difficult were it not for the support of the fellow postgraduates and friends at Brunel University that I met during my academic journey.

I also would like to thank the Division of Psychology at the Department of Life Sciences in the newly formed College of Health and Life Sciences for providing the funding that allowed me to conduct this research, and to also attend conferences and meet numerous inspiring researchers. I would like to express my gratitude to the technicians for their technical support, and to all the participants who took part in my research. Finally, I would like to thank Brunel University in general for providing me with an Isambard Research Scholarship that gave me the opportunity to embark on this incredible, intellectual journey.

Table of Contents

List of Figures

List of Tables

Chapter 0

Speech Communication

The aim of daily spoken communication is to convey information in an understandable way to listeners. When people talk to each other, accurate speech perception and comprehension are the main aims. The intelligibility of speech involves the speaker, the medium and the listener. Factors that can influence intelligibility on the listener's side are the degree to which listeners can hear well: consequently there is a continuum from normal-hearing people on one end to severely hearing-impaired and deaf individuals on the other end. Factors that can affect intelligibility concerning the medium include the extent of noise there is within the medium, such as complete silence, background noise or reverberation in the environment of a conversation. The perception of speech can become degraded in the presence of background noise or when listeners require more acoustic information due to hearing impairments or nonfamiliarity with the target language (Payton, Uchanski, & Braida, 1994; Uther, Knoll, & Burnham, 2007).

If listeners experience perceptual difficulties (because they are not native speakers or because the listeners are hearing-impaired or because of the presence of noise), speakers will assume a style of speech that is known as 'clear speech'. Various researchers have used the term 'clear speech' to denote speech directed at deaf listeners, foreign listeners or infant listeners, all of which, have exaggerated acoustic-phonetic components such has loudness or speech rate compared to 'normal' speech (Ferguson, 2012; Fernald, Taeschner, Hirsh-Pasek, & Jusczyk, 1989; Kuhl et al., 1997; Picheny, Durlach & Braida, 1986; Schum, 1996).

Thus, another determinant of speech intelligibility is the clarity of speakers' speech sounds. It has been proposed that speakers make acoustic alterations to their speech depending on their environment and audience to ensure that their speech is discriminable (Lindblom, 1990). This viewpoint is the basis of Lindblom's H&H theory, according to which the production of speech is adapted to listeners' needs.. For example, it has been shown that speakers produce clear speech and not casual speech when speaking to hearing-impaired listeners in quiet (Picheny, Durlach, & Braida, 1985) and to normal-hearing listeners in the presence of noise (Uchanski, 1988; Uchanski, Choi, Braida, Reed, & Durlach, 1996) because clear speech is more intelligible. It has been demonstrated that speakers, when generally instructed to produce clear speech, differ largely in their capability to generate effective clear speech (Ferguson, 2004; Gagne, Masterson, Munhall, Bilida, & Querengesser, 1994).

It has also been shown that whether speakers are instructed to talk as if talking to hearing-impaired listeners or non-native listener or infants or in the presence of noise yields acoustic-phonetically dissimilar alterations in pitch and the shape of pitch contours (Ferguson, 2004; Fernald& Simon, 1984; Gagne et al., 1994; Knoll, Scharrer & Costall, 2009; Lombard, 1901; Uther et al., 2007; Wassink, Wright & Franklin, 2007). For example, regarding child-directed speech (CDS), even though it was shown that speech to actual and imaginary children results in higher pitch, larger vowel duration, as well as larger F1 and F2, differences between the actual and imaginary interaction were found in the perturbations in fundamental period amplitude and the association between periodic and aperiodic aspects in the speech signal (Schaeffler, Kempe, & Biersack, 2006). It is therefore evident that clear speech produced within a communicative setting with an actual listener is closer to an everyday interaction between a speaker and listener and is therefore more authentic than clear speech produced upon instruction with no communicative intent, and with only a speaker but no listener.

Although one study found that clear speech is more intelligible than conversational speech for both normal-hearing and hearing-impaired listeners in both noisy, reverberant and combined environment (Payton et al., 1994), it is still not well established which acoustic features of clear speech are the most responsible for enhanced intelligibility and comprehension (Ferguson, Poore, & Shrivastav, 2010; Ferguson & Kewley-Port, 2002; Ferguson & Kewley-Port, 2007). The next section will look at the acoustic-phonetic feature of clear speech in the auditory mode that by the majority of previous studies has been suggested to contribute to higher speech intelligibility in clear speech. That feature is expanded vowel space.

0.1 Expanded vowel space and speech perception

Past research has demonstrated clear speech with interlocutors with different linguistic needs. These include speech directed at infants (Kuhl, Andruski, Chistovich, Chistovich, & Kozhevnikova, 1997), children (Liu, Tsao, & Kuhl, 2009), foreigners (Uther et al., 2007), parrots (Xu, Burnham, Kitamura, & Vollmer-Conna, 2004), and computers (Burnham, Joeffry, & Rice, 2010). In all these types of studies, expanded vowel space (measured by F1/F2) was reported as a feature of clear speech. An association between increased overall speech clarity and extended vowel space was found (Bradlow, Torretta, & Pisoni, 1996), and also a correlation could be established between mothers' use of articulatory stretched vowel space and their infants' speech perception capabilities (Liu, Kuhl, & Tsao, 2003). However, this evidence is correlational and not sufficient to establish causation. Further research is necessary to determine whether a hyperarticulatory speech style helps listeners to understand speech better, both perceptually and cognitively. Although one study has indicated that infants' word recognition might have been facilitated by expanded vowel space to a certain degree (Song, Demuth, & Morgan, 2010), to date there is no clear-cut evidence that shows exaggerated vowel space to be unambiguously beneficial for infants' and foreigners' learning of words. The present thesis therefore aims to address the issue whether vowel hyperarticulation improves speech intelligibility by focusing on speech to foreigners (FDS) and to infants (IDS).

Only a small number of studies focused on foreigner-directed speech (FDS), sometimes known as 'Foreignese' (Freed, 1978) produced in interactions with a communicative intent (e.g. Kangatharan, Uther, Gobet, 2012; Knoll &Scharrer, 2007; Sankowska, Lecumberrri & Cooke, 2011; Uther et al., 2007). FDS has been shown to be acoustically modified compared to speech directed at native speakers (Uther et al., 2007). It has been suggested that these differences may be due to their didactic needs rather than emotional needs in the target language (Bradlow & Bent, 2002; Uther et al., 2007). It has also been further argued that the needs of foreigners in speech might be distinguishable to a certain extent from the needs of the hearing-impaired or those listening in noise. As a result of the differing linguistic needs for instance, FDS can be considered to be different in quiet from other kinds of speech such as speech to hearing-impaired listeners or Lombard speech because FDS was observed to include clear speech modifications such as vowel space exaggeration (Uther et al., 2007).

By contrast, Lombard speech in noisy environments normally involves high pitch, long vowel length and no vowel space exaggeration (Cooke & Lu, 2010; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988) and in quiet, Lombard speech is less intelligible than FDS (Sankowska et al., 2011). However, little is known about whether vowel space exaggeration in FDS actually helps foreigners with their intelligibility of speech. Similarly, although acoustic modifications in IDS are well-documented (Fernald & Simon, 1984; Grieser & Kuhl, 1988; Kitamura, Thanavishuth, & Burnham, 2001; Kuhl et al., 1997; Stern, Spieker, Barnett, & MacKain, 1983; Xu & Burnham, 2010), no human listening experiments have been conducted so far that would clearly show that vowel space exaggeration in IDS actually contribute to increased intelligibility of speech.

0.2 Rationale and research questions underlying present research

There were four main motivations behind this thesis. First, previous research has not explored the circumstances under which vowel hyperarticulation is elicited in speech to foreigners. Specifically, they did not separate the effects of ‘foreign appearance’ vs. ‘foreign accent’ (Snow, van Eeden, & Muysken, 1981; Uther et al., 2007). Hence, the first research question is whether the physical appearance or accent of the interlocutor results in independent effects on resultant hyperarticulation by a native speaker.

Second, past research on infant-directed speech (IDS) provided only circumstantial and not direct evidence linking linguistic discrimination and hyperarticulation (Liu et al., 2003). The second research question of the present thesis thus addresses the issue whether the acoustical modifications in IDS are uniquely didactic and distinguishable from other forms of clear speech.

Third, according to Ferguson and Kewley-Port (2007), it was established that there must be an extended $F1$ range and front vowels must have larger $F2$ values for normal-hearing listeners to perceive increased vowel intelligibility. However, it is not clear whether vowel hyperarticulation improves speech intelligibility at word level. Therefore another research question is whether hyperarticulation in IDS and FDS enhances speech clarity.

Fourth, recent evidence has suggested that clear speech might provide an advantage only for listeners who have had extensive exposure to the sound structure of a second language (L2) and who have the phonetic experience to produce L2 such as native (L1) speakers and early non-native (L2) learners of L2 (Bradlow & Bent, 2002; Smiljanic & Bradlow, 2011). Since, no research has focused on finding out whether vowel space expansion provides L1 listeners, early L2 learners and late L2 learners with enhanced speech comprehension, the fourth research question in this thesis is what effect vowel space exaggeration will have on L1 listeners', early L2 learners' and late L2 learners' speech comprehensibility in both quiet and adverse situations.

0.3 Organization of thesis

The first chapter in this thesis will look at phonemes and their phonetic properties. It will then critically assess how specific acoustic properties of speech such as fundamental frequency, vowel formants and intensity play a role in speech communication. The second chapter will evaluate four models of spoken language acquisition. These models view phonetic speech perception in infant development from different perspectives and include the theories proposed by Kuhl (Kuhl et al., 2008), Best (Best & McRoberts, 2003), Flege (Flege, 2003a) and Werker (Lalonde & Werker, 1995).

The third chapter will begin with the Hypospeech and Hyperspeech (H&H) theory that looks at how speech perception and production change as a function of the interactive and communicative context. After evaluating the H&H theory, the chapter will focus on speech modifications aimed at different normal hearing audiences that can be accounted for by the H&H theory.

The fourth chapter contains Experiment 1, which explored whether appearance and speech separately affect native speakers' hyperarticulation. This experiment employed a 3x2x2 mixed design. Fifty-two White British adult speakers communicated with one of four different confederate groups (2 types of appearance x 2 types of accent) to solve three modified versions of the DiapixUK tasks. Results indicated that accent but not appearance had an effect on native speakers' production of vowels. Specifically, vowel space was significantly larger in speech directed to foreign-accented individuals than to individuals with native accent irrespective of their physical appearance.

The fifth chapter presents Experiment 2 that aimed to find out whether acoustical modifications used in IDS are uniquely didactic and distinguishable from other forms of clear speech such as clear read speech. This experiment focused on the collection of speech samples produced by eleven mothers interacting with their infants, and in another condition where they were asked to read sentences aloud in a clear voice. This experiment used a 3x2 within-subjects design. The acoustic measure of vowel space was compared across conditions. Results showed that mothers' articulatory vowel space in speech to infants was significantly more expanded than their vowel space expressed when reading sentences. The results therefore support expanded Kuhl's Native Language Magnet (NLM) Theory (enhanced) (Kuhl et al., 2008), which suggests that vowel space expansion in IDS serves to overemphasise important phonetic contrasts.

The sixth chapter dealt with Experiments 3 and 4 of the thesis: Experiment 3 looked at whether vowel space exaggeration elicited in speech to foreign-sounding interlocutors in Experiment 1 leads to a better speech clarity compared to speech to native sounding interlocutors. Experiment 4 looked at whether vowel space exaggeration elicited in IDS in Experiment 2 leads to a better speech clarity compared to clear read speech. In Experiment 3, 21 native listeners rated the speech samples of speech to foreign vs native sounding interlocutors in a transcription task, confidence rating, goodness rating and a clarity rating. In Experiment 4, they rated the speech samples of IDS vs read speech in a transcription task, confidence rating, goodness rating and a clarity rating.

Results from clarity ratings showed that native speakers rated infant directed speech samples as clearer than read speech samples. Moreover, native speakers rated speech samples directed to foreign sounding interlocutors as clearer than the speech samples directed to native sounding interlocutors. These results are in line with the H&H theory, which suggests that hyperarticulated speech allows phonetic units to be more easily perceived as being acoustically distinct (Lindblom, 1990).

The seventh chapter will comprise Experiment 5 that assessed if expanded vowel space has an improved cognitive effect on speech comprehensibility for both native and non-native listeners of British English in quiet and in noise conditions. This experiment used a 2x3x4 mixed design and indicated that both native and non-native speakers' speech comprehension benefit from vowel space expansion in quiet and in noise. This experiment supports the H&H theory, which suggests adults alter their speech to provide the listener with sufficient information to make speech comprehension possible (Lindblom, 1992). The eighth chapter integrates the observations and findings made in the last four chapters before evaluating them and discussing implications for future research.

Chapter 1

Physical and acoustic properties of speech

Chapter 1 provides an overview of the physical characteristics of vowels and in particular, those features most relevant to understanding Experiment 1 and Experiment 2 presented in Chapters 4 and 5 respectively. Chapter 1 starts by giving a short introduction to phonemes, and brief overview of consonants. The chapter then deals with vowels and acoustic properties such as pitch, formants and intensity, which represent the dependent measures in Experiment 1 and Experiment 2 presented in Chapters 4 and 5 respectively.

1.1 Phonemes: the basic building blocks of speech

The sounds of vowels and consonants are realized as phonemes, which are considered the smallest specific units of sound in a language (Raphael, Borden, Harris, 2007). Phonemes are characterized by symbols of the International Phonetic Alphabet (IPA) and are placed in between two slashes (e.g. /p/). The role of phonemes within a language is to specify a meaningful difference: the words ‘pit’, ‘pet’, ‘pat’, ‘put’ and ‘pout’, for example, can be discriminated through a single phoneme (i.e. the vowel) (Moats, 2006). If a phoneme, such as a consonant, differentiates a pair of words, this is known as a minimal pair: for example, the initial consonant phoneme in the minimal pair of ‘pat’ and ‘bat’ phonetically differentiates the two words, thereby contributing to their distinct meanings (Raphael et al., 2007). Variants of phonemes are known as allophones: the sound of /p/ in ‘pie’, for instance, varies from the sound of /p/ in ‘top’ because of the aspirated /p/ in ‘pie’ and the non-aspirated /p/ in ‘top’ (Raphael et al., 2007).

1.1.1 Consonants

Consonants are produced by narrowing a part of the vocal tract (constriction) (Raphael al., 2007). In fact, the location and the extent of the constriction specify which consonants are produced (Raphael et al., 2007). The majority of English consonants are categorised using the three articulatory parameters: voicing, place of articulation, and manner of articulation (Ladefoged, 2005). However, because the focus of the experimental studies in the thesis is on the production and perceptual processing of vowels, the remainder of this chapter attends to the physical properties of vowels.

1.1.2 Vowels

The generation of vowels involves the vibration of vocal folds, which is why vowels are voiced sounds (Raphael et al., 2007). As the airstream flows through the vocal tract, the articulators do not get in contact. Alterations in the form and size of the oral cavity of the vocal tract, caused by modifications in palate, tongue, teeth and lip positions, result in dissimilar resonances whereby dissimilar vowels are produced (Raphael et al., 2007).

The peaks of resonances within the vocal tract are known as formants. These consist of frequencies measured in Hertz (Hz) and correspond with the peaks in the spectrum of a vowel (Ladefoged, 1996). Accordingly, the first and second formants of each vowel are considered the most essential to qualitatively differentiate between vowels

(Hillenbrand & Nearey, 1999). Because each vowel has a different set of formant frequencies, they are phonemically differently classified. Information on the formant frequencies of vowel sounds ascertains the quality of vowels. It helps listeners towards vowel distinction, thereby contributing to the intelligibility of speech (Carlson, Granström, & Fant, 1970; Peterson & Barney, 1952). The first formant ($F1$) has the lowest frequency and a range between 150-850Hz (Ladefoged, 1996). On a physiological level, it represents the first resonance peak of the vocal tract. The second formant $F2$ has a range between 500 and 2500 Hz and represents the second resonance peak of the vocal tract.

As vowels can be produced by varying tongue height, tongue frontness/ backness and lip rounding, the first and second formant frequencies for each vowel differ depending on the tongue position and the shape of the lips. Accordingly, $F1$ is smaller for close (high) vowels, and larger for open (low) vowels, such as the vowel /a/, which requires a wider opening of the jaw. $F2$ is smaller for back vowels, and larger for front vowels, such as /i/, and is considered to be responsive to the tongue shape (Benade, 1976; Sundberg, 1977). The information on the first and second formants depending on tongue height and tongue frontness/ backness is summarised in Figure 1.1.

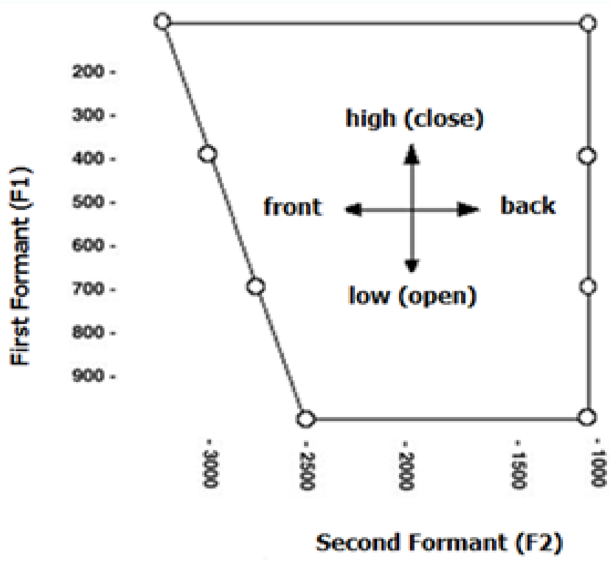


Figure 1.1: The vowel quadrilateral with the main parameters tongue height (high to low) and tongue advancement (front to back) (modified from https://notendur.hi.is/peturk/KENNSLA/02/TOP/VowelSpace.html).

The space in which vowels are characterised according to tongue height and frontness/backness is represented in form of the IPA vowel chart (Figure 1.2). The IPA vowel chart contains both primary and secondary cardinal vowels. The left hand side of Figure 1.2, for instance, illustrates the corner vowels (primary cardinal vowels) (numbered from 1-5) of the IPA chart that are evenly spaced regarding auditory quality and that are used to denote vowels in a standardised manner. The right hand side of Figure 1.2 demonstrates the secondary cardinal vowels (numbered from 9-13) that are used to differentiate between vowels that are created using different extents of rounding the lip.

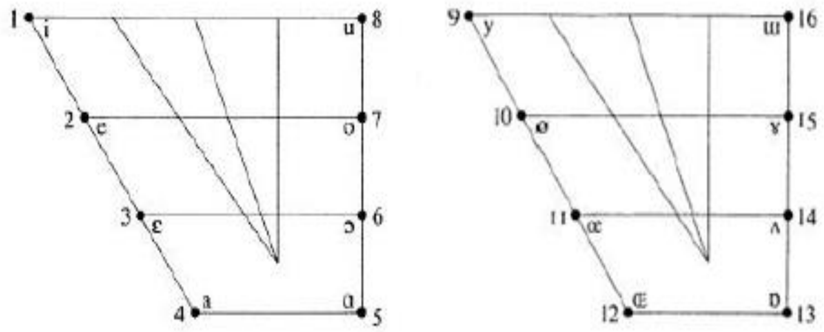


Figure 1.2 Primary and secondary cardinal vowels from the IPA vowel chart (from www.sussex.ac.uk/linguistics/documents/q1027_lecture_3.pdf).

Figure 1.3 demonstrates both primary and secondary cardinal vowels within a vowel quadrilateral according to the IPA.

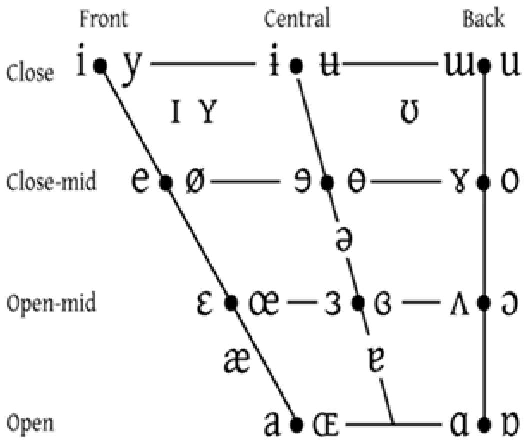


Figure 1.3: Cardinal vowel chart with both primary and secondary cardinal vowels according to the IPA (from http://www.phonetics.ucla.edu/course/chapter1/vowels.html).

The third formant *F*3 has been suggested to be of no substantial importance for the identification of vowels and is considered to show responsiveness to the tip of the tongue while the fourth and fifth formant *F*4 and *F*5 have been implicated in speakers' timbre (Sundberg, 1970). The importance of the different formants for the perception of speech sounds varies according to the type of speech sound: for instance, while the third formant is crucial to perceptually differentiate between the consonants /r/ and /l/ (Iverson et al., 2003), the first and second formants are important for the perception of vowels. The average values for the first and second formant for British English vowels can be viewed in Table 1.1 (Hawkins & Midgley, 2005).

Table 1.1: Means of the first formant F1and second formant F2 (in Hz) of British English monophthongs as articulated by British English adult speakers (from Hawkins & Midgley, 2005).

Vowel	***F*1**	***F*2**
/ɑː /	604 Hz	1040 Hz
/a/	917 Hz	1473 Hz
/ʌ/	658 Hz	1208 Hz
/ɒ/	484 Hz	865 Hz
/ɔː /	392 Hz	630 Hz
/e/	600 Hz	1914 Hz
/ɜː /	494 Hz	1373 Hz
/iː /	276 Hz	2338 Hz
/I/	393 Hz	2174 Hz
/uː /	289 Hz	1616 Hz
/ʊ/	413 Hz	1285 Hz

1.2 Acoustic properties of phonemes

1.2.1 Pitch and pitch changes

The frequency of a produced sound is construed as pitch by the brain. The more rapidly the changes in air pressure are created by the produced sound, the more increased is pitch perceived. Frequency and pitch are not linearly related because as frequencies increase, a more sizeable alteration in frequency is needed to cause a modification in perceived pitch (Raphael et al., 2007).

A sound can be represented by means of a waveform, which is characterised by a cycle (Figure 1.4) (Raphael et al., 2007). The time taken to complete one cycle is known as period (Figure 1.5) (Raphael et al., 2007). Thus, frequency is the number of cycles per second that a wave goes by a specific point and is measured in Hertz (Hz) (Ladefoged, 2005). The horizontal length of a single wave cycle is known as wavelength and the altitude of a wave is designed as amplitude (Figure 1.6) (Raphael et al., 2007).

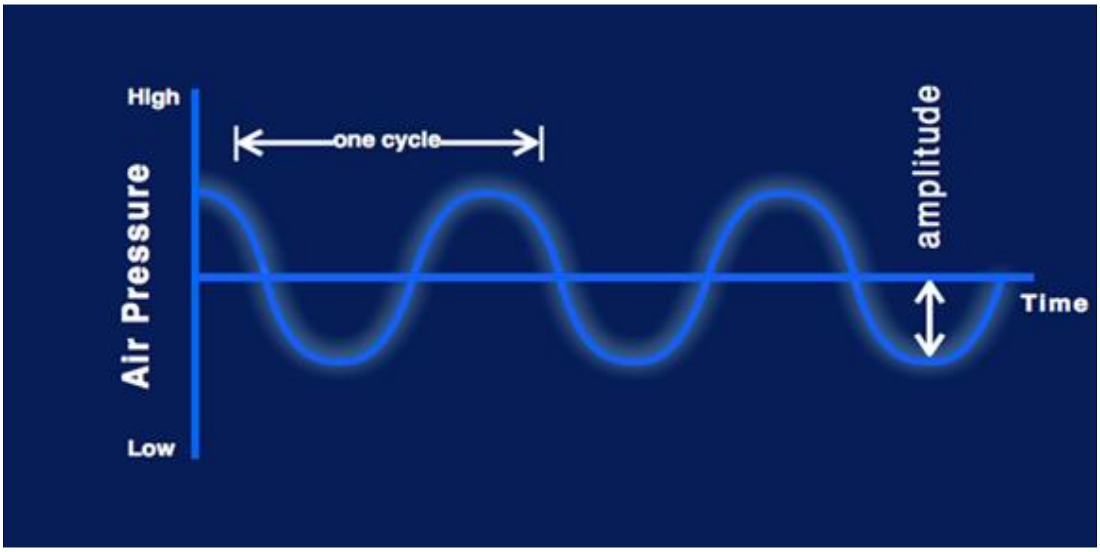


Figure 1.4: Representation of sound through a simple waveform with the length of one cycle being visualised by the white arrow (modified from ThinkQuest, http://library.thinkquest.org/19537/).

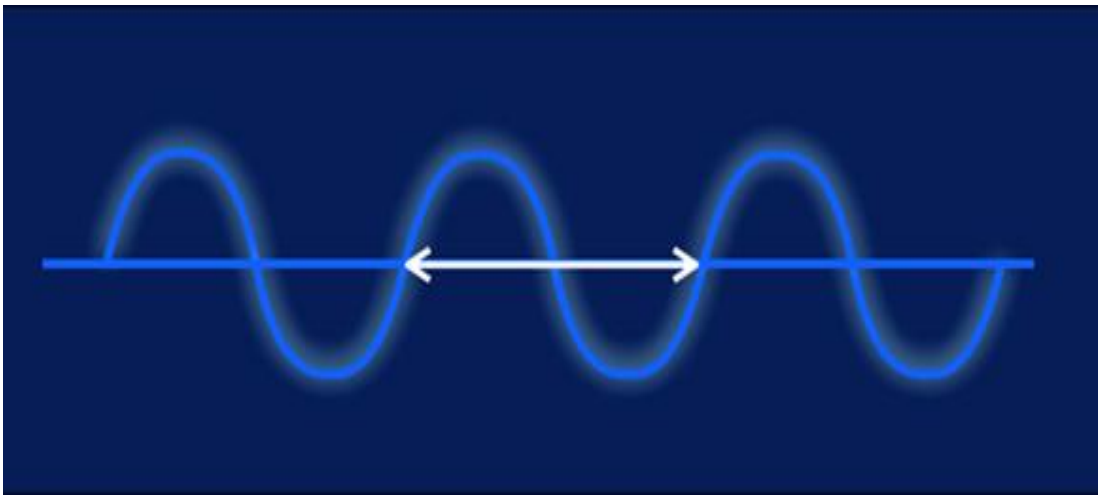


Figure 1.5: The period (as indicated by the white arrow) is the time that is needed for one wavelength to go by a particular point (time on the x-axis and amplitude on the y-axis) (modified from Thinkquest, http://library.thinkquest.org/19537/).

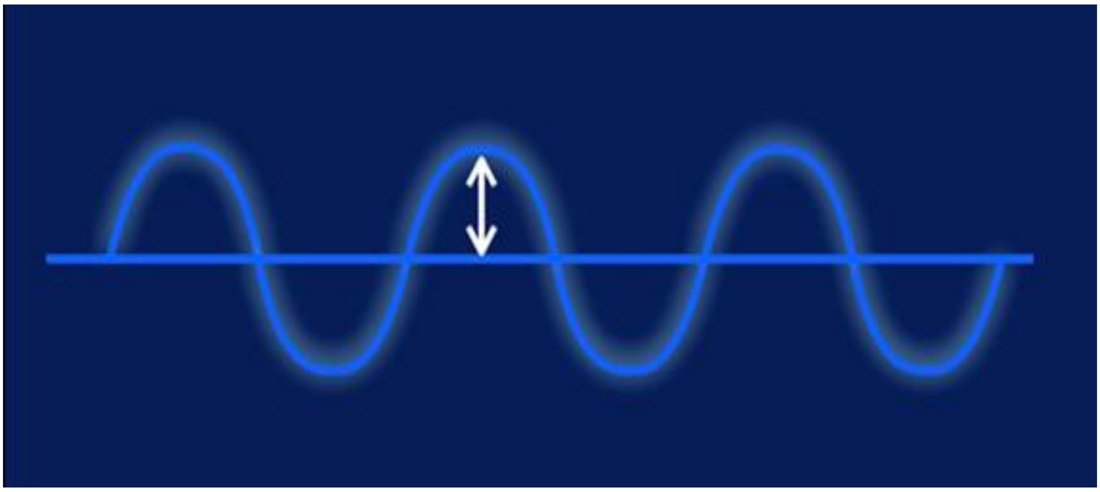


Figure 1.6: The amplitude (as indicated by the white arrow) here represents the highest point of the sound wave (time on the x-axis and amplitude on the y-axis) (modified from Thinkquest, http://library.thinkquest.org/19537/).

A sine wave is the simplest waveform due to its association with solely one frequency (Figure 1.7). It can be produced through a pure tone that vibrates in simple harmonic motion. A complex tone can be derived by adding two sine waves of different frequencies together.

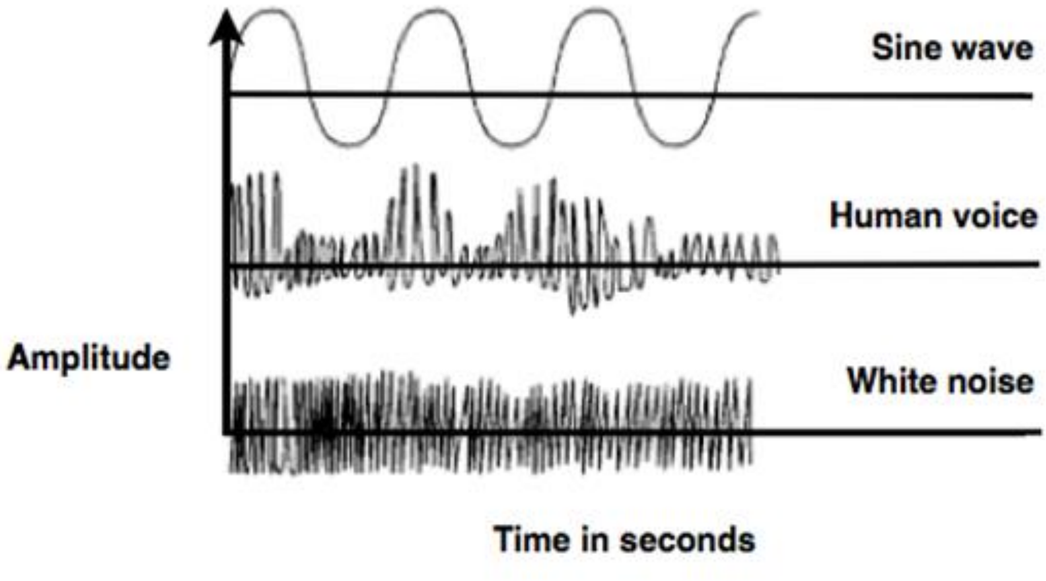


Figure 1.7: Waveforms of sine wave, human voice and white noise with time on the x-axis and amplitude of sound pressure on the y-axis (modified from http://flylib.com/books/en/1.500.1.12/1/#_/term_).

On the other hand, the human voice is a complex tone as it encompasses multiple frequencies (Raphael et al., 2007). One's perception of a talker's pitch usually relies on the lowermost frequency in the talker's voice that is the fundamental frequency *F*0. Changes in *F*0 can lead to different intonation patterns of utterances (Raphael et al., 2007). Due to anatomical differences, men and women differ in *F*0 when producing speech: women have a *F*0 of about 220 Hz whereas men have a *F*0 of 120 Hz (Diehl, Lindblom, Hoemeke, & Fahey, 1996). Although human auditory sensitivity encompasses a range from 20 to 20,000 Hz, 1000 to 4000 Hz is the range of frequencies that the healthy ear is best tuned for since its range includes important information about human interaction as for example about speech (Raphael et al., 2007).

In contrast to speech communication in which periodicity usually is an intrinsic characteristic, white noise can be considered to be an aperiodic type of complex wave (Raphael et al., 2007; Yrttiaho, May, Tiitinen, & Alku, 2011). Due to the absence of $F0$ in white noise, there is no sensation of pitch and there is no pattern of vibration that repeats itself either (Raphael et al., 2007).

1.2.2 Role of pitch and pitch changes

The use of increased mean pitch and expanded pitch range in infant-directed speech (IDS) has been widely reported (e.g. Ferguson, 1964; Stern et al., 1983). It has been shown that mean pitch and pitch range are larger when mothers interact with their newborn infants than when they speak to adults. The intonation of pitch (its rise and fall), which is also known as pitch excursion or contour, is also broader, and pitch range larger in mothers' communication with their newborn infants than in conversation with adults (Fernald & Simon, 1984). For example, it was reported that pitch in adult-directed speech (ADS) had a range between around 90Hz to 300Hz while pitch range in IDS ranged from 90Hz to 800 Hz (Fernald, 1985).

It has been argued that overstressed pitch excursions in IDS might provide infants with considerably significant auditory input (Fernald, 1984). This is in line with prior research on vocal affect in which exaggerated pitch and pitch range that is specific to IDS was observed to improve the verbal expression of positive vocal emotion toward infants (Scherer, 1981). Consequently, pitch range has been considered to be acoustically associated with affect (Scherer, 1986). Exaggerations in pitch intonation were also demonstrated to increase infants' stimulation and to guide infant's attention (Fernald, 1991; Stern, Spieker, & MacKain, 1982).

More recent research has provided evidence showing that mean pitch and pitch range are adapted during the first year of infancy (e.g. Kitamura & Burnham, 2003). However, the relation between prosodic aspects and alterations associated with age is not clear (Kitamura & Burnham, 2003; Stern et al., 1983). For example, when speaking to infants of six months and twelve months of age, higher mean pitch was used by mothers in order to express articulations with positive emotions (Kitamura & Burnham, 2003). However, when talking to infants of nine months, speech by mothers was observed to include a noticeable decrease in mean pitch (Kitamura & Burnham, 2003). In addition, mothers were shown to use utterances of more directive nature as well as larger pitch range and reduced positive vocal affect in speech to nine months old infants (Kitamura & Burnham, 2003). This result has been explained by suggesting that around 9 months, infants' phonetic speech perception becomes language-specific (Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993), and infants are able to appreciate uncomplicated directives (Hubley & Trevarthen, 1979). Consequently, decreased vocal emotion would enable infants to focus on processing and analysing spoken language while exaggerated vocal emotion would distract them from doing so (Lacerda, Sundberg, Andersson, & Rex, 1995).

The idea that infants become more attuned to the linguistic aspect of utterance is supported by research showing that Dutch and American infants of nine months of age showed an inclination to listen for an extended duration to words in their mother-tongue (Jusczyk et al., 1993). Their increased responsiveness to the phonetic and phonotactic characteristics of their native languages is illustrated by the observation that absence of these phonetic and phonotactic characteristics through low-pass filtering led them not to prefer the native language over a non-native language (Jusczyk et al., 1993). This sensitivity to sound patterns of one's native language has been argued to be in line with the decreased responsiveness to particular non-native contrasts that is reported at about nine months (Best, 1991).

Other investigations of age-related alterations in the prosodic aspects of IDS include a broader pitch range specifically used for infants at the age of four months (Stern et al., 1983). By contrast, broader pitch range was not observed in newborn or infants older than a year old, which has been argued to occur because larger variety in pitch might be required to support four-month old infants' verbal articulations (Stern et al., 1983). Studies on prosodic characteristics of IDS in other language systems could also demonstrate the use of increased pitch and raised pitch range in tonal languages, such as in Mandarin IDS compared to CDS (Liu, Tsao, & Kuhl, 2009), as well as the use of higher pitch and pitch range in Mandarin IDS compared to ADS (Grieser & Kuhl, 1988).

The use of heightened pitch and increased pitch excursions has also been demonstrated in Thai IDS (Kitamura, Thanavishuth, Burnham, & Luksaneeyanawin, 2002) although mean pitch and pitch range were reported to be lower in Thai IDS than Australian English IDS. This observation on pitch features has also been made in other tonal languages such as Mandarin (Grieser & Kuhl, 1988) and Japanese (Fernald et al., 1989), which has been considered to be due to restrictions in tonal languages of using pitch to indicate lexical differences (Kitamura et al., 2002). Although in the Mandarin language, alterations in pitch for lexical tones in syllables were not reported to be influenced by age during infants' first year (Liu et al., 2007), and heightened pitch might have a different social function in different cultures (e.g. Ingram, 1995), it can be argued according to the majority of IDS studies that the increased acoustic properties of IDS can be generalised (e.g. Burnham et al., 2002; Fernald & Kuhl, 1987; Fernald et al., 1989; Grieser & Kuhl, 1988; Kuhl et al., 1997).

1.2.3 Formants

As explained above, formants are the component frequencies that relate to the peaks in the sound spectrum of vowels (Benade, 1976; Fant, 1960). Vowels are considered to be formed by means of the formant frequencies $F1$ and $F2$. Vowel sounds differ in their formant frequencies, which is why vowel sounds can be distinguished from one another (Benade, 1976). Because this thesis deals with the hyperarticulation of vowels, which is the physical expansion of vowel formant space (measured by changes in $F1$and $F2$), this section will cover $F1$ and $F2$ only (Kuhl et al., 1997).

1.2.4 Vowel Space

If a waveform of a vowel is cross-sectioned, one can see $F1$ and $F2$ as the summits of the most extreme frequencies of the waveform. The first part on the left of Figure 1.9, for instance, illustrates the vowel /eɪ/ in the word “bait”, while the second part on the right of Figure 1.8 shows its cross-section including the peaks $F1$ and $F2$ as the most extreme frequencies of the waveform (Wallace, 2010).

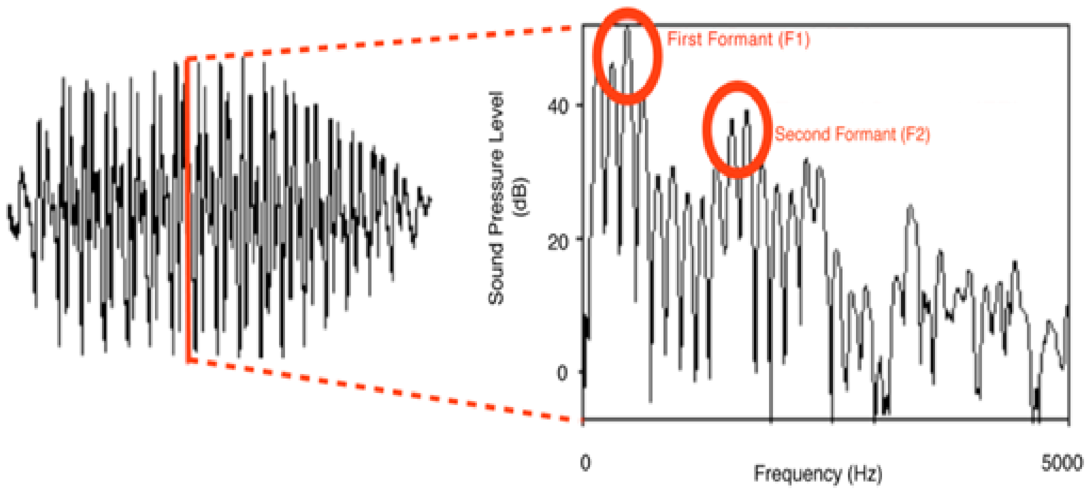


Figure 1.8: The waveform of the vowel in “bait” on the left and its cross-section is demonstrated on the right (Wallace, 2010).

Figure 1.9 demonstrates that the values of these peaks indicate a correlation with the magnitude of the vocal tract cavity in the anterior and posterior parts of the mouth (Wallace, 2010). Vowel space can be measured if the first formant is charted against the second formant.

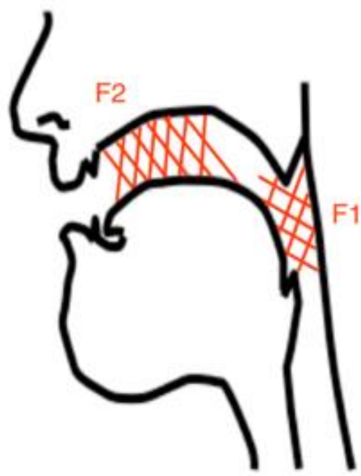


Figure 1.9: The correlation of F1 and F2 with vocal tract cavity size (Wallace, 2010).

If one, for example, analyses the vowel space of the vowels in the words “boot”, “beat”, “bought” and “bat”, the vowel space physically extends into a trapezoidal form indicating that the four vowels /u:/, as in “boot”, /i:/, as in “beat”, /ɔ:/, as in “bought”, and /æ/, as in “bat”, are acoustically different (Figure 1.10) (Wallace, 2010).

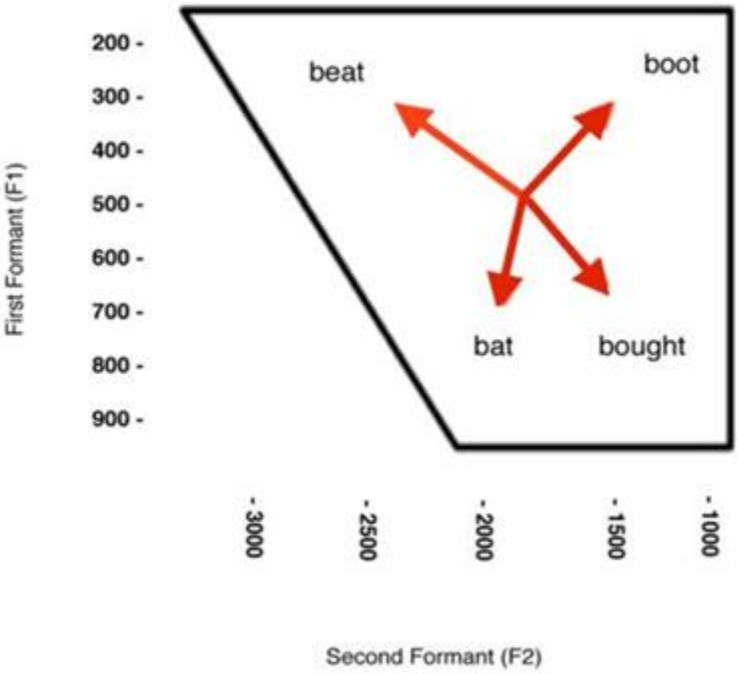


Figure 1.10: The physical extension of vowel space (Wallace, 2010).

A vowel triangle area (VTA) can be obtained if one measures the vowel formant space of three vowels. For example, Figure 1.11 illustrates the three vowels /ɜ:/, as in “shirt”, /a:/, as in “shark”, and /ɔ:/, as in “shorts”.

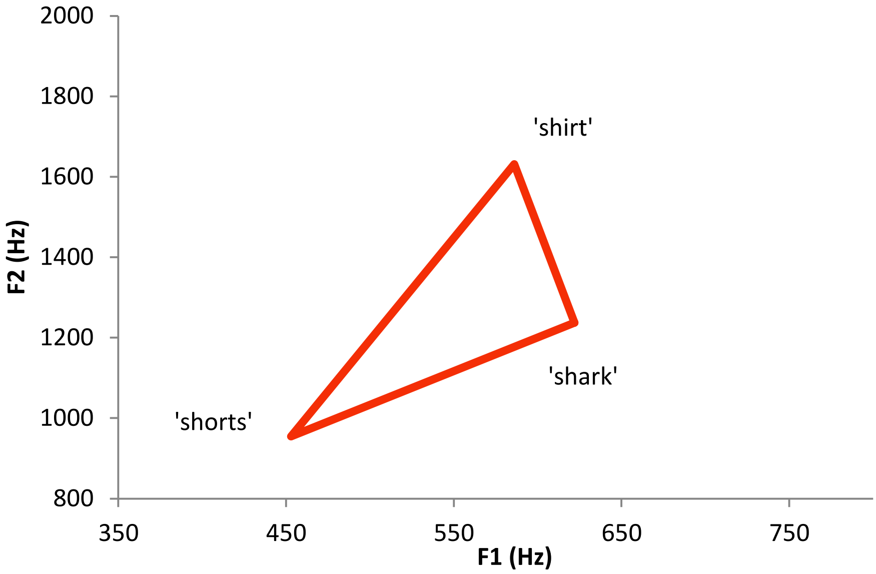


Figure 1.11: Hyperarticulation of three vowels /ɜː/, as in “shirt”, /aː/, as in “shark”, and /ɔː/, as in “shorts” as indicated by differences in F1 and F2 (Kangatharan, 2009).

1.2.5 Role of formant frequencies in vowel identification

As previously explained, *F*1 is inversely proportional to vowel height since a higher first formant yields a lower vowel such as /a/ and /ɑ/ whereas high vowels such as /i/ and /u/ have small *F*1 values. The second formant is associated with vowel fronting and has high values for front vowels such as /a/ and /i/ while back vowels such as /u/ and /o/ have low second formants.

The first two formants have been reported to be highly associated with listeners’ perception of the quality of vowels. Peterson and Barney (1952), for example, measured the formant frequencies of 76 participants as they generated ten vowels in a /hVd/ (V=vowel) context. They measured formants at one time slice regarded as being steady state, which is that part of the vowel centre, in which formants are considered to stay static for a continued period. They found the measured formant frequency values to be highly correlated with the expected vowel (Peterson & Barney, 1952). It is generally acknowledged that formant frequency analysis can be carried out using the standard approach of measuring the stationary steady-state portion of the vowel (Plichta, 2010), which has been used in recent speech studies in which formants are extracted at the midpoint of spectral target vowels (e.g. Hunter & Kebede, 2012; Krause & Braida, 2004; Lam & Kitamura, 2012; Moon & Lindblom, 1994; Picheny et al., 1986; Uther et al., 2007).

The stationary steady-state portion of vowels has been previously reported to provide listeners with sufficient information for listeners to identify vowels (Hillenbrand & Gayvert, 1993a), In addition, the notion of vowel duration and formant movement (when formants are measured at two positions in time (Hillenbrand, Getty, Clark, & Wheeler, 1995)), as being important cues for vowel identification was also evaluated by Neel (1998). After testing vowels that were resynthesized including and excluding vowel duration and formant movement for identification by elderly listeners with hearing-impairment and by listeners with no hearing-impairments, it was observed that both listener groups performed better when stimuli had both adequate vowel duration and dynamic formant information than when they solely had spectral target information. While for normal-hearing listeners, formant frequency information was considered to represent primary cues for vowel identification, vowel duration was regarded to be a secondary cue (Strange, 1989; Summers & Leek, 1992).

Recent evidence suggested that when speaking styles are influenced, static spectral targets, information on formant movement and vowel duration all have an important part in the perceived clarity of vowels for normal-hearing listeners (Ferguson & Kewley-Port, 2002). Ferguson and Kewley-Port (2002) observed from previous research and from their own laboratory pilot data that vowels, when spoken clearly (clear speech), differ from vowels that are spoken in a conversational speech style in the way that vowels in clear speech are of longer duration, are at the furthest positions in vowel space, and have more dynamic formant movement than vowels in conversational speech. Their study aimed at evaluating the association between the acoustic characteristic of vowels and their recognition in clear and casual speech for two groups of listeners: hearing-impaired listeners and listeners with no hearing-impairment (Ferguson & Kewley-Port, 2002). Their study used naturally elicited speech to test (among others) whether vowels are clearer in clear speech than in conversational speech for both listener groups and whether vowels in both speaking styles will vary in their acoustic characteristics (Ferguson & Kewley-Port, 2002). Monosyllabic words were extracted from sentences that were recorded in the two speaking styles before being presented to both listener groups in a background of 12-talker babble (Ferguson & Kewley-Port, 2002).

It was found that the steady-state portion of vowels and formant movement, and also the duration of vowels might represent primary cues for vowel recognition for both listener groups. Nonetheless, the use of the three acoustic characteristics varied highly for single vowels for both listener groups indicating that the use of these characteristics for vowel recognition may be changed through impaired hearing (Ferguson & Kewley-Port, 2002). According to the researchers, the finding that both static spectral targets and dynamic formant movement emerged as primary indicators for vowel perception of normal-hearing listeners, illustrates the requirement of both kinds of cues to be present in order to recognise English vowels (Ferguson & Kewley-Port, 2002). This is supported by research in which it was found that access to both cues can lead to more than 90% recognition accuracy (Hillenbrand & Nearey, 1999). In addition, duration of vowels was observed to be longer in clear speech than in conversational speech. This finding has been supported by prior research in which duration was identified as an essential cue for vowel identification (Hillenbrand, Clark, & Houde, 2000).

Nonetheless, the researchers argue that their findings concerning the role of steady-state portions of vowels, formant movements and vowel duration in vowel identification cannot be generalised yet as their findings can benefit from replications with other speakers (Ferguson & Kewley-Port, 2002). Therefore, it can be argued that both steady-state portions of vowels and formant movements seem to be involved in the identification of vowels. The importance of steady-state portions of vowels and vowel duration for vowel intelligibility has been supported by another recent study, which made use of variability observed in a group of speakers (Ferguson & Kewley-Port, 2007).

1.2.6 Intensity

The magnitude of a sound wave is usually assessed in terms of intensity, which is measured in decibel (dB) (Ladefoged, 2005). With increasing intensity, the sound wave is perceived as subjectively louder. Similar to frequency and pitch, there is no linear relation between intensity and loudness since the perception of loudness has been observed to rise in a more decelerated manner than the concrete rise in intensity. Intensity has been observed to increase under noisy conditions (Bond, Moore, & Gable, 1989; van Summers et al., 1988) or situations in which speech is conveyed in a shouting manner (Rostolland, 1982a).

1.2.7 Role of intensity

Previous research on the role of intensity reported that amplitude as a main acoustic correlate of intonation in form of loudness (Fry, 1968), is unlike pitch, not involved in drawing and sustaining attention in IDS (Fernald & Kuhl, 1987). However, this does not mean that intensity plays no role in IDS since previous research (Fernald & Kuhl, 1987) normalised speech samples for effectiveness to be, for example, between 0 and 68 decibels. The complete dynamic variety of intensity that mothers employ when addressing their infants, such as whispering (de Boer, 2011), was not considered. Nonetheless, it has been suggested that lower volume in form of whispering can be used in a calming or lively way in IDS (de Boer, 2011; Fernald & Simon, 1984). When used to soothe infants, the use of lower volume in IDS is exemplified through lullabies (de Boer, 2011). At the moment, it is not certain whether patterns of intensity in IDS are as unique as the manner in which pitch is modulated in IDS.

The comparison of vocal intensity between clear and conversational speech (Picheny, Durlach and Braida, 1986), showed that intensity was five to eight decibels higher in clear speech than in conversational speech. This rise in vocal intensity between conversational and clear speech has been recently supported by Ferguson et al.'s results (Ferguson, Poore, Shrivastav, 2010).

Another study that investigated the role of intensity amplitude in spoken word identification has observed that variability in overall intensity does not reduce recognition of spoken words (Sommers, Nygaard, & Pisoni, 1994). It was argued that spoken word recognition will be reduced only when acoustic-phonetic variability changes the acoustic characteristics that are relevant to phonetically perceive speech

(Sommers et al., 1994). Considered to merely represent the difference in speaker's distance to the listener, variability in overall amplitude is argued not to influence those acoustic characteristics that lead to speech perception (Sommers & Barcroft, 2006).

Research that compared intelligibility between clear, loud and conversational speech styles at slow, normal and fast speech rates found that the intelligibility of speech with increased intensity is higher than the intelligibility of conversational speech at normal speech rate (Krause & Braida, 2002). Specifically, it was observed that even if loud speech revealed a lesser speech intelligibility advantage than clear speech, it was reported to be more intelligible than conversational speech at normal speech rate (Krause & Braida, 2002). Recent research has also shown that speakers increase intensity in clear speech compared to in conversational speech, and show awareness when regulating the intensity of their speech output (Ferguson et al., 2010).

However, not every speaker was shown to attain an intelligibility benefit with loud speech at normal speech rate suggesting that apart from clear speech, neither conversational nor loud speech could yield an intelligibility benefit that is as sizeable or reliable as that of clear speech (Krause & Braida, 2002). Nonetheless, although clear speech was found to be louder between 5 to 8 decibels compared to casual speech (Picheny et al., 1986), the observed intensity differences between clear and casual speech styles have been argued not to represent an aspect that might account for increased intelligibility of clear speech. This is because clear and casual speech materials were equated for amplitude in previous speech intelligibility studies (Uchanski, 2005).

Since the focus of this thesis relies on the question “Does expanded vowel space contribute to the improved understanding of spoken speech?”, Chapter 1 considered the role of formant frequencies in identifying vowels. It also discussed the role of pitch, pitch range and of intensity amplitude in speech communication as these will represent the dependent measures in Experiment 1 and Experiment 2 that are presented in 4 and 5 respectively. The next chapter will deal with theories in the spoken language acquisition of a first and second language since Experiment 5 presented in Chapter 7 will address whether compared to native English learners, non-native learners of English with varying proficiencies in English will benefit in their understanding of spoken English from speech with modified vowels.

Chapter 2

Developmental theories of spoken language acquisition

Chapter 2 will cover theories of spoken language acquisition of first language (L1) and second language (L2) in order to offer an explanation of how non-native speakers of English with differing proficiencies in English are able to process speech with vowel hyperarticulation at word level compared to native speakers. This question will be addressed in Experiment 5 presented in Chapter 7.

2.1 Phonetic speech perception during the first year of life

Infants' speech perception abilities have been observed to undergo a twofold modification towards the end of the first year of life (Kuhl et al., 2008). While infants early in life seem to be able to differentiate between practically all phonetic units in the languages of the world (Lasky, Sydral-Lasky & Klein, 1975; Trehub, 1976), infants have been reported to adjust to the information of their native language formation around the age of nine months (Werker & Tees, 1984a).

Specifically, to find out at what age infants' speech perception capacities start to become comparable to adults' abilities with the same linguistic background, Werker and others conducted cross-sectional and longitudinal studies that showed that while six to eight months old English infants could differentiate non-native consonant contrasts easily (for example contrasts in Hindi and Salishan languages), ten to twelve months old infants were found to have difficulty with this task (Werker & Tees, 1984a; Werker & Lalonde, 1988). Half of eight to ten months old infants were reported to differentiate the non-native contrasts while the other half did not (Werker

& Lalonde, 1988; Werker & Tees, 1984a). Similar results were observed when contrasts were artificially produced, when three Zulu click contrasts were included or when the task was completed with a Salish contrast but using a different procedure (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Best, 1989; Werker & Lalonde, 1988).

Similarly, regarding vowel perception, it was found that by the age of six months, infants are able to differentiate within vowel categories in their L1 as they have an internal formation of vocalic categories in L1, thereby demonstrating the effect of infants' L1 by that age (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). Nonetheless, by approximately the age of ten months, infants' capability to discriminate between non-native vowel categories has been reported to decrease (Polka & Werker, 1994; Werker & Polka, 1993).

As a result of infants' increasingly improving perception of their native language and their gradually decreasing perception of non-native languages towards the end of their first year of life, the discrimination between non-native phonetic units is experienced to be more problematic by adulthood (Best, McRoberts & Goodell, 2001; Werker & Lalonde, 1988). This observation has been shown by studies with both adults and infants (Werker & Tees, 1984b; Werker & Logan, 1985; Kuhl et al., 2006; Tsao, Liu & Kuhl, 2006). For example, in studies where the ability to discriminate non-native contrasts (e.g. in Hindi) between infants and adults was compared, it was shown that infants performed significantly better than adults (Trehub, 1976; Werker, Gilbert, Humphrey & Tees, 1981).

Werker and colleagues, for instance, demonstrated that even a brief training could not improve adults' discrimination of the non-native contrasts (1981). Phonologically, adults have been reported to encounter problems when learning a secondary language since their capacity to perceive phonetic distinctions in the new language might be affected by their native-language categories (Flege, 1989). Specifically, after acquiring a secondary language after puberty, talkers have been observed to articulate it using an accent that is characteristic of their mother tongue (Flege, 1993). Regarding evidence for infants, studies have shown that by the end of the first year, English infants were not able to differentiate non-native contrasts in the Mandarin and Hindi language (Tsao, Liu, Kuhl & Tseng, 2000; Best, McRoberts, LaFleur & Silver-Isenstadt, 1995). Similarly, Japanese infants were observed to not be able to discriminate the American English [r]-[l] contrast (Kuhl et al., 1997).

2.2 Four accounts of spoken language acquisition

This chapter looks at four developmental models of spoken language acquisition that are regarded the leading models in present language perception research. These models were selected because they propose different explanations underlying the decline in L2 speech perception and the improvement in L1 speech perception that can be observed in infants between 6-12 months (Werker & Tees, 1984a). These developmental models are Werker's Cognitive Theory of spoken language acquisition (Diamond, Werker, & Lalonde, 1994; Lalonde & Werker, 1995), Flege's Speech Learning Model (SLM) (Flege, 1988b; Flege, 2002; Flege, 2003a; Flege, 1992a; Flege, 1995), Best's Perceptual Assimilation Model (PAM) (Best, 1994a, b, 1995; Best, McRoberts & Sithole, 1988; Best & McRoberts, 2003) and Kuhls' Native

Language Magnet theory (NLM) (Kuhl, 1991; Kuhl, 1993; Kuhl, 2000a Kuhl et al., 2006; Kuhl et al., 2008).

These models look at phonetic speech perception in infant development from different perspectives. Both Kuhl's NLM model and Werker's cognitive account of spoken language acquisition adopt general-mechanism approaches. However, the NLM model ascribes infants' phonetic speech perception skills in L1 at 7.5 months to neural commitment to L1 and their skills in L2 to uncommitted neural circuitry (Kuhl et al., 2008). By contrast, in Werker's account changes in infants' phonetic speech perception abilities in L1 and L2 are attributed to emerging primary cognitive capacities that are considered to cause improved object search, navigation and categorisation abilities in infants at about 10 months old (Lalonde & Werker, 1995). By comparison, Best's PAM is a phonetic account of early perceptual development. It proposes that phonetic units are perceived with regards to the articulatory gestures that cause them. It suggests that the decline in L2 perception occurs when not different but the same articulatory organs are involved in phonetic contrasts (Best & McRoberts, 2003).

Flege's SLM is different from the other models because it does not specify whether general or specialised mechanisms are involved in speech perception. It mainly deals with L2 and how phonological segments for L2 are acquired. SLM focuses on single phonemes and not on contrasts as PAM does. It suggests that new L2 phonetic categories are created when phonetic dissimilarities between an L2 phone and the closest L1 phone are observed (Flege, 1995). Although all four models propose different mechanisms that may account for infants' decline in L2 speech perception and improvement in L1 speech perception, they can be seen as complementing each other.

2.3 Werker's Cognitive Theory of spoken language acquisition

A model of developmental change in the phonetic perception of infants aged between six months and twelve months of age was proposed by Werker (Diamond et al., 1994; Lalonde & Werker, 1995; Werker & Pegg, 1992). Similar to the NLM (Kuhl, 1993, 1994) but in contrast to Best's PAM and Flege's SLM, Werker's cognitive account is a theory based on general auditory mechanisms (Werker & Pegg, 1992). According to this view, alterations in the identification of speech from a non-native category at approximately ten months of age can be attributed to the development of overall, domain-general cognitive capabilities such as object search, categorization and navigation (Diamond et al., 1994; Lalonde & Werker, 1995).

Specifically, according to Werker, non-native speech contrasts can still be differentiated in adulthood because the alterations that occur in non-native speech perception between ten to twelve months of age do not imply the complete loss of being able to discriminate non-native contrasts (Lalonde & Werker, 1995). Specifically, it has been suggested that when, for example, a more responsive procedure is used or when listeners are sufficiently trained, adults are able to differentiate non-native contrasts that they seldom heard before and which were not easy to differentiate (Logan, Lively, & Pisoni, 1991; Werker & Tees, 1984b). Nevertheless, their performance was infrequently reported to be as good as that of native talkers (Polka, 1991; Strange, 1986; Werker & Tees, 1984b; Werker & Logan, 1985).

Nonetheless, it has been shown that certain contrasts, such as a non-native Zulu click contrast, can be differentiated by English adults and older infants (Best et al., 1988). This indicates that that contrast is not influenced by the perceptual alteration caused by one's L1 phonology because the contrast sounds deviate from the sounds usually used in English (Best et al., 1988). It has therefore been argued that general auditory mechanisms might process such non-native contrasts, indicating that even if infants and adults did never hear this sound contrast used in their L1, this absence of listening does not contribute to decreased perceptual sensitivity (Lalonde & Werker, 1995). Therefore, the developmental alterations in infants' responsiveness to non-native consonants have been argued to actually represent a reorganization of early phonetic sensitivities and not a loss of these sensitivities (Werker, 1989).

This notion of general cognitive skills affecting speech perception is consistent with research (Morgan, 1990; Morgan & Saffran, 1995). Infants before the age of nine months have been reported to be able to discover alterations in a series of syllables or in stress pattern but not if correlations between the two are interrupted (Morgan & Saffran, 1995). Around the age of nine months, infants organise distributional and rhythmic information from multisyllabic items and it is around this time that they seem to first discover and make use of correlations between a series of syllables and stress pattern (Morgan & Saffran, 1995). It has therefore been claimed that once infants acquire more general cognitive capabilities, which usually emerge around nine months, they will likely be able to combine distributional and rhythmic properties in multisyllabic stimuli (Morgan & Saffran, 1995).

Therefore, Lalonde and Werker (1995) aimed to evaluate the associations between alterations in non-native consonant sensitivity and alterations in other fields of cognitive abilities. They assumed that a relation between age-related alterations in speech perception and developmental alterations in cognitive skills would indicate that the reorganisation of infants' phonetic sensitivities occurs at the same time as the essential developmental changes in other general-domain cognitive abilities. After recruiting a sample of forty eight ten months old infants, half of which according to prior investigations were expected in the cross-language consonant contrast task to exhibit non-native contrast sensitivity as opposed to the other half of infants, Lalonde and Werker compared infants' performance of discriminating non-native consonant contrasts with their performance on two non-linguistic tasks (1995). The non-linguistic comparison tasks included a visual categorisation task and an object search task. This comparison enabled Lalonde and Werker (1995) to find any signs of developmental consistency in infants' performance across all three tasks.

It was argued that consistency in performance across tasks within infants would support the notion that developing non-linguistic mechanisms might contribute to the developmental rearrangement of infants' discrimination of non-native consonant contrasts (Lalonde & Werker, 1995). It was found that the infants' speech perception task performance of discriminating non-native consonant contrasts was strongly related to their skills on the object search task and to the visual categorisation task. Rather than being the result of age-related effects, it was argued that the observed alterations in the performance across tasks indicate that a general type of cognitive abilities might contribute to the developmental alterations in infants' speech perception close to the conclusion of infants' first year (Lalonde & Werker, 1995).

Therefore, consistent with Werker's cognitive model of spoken language acquisition, the developmental synchrony seen in infants' performance on both the speech and the two non-linguistic tasks seems to indicate the effect of a broad type of cognitive abilities on infants' alteration in their speech perception. It is thus argued that these results suggest that domain-general cognitive abilities affect age-related alterations in infants' performance of the speech perception task used in Lalonde and Werker's study (1995).

Nonetheless, despite the observed strong synchrony among infants' cross-language speech perception task and their competence on the two non-linguistic tasks, the relationship is not absolute (Lalonde & Werker, 1995). Moreover, it might be that reported deviations from the consistent pattern in infant's performance across tasks might not reflect simple inaccuracies in measurement but indicate notionally profound individual differences and different developmental directions (Lalonde & Werker, 1995).

In addition, it can be argued that the finding in Lalonde and Werker's study (1995) only applies to infants' cross-language perception of consonant contrasts and not vowels. Therefore it cannot be claimed that widely based cognitive skills might influence infants' discrimination of non-native vowel contrasts, especially since discrimination of vowels occurs earlier than that of consonants (Kuhl et al., 1992). Thus, to arrive at the same conclusion for infants' performance on a cross-language vowel contrast task, the task used in Lalonde and Werker's study (1995) would have to be conducted with non-native vowel contrasts.

Also, it can be argued that Werker's developmental model of spoken language acquisition can benefit from more research that can investigate the degree to which more general advances in infants' capacities affect the association between the larger process of language learning and age-related alterations in speech perception (Lalonde & Werker, 1995). Specific issues include questions such as what the elements are, which contribute to infants becoming native listeners, and that depend on changing cognitive competencies, and questions on the ways in which competencies help infants become purposeful users of language (Lalonde & Werker, 1995).

Nonetheless, more recently, it has been shown that, consistent with Werker's developmental account of spoken language acquisition, developmental alterations in infants' responsiveness to non-native consonants actually represent a reorganization of early phonetic sensitivities and not a loss of these sensitivities (Kuhl et al., 2006; Tsao et al., 2006). Moreover, it has been shown that poor abilities of discriminating non-native contrasts by infants of eleven months of age were related to improved performance on cognitive control abilities (Conboy, Sommerville, & Kuhl, 2008). This link has been therefore suggested to indicate that infants' developing domain-general skills help to differentiate between essential and non-essential input and to ignore non-essential information (Conboy et al., 2008).

2.4 Flege's Speech Learning Model (SLM)

The Speech Learning Model (SLM) by Flege (1988b, 1992a, 1995, 2002, 2003a) is another model that, in addition to explaining how speech perception is shaped by L1 acquisition, deals with the influence of L1 on L2. Previous research suggested that the process of learning L2 sounds is shaped by speech sound patterns of L1, implying that certain L2 sounds might be perceived by non-native L2 speakers differently from native L1 speakers (Weinberger, 1990; Wode, 1978). For example, the phonetic symbol /θ/ for the consonant sound 'th', as in the word 'thing,' does not exist in the Russian and Japanese languages. However, because these languages have the consonant sounds /s/ and /t/, Russian learners of English tend to articulate /θ/ as /t/ while Japanese learners of English tend to articulate /θ/ as /s/ (Weinberger, 1990).L1 and L2 speakers have been numerous times demonstrated to differ in their perception on the segmental and word level (Flege & Eefting, 1986; Flege & Hillenbrand, 1987; Koster, 1987; Miyawaki, Strange, Verbrugge, & Liberman, 1975). Although L2 learners' perceptual processing of L2 sounds can be trained to become more accurate, it does not lead to native like perception of L2 sounds (Flege & Mackay, 2004; Strange, 1992).

SLM aims to explain restrictions linked to age that affect L2 learners' capacity to generate L2 sounds like native speakers. The SLM states that speech perception adjusts to contrastive phonetic aspects of L1 when listeners learn L1 (Flege, 1995) and that adult learners of L2 make use of the same capacities to acquire speech sounds in L2 that infants and children employ to learn L1 (Flege, Schirru, & MacKay, 2003). This involves both the accurate perception of characteristics of speech sounds in L2 as well as the creation of L2 phonetic categories. The model argues that for L2 sounds to be generated accurately, perceptual objectives are needed, by which the acquisition of L2 sounds on a sensorimotor level can be directed with accuracy.

The SLM assumes the existence of phonetic categories in which language-specific characteristics of phonemes are stored (Flege, 1995). These categories are considered to develop throughout one's lifetime when created during childhood for L1 speech sounds. Phonetic categories are regarded to indicate the aspects of all speech sounds in L1 or in L2 that are classified to implement each phonetic category. Phonetic categories for L1 and L2 are considered to share the same phonological space, which is why an effort by bilinguals is necessary to differentiate between L1 and L2 phonetic categories (Flege et al., 2003). It is argued that if L1 phonetic categories are well established, they will strongly attract L2 speech sounds and that they are more likely to block the creation of new categories for L2 speech sounds if L2 sounds are recognised as instances of an L1 category (Flege et al., 2003; Walley & Flege, 2000).

If a new category is not created for a L2 sound although it sounds audibly dissimilar from the nearest L1 speech sound, this is known as 'category assimilation' (Flege et al., 2003). The model proposes that in such situations the development of an amalgamated category occurs over time that incorporates the phonetic characteristics of the L1 and L2 speech sounds that are perceptually connected (Flege et al., 2003).

If L2 speech sounds are perceived by L2 learners to be phonetically dissimilar from the nearest speech sounds for L1, new categories for L2 speech sounds are more likely to be created (Flege, 1995; Flege et al., 2003). In such a case, L2 speech sounds will be articulated according to these phonetic categories in the end. The accuracy of a certain L2 sound that is articulated will depend on the correspondence between the L2 learners' newly created phonetic category for that L2 sound and the native speakers' phonetic category for that sound (Flege, 1995).

According to the SLM, the age of learning (AOL) a second language plays an essential role in L2 acquisition (Flege et al., 2003). It is, for instance, argued that with L2 learners' age, it is less probable for them to phonetically distinguish between L1 and L2 sounds and to generate L2 sounds with accuracy. This is because L2 learners who begin L2 acquisition subsequent to the end of the critical period (after around 12 years of age) are, due to neurological limitations, considered to produce L2 with a more noticeable accent than L2 learners who acquired L2 early between 3-12 years (Flege et al., 2006; Lenneberg, 1967; Moyer, 1999). Moreover, it is less likely for them to differentiate between L2 sounds that do not exist as non-contrasts in L1. Consistent with SLM, it was found that the English vowel /æ/ is easier to differentiate from the nearest German vowel /ε/ for German children than adults (Weiher, 1975). This indicates that for German children a larger perceptual distance between these vowels enabled them to differentiate them phonetically (Flege, 1995).

Similarly, in case of immigrants' entrance in a mainly L2-speaking country, the age of arrival (AOA) has been viewed to indicate the age of their first contact with L2 (Flege, 1992a; Flege, MacKay, & Meador, 1999a; MacKay, Meador, & Flege, 2001). Accordingly, it was shown that compared to L2 speakers who arrived in the L2-speaking country as young adults, L2 speakers who immigrated as children were better in producing L2 consonants (Flege, Munro, & MacKay, 1995b) and L2 vowels (Piske, Flege, MacKay, & Maeador, 2002), and perceiving L2 consonants (Yamada & Tohkura, 1995). A study that employed a discrimination task to look at L2 learners' vowel perception and also took into account the frequency of L1 use of L2 learners, found that in contrast to early L2 speakers who used L1 scarcely, those L2 speakers who were early L2 learners and had a higher use of L1, differed from L1 speakers in vowel perception (Flege & Mackay, 2004). It was concluded that early learning of L2 does not necessarily mean that it will lead to native like L2 vowel perception. At the same time, late L2 learning does not prevent the perception of L2 vowels that functionally is similar to native like perception of L2 vowels (Flege & MacKay, 2004).

If L1 and L2 sounds are perceptually related, they are known as diaphones (Weinreich, 1957). According to SLM, one phonetic category is used to process diaphones and the event of ‘equivalence classification’ occurs, which will prevent the formation of a phonetic category for the L2 sound. After a while, when being produced, the perceptually associated L1 and L2 sounds will sound comparably (Flege, 1995). This has been supported by a previous study that showed that those voice onset times (VOT) for consonantal stops in L1 by bilinguals were similar to those they produced for consonantal stops in L2 (Flege, 1987a).

According to the SLM, allophones in L2 within certain positions are perceptually associated to the allophone in L1 that is positionally nearest (Flege, 1995). When L1 and L2 sounds are related in such way, they are diaphones (Weinreich, 1957). Support for this is provided by research that showed that L2 learners are able to acquire the phonetically distinct allophones of the English speech sounds /l/ and /ɹ/ at dissimilar speeds to various levels (Strange, 1992). L1 Japanese speakers, for instance, have problems to perceive and produce the two English liquid consonant speech sounds /l/ and /ɹ/ because only one liquid consonant phoneme exists in the Japanese language, which is a consonantal variation between the allophones [ɭ] and [ɾ]. It was shown that when the English liquid consonants were positioned at the end of words than at the beginning, native Japanese speakers' perception and production of English liquid consonants demonstrated more accuracy (Strange, 1992).

If an L2 sound and the perceptually nearest L1 sound are perceived to be phonetically different to a large extent, these sounds are more probable of being perceived as being phonetically different (Flege, 1995). It has, for instance, been considered that because the sound /r/ in the Japanese language is perceptually nearer to the sound /l/ in English than /ɹ/ in English, the majority of Japanese learners of English will phonetically differentiate /r/ in Japanese better from /ɹ/ in English than from /l/ in English (Flege, 1995; Takagi, 1993).

The SLM proposes that because L1 and L2 categories are present in the same phonological space for an L2 learner, L1 and L2 vowels are scattered in order to sustain auditory contrast in the space (Flege et al., 2003). It is therefore argued that the phonetic category that an L2 learner creates for an L2 vowel is drawn away from the closest L1 speech category. This is known as 'phonetic category dissimilation' (Flege et al., 2003). This means that the L2 vowel thus is different from the phonetic category for that L2 sound in a native speaker. This was supported by a study in which native speakers of Italian who learned English before or after the age of twelve assessed their own capacity to articulate Italian and English (Flege, Munro, & MacKay, 1995a). It was found that while those who learnt English after the age of twelve stated to articulate Italian better than English, the opposite pattern was observed for those who learnt English before the age of twelve (Flege et al., 1995a).

The SLM also argues that the phonetic category for L2 created by an L2 speaker might be different from that of an L1 speaker if the phonetic category for L2 of the L2 speaker is based on dissimilar features than that of the L1 speaker (Flege, 1995). This is considered to be the case when features, not used in L1, differentiate a speech sound in L2 from other L2 speech sounds. This implies that this L2 sound for which a new phonetic category is created by an L2 speaker may not be identical in production to the one generated by the L1 speaker. Therefore the SLM proposes that in L2 speakers L1 and L2 systems are continually in use (Flege, 1995).

However, a weakness of the SLM is that certain hypotheses cannot be tested very easily due to limited availability of appropriate measurements (Flege, 1995). The SLM, for instance, proposes a larger perceived distance of an L2 speech sound from the nearest L1 sound leading to a higher probability of the creation of a distinct category for the L2 speech sound. It has also been argued that this perceived phonetic distance is slighter when L2 is learned sooner for the creation of an L2 category (Flege &MacKay, 2004). Nonetheless, no objective measuring instruments exist at the present that can help assess the extent of perceived L2 learners' perceived phonetic distance of an L2 sound from the nearest L1 sound (Flege, 1995). Moreover, it is not clear what system of measurement L2 speakers make use of to evaluate their perceived cross-language phonetic distance between sounds in L1 and L2 (Flege, 1995).

In addition, there is contradictory evidence regarding the proposition of a higher accuracy of producing L2 speech sounds that are non-existent in L1 compared to L2 sounds that do exist in L1 (Flege, 1988b). Although such a difference in accuracy would be evidence of an appropriate differentiation between novel L2 sounds versus similar L2 sounds in L1, it has been shown that even very proficient L2 speakers showed inaccurate articulation of some L2 sounds that do not exist in L1 (Munro, Flege, & MacKay, 1996).

2.5 Best's Perceptual Assimilation Model (PAM)

Best's Perceptual Assimilation Model (PAM) (Best et al., 1988; Best, 1993; Best, 1994a, 1994b; Best & McRoberts, 2003) was proposed to address the conceptual weaknesses of Werker's model, SLM and NLM that cannot account for those variations in the discrimination of L2 sounds that are not due to existent or non-existent phonetic features in L1 (Best et al., 2001).

In contrast to the other models, Best's PAM includes ideas from the phonological theory of articulatory phonology, according to which gestures represent descriptions of articulatory incidents (Browman & Goldstein, 1992; Fowler, 1986). Gestures are determined by place and manner of articulation as well as by articulatory organs (Best, McRoberts, & Goodell, 2000). This agrees with PAM's direct realist view that the information that listeners discover in speech is about the articulatory gestures by the articulators that produced the speech signal (Best, 1995; Fowler, Best, & McRoberts, 1990). According to PAM, L2 sounds are perceived depending how similar or different they are to L1 sounds they are articulatorily nearest to. Perceived articulatory characteristics including articulatory organs, place and manner of articulation are considered to assist listeners in perceiving how similar and different L2 sounds are from L1 sounds (Best, 1994a). The distance perceived between the L2 sounds and the nearest L1 sound results in differences in discriminability.

According to PAM, during the first half of their first year, infants discover general, linguistically non-specific articulatory patterns in both L1 and L2 (Best & McRoberts, 2003). It is during the second half of the first year that infants exhibit language-specific influences when discriminating L2 contrasts. Because of perceptual learning, they are considered to identify articulatory patterns that they have familiarised themselves in L1. By 10-12 months, infants' perception of speech adjusts to L1 articulatory-phonetic patterns, especially to L1 arrangements of gestures (Best & McRoberts, 2003). While before this perceptual adjustment, infants have universal responsiveness to perceive basic dissimilarities between individual gestures (e.g. opening of tongue tip versus opening of lip), as a result of this perceptual adjustment, infants are considered to observe how gestures are integrated into L1 arrangements (e.g. opening of lip with changed position of the laryngeal organ to perceive how different consonant contrasts are created). Having just adjusted to phonetic-articulatory patterns, infants are considered to undergo truly phonological adjustment to the L1 system of minimal contrasts and phonological alternations not during this period but later in development (Best, 1994a, 1994b, 1995).

It is argued that infants' responsiveness to those phonetic characteristics that operate contrastively to enable infants to differentiate between words in L1 eventually influence their discrimination skills in speech (Best, 1994a). When infants encounter problems in differentiating between L2 contrasts, this is considered to be caused by the articulatory resemblance between particular L2 and L1 categories (Best & McRoberts, 2003).

PAM argues that phonetic similarities that listeners discern between the L1 phonological system and L2 sounds will determine the discrimination of L2 sounds (Best, 1999). There are three ways in which listeners are considered to perceptually assimilate L2 sounds. First, L2 sounds can be allocated to the L1 phonological category as an acceptable or a deviant exemplar. Second, L2 sounds can represent uncategorised sounds and fall between L1 categories into the open phonetic space. Third, L2 sounds can be situated completely outside of the L1 phonological system (Best, 1995).

The manner in which L2 sounds are assimilated determine how L2 phonetic contrasts are discriminated (Best, 1995). The pairwise assimilations that emerge as a result are considered to include at least the following types of assimilation: First, if two L2 contrast sounds are phonetically similar to two different L1 phonemes and assimilate to them individually, this is known as ‘Two Categories’ (TC). This produces very good discrimination as two L2 sounds are separated by L1 phonological boundaries. In TC assimilation, L1 phonology therefore supports discrimination (Best, 1994a, 1995).

Second, if two L2 sounds assimilate to a single L1 phoneme strongly or weakly to an equal extent, this is known as Single Category assimilation (SC) (Best, 1995). This assimilation produces poor discrimination because when two L2 sounds assimilate to the same L1 phoneme, L1 phonology obstructs discrimination (Best, 1994a, 1995). This is illustrated by native adult Japanese learners of English who show weak discrimination of the English liquid phonemes /ɹ/-/l/ (Goto, 1971; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975). Because /ɹ/-/l/ are phonemically non contrastive in Japanese (Tsushima et al., 1994), native Japanese speakers tend to show the assimilation of the English /ɹ/-/l/ as weak instances of the individual Japanese phoneme /r/, thereby producing poor discrimination (Miyawaki et al., 1975; Takagi & Mann, 1995; Vance, 1987). Support for TC assimilation was provided by a study that tested English listener's perception of Hindi dental-retroflex stop contrasts (Polka, 1991). It showed that TC-type assimilations were related to higher discrimination scores than SC-type assimilations, which is consistent with PAM (Polka, 1991).

The third assimilation type is the assimilation of two L2 sounds to one L1 phoneme, with one L2 sound matching up better with the L1 phoneme than the other. This is known as 'Category Goodness' difference (CG) (Best, 1995). Although this assimilation produces good discrimination, it is not as good as for TC assimilation because both L2 sounds assimilate to the same L1 sound in CG. The ability to discriminate these L2 sounds is therefore obstructed by L1 phonology (Best, 1994a, 1995). Evidence for better discrimination for CG than SC assimilation was shown in a study that tested native English listeners' perception of a Farsi stop contrast and a Salish contrast (Polka, 1992). It was found that English listeners assimilated the Farsi stop consonants as a CG contrast while assimilating the Salish consonant as a SC contrast. Consistent with PAM, they showed better discrimination for the CG contrast than the SC contrast (Polka, 1992).

Regarding the assimilation of vowel contrasts, it was found in line with PAM that American adult listeners assimilated a Norwegian vowel contrast as a SC category, and showed low accuracy in the discrimination of that contrast. Their assimilation of a Thai vowel contrast as CG difference assimilation produced good discrimination.

Similarly, their assimilation of another Norwegian vowel contrast and two French vowel contrasts as TC assimilations showed very good discrimination (Best, Faber, & Levitt, 1996).

Further support for a better discrimination of TC contrasts and CG contrasts than for SC contrasts was shown by a study that native English listeners' perception of Zulu and Tigrinya consonant contrasts (Best et al., 2000). Specifically, regarding the Zulu consonants, it was observed consistent with PAM that the listeners assimilated L2 lateral fricatives and velar stops as a TC assimilation and CG assimilation respectively, with the majority of listeners showing SC assimilation for L2 bilabial stops (Best et al., 2000). Similarly, regarding discrimination between contrasts it was reported in support of PAM that listeners demonstrated better discrimination for lateral fricatives than velar stops and better discrimination for velar stops than bilabial stops (Best et al., 2000). The pattern for TC assimilation and its related high discriminability was also found with the Tigrinya consonant contrast (Best et al., 2000).

A fourth assimilation type is when two L2 sounds consist of an uncategorised sound (an L2 sound that is approximately comparable to at least two L1 phonemes) and a categorised sound (an L2 sound as an exemplar of an L1 phoneme with a goodness of correspondence that can vary from strong to weak) (Best, 1995). This represents an Uncategorised-Categorised (UC) pair. Similar to TC contrasts, UC contrasts are considered to produce good discrimination since each of the L2 sounds in the UC contrast are on opposing areas of an L1 phonology boundary. When two L2 sounds include two uncategorised sounds, they represent an Uncategorised-Uncategorised

(UU) pair, for which discrimination is not influenced by L1 phonology. Therefore, depending on how comparable the L2 sounds are perceived to be both to proximate L1 phonemes and to each other, their discriminability is considered to vary from poor to good (Best, 1994a, 1995). Evidence for both UC and UU contrasts have been provided by a study that tested native adult Japanese listeners' perception of English consonants (Guion, Flege, Akahane-Yamada, & Pruitt, 2000).

It was shown native Japanese listeners assimilated the English /ɹ/-/l/ contrast as an UU contrast because both /ɹ/ and /l/ come in between two specific L1 categories in Japanese (Guion et al., 2000). Consistent with PAM, discrimination of the UU contrast was poor. The /ɹ/-/w/ and /s/-/θ/ contrast were assimilated as UC contrasts. However, good discrimination was only shown for /ɹ/-/w/. In contrast, /s/-/θ/ was poorly discriminated as /θ/ was perceived as positioned between to L1 sounds. As this is not consistent with PAM's prediction for UC contrasts, this indicates that PAM needs to be revised in order to be able to better explain cases in the discrimination of uncategorised versus categorised L2 sounds in which the uncategorised sound is found nearby to the categorised sound (Guion et al., 2000).

Lastly, if two L2 sounds cannot be assimilated to any L1 phoneme, they are known as 'Non-Assimilable' sounds (NA). They are perceived as non-linguistic sounds and thus produce quite good discrimination, which is subject to how different they are perceived as non-linguistic sounds. Best et al. (1988), for instance, showed that English adult listeners perceive non-native Zulu click contrasts as non-linguistic sounds due to their inability to assimilate them to English consonants. Consistent with

PAM, English listeners' discrimination of the Zulu click contrasts was very good (Best et al., 1988).

Recently, due to the emphasis of PAM on articulatory gestures, it has been combined with the articulatory organ (AO) hypothesis by Goldstein (Best & McRoberts, 2003; Studdert-Kennedy & Goldstein, 2003), which has been created to theoretically expand the theory of articulatory phonology (AP) and which states that primary articulatory organs such as the larynx and lips that contribute to the generation of a word are what infants note when they perceive the word. According to AO, the discrimination of a minimal phonetic contrast that is differentiated by two dissimilar gestures elicited by the same primary articulator is known as a within-organ contrast while the discrimination of a minimal contrast that is differentiated by one gesture made by dissimilar articulators is known as a between-organ contrast. Because gesture information such as the rate or exact location of a gesture may not be able to be identified by infants, infants are considered to find the discrimination of within-organ contrasts more problematic than that of between-organ contrasts (Best & McRoberts, 2003).

According to the PAM/AO, there will be an earlier and more notable decline in 10-12 month old infants' discrimination of L2 within-organ contrasts than that of L2 between-organ contrasts when the L2 sounds in the L2 within-organ contrasts are perceived to belong to a L1 phonetic category (Best & McRoberts, 2003). This prediction was tested in a study with 10-12 month old infants' discrimination of three Zulu contrasts as within-organ contrasts and of a Tigrinya contrast as a between-organ contrast, compared to that by 6-8 month old infants (Best & McRoberts, 2003). It was shown that while the 6-8 month old infants were able to discriminate all three L2 contrasts, the 10-12 month old infants displayed a decline in the discrimination of the L2 contrasts that included one articulatory organ. Moreover, consistent with PAM/AO it was observed that both infant groups discriminated the L2 between-organ contrast (Best & McRoberts, 2003).

However, there are some observations that cannot be accounted for by PAM/AO. For example, research findings by other researchers on contrasts in which two L2 sounds are assimilated to a CG pattern seem to be inconsistent with PAM's assumption for CG assimilations (Best, 1995; Polka & Werker, 1994; Polka & Bohn, 1996). While Polka and Bohn (1996), for instance, reported in line with PAM good discrimination of two CG vowel contrasts, another study showed a lack of discriminability of two other CG vowel contrasts (Polka & Werker, 1994). This indicates that PAM requires further development to account for infants' phonetic speech perception by including a broader variety of non-native contrasts (Polka, Colantonio, & Sundara, 2001).

Moreover, it has been observed that for certain L2 consonant contrasts, such as Zulu bilabials, a remarkable individual variation exists in the way these contrasts are assimilated (Best et al., 2000). Since phonetic characteristics of L2 sounds can be similar to diverse L1 sounds on many levels and the focus on particular levels can vary among different listeners, the PAM/AO might need to be developed further to take this into account in future studies (Best et al., 2000).

In addition, although there is research that suggests a biological specialisation of the left hemisphere for discerning particularly linguistic information in speech (Best &

Avery, 1999), the PAM/AO does not specify the kind of biological specialisation that may assist listeners to discover articulatory information in speech and that particularly may help infants' adjustment to L1. Therefore more evidence is needed for the PAM/AO in order to find out how biological specialisation for speech might lead to infants' L1 adjustment before the beginning of infants' second year of life (Best & MacRoberts, 2003).

2.6 Kuhl's Native Language Magnet theory (NLM)

Kuhl's native language magnet theory is another account that attempts to explain the developmental change in infants' phonetic perception between six and twelve months of age (Kuhl, 1994; Kuhl, 1998; Kuhl, 2000a, 2000b; Kuhl et al., 2008). The model consists of three developmental phases. Infants in the first phase are considered to be able to discriminate between practically all the sounds of all languages in the world due to general auditory mechanisms (Kuhl, 1991b). Phase two is concerned with phonetic representations that are created as result of infants' responsiveness to distributional and perceptual characteristics in linguistic information (Kuhl, 2000b). The phonetic representations are therefore particular to infants' ambient language and indicate the distributional characteristics of sounds in their specific language (Kuhl, 1998).

Infants' perception is considered to become distorted through experience of their ambient language (Kuhl et al., 1992). It is argued that as result of early language exposure, the perceived distances in the acoustic space that are considered to lie beneath phonetic distinctions are altered or warped through experience (Kuhl, 1994). Those phonetic representations that are most frequently stimulated through experience are regarded to operate as phonetic prototypes and exert perceptual magnet effects

(Kuhl, 1991a). The nearby perceptual space around them is considered to be reduced since the region around them has decreased discrimination sensitivity in contrast to regions surrounding nonprototype members of the category (Kuhl, 1991a). Therefore, the discrimination of phonetic variation is more problematic in the region surrounding prototypes than surrounding weak examples of the same category.

Due to the distorting consequence of the perceptual magnet effect on perception, infants' phonetic capacities in L1 perception during the third developmental phase are considered to be supported while infants' phonetic capacities in L2 speech perception decrease (Kuhl et al., 2008). The proposition of infants' responsiveness to distributional frequencies in L1 phonetic patterns was supported by a cross-language study in which six months old infants displayed more responsiveness to distributional frequencies in L1 than to those in L2 (Kuhl et al., 1992). This sensitivity has been suggested to instigate the emergence of prototypes in the first half of their life (Kuhl et al., 1992). The development of these prototypes has been argued to have consequently changed infants' phonetic perception and to have contributed to their perception of L1 vowels at the age of six months (Kuhl, 1993). Six months is the earliest age at which phonetic perception was observed to be influenced by experience of the ambient language (Kuhl et al., 1992). Evidence of phonetic learning by 6-8 months old infants after having been exposed to syllables from an eight-stimulus continuum (Maye, Werker, & Gerken, 2002), and of phonetic learning by infants of about 9 months who learnt L2 contrasts after having been exposed to L2 (Kuhl, Tsao, & Liu, 2003) further support the importance of infants' sensitivity to distributional and perceptual characteristics in ambient speech to perceptual learning.

The organization of phonetic categories in acoustic space as proposed by NLM was upheld by research on phonetic prototypes. A prototype here is referred to as an appropriate example of a category (Kuhl & Iverson, 1995; Rosch, 1975). Accordingly, it was shown that adults were proficient at recognising the prototypes of phonetic categories in L1 (Grieser & Kuhl, 1989; Kuhl, 1991a). The results showed that phonetic prototypes operated in a unique manner as they acted as perceptual magnets for other sounds within a category (Grieser & Kuhl, 1989; Kuhl, 1991a). Specifically, the prototype (P) is considered to draw other category members to itself and therefore have a magnetising influence. As a result, the perceived distance between P and other category members shrink (Kuhl, 1993a). In contrast, nonprototypes (NP) are weak examples from the same category that do not operate in this manner.

The NLM proposes that a prototype vowel sounds more comparable to its variants than the nonprotype sounds to its variants even in case of identical acoustic distance (Kuhl, 1991a). It is therefore argued that in order to hear a difference between a prototype and its variants, one has to move farther away from the prototype (Kuhl & Iverson, 1995). The pulling effect of the prototype was reported by a study that assessed adult and six months old infants' capacity to differentiate from NP and P and their respective variations (Kuhl, 1991a). Specifically, when adults and infants listened to a prototype of a phonetic classification and compared it with sounds that were ambient to the prototype in the same acoustic space, it was shown that P was associated to its variations more frequently than was NP, thereby indicating a strong magnet effect (Kuhl, 1991a).

The finding of decreased responsiveness to acoustic differences within the area surrounding prototypes was supported by studies that, for instance, used signal detection methods and theory to evaluate the perceptual magnet effect with a response bias-free measurement (Iverson & Kuhl, 1995; Sussman & Lauckner-Morano, 1995). Moreover, research that assessed the distortion of perceptual space of phonetic categories using multidimensional scaling and that showed shrunk perceptual distance in the area of the ideal instances, and expanded perceptual distance in the area of the least optimal instances, provided further evidence for the perceptual magnet effect (Iverson & Kuhl, 1995; Kuhl & Iverson, 1995).

In addition to infants' discovery of distributional and perceptual patterns in their ambient language, the NLM claims that overemphasised crucial phonetic differences in infant-directed speech (IDS) contribute to improved L1 phonetic perception and reduced L2 phonetic perception (Kuhl et al., 2008). This has been supported by research that reported mothers to differentiate between acoustic cues of phonetic units in an overarticulated manner in IDS (e.g. Bernstein-Ratner, 1984; Burnham, Kitamura, & Vollmer-Conna, 2002; Uther et al., 2007). Exaggeration of differences between phonetic units was observed for both vowels (Kuhl et al., 1997), consonants (Englund, 2005) and it has been suggested to make speech more intelligible (Liu, Kuhl, & Tsao, 2003) and contribute to infants' learning of words (Song et al., 2010).

A further critical aspect of the NLM is the notion of native language neural commitment (NLNC) according to which 6-12 month old infants' neural tissue is modified through their acquisition of statistical regularities and acoustic patterns within L1 (Kuhl et al., 2006, 2008). The neural connections that are sensitive to the regularities in L1 are consequently reinforced whereby intricate patterns such as words in L1 that are based on originally acquired regularities in the ambient language are learnt (Kuhl et al., 2008). In contrast, the sensitivity to alternative patterns that are not in line with regularities that have been acquired before within the ambient language, such as L2 patterns, will be reduced. Evidence was provided by research that showed raised neural responses to L1 consonant contrasts and reduced responses to L2 contrasts in 7 to 11 months old infants (Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005) and by research that showed raised neural responses to L1 vowels and reduced responses to L2 vowels in 6 to 12 months old infants (Cheour et al., 1998).

Neural commitment is also considered to influence the acquisition of a new language as an adult (Kuhl et al., 2008). Accordingly, the effect of linguistic experience on L2 sound processing was demonstrated by magnetoencephalography (MEG) research that indicated that neurally more time and more brain supplies are involved when adults process L2 sounds (Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005). Nonetheless, it was also shown that overemphasising phonetic units similar as in IDS can assist in adults' acquisition of L2 contrasts (Iverson, Hazan, & Bannister, 2005; Vallabha & McClelland, 2007).

According to the NLM, early phonetic learning is affected by social communication (Kuhl et al., 2008). Research, for instance, that compared English infants' discrimination scores after they have been exposed to Mandarin syllables through different types of sources such as video - or audiotape as well as through a natural language learning condition, demonstrated that infants who experienced Mandarin in a naturalistic language learning condition had higher discrimination scores than infants who had been exposed to the same Mandarin material but in a video-and audiotaped manner (Kuhl et al., 2003).

NLM also states that the connection between speech perception and production is established in infants between 6-12 months old and is created as result of their perceptual exposure and acquired mapping between perceived and produced speech (Kuhl et al., 2008). Accordingly, linguistic experience is considered to enable infants to learn at a sensory level, which then will advance their motor development when infants start associating their own verbalisations to the articulatory activities that initiated these and when the infants replicate the sounds they hear (Kuhl et al., 2008). Evidence from MEG research, for instance, demonstrated that the perceptual processing of syllables causes the brain region for speech production to activate progressively in infants between 6 to 12 months of age (Imada et al., 2006).

The NML proposes that infants' later language development is related by their early perceptual skills in L1 and L2. In a study that applied the head-turn method to assess infants of 7 months on an L1 and L2 consonant contrast, Kuhl and colleagues (2005b) showed that improved L1 perception at this age can be used to forecast faster language growth in infants aged between 14 and 30 months. By contrast, improved L2 perception at the age of 7 months was shown to be a predictor of decelerated language growth (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005b). This finding has been supported by a more recent study that used electrophysiological methods in which 7.5 months old infants' discrimination of L1 and L2 consonant contrasts was found to forecast their linguistic capacity two years afterwards (Kuhl et al., 2008).

Recently, the NLM has been expanded (NLM-e) and is considered to include four phases (Kuhl et al., 2008). In the first phase, infants are considered to be able to crudely distinguish between phonetic units of all languages of the world due to general auditory mechanisms. This phase is also characterised by directional asymmetries that are consistent across age and culture (Kuhl et al., 2006). In phase 2, infants' perception is modified in that their speech perception skills improve in L1 and decrease in L2 (Kuhl et al., 2006). Phase 2 also involves the link between speech perception and production as infants start vocalising and replicating the sounds they hear (DePaolis, 2005). Their phonetic learning is facilitated through social communication as infants start developing an increased awareness of their social surroundings (Tomasello, 2003).

In phase 3, infants are regarded to enhance abilities as result of their increased L1 perceptual capacities that will assist them in the learning of words: these abilities include their observations of phonotactic rules, components that are comparable to words and phonetic information in initially acquired words (Ballem & Plunkett, 2005; Mattys, Jusczyk, Luce, & Morgan, 1999; Newport & Aslin, 2004). In phase 4, relatively consistent neural representations are considered to have been established through exposure to the ambient language due to neural commitment. Infants are considered to be able to acquire many languages because of not entirely developed neural networks (Maye et al., 2002). By contrast, the exposure of a new language to adults is not considered to mechanically develop new neural formations in them due to their steady neural representations. The acquisition of L2 is thus regarded to rely on the extent of neural commitment (Kuhl et al., 2008).

Nonetheless, there are some observations that cannot be explained by NLM-e. It has been, for instance, claimed that the perceptual magnet effect may show consistently for different listener groups (Lively, 1993). Moreover, although according to the NLM-e model, particular perceptual boundaries will be withdrawn when infants acquire the perceptual magnet effect, studies on adults have indicated that boundaries might not vanish (Kuhl & Iverson, 1995; Logan et al., 1991; MacKain et al., 1981). It was, for example, shown that the ability to differentiate contrasts in a foreign language can be enhanced as result of persistent training (Logan et al., 1991). The research results indicated that the sensory capacity for contrast differentiation might not be able to be changed through perceptual modification owing to experience with language. However, it was suggested that linguistic experience might probably cause an alteration on the dimension of attention and also memory (Logan et al., 1991).

In addition, although the NLM-e position that infants' phonetic perception is influenced by speech input can be regarded to be generally supported by previous research, it has been, however, argued that experience with dissimilar phonetic characteristics might not be sufficient to enable infants to create phonological categories (MacKain, 1982). It has been, for example, argued that not the sounds in one's surrounding language but that the linguistically meaningful distinctions might be crucial in shaping the complete perceptual system (Best et al., 2001).

Chapter 2 has covered four developmental models of spoken language acquisition for L1 and L2. These are Werker's Cognitive Theory of spoken language acquisition, Flege's SLM, Best's PAM, and Kuhl's NLM. The final chapter (Chapter 9) will return to one of these models to find out which model is most applicable to explain how non-native speakers of English with differing proficiencies in English process speech with vowel hyperarticulation at word level compared to native speakers. The next chapter will cover the H&H theory, which can be considered to theoretically underlie different types of speech, such as foreigner-directed speech and pet-directed speech, all of which have been observed to contain vowel hyperarticulation.

Chapter 3

Communicative aspects of speech

Chapter 3 deals with speech to different types of listeners that involves vowel hyperarticulation. The chapter starts with the H&H theory that can be considered to theoretically best underpin the observation of vowel hyperarticulation in speech to certain interlocutors. Different speeches to interlocutors such as infants, foreigners, verbal pets and computers are outlined that involve vowel hyperarticulation. It is pointed out that despite much research on foreigner-and infant-directed speech, which are covered in Chapters 5 and 6 respectively, it is still not clear whether vowel hyperarticulation causally leads to increased intelligibility for these interlocutors. This issue is addressed in Experiments 3 and 4 presented in Chapter 7. Towards the end, Chapter 3 highlights an alternative view that proposes that changes in speech production can arise from affective, and not necessarily didactic motivations.

3.1 Hypospeech and Hyperspeech (H&H) theory

3.1.1 Relationship between speaker, audience and environment

Most influential theories of speech perception and production are focused on the underlying physical / perceptual mechanisms of speech without specific reference to communicative context. However, another framework (Hypospeech & Hyperspeech (Lindblom, 1990)) focuses on how speech perception and production changes as a function of the interactive and communicative context.

According to Lindblom's (1990) Hypospeech & Hyperspeech theory (H&H theory), the speech signal is not invariant but instead is a result of listener-dependent modifications. A talker’s task to produce speech can show both temporary and permanent variability and is considerably influenced by the listener and the speaking environment. Speakers are therefore considered to verbally adjust to the perceptual needs of their audience and to the environmental speaking conditions (Lindblom, 1990). This is, for instance, demonstrated by the observation of vowel hyperarticulation in speech directed to infants (IDS) who are considered to require atypically clearer speech input than adults (ADS) (Fernald et al.,1989; Grieser & Kuhl, 1988; Uther et al., 2007).

The perception of speech entails the process of discriminating between items that are in one’s lexicon, i.e. the differentiation between lexical items, such as words, that can be distinguished through phonemes, such as between ‘bat’ and ‘pat’ (Lindblom, 1990). The access to these lexical items is considered to depend on the distinct nature of an acoustic stimulus. One’s distinction of one item from contending lexical items is achieved to the extent to which speakers can articulate lexical items distinctly (Lindblom, 1990). By this means, the aim of increased intelligibility is achieved through maximised discriminability within the limitations of a communicative situation (Lindblom, 1990).

Speakers adapt their speech using information on their environmental listening circumstances and on their audience’s background knowledge to evaluate the audience’s information. This information, which does not include speech signal related information, is known as signal-complementary processes. Speakers are considered to make use of signal-complementary processes to change their speech accordingly to enhance communicative efficacy (Lindblom, 1990). As a result of speakers' assumption that their speech output must be sufficiently contrastive to enable lexical access, the audience's discrimination of lexical items, and therefore lexical access, can be viewed to be facilitated by signal-complementary processes.

3.1.2 Hypo-and hyperspeech

Lindblom considered hypospeech to occur in a type of communicative situation in which light demands are put on the articulators, as for instance in a silent room as this presents minimal constraints to produce speech (1990). Hypospeech is an economical mode of behaviour and is regulated by making sure that lexical items can be sufficiently differentiated so that speakers do not overly economise and their speech becomes unintelligible (Lindblom, 1990). By comparison, speakers use hyperspeech when the demands placed on the articulators (known as system-oriented control) are substantial. This is, for example, the case when speech is produced in a noisy environment. The purposeful, communicative character of speech that is described by articulators is known as output-oriented control.

Because the articulators can be malleably rearranged to meet the restrictions of the communicative environment, the purpose-driven speech motor regulation exemplifies plasticity in form of hyperspeech (Lindblom, 1990). Depending on the requirements of the communicative situation, speakers talk along the scale from hypospeech to hyperspeech (Lindblom, 1990). As a speaker moves from hypospeech to hyperspeech, the articulatory gestures show less temporal overlap and show rise in both duration and amplitude (Lindblom, 1990). Listeners' articulatory and acoustic patterns consequently rely less on context in hyperspeech than in hypospeech (Lindblom, 1990).

Moreover, while hypospeech contains coarticulation and vowel and consonant reduction, hyperspeech includes vowels and consonants that are aimed at being articulated as accurately as possible (Lindblom, 1990). The aspects of plasticity and economy enable signals to be generated that are sufficiently informative (Lindblom, 1990). The aspect of economy has been empirically supported in several animal studies in which animals were observed to have a variety of behavioural choices and to nonetheless move in a way that caused them a minimal effort of physical power and that corresponded with the task (Alexander, 1988; Hoyt & Taylor, 1981). In contrast, plasticity is the consequence of a regulation that is output-oriented and through which animals have been observed to achieve the same motoric aim in various contexts such as under different constraints (Granit, 1977; Lashley, 1951).

3.1.3 Signal-plus-knowledge hypothesis

The H&H theory proposes that speech perception deals with partial information because auditory objects are identified when the signal is modulated through what is stored in the recipient's memory (Lindblom, 1996). According to this signal-plus-knowledge hypothesis, auditory percepts are the result of both the physical signal and the recipient's knowledge (Lindblom, 1996). This is regarded to be applicable to both lower and higher levels of language processing.

Empirical support for the proposed signal-plus-knowledge hypothesis is provided by Kluender, Diehl, and Killeen (1987) in whose study Japanese birds were taught in the first experimental part to differentiate between syllables that included or excluded /d/ by pecking when hearing the syllables /dVs/ but not /gVs/ and /bVs/, where V stands for vowel. The effective completion of the task with four different vowels was considered to have occurred due to learning, with the birds having formed a store of the auditory parameters of the single stimuli in their memory (Lindblom, 1996). The acoustic demarcation between /d/, /g/ and /b/ was considered to have been neurally supported through the creation of neural networks in the bird brains (Lindblom, 1996). In the second experimental part, the same discrimination task with dissimilar vowels was used and the birds completed the task successfully (Kluender et al., 1987). Lindblom (1996) argued that the result can be accounted for by the birds having had associated the auditory parameters of the new stimuli with the stored input from the previous stimuli. According to Lindblom (1996), the task used in Kluender et al.'s (1987) study might be considered to include an interplay of signal-plus-knowledge. The birds are seen as having had made judgements based on not simply the physical signal but also by using phonetically organised knowledge (Lindblom, 1996).

According to the H&H theory, speech production is a changeable verbal response to a mutable communicative environment. The role of the speech signal is to provide the listener with sufficient information to make speech comprehension possible (Lindblom, 1992). Thus, speech production can be regarded as a listener-oriented modification with the goal to ensure lexical distinctiveness and therefore sufficient contrast.

The idea of speech production being an adaptive response has been empirically illustrated by the observation of clear speech, which includes the interaction between the aims of meeting the listener's linguistic needs on the one hand and simplifying speech production on the other hand (Payton et al., 1994; van Summers et al., 1988). Studies on clear speech support the claims of the H&H theory that the speaker's task varies with the listener's needs and the communicative situation and that the speaker can segmentally adapt speech output. Studies on infant-directed speech (IDS) and foreigner-directed speech (FDS), for example, have illustrated consistently with claims of the H&H theory that adults modify their speech (Ferguson, 1977; Kuhl et al., 1997; Uther et al., 2007) to generate a speech signal that maximises discriminability and therefore intelligibility (Kuhl et al., 1997; Liu et al., 2003; Song et al., 2010;).

Specifically, while speech to infants was reported to include speech modifications such as decelerated rate of speech, increased pitch and extended vowel space (Fernald & Simon, 1984; Grieser & Kuhl, 1988; Kuhl et al., 1997; Stern et al., 1983), speech directed to foreigners was found to involve the reduction of grammatical difficulty levels, simplified negations, simplifications of sentences and increased vowel space among others (Campell, Schlue, & Vander Brook, 1977; Gaskell, Cambell, & Vander Brook, 1977; Hatch, Shapira, & Gough, 1975; Meisel, 1977; Uther et al., 2007). Moreover, speech to children (CDS) was observed to include expanded vowel space, decreased grammatical intricacy, rise of reiterations, elucidations and further explanations (Biersack, Kempe & Knapton, 2005; de Paulo & Coleman, 1986; Liu et al., 2009). Thus, it can be said that both IDS, FDS and CDS display common verbal adaptations such as expanded vowel space, which as a part of clear speech can be seen in consistency with the H&H theory, as a verbal adaptation to meet listeners' linguistic needs.

Additional support for the H&H theory is provided by other examples of verbal adaptations such as the observed changes in speech to elderly people that is louder and slower in nature (Kemper, Ferrell, Harden, Finter-Urczyk, & Billington, 1998) while speech to hearing-impaired individuals includes a rise in duration, volume and the change from casual to clear speech articulatory patterns (Picheny et al., 1986). Another example is that in the presence of environmental noise, speakers will aim to preserve a consistent degree of speech clarity by putting in more verbal effort (Draegert, 1951; Hanley & Steer, 1949; Lane & Tranel, 1971).

Moreover, a rise in pitch and overall amplitude under environmental noisy situations and the longer duration of sentences in noise conditions than of sentences in quiet conditions are, for example, some of the reliable alterations of a speech register called the Lombard reflex (Dreher & O'Neill, 1957; Pisoni, Bernacki, Nusbaum, & Yuchtman, 1985; Van Summers, et al., 1988). Thus, it can be said that these studies provide supporting evidence of the notion of the H&H theory that a talker's speech is substantially determined by the listener and the speaking environment and that a talker's verbal adaptations to the communicative situation involves the generation of a speech signal that can vary on a continuum from hypospeech to hyperspeech (Lindbom, 1990, 1996).

The H&H theory was also tested by Lively, Pisoni, Summers and Bernacki (1993) who investigated the effect of high cognitive workload on verbal adaptations in speech production. They theorized that when speakers are required to complete an attention-demanding task whilst speaking, they are likely to heighten system-oriented control to preserve speech clarity (Lively et al., 1993). It was therefore hypothesised that, compared to the speech produced in the control condition; speech generated under workload would adjust to the workload environment and will be of shorter duration so that the workload task could be focused on. Specifically, speech articulated during the workload condition was considered to exemplify hyperspeech while the speech generated in the control condition was considered to include hypospeech (Lively et al., 1993). In a perceptual identification task, it was further tested whether the sentences articulated during the workload condition were perceived by listeners to be more intelligible than those generated in the control condition (Lively et al., 1993).

In the first part of their experiment, five male native speaking participants were asked in the workload condition to articulate test sentences whilst completing a visual tracking task. In the control condition, they were instructed to speak the sentences (Lively et al., 1993). It was found that in contrast to the control condition, speakers showed the propensity to generate sentences of shorter duration, with larger amplitude and more variability in amplitude in the workload condition (Lively et al., 1993). Nonetheless, workload was not found to lead to a significant modification in the first three formant frequencies of the vowels by the five speakers (Lively et al., 1993).

In the second part of their experiment, vowel tokens were taken out from a section of each sentence in the workload and control conditions and presented in a perceptual identification task to listeners (Lively et al., 1993). It was found that listeners perceived the vowel tokens in the workload condition as more intelligible than those in the control condition. Specifically, intelligibility was observed to improve two to five per cent from the control to the workload condition for three of five speakers (Lively et al., 1993). The three speakers who produced sentences of more intelligibility in the workload condition were also reported to produce the sentences under workload that demonstrated reliable elevations in amplitude and amplitude variability (Lively et al., 1993).

Thus, the main aspect controlling intelligibility was suggested to be generated by changes in amplitude and amplitude variability. Despite absent changes in formant frequencies, it was argued that the verbal adjustments under the workload reflect laryngeal and sublaryngeal adaptations in speech and changes in the absolute timing of articulatory movements (Lively et al., 1993). The occurrence of a small but significant increase in intelligibility across conditions was accounted for by the use of isolated vowels for the perceptual task and the observation that the task demand was not substantial. Lively et al. (1993) therefore argued that their study supports the idea by Lindblom (1990) that speakers modify their speech to meet the perceptual needs of their environment with the aim to increase discriminability and therefore intelligibility.

However, one of the limitations of the H&H theory is the fact that the term of ‘sufficient contrast’ has not been defined. It is argued that the purpose of listener-oriented modifications of speech patterns is to sustain sufficient contrast to simplify lexical access (Lindblom, 1990). Nonetheless, no information is given about the manner in which sufficient contrast can be specified. Moreover, the H&H theory assumes the absence of acoustic invariance (Lindblom, 1990). Nevertheless, the theory does not propose testable hypotheses through which the lack of signal invariance can be quantitatively assessed (Diehl & Lindblom, 2004).

Thus, it can be said that the H&H theory can be considered to be able to account for inter-and intra-talker variability and to integrate variability instead of searching for invariance. Nonetheless, the H&H theory still has to account for the observation why studies that focus on invariant physical correlates of linguistic units have not been effectual in specifying the invariant correlates of these units. Moreover, it does not provide hypotheses on which the lack of signal invariance can be empirically established (Diehl & Lindblom, 2004). The following sections will look in more detail at verbal modifications that are made in speech to different audiences and that can be accounted for by the H&H theory (Lindblom, 1990, 1992, 1996).

3.2 Infant-directed speech (IDS)

Speech directed at infants (IDS), also known as motherese, differs from adult-directed speech (ADS) since IDS involves acoustic changes such as raised mean pitch, exaggeration in pitch contours, decelerated speech rate, and increased emotional affect (Andruski & Kuhl, 1996; Kitamura & Burnham, 2003; Kuhl, 2004; Stern et al., 1983; Trainor & Desjardins, 2002). Moreover, IDS includes modifications such as long pauses, brief sentences, additional reiterations, and exaggerated vowel space (Fernald & Simon, 1984; Fernald & Kuhl, 1987; Fernald et al.,1989; Grieser & Kuhl, 1988). The following sections will look more closely at those speech modifications that have been proposed to contribute to infants' learning of words. These speech modifications are vowel space exaggeration, decreased speech rate, and increased pitch.

3.2.1 Vowel space exaggeration in IDS

Infant-directed speech contains phonetically exaggerated properties that have been correlated with speech intelligibility in IDS (Burnham et al., 2002; Kuhl et al., 1997;

Liu et al., 2003; Uther et al., 2007). In particular, expanded vowel space (as measured by *F*1/*F*2 space) has been shown to represent very reliable measurements of overall clarity in speech, which has been regarded to help infants to focus on acoustically crucial information (Bradlow et al., 1996; Turner, Tjaden, & Weismer, 1995).

Research on infants' acquisition of phonetic features of their native language has shown that IDS show hyperarticulated vowels (Kuhl et al., 1997; Liu et al., 2003). Kuhl et al. (1997), for example, have analyzed the acoustic qualities of IDS directed at infants from Sweden, Russia and the United States. They have observed that the vowels /i/, /a/ and /u/ in IDS were acoustically more exaggerated than in ADS with the consequence of the vowel space acoustically being "stretched" (Kuhl et al., 1997, p. 684). This is illustrated in Figure 3.1, with the vowel triangles for /a/, /i/, and /u/ produced in IDS marked by solid circles, and those produced in ADS marked by open circles.

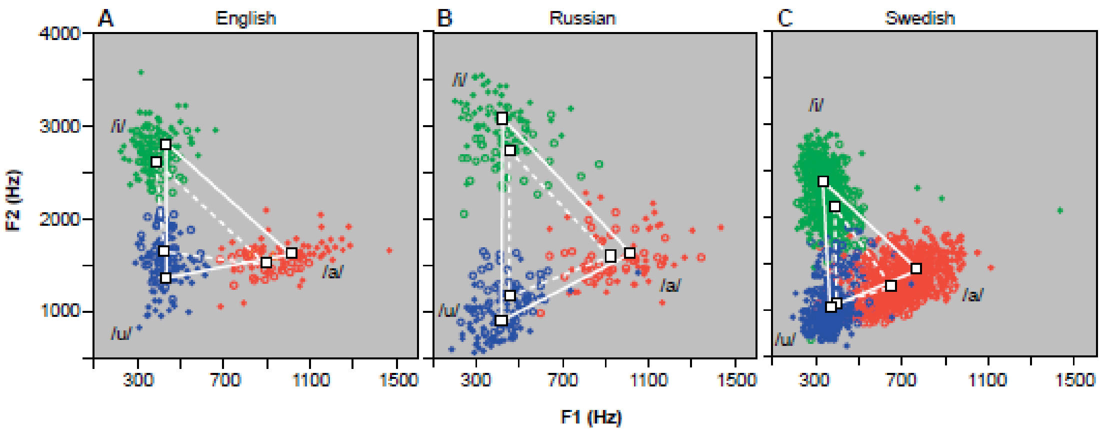


Figure 3.1: Vowel triangles formed by the "point" vowels, /i/ (green), /a/ (red), and /u/ (blue), in infant-directed (solid circles) and adult-directed (open circles) speech in three languages—English, Russian, and Swedish. Each data point represents the coordinate of the first two formant frequencies of a vowel. A universal stretching of the vowel triangle is observed in infant-directed (solid line) relative to adult-directed (dashed line) speech. Adapted from Kuhl et al. (1997).

This occurrence of vowel hyperarticulation in IDS, which was observed across diverse languages, might help infants to learn phonetic units (Kuhl et al., 1997). Similarly, adults have experienced hyperarticulated vowels as improved examples of vowel categories (Iverson & Kuhl, 1995; Johnson, Flemming, & Wright, 1993).

Regarding the effect of exaggerated vowel space on infants' speech processing capacities, Liu et al. (2003) reported an association between the qualitative speech effort by mothers and infants' performance on phonetic speech perception. In this study, the extent of vowel space area, which was produced in speech by mothers, was correlated with infants' phonetic discrimination abilities (Liu et al., 2003). It was observed that the use of more sizeable vowel space by mothers was significantly linked to a better linguistic discrimination performance by infants (Liu et al., 2003).

Another study analysed how IDS acoustic characteristics such as expanded vowel space influenced 19-month-old infants' word recognition by applying the preferential looking method (Song et al., 2010). It was found that infants' recognition of words improved when vowels were hyperarticulated. However, as this finding was only based on the time course measures of infants' looking to test and distractor stimuli when listening to typical IDS and altered IDS (Song et al., 2010), no findings of any difference in total looking time or latency of the first look to the test stimuli as a purpose of vowel space were made. There is therefore no clear evidence of exaggerated vowel space being definitely beneficial for infants' learning of words.

3.2.2 Changes in speech rate

In previous studies, decelerated speech rate has regularly been observed to be an IDS feature (Cooper & Aslin, 1990; Fernald & Simon, 1984; Fernald et al., 1989; Stern et al., 1983). It was shown not only to be an aspect of IDS produced by mothers who talked to their actual infants (Fernald &Simon, 1984) but also to be a characteristic of speech produced by female adults in imaginary IDS (Cooper & Aslin, 1990). In both actual and imaginary conditions of IDS production, speech rate was slowed down in speech to infants more than to adults (Cooper & Aslin, 1990; Fernald & Simon, 1984).

One study (Cooper & Aslin, 1990), tested whether newborns and 1-month old infants displayed longer looking time at a visual item when their looking elicited IDS or speech to adults. Making use of a modified version of the visual-fixation-based auditory-preference procedure, the experimenters found both newborns and 1-month old infants to have a more prolonged looking period when IDS was generated by their looking than when ADS was produced (Cooper & Aslin, 1990). It was therefore suggested that this result could be attributed to the significantly longer duration of sentences in IDS than in ADS. This indicates that the overemphasised aspect of longer mean duration of sentences in IDS is an aspect of IDS that plays a role in newborns' preference of IDS over ADS (Cooper & Aslin, 1990).

Research on the effect of decelerated speech rate on infants' capacities of processing speech has also shown that decelerated stimuli are more beneficial for infants' word recognition (Zangl, Klarman, Thal, Fernald, & Bates, 2005), listening comprehension of sentences (Nelson, 1976) and learning of words (Song et al., 2010). In one study, it was revealed, for example, that infants between one to about three years of age showed higher accuracy for words in a listening word identification task when the words were not modified compared to when the words were modified to be twice as quick as the non-modified words (Zangl et al., 2005). More recently, it was shown in

Song et al.'s study that 19-months old infants' enhanced capacity to recognise words was contributed to by decelerated speech rate, in addition to exaggerated vowel space (2010). Song et al. (2010) therefore suggested that a decelerated speech rate is crucial for infants' improved learning of words because it assists in partitioning between speech sounds and therefore give infants instructive linguistic information.

3.2.3 Changes in mean pitch and pitch range (F0)

The observation of broader range in *F*0, more increased *F*0 and also positive emotion in IDS is well documented and their use has been suggested to attract and maintain infants' attention, and to express emotion (Burnham et al., 2002; Fernald, 1989; Kitamura & Burnham, 1998, 2003; Kuhl et al., 1997; Stern et al., 1982). Studies have also shown that broader range in *F*0 and more increased *F*0 in IDS are observable in tonal and nontonal languages (Fernald et al., 1989; Grieser & Kuhl, 1988; Xu, Burnham, & Kitamura, 2007).

Alterations in pitch range have also been observed to be specific to infants' age (Kitamura & Burnham, 2003; Stern et al., 1983). One longitudinal experiment, for instance, found mothers to overemphasise pitch range more when directing speech to infants of four months than to older infants or to new-borns (Stern et al., 1983). This finding was argued to be due to the necessity of a more varied pitch to motivate infants at four months to vocally express themselves while for older infants, a larger pitch range, as typical of IDS, is not principally used to sustain their attention or stimulate their verbal articulation (Stern et al., 1983). In contrast to four months old infants, a further decreased extent of pitch range was adequate to maintain new-borns' attention (Stern et al., 1983).

In another study, it was reported that in speech to nine months old infants, mothers communicated instructively with further reduced mean pitch and more varied pitch range, while in speech to infants at six and twelve months, more increased mean pitch was utilised to express positive emotions (Kitamura & Burnham, 2003). Although these experiments seem to suggest that infants' developing language capacities are matched by prosodic aspect of IDS, modifications in IDS regarding sentence structure were reported not to be associated with infants' age (Kavanaugh & Jirkovsky, 1982), indicating that it is not clear whether IDS is overall adjusted to infants' language abilities at a certain age.

Recent research by Song et al. (2010) looked at the effect of broad pitch range on infants' capacities of speech processing and found that compared to slow speech rate and exaggerated vowel space, broad pitch range did not contribute to infants' enhanced recognition of words. According to this study, exaggerations in pitch range do not provide infants with linguistic details that are crucial for the learning of words. Although alterations in pitch range has been suggested to possibly affect infants of tonal languages such as Mandarin Chinese, it was concluded consistent with prior research (Scherer, 1986) that in non-tonal languages such as English broad pitch range is used to attract infants' attention (Song et al., 2010).

3.3 Pet-directed speech (PDS)

Speech directed to pets has been observed to share common aspects with CDS and IDS (Hirsh-Pasek & Treiman, 1982). For example, it was shown that in contrast to ADS but similar to CDS, speech to dogs had a mean length of words of about four words (Hirsh-Pasek & Treiman, 1982; Snow, 1977) and included more instructive commands than declarative statements (Newport, Gleitman, & Gleitman, 1977). Moreover, speech to dogs contained frequent repetitions (Cross, 1977; Snow, 1972), higher portion of grammatically correct statements, simple sentences (Newport et al., 1977), and the use of diminutives (Hirsh-Pasek & Treiman, 1982).

Recently, it was observed in similarity to CDS (Liu et al., 2009), speech to verbal pets such as parrots contains hyperarticulated vowels (Xu et al., 2004). Given the fact that parrots, as for example the African Grey Parrot, have a propensity and capacity to imitate speech (Pepperberg, 2002), this finding suggests that one circumstance under which hyperarticulated speech seems to occur is when the interlocutor is regarded to be capable of speech (Xu et al., 2004).

Regarding changes in pitch and affect in speech to pets such as dogs it was found that, mean pitch and affect was higher in PDS and IDS than ADS (Burnham et al., 2002). It was, however, observed that only IDS was accompanied by vowel hyperarticulation and not PDS, implying that exaggerated vowel space is an instructive instrument (Burnham et al., 2002). This finding seems to agree with the above finding that vowel hyperarticulation might occur in PDS if pets are able to verbalise with adult speakers to a certain extent (Xu et al., 2004).

However, Kim, Diehl, Panneton and Moon (2006), who compared mothers' speech to puppies (PupDS) with the same mothers' speech to their infants, reported that hyperarticulation was expressed to puppies and that this hyperarticulation was not significantly different to that in IDS. This indicates that not educational desires but emotional articulacy might be expressed through vowel hyperarticulation (Kim et al., 2006). Nonetheless, this research had a sample size of six mothers and has not compared PupDS with speech directed at adult dogs. Therefore, the purpose of vowel hyperarticulation may not be exclusively attributed to emotional articulacy. It can thus be said that changes observed in PDS seem to have the purpose to draw and sustain attention and to express positive affect.

More recently, a study that compared ADS, IDS and speech to dogs and parrots (Xu, Burnham, Kitamura, & Vollmer-Conna, 2013) has supported the important role of speakers' assumption about the interlocutors' linguistic competence in order for vowel hyperarticulation to occur. It was found that the extent of vowel hyperarticulation linearly rose from ADS and speech to dogs to speech to parrots and then IDS (Xu et al., 2013). This finding implies that the extent to which vowel hyperarticulation occurs can be associated to the listeners' real or assumed linguistic capability (Xu et al., 2013). Thus, the observed changes in PDS can be said to serve to convey positive emotion and engage the attention of the pet.

3.4 Foreigner-directed speech (FDS)

In contrast to PDS and IDS, interactions in FDS are considerably influenced by the speakers' dissimilar cultural backgrounds, their experiences and assumptions they have about particular people and cultures (Storti, 1999). Knowledge about how understanding between different cultures can be improved can overall be regarded as being particularly useful with regard to multicultural societies since they involve daily interactions between individuals of different ethnicities. Verbal communication in FDS can therefore be considered to occur in multicultural societies anytime and anywhere including schools and workplaces (Snow et al., 1981).

Past research on speech adaptations in FDS focused on foreigners with both a foreign appearance and a foreign accent (Ferguson, 1975; Snow et al., 1981; Littleford, Wright & Sayoc-Parial, 2005; Uther et al., 2007). Nonetheless, these previous investigations do not consider the possibility that multicultural societies can include individuals who physically appear foreign but linguistically are as competent as native speakers. Consequently, similar to the possibility that native speakers might create linguistic and cultural misconceptions when speaking to individuals with native accents and foreign appearances, native speakers might possibly also produce misconceptions when talking to foreigners, who came to a country as adults and therefore have foreign accents but native physical appearances.

Thus, it can be said that previous FDS studies did not differentiate between foreigners' appearance and accent and their separate and joint effects on native English speakers' speech addressed to them. It is possible that foreign-looking and native-sounding foreigners might cause native speakers to modify their speech in a way that is different to native speakers' speech to foreigners who sound foreign but look native. At present, it is not known whether appearance and accent separately cause acoustic alterations in native speakers' speech or if they contribute to the same acoustic effects. Moreover, little research has been conducted to find out if interactions with native looking and foreign sounding foreigners (NLFS), and with foreign looking and native sounding foreigners (FLNS) might give rise to possible communicative problems.

Past research on foreign looking and foreign sounding foreigners (FLFS) has suggested that difficulties in verbal interaction that involve foreign looking and foreign sounding foreigners may possibly cause harm at the workplace. Recent research has revealed that, for example, more attention from listeners is needed when speakers speak in Asian-accented English speech than in English with standard American accents (Hosoda, Stone-Romeko, & Walter, 2007). This might be due to the difficulty to comprehend foreign accented speech. According to Hosoda et al. (2007), this result indicates the very crucial implication that if individuals have restricted attention because they are, for example, completing an intricate task or operating on several tasks, information from individuals with foreign accents might not be correctly dealt with. This may result in misinterpretation between the native listener and the foreign-accented speaker, which may lead to some damage of workplace relationships and to reduced productivity.

This might also affect the manner in which Asian-accented speakers are perceived and assessed (Hosoda et al., 2007). According to Ryan, Hewstone, and Giles (1984), foreign-accented English speakers may come into contact with decreased chances for social progress since they are considered not to be as potent and good communicators as English speakers with native accents. Particularly, it has been shown that a bias towards foreign-accented speakers exists among employers (Sato, 1991). Moreover, speakers with foreign accents have been reported to be relegated by native speakers because of their foreign accent (Anisfield, Bogo, & Lambert, 1962).

Nonetheless, it is important to note that foreign interlocutors' linguistic needs might not be the only aspect that might motivate native English speakers to modify their speech. Recent research, for instance, demonstrated that although native English speakers talked more clearly by using overemphasised vowels to foreigners with foreign appearance and foreign accent than they did to native English interlocutors, native English speakers indicated more negative affect in speech to foreign interlocutors than to native English interlocutors (Uther et al., 2007). The observed clear speech to the foreign interlocutors might have been caused by native English speakers' realisation of foreign interlocutors' linguistic needs while alternatively the result of negative emotion expressed at foreigners may imply native speakers' exasperation at the necessity to alter their speech (Uther et al., 2007). On the other hand, it was argued that the result of negative affect might reflect the holding of unfavourable positions such as prejudice (Uther et al., 2007). Prejudice has been described to among others include the expression of negative affect towards an individual of a group due to its belonging to that group (Brown, 1995).

3.4.1 Linguistic Adaptations in FDS

Speakers' linguistic ability can be regarded as crucial in an intercultural interaction. This is because difficulties in understanding the language the communication is based on can represent obstacles that may affect the way native English speakers talk to foreigners (Snow et al., 1981). Past research on linguistic adaptations in interethnic speech has focused on speech directed to foreigners with both foreign appearances and foreign accents (Ferguson, 1975; Snow et al., 1981; Warren-Leubecker & Bohannon III, 1982).

The characteristics of FDS related to the patterns of syntax and associated with single words have been investigated with the aim of discovering specific modifications that describe FDS (Ferguson, 1975; Warren-Leubecker & Bohannon III, 1982). It was shown that FDS in hypothetical situations includes the reduction of grammatical difficulty levels, and single words, which the vocabulary of the spoken language is composed of (Ferguson, 1975). Specifically, to describe FDS, Ferguson instructed students to state how they would phrase particular sentences or clarify notions to foreigners whose first language is not English (1975). It was found that easier grammatical constructions were applied that were similar to those used in speech directed to children (Ferguson, 1975).

While Ferguson's study (1975) explored some potential consequences of speakers' previous expectations of listener competence in FDS, research by Warren-Leubecker and Bohannon III (1982) employed a paradigm using natural conversation in which speakers' expectations of listener ability and the responses showing the listeners' understanding were evaluated in terms of their capability to lead to simplified FDS. It was discovered that verbal reactions by foreign participants indicating their speech comprehension during FDS will normally cause simplified FDS irrespective of speakers' early anticipation of the listener's ability (Warren-Leubecker & Bohannon III, 1982). Additional research on natural FDS has shown that native speakers modify their speech by making use of simplified negations and by reproducing foreigners' errors (Hatch et al., 1975; Meisel, 1977) while other studies reported an absence of errors in simplified FDS sentences or did not observe linguistically typical alterations in FDS at all (Campell et al., 1977; Gaskell et al., 1977).

With respect to these contradictory results, Snow et al. (1981) have argued that one has to regard the aspect of native speakers' familiarity with regular contact with foreigners that can shape the speakers' personality and consequently affect the way they approach non-natives. Snow et al. (1981) also suggested that the foreigners' apparent social status as well as their competence in their non-native language, on which FDS is based, might influence aspects of FDS. These views can be supported by these researchers who found that when native speakers had frequent contact with foreigners, a significantly positive correlation existed between the native speakers' inclination to make grammatical, lexical and syntactic alterations and the foreigners' linguistic errors (Snow et al., 1981). Consequently, it can be said that previous research has identified the aspects determining the flow of FDS: the foreigners' linguistic abilities in their non-native language and the frequency of contact of native speakers with foreigners. The subsequent section will look at studies on the acoustic characteristics of FDS.

3.4.2 Acoustic characteristics of FDS

Investigations on the acoustic characteristics of FDS have shown that speakers apply acoustic-phonetic changes in FDS (Scarborough et al., 2007). These adjustments are acoustically different from speech to native speakers and consistent with alterations observed in listener-directed comprehensible speech such as speech to people with a damaged auditory function (Picheny et al., 1986; Scarborough et al., 2007; Uther et al., 2007). Regarding linguistic necessities, FDS has qualitative similarities with IDS or CDS beyond the age of three including increased number of repetitions. Similarities between CDS and FDS also include decreased grammatical intricacy, rise of reiterations, elucidations and further explanations (Biersack et al., 2005; De Paulo & Coleman, 1986).

Nonetheless, referring to pitch maxima and minima, FDS and CDS are regarded to be dissimilar (Biersack et al., 2005). This can be illustrated by Biersack et al.'s experiment (2005) that aimed to discover if FDS arises from speech to children. The experimenters asked twelve female participants to do a referential communication task directed to a hypothetical child, a hypothetical foreigner and a hypothetical grown-up. Pitch range and pitch maxima were found to be higher in CDS than in FDS. The notion that in terms of pitch contours, FDS is dissimilar to CDS and IDS has been well supported and it has also been found regarding ascending and descending patterns of pitch alterations that FDS is similar to ADS (Knoll, Uther, MacLeod, O'Neill, & Walsh, 2006).

One aspect that FDS, CDS and IDS have in common, however, is expanded vowel space (Burnham et al., 2002; Kuhl et al., 1997; Liu et al., 2009; Uther et al., 2007). The idea that native speakers might make use of hyperarticulated vowels in speech to foreigners (Burnham et al., 2002) was tested by Uther et al. (2007) who showed that in natural interactions with foreigners and infants, native British English speakers produced larger vowel formant space than in communication with native English speakers. Vowel length in FDS did not differ significantly from that in IDS and ADS. Similar to prior research (Biersack et al., 2005), pitch was low in both FDS and ADS compared to high pitch in IDS (Uther et al., 2007).

Moreover, while more negative emotion was expressed in FDS than ADS, more positive emotion was conveyed in IDS than ADS (Uther et al., 2007). The finding of negative affect in FDS compared to ADS is similar to the outcomes of investigations that compared vocal affect toward standard American English speakers with the affect to speakers with foreign accents (Cargile & Giles, 1997; Hosoda et al., 2007; Ryan & Bulik, 1982; Sebastian, Ryan, Keogh, & Schmidt, 1980). Accordingly, Hosoda et al. (2007) for example reported more negative affect toward English speakers with Asian accents than to English speakers with standard American accents.

These results therefore illustrate that articulatory stretched vowel space does not rely on affective alterations in speech as it can transpire in the presence of both increased negative affect and increased positive affect (Uther et al., 2007). At the same time, this study supports the idea that increased vowel space might be of didactic use for both native learners (infants) and non-native learners (foreigners) of English (Uther et al., 2007).

Thus, it can be said that the observed changes in FDS seem to serve to simplify speech to foreigners by making phonemes more distinct from each other through vowel hyperarticulation. However, little is known about whether comparable changes in speech can be observed if foreigners have foreign appearance and native accents and if they have native appearance and foreign accents.

3.5 Speech to computers: Human-Computer Interaction (HCI)

Vowel hyperarticulation has not only been reported in IDS, FDS and or speech aimed at parrots but also in human-computer interaction (HCI) (Burnham, Joeffry, & Rice, 2010; Oviatt, 1998; Oviatt, Levow, Moreton, & MacEachern, 1998; Oviatt & Coulson, 2003; Stent, Huffman, & Brennan, 2008). In case of errors occurring in HCI, the automatic speech recognition (ASR) system cannot recognise information or substitutes it with an incorrect word, humans are understood to make corrections in order to ensure that computers comprehend them (Burnham et al., 2010).

In order to make HCI as comparable to human-human interaction (HHI) as feasible, different types of mending that occur in HCI have been explored. Accordingly, an increase in the length of utterances, a reduction in speech rate, and a rise in the occurrence and periods of pauses have been reported in HCI (Oviatt et al., 1998; Stent et al., 2008). Children have also been observed to use these speech modifications as well as heightened amplitude in HCI (Oviatt et al., 1998). According to Oviatt (1998), the Computer-elicited Hyperarticulate Adaptation Model (CHAM) can in two phases explain methodical alterations in human speech when errors are dealt with in an interactive setting.

In the first stage, speech modifications include long periods of speech and pauses and are moderately related to hyperarticulation (Oviatt, 1998). Hyperarticulation is considered to arise in the first stage when the rate of errors is small. On the contrary, speech modifications in the second stage are related to more intense hyperarticulation when there is a large rate of system error (Oviatt, 1998). Alterations in speech can affect articulation, pattern of intonation, amplitude and fundamental frequency ($F0$) (Oviatt, 1998). The CHAM model predicts, among other predictions, that there will be comparable adaptations to dissimilar kinds of system recognition error in speech (Oviatt, 1998).

Moreover, it is predicted that during global amendments, hyperarticulated modifications will be used for a whole statement while in the course of focal improvements, hyperarticulated adaptations will relate to a single syllable or term in a larger statement (Oviatt, 1998). These predictions have been supported by empirical results. Alterations in articulation such as reduced verbal dysfluencies and the amplified number of hyper-clear phonological qualities, for example, represented a reasonably important quality of hyperarticulated modification that have been reported

(Oviatt et al., 1998). Particularly, speech directed to computers was not only characterized by reduced speed of speech and by increased distinction of words but also by more intentional and improved specifications of cues indicating phonetic distinctiveness (Oviatt et al., 1998). These speech modifications were clearly observable during a high error rate.

3.5.1 Linguistic theory of CHAM model

The linguistic theory, from which the CHAM model theoretically originates, is Lindblom's H&H theory (Lindblom, 1990; Lindblom et al., 1992). In agreement with this linguistic theory on hyperarticulation (Lindblom, 1990) and the second stage of the CHAM model (Oviatt, 1998), Burnham et al. (2010) have shown by comparing HCI and ADS that an increased computer error rate in HCI brings about higher vowel duration and $F0$ range and also expanded vowel triangle areas. Accordingly, compared to ADS, HCI showed an expansion of vowel space whereby vowels were hyperarticulated towards the computer. However, although the vowel duration was more extended in HCI than in ADS, particularly after correcting the computer, a larger $F0$ was not observed in HCI. This led Burnham et al. (2010) to suggest that HCI is comparable to FDS in this regard since mean $F0$ has been reported not to increase in FDS either. Thus, HCI research seems to suggest that the changes seen in HCI seem to serve to improve humans' interaction with computers.

3.6. Summary

The table below represents a summary of discussed changes in the acoustic-phonetic components for different speech registers.

Table 3.1: Summary of the acoustic-phonetic components for different speech registers.

Speech registers	Acoustic-phonetic components: Vowel space expansion	Pitch	Vowel length	Loudness	Speech rate
Infant-Directed Speech (Fernald et al., 1989; Kuhl et al., 1997)	Yes	High	Long	Not tested	Slow (resulting in longer vowels)
Infant-Directed Speech with hearing loss (Lam & Kitamura, 2010)	No	High	Long	Not tested	Slow
Infant-Directed Speech with simulated hearing loss (Lam & Kitamura, 2012)	No	High	Long	Not tested	Slow
Child-Directed Speech (Garnica, 1977; Liu et al., 2009)	Yes	High	Long	Whispered to younger children	Not tested
Child-Directed Speech (with speech perception deficits) (Bradlow et al., 2003)	Yes	High	Not tested	High	Slow
Foreigner-Directed Speech (real) (Scarborough et al., 2007; Uther et al., 2007)	Yes	Low	Long	Not tested	Slow
Foreigner-Directed Speech (imaginary) (Biersack et al., 2005; Scarborough et al., 2007)	Yes	Low pitch range	Long	Not tested	Slow
Pet-Directed Speech (verbal pets) (Xu et al., 2004; Xu et al., 2013)	Yes	High	Not tested	Not tested	Not tested
Pet-Directed Speech (non-verbal pets) (Burnham et al., 2002; Xu et al., 2013)	No	High	Not tested	Not tested	Not tested
Computer-Directed Speech (Burnham et al., 2010; Oviatt, MacEachern, & Levow, 1998)	Yes	High pitch range	Long	Not tested	Slow
Speech to elderly (normal hearing) adults (Helfer, 1998; Picheny et al., 1986; Schum, 1996)	Yes	Low	Long	High	Slow
Speech to hearing-impaired adults (Ferguson & Kewley-Port, 2002; Payton, Uchanski, & Braida, 1994;. Picheny, Durlach, & Braida, 1985; Picheny et al., 1986; Uchanski, Choi, Braida, & Durlach, 1996)	Yes	Low	Long (tense vowels)	High	Slow
Speech to hearing-impaired, elderly adults (Ferguson, 2012; Picheny et al., 1986; Schum, 1996)	Yes	Low	Long	High	Slow
Adult-directed speech (Uther et al., 2007)	No	Low	Medium	Not tested	Not tested

Shouted Speech (Rostolland & Parant, 1974; Rostolland, 1982a; Rostolland, 1982b)	**No (only increase in F1 frequency)**	**High**	**Long**	**High**	**Not tested**
Lombard speech (in noisy environment) (Cooke & Lu, 2010; Lombard, 1911; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988)	**No**	**High**	**Long**	**High**	**Slow**

3.7 Sociolinguistic approaches to variation in speech style

While the H&H theory and related theories such as the CHAM model focus on phonetic variation, and speaker adjustments in terms of the potential need for clear signal information, (Lindblom, 1990), other approaches such as sociolinguistic perspectives focus on stylistic variation, and emphasise the social context as triggers for variation in speech (e.g. Bell, 1984; Bell & Johnson, 1997; Giles, 1973; Giles & Coupland, 1991; Labov, 1966). 'Style' is defined as a collection of linguistic variations, made at phonological, lexical and syntactic levels by a speaker, that are considered to hold dissimilar social meanings (Bell, 1997). Social meanings can refer to individual characteristics or membership within groups, and are considered in speech to be conveyed through linguistic variation. Stylistic variation in speech thus indicates intergroup and interpersonal connections (Bell & Johnson, 1997). This section will outline and evaluate two influential sociolinguistic approaches: first the accommodation theory (Giles, 1973; Giles & Coupland, 1991), which looks at variation in speech from a social psychological perspective, and is considered to be one of the leading social theories of social interaction (Littlejohn & Foss, 2005). The second approach is audience design (Bell, 1997, 2001), which is regarded in sociolinguistics the most common view with regards to style (Eckert & Rickford, 2001).

3.7.1 Accommodation theory

The accommodation theory emphasizes communicators' emotions and aims that pervade the speakers' speech output and its construal through listeners (Giles, 1973; Giles & Coupland, 1991). Formerly known as the Speech Accommodation Theory (SAT), it derives from theories of social psychology including similarity-attraction theory, causal attribution theory, social-exchange theory, and the theory of intergroup distinctiveness (Giles & Coupland, 1991). The similarity-attraction theory, which states that attraction between two individuals will be stronger if the resemblance in their viewpoints is higher (Byrne, London, & Rees, 1968), contributed to the notion within accommodation theory that interlocutors will converge through verbal and non-verbal behavior to indicate and sustain mutual support if they have similar attitudes and convictions (Giles & Coupland, 1991). The social-exchange theory regards individuals to evaluate rewards and costs of their prospective behavior before taking action (Homans, 1958; Thibault & Kelley, 1959), and led to the idea that interlocutors use speech convergence because it might cause them more rewards, including higher social appreciation, than costs (Giles & Coupland, 1991). The causal attribution theory states that interlocutors assess a speaker's motives that are attributed to cause the speaker's action (Heider, 1958). This contributed to the notion that speakers will be perceived as negative, when the speaker's act of convergence was regarded to be caused by outside pressures (Giles & Coupland, 1991).

The theory of intergroup distinctiveness considers contact with dissimilar groups to cause individuals within a group to compare their group with other groups based on skills and achievements (Turner & Tajfel, 1986). Because language represents an essential aspect of group identity (Sachdev & Bourhis, 1992), accommodation theory assumes that a group will show divergence in speech to another group in an attempt to verbally distinguish itself and to preserve group identity (Harwood, Giles & Palomares, 2005). SAT was later renamed Communication Accommodation Theory (CAT) in later revisions of SAT so that it encompassed sociolinguistic and nonverbal aspects of a communicative situation in addition to the verbal aspects of an interaction. In terms of the present thesis, focus will primarily be given to the most recent version of the accommodation theory, i.e. CAT.

3.7.1.1 Communication Accommodation Theory (CAT)

CAT addresses interlocutors' initial orientation when approaching an interaction that includes their previous experiences with others in an interpersonal and intergroup context (Dragojevic & Giles, 2014) and in the existing socio-historical context (Giles & Ogay, 2006), and also addresses socioliguistic and affective aspects, identity processes, such as being a member of a group, and also nonverbal characteristics, such as smiling, within a communicative situation (Coupland, Mulac, Bradac & Johnson, 1988). The two main accommodation strategies within CAT are termed convergence and divergence (Giles & Clair, 1979; Giles & Coupland, 1991).

Within CAT, convergence is the adoption of the interlocutor's manner of communicating, which is caused by the speaker's wish to increase reciprocal understanding and communicative effectiveness, and to decrease social distance (Coupland et al. 1988). Speech modifications can include changes in pronunciation or intonation (Giles & Clair, 1979; Giles & Coupland, 1991). In divergence, an interlocutor is considered to emphasize and sustain their individual or group identity by differentiating themselves verbally and nonverbally from other interlocutors

(Coupland et al., 1991). Speakers who converge are seen as positive, and competent and supportive communicators, compared to diverging communicators (Soliz & Harwood, 2006), who are considered to obstruct mutual agreement, and reduce interlocutors' interactional appreciation, and the positive ascriptions about the speakers (Giles & Ogay, 2006). It is regarded to be possible that a speaker can diverge linguistically with an interlocutor and converge psychologically at the same time, and vice versa (Thakerar, Giles & Cheshire, 1982). Thus, interpersonal and intergroup aspects as well as aspects related to identity and attitude can affect, and be influenced by accommodation and nonaccommodation in communication (Soliz & Giles, 2014).

3.7.1.2 CAT as an explanation of communicative variation in native speakers' interaction with non-native speakers

Native speakers' communicative aims can include communicative efficacy and reciprocal comprehension (Coupland et al., 1988; Ferguson, 1971, 1975; Zuengler, 1991), which can be shown in form of variations in speech, described as 'foreigner talk' (FT) (Ferguson, 1975), such as decelerated speech rate, use of high-frequency words and reduced use of contractions when the communication is instantaneous and deals with intricate topics (Long, 1983; Snow et al., 1981, Zuengler, 1991). Another communicative aim includes obtaining non-native speaker's social approval (Coupland et al., 1988), which occurs when native speakers use FT to encourage non-native speakers' linguistic exertions (Clyne, 1981; Evans, 1987). Native speakers' maintenance of their distinctiveness can be another aim (Zuengler, 1991), and can be shown by native speakers' disinterest in supporting communicative effectiveness while upholding ethnocentric views (Clyne, 1981; Perdue, 1984; Zuengler, 1991).

CAT argues that native speakers' communicative aims can be modified by their perception of non-native interlocutors' linguistic ability (Coupland et al., 1988) because in case of low ability, native speakers use FT more to increase reciprocal understanding (Coupland et al., 1988; Kleifgen, 1985; Zuengler, 1991). Communicative aims can also be modified by their perception of non-native interlocutors, which can alter over the course of the interaction (Coupland et al., 1988). For example, if ethnical or cultural dissimilarities are regarded as a threat, native speakers might choose not to utilize FT to maintain distinctiveness, while those, who aim to gain social approval, may use FT to disregard ethnical or cultural dissimilarities (Zuengler, 1991).

If native speakers adapt their speech to make it more similar to what they perceive non-native speakers' speech to be, and consequently make changes in their choice of words, articulation and speaking rate, they display convergence in speech. However, when native speakers hold ethnocentric views towards non-native speakers (Coupland et al., 1988), they use maintenance that is characterized by an absence of FT (Perdue, 1984). The strategy of divergence is used if native speakers notice that their speech maintenance is not making them sufficiently distinctive from the non-native speaker. Similar to maintenance, divergence does not include FT but it makes sure that a larger speech difference exists (Zuengler, 1991). If native speakers wish to stress that they assume roles different from those taken up by non-native speakers, they are considered to use the speech style of complementarity (Coupland et al., 1988). Complementarity can include or exclude FT. If native speakers view their interactions with non-native communicators as simple non-foreigner-foreigner interactions, they may make use of standard speech if that is the style that they anticipate non-native interlocutors to use (Janicki, 1986).

3.7.1.3 Strengths and limitations of CAT

Since its early beginning as SAT, the CAT has conceptually grown into a framework that focuses on social communication between individuals from dissimilar cultures or groups at interpersonal and intergroup level (Giles & Soliz, 2014). By including a broader variety of communicative behaviors, CAT has demonstrated its interdisciplinary usefulness in contextually diverse situations and studies with many methodological paradigms (Coupland & Jaworski, 1997). For instance, the CAT has been recently used to investigate the communication between non-human primates (Candiotti et al., 2012), and to explore human-to-computer interfaces (Tomko & Rosenfeld, 2006). However, although the main advantage of CAT is that it can be adjusted to diverse communicative contexts, it is evident that investigators' description of accommodative behaviours and the variety of behavior studied depend on the context of their investigation (e.g. overall evaluation of accommodation or study of different types of accommodative behavior) (Soliz & Giles, 2014). This indicates that this variation in how investigators conceptualise accommodative behavior, and the intricacy of communicative behavior studied can obscure the principles of CAT unless investigators provide clear information on how their description of accommodative behaviours connects to the principles of CAT (Soliz & Giles, 2014). Another weakness of CAT is that it mainly is useful when applied to the subject matter and consequences of an individual communicative interaction because CAT does not attend to how accommodative behavior in a collection of communicative interactions at interpersonal or intergroup level might alter over time (Gasiorek, Giles, Soliz, 2014).

3.7.2 Audience Design

Similar to CAT, the framework of audience design is another approach according to which variations in an individual's style of speaking can be understood as speech production that is designed by speakers in relation to their audiences (Bell, 2001). However, in contrast to CAT but similar to the H&H theory, the notion of audience design addresses modifications in speech style at a linguistic level. According to audience design, which can occur in both multilingual and monolingual settings, a speaker is considered to show the refined capacity to design different styles for a dissimilar variety of dissimilar audiences (Bell, 1997, 2001). The speaker chooses a style mainly as a response to the audience to verbalize either affinity with or distance to them. The framework proposes that style is interactive in character, and originates its significance from the relation of linguistic aspects to specific social groups. Similar to the idea of convergence in CAT, the change in a speaker's style to become more similar to the interlocutor is considered to overall establish audience design.

Thus, a speaker's stylistic variation in speech is viewed to originate from and to reflect variation between a speaker and interlocutor at the social level (Bell, 1997, 2001). Audience design also argues, in contrast to CAT, that in a social interaction, speakers can show affinity with or distance to absent reference groups whilst speaking to present interlocutors (Bell, 1997, 2001). Empirical evidence for audience design was provided by research that demonstrated style shifting in African American speakers (Baugh, 1979; Fasold, 1972), such as by an African American teenager who showed stylistic variations in her vernacular English that were affected by topic and listener type (Rickford & McNairKnow, 1994). Although audience design has been recently expanded in terms of the variety of linguistic analysis, it has been criticized for the lack of regard it has for the possible part that attention might have in speech (Labov, 2001), and is considered to have difficulties explaining which exact aspects of an audience speakers are reacting to in an interaction (Bell, 2006).

3.8 Speech – listener-oriented or speaker-oriented?

Thus, it can be said that in contrast to the H&H theory, sociolinguistic theories consider speech not to be determined by the needs of the listener but they view speech to be more of an interaction between what is practical for the speaker and what social aspects may ascertain whether they behave in accommodative behavior or nonaccommodative behavior. However, such an assumption makes it look as if speech can only be either listener- or speaker oriented, and it therefore ignores the dynamic character of how and why speech is used (Raphael et al., 2007).

Whether an individual speaks to another person by changing their speech to meet the listeners' linguistic needs because they are language learners, hearing-impaired, or because the communication takes place in a noisy environment, or whether changes in speech occur to increase integrity or distance or maintain one's identity depends on multiple variables that can interact with each other, and on substantial alterations in speech that can vary from context to context by speakers (Giles & Coupland, 1991). These variables can include the relationship between the speaker and interlocutor that can qualitatively change over the course of time. Accordingly, if a speaker is a native speaker and is meeting a high-proficient interlocutor for the first time, the native speaker may first assess how well the produced speech in the language they are conversing in (L2) is perceived by the non-native interlocutor. Thus, it may be that at the beginning the native speaker might need to adjust their speech to the linguistic needs of the non-native interlocutor as suggested by the H&H theory (Lindblom, 1990).

After having become accustomed to each other's articulatory speech patterns over the course of several conversations, the native speaker may now know how to shift their speech appropriately to signal convergence, divergence, or maintenance while keeping speech at an intelligible level. Consequently, while the native speaker might have initially used vowel space expansion, slow speech rate, and high intensity range to ensure intelligible speech and sustain the non-native interlocutor's attention, after having talked to the non-native interlocutor several times over a period of time, the native speaker might change their speech towards the interlocutor by using a larger variation in speech rate, low intensity and occasionally vowel space expansion to refer to or introduce new words into the conversation.

Thus, it can be suggested that over time, native speakers may change their attitude toward the ethnic group that the interlocutor belongs to, and as CAT suggests, might show social and psychological convergence, or divergence depending on how the relationship progresses from the level of acquaintance to friendship, and how well the speaker dis/likes the interlocutors' personality, and whether the context in which the interaction occurs again will be at the interpersonal, intragroup or intergroup level, and finally whether the level of contents depth increases over the course of several conversations.

Therefore it seems that speech appears to alter its role and undergo a change in its importance from the first time it is used between a native speaker and non-native interlocutor compared to when it is used at a later point by which the initial linguistic convergence in speech may be accompanied later by social and psychological convergence or divergence as the quality of a relationship between the native speaker and non-native interlocutor progresses or changes. It can therefore be said that whether the H&H theory or CAT/ or audience design seem appropriate to explain variations in speech, depends on the quality of the relationship between two communicators and how its changing nature interacts with situational factors (i.e. whether the setting is at interpersonal or intergroup level), and the context in which speech is produced (i.e. contents), and the aspects related to communicators' personal background and their experiences (e.g. personality, intentions, beliefs) (Giles &Coupland, 1991).

Chapter 3 covered the H&H theory by Lindblom, according to which the speaker's task varies with the listener's needs and the communicative situation. This theory can be considered supported by the observation of speech modifications such as vowel hyperarticulation in speech to infants and foreigners. This theory therefore can be regarded the most suitable that can account for these observations as it suggests that speakers segmentally adapt speech output depending on the communicative context. Speech modifications in interaction with infants and foreigners will be experimentally addressed in Chapters 5 and 6 respectively. The claim of the H&H theory that adults modify their speech to generate a speech signal that maximises discriminability and therefore intelligibility, will be tested in Experiments 3 and 4 presented in chapter 7, in which the intelligibility of IDS and of speech to foreign-accented listeners will be evaluated. Chapter 3 therefore covered the acoustic characteristics of IDS and speech to foreigners in detail while outlining speech modifications in speech to other interlocutors that also involve vowel hyperarticulation.

Chapter 4

An examination of the effects of physical appearance and accent on vowel hyperarticulation

4.1 Vowel Hyperarticulated Speech

Speakers know to adapt their speech to meet the needs of certain situations: for example, one might attempt to increase loudness in a noisy environment (Cooke & Lu, 2010) or, one might exaggerate the aspect of pitch in speech to infants (Kuhl et al., 1997). One type of adjustment in speech is vowel hyperarticulation, which has been reported to occur in infant-directed speech (IDS) (Kuhl et al., 1997; Burnham et al., 2003; Uther et al., 2007) and also foreigner-directed speech (FDS), (Uther et al., 2007). Uther et al. (2007) compared the way British English individuals communicate to infants (IDS) that learn English as their first language, to adult second language speakers of English and to White British adults as native English speakers. The analysis of the target vowels /a:/, /i:/ and /u:/, which were elicited by using the target words 'shark', 'sheep' and 'shoe', showed that in vowel triangle area (VTA), FDS did not differ from IDS (Uther et al., 2007). FDS and IDS had larger formant space than ADS. This result therefore suggests that phonetic alterations in IDS are comparable to those in FDS because both foreigners and infants are considered to represent audiences that are understood to need linguistic support.

With regard to pitch, which is measured in Experiment 1, it was shown that in contrast to FDS and ADS, IDS showed the highest pitch while FDS and ADS were significantly comparable (Uther et al., 2007). This result is consistent with previous research (Biersack et al., 2005) in which speech to children showed the highest pitch whereas FDS and ADS were similar in pitch. With respect to vocal affect, Uther et al.

(2007) observed similar to previous findings that IDS had significantly more positive affect than ADS and that ADS had significantly more positive affect than FDS. Increased pitch and high positive vocal affect therefore cannot represent distinct alterations in language. High positive vocal affect is not a change that can be regularly associated with vowel hyperarticulation due to negative vocal affect in FDS and positive vocal affect in IDS (Uther et al., 2007). Regarding other acoustic features in FDS such as intensity and speech rate, which are measured in Experiment 1 as well, it was found that speech it was found that speech was slower in FDS than in speech to native-accented speakers (Scarborough et al., 2007; Sikveland et al., 2006). This was observed regardless of whether the interactions were imaginary or real (Scarborough et al., 2007). Intensity is also considered to be higher in FDS than to native speakers (Lipski, 2005).

4.2 The purpose of vowel hyperarticulation

The presence of vowel hyperarticulation in FDS appears to be driven by the listeners' perception of the linguistic need of the interlocutor (Uther et al., 2007). It has therefore been proposed that acoustic hyperarticulation might serve a didactic purpose for foreigners (Uther et al., 2007).

However, little is known about whether native speakers' speech at native looking and foreign sounding individuals on the one hand is different from their speech atforeign looking and native sounding individuals, on the other hand. Past research on different speech registers used with foreigners has focused on those with both a foreign appearance and foreign accent (Littleford, et al., 2005; Sankowska, Garcia-Lecumberri, Cooke, 2011; Snow et al., 1981; Uther et al., 2007) (foreign looking and foreign sounding: FLFS). To this end, the first experiment sought to separate the variables of looking and sounding foreign in order to answer the research question on what effect drives hyperarticulation in FDS: appearance or accent or both.

In this experiment, there are four kinds of interlocutors: those who both look and sound native (native looking and native sounding: NLNS), those who appear foreign but linguistically sound like native speakers (foreign looking and native sounding: FLNS), those who appear native but sound foreign (native looking and foreign sounding: NLFS) and those who both look and sound foreign (foreign looking and foreign sounding: FLFS). There is a strong theoretical and practical importance in investigating these variables separately so that it can be ascertained whether speech registers are most affected by either interlocutor' appearance or accent. This experiment investigated whether the physical appearance or accent of the interlocutor results in independent effects on eliciting hyperarticulation of a native speaker. This experiment specifically looked at the tense vowels /a:/ (as in 'car'), /u:/ (as in 'blue') and /i:/ (as in 'beach') because these corner vowels represent the furthest points in a speaker's articulatory vowel working space and therefore have the furthest formant frequencies in acoustic space (Cooke et al., 2013; Lindblom, 1990).

Based on theory (H&H theory by Lindblom (1990)) and previous evidence (Bond & Moore, 1989; Burnham et al., 2003; Lipski, 2005; Sankowska et al., 2011; Uther et al., 2007), it is hypothesized that exaggerations will occur in phonemically relevant features such as vowel hyperarticulation, speech rate, intensity, vowel length, and word length in speech to foreign-sounding interlocutors regardless of appearance compared to speech to native sounding interlocutors irrespective of their appearance. This is because native speakers will enhance speech to interlocutors who require linguistic clarifications, such as foreign sounding interlocutors, which therefore will include more acoustic exaggerations than speech to native sounding interlocutors (Lindblom, 1990; Scarborough et al., 2007). However, no significant differences will occur in overall pitch across speakers irrespective of interlocutors' accent and appearance because of absent theoretical underpinnings of a pitch change in native speakers' speech with interlocutors with varying accent and appearance (Uther et al., 2007).

It was not clear from previous studies whether changes in acoustic aspects such as vowel space in speech to foreign interlocutors were due to speakers' didactic purposes or due to unconscious prejudice. To this end, speech samples collected in interaction with native and foreign interlocutors will be rated for positive emotion, negative emotion, and encouragement of attention by a separate group of White British English speakers. Based on prior research (Uther et al., 2007), vocal affect in speech to foreign sounding interlocutors might probably express more negative emotion than speech to native sounding interlocutors. Similarly, vocal affect in speech to interlocutors with native accent and native appearance might probably express more positive emotion than speech to foreign sounding interlocutors with foreign appearance (Uther et al., 2007). Thus, it is hypothesized that the acoustic change of vowel space expansion in native speakers' speech to foreigners might be motivated by native speakers' implicit and explicit attitudes towards foreigners. Regarding appearance, one might expect foreigners' foreign appearance to evoke negative emotion in native speakers (Uther et al., 2007).

4.3 Methods

4.3.1 Participants

Fifty-two White British speakers between 18 and 35 years were asked to communicate with one individual from four different speaker groups (White British individuals, speakers of White European ethnicity with native White British appearance and foreign accent, speakers of Asian (Indian/Pakistani or Bengali) ethnicity with foreign appearance and native accent, and speakers of Asian ethnicity with foreign appearance and foreign accent). Participants were recruited from the student population of Brunel University.

4.3.2 Design

This study used a 2 (interlocutor's accent: native, foreign) x 2 (interlocutor's physical appearance: native, foreign) x 3 (three target vowels: /a:/, /u:/ and /i:/) mixed design. Therefore there were four different types of interlocutors: NLNS (native looking and native sounding), NLFS (native looking and foreign sounding), FLNS (foreign looking and native sounding), and FLFS (foreign looking and foreign sounding). The dependent variable was the degree of hyperarticulation in the target words in which one of the three target vowels was present. Other dependent variables were mean $F0$, mean intensity, mean vowel and word length.

4.3.3 Materials

For the purpose of eliciting the tense target vowels /a:/, /u:/ and /i:/, the words 'car', 'blue' and 'beach' were chosen as specific target words. To facilitate the elicitation of these target vowels from the native speakers, three "Spot-the-difference" (Diapix) tasks were used. These tasks were modified versions of the tasks developed by Baker and Hazan (2009). The first picture depicted a beach scene, the second a farm scene and the third picture a street scene (See Appendix I). A digital voice recorder Edirol R-09HR by Roland (sampling rate: 44.1 kHz) was used to record all verbal interactions. Each interaction was recorded as a mono 16-bit file in wav format. Two measures of explicit prejudice were completed by native speakers that included the Modern Racism Scale (MRS) (McConahay, 1986), and the Resource Allocation Task (RA) (Tajfel et al., 1971).

The MRS consists of six items and gauges racism through six questions that evaluates individuals' racialist attitudes directed at Blacks (Walker & Jussim, 2002). It is considered to be a reliable measurement of explicit prejudice. The RA consists of participants allocating resources such as money to their own group members and to members of another group (Tajfel et al., 1971). Participants have been reported to show ingroup bias since they were observed to allocate more resources to members of their group than to members of the outgroup (Mullen, Brown & Smith, 1992; Tajfel et al., 1971). Thus, the RA task has been considered to demonstrate that participants' allocation indicates ingroup-favouritism and discrimination against the outgroup (Mullen et al., 1992; Tajfel et al., 1971).One question from the Resource Allocation was, for example, "When you had the decision to distribute £400 to your group and a White European group, how much money would you give to each group?" (see Appendix K).

Communication studies showed that explicit prejudice normally represents a prediction of more intentional behaviours such as oral sociability while more spontaneous conduct, such as nonverbal sociability has been forecasted by implicit prejudice (Dovidio, et al., 2002). The most established implicit measurement tool is the Implicit Association Test (IAT) constructed by Greenwald et al. (1998), which was administered to native speakers.

In a vocal affect rating short samples from native speakers' speech to the 4 different types of interlocutors from the Diapix task interactions were rated for positive and negative affect, and encouragement of attention by a separate group of 20 native English speakers on a scale from 1 to 4 (1 = very positive; very negative; high encouragement of attention and 4 = not positive at all; not negative at all; low encouragement of attention) (see Appendix J).

4.3.4 Procedure

In each half an hour audio-recorded interaction, a White British English speaker and an interlocutor were seated opposite each other and each participant received a folder with three pictures, each illustrating a different scene. For each scene, there were 13 differences between the picture that one participant received and the picture of their partner interlocutor. The differences included an absent object or an alteration to one of the objects on the picture. Participants were instructed to work together to verbally find out the differences between their pictures. The task lasted about ten minutes and was used for all three pictures. Demographic and linguistic background information was collected from each participant (see Appendix L). At the end, native speakers completed the MRS, RA questionnaires and the IAT task in a random order.

A vocal affect rating was conducted separately from the Diapix task in which short samples from native speakers' speech to the 4 different types of interlocutors were rated for positive emotion, negative emotion, and encouragement of attention by a separate group of 20 native English listeners. The rating was presented using E-Prime software. The listeners rated on a scale from 1 to 4 the extent to which each speech sample expressed positive emotion, negative emotion, and encouragement of attention (1 = very positive; very negative; high encouragement of attention and 4 = not positive at all; not negative at all; low encouragement of attention).

4.3.5 Data analysis

This experiment employed a mixed ANOVA for the analysis of the acoustic measures (mean vowel triangle area, mean $F0$, mean intensity, mean vowel and word length), with accent (native/ foreign) and appearance (native/ foreign) representing the between-subjects factors, and with vowel (three corner vowels /a:/, /uː/ and /iː/) representing the within-subjects factor (2x2x3 mixed design). A Pearson correlation was used to test the presence of a correlation between vowel space and explicit/ implicit attitude measurements (MRS, RA tasks/ IAT). A MANOVA was used to analyse whether speech rate (number of syllables, words, pauses, length of pauses) and vocal affect ratings (positive and negative emotion, encouragement of attention) differed across the different appearance and accent conditions. Initially 150 target words belonging to one of thirteen target vowels were recorded from native speakers during the completion of the Diapix task (see Appendix I). The vowels /a:/, /i:/, /u:/, /i/, /e/ and /ɒ/ were chosen from the target words “car”, “beach” ,“blue”, “pink”, “red” and “shop” respectively due to sufficient tokens of each of them in each interlocutor condition.

4.4. Results

Table 4.1 below provides an overview of the results for different dependent variables from the mixed ANOVA. These dependent variables that are presented in the first column include mean vowel space, mean pitch, mean pitch range, mean intensity, mean vowel and mean word length. Central to answering the research question of whether accent or appearance contributes to vowel hyperarticulation are the next two columns. These columns present the main effects of accent and appearance while the last column presents the interaction between accent and appearance.

Table 4.1 Summary table of results from mixed ANOVA for acoustic dependent variables. Results are indicated as significant (p< .05) or as not significant (ns).

	Main effects		Interaction
Dependent variables	accent	appearance	accent x appearance
Mean vowel space	p< .05	ns	ns
Mean pitch of vowels	ns	ns	ns
Mean pitch range of words	ns	ns	ns
Mean intensity of vowels	ns	ns	p< .05
Mean intensity of words	ns	ns	p< .05
Mean vowel length	p< .05	ns	ns
Mean word length	p< .05	ns	ns

4.4.1 Acoustic measures

4.4.1.1 Vowel triangle area of target vowels

A mixed ANOVA (2x2x3 mixed design) was used to analyse the effects of appearance and accent on vowel triangle area and it showed that accent significantly differed across conditions (F (2, 40) = 61. 698; p < .05; η^2_p = .755). There was no main effect of appearance. There was no significant accent by appearance interaction.

The vowel triangles with the formant frequencies from the vowels of the target words 'car', 'blue' and 'beach' are shown in Figure 4.1. The mean areas from the vowel triangles are shown in Figure 4.2.

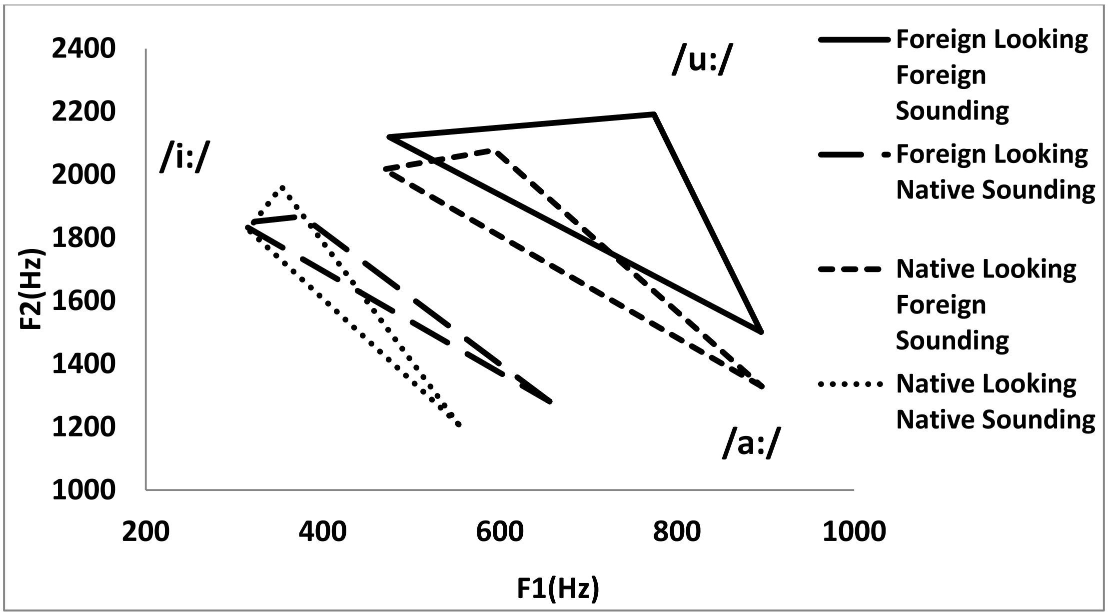


Figure 4.1: Hyperarticulation of target vowels in foreign looking foreign sounding condition and native looking foreign sounding condition and absence of hyperarticulated target vowels in native looking native sounding condition and foreign looking native sounding condition.

A comparison between the foreign sounding conditions and the native sounding conditions revealed that vowel space was significantly larger for the foreign sounding conditions than the native sounding conditions. This clearly shows that the acoustic modification of vowel hyperarticulation is observed in speech to foreign-accented foreigners irrespective of whether their appearance is native or foreign. This finding therefore supports the hypothesis that native speakers will hyperarticulate vowels in speech to interlocutors who require linguistic clarifications, such as foreign sounding interlocutors compared to native sounding interlocutors irrespective of their appearance.

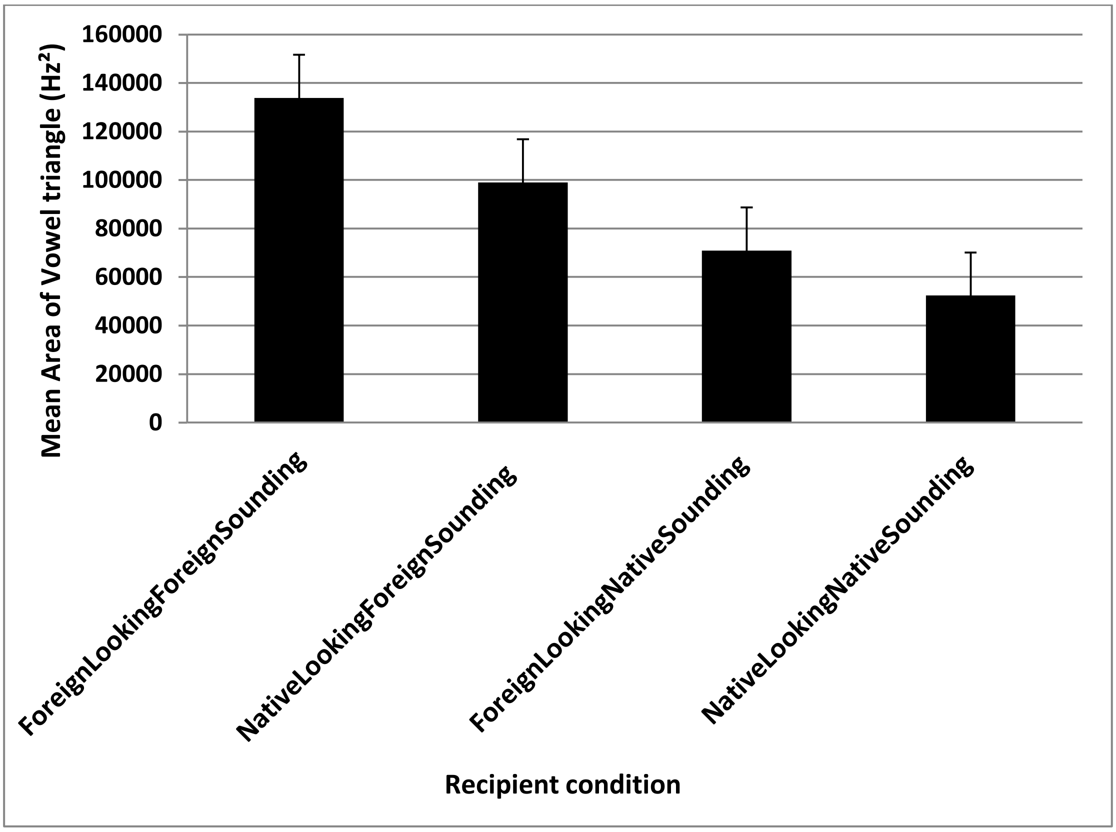


Figure 4.2: Mean area calculated from the vowel triangle in F2/F1 space in the foreign looking foreign sounding condition, foreign looking native sounding condition, native looking foreign sounding condition and native looking native sounding condition.

4.4.1.2 Fundamental frequency (F0) – mean of vowels and range of words

Unsurprisingly, for the mean pitch of vowels, there was neither a significant main effect of accent or appearance nor an interaction between accent and appearance. As expected, for the mean pitch range of words, there was neither a significant main effect of accent or appearance nor a significant interaction between accent and appearance. These nonsignificant results for mean *F*0 and range are as theoretically expected because changes in mean *F*0 were previously not observed to depend on changes in interlocutors' accent or appearance. This finding therefore supports the hypothesis that there will be no significant differences in overall pitch in native speakers' speech to native-and foreign sounding interlocutors regardless of appearance.

4.4.1.3 Mean intensity of vowels and words

This section is going to assess only the mean intensity of vowels and words because mean intensity range was not reported in prior literature to be modulated when speaking to interlocutors who differ in accent and appearance. Unsurprisingly, for both the intensity of vowels and words, there were no significant main effects of accent or appearance. However, as Figure 4.3 for the intensity of vowels and as Figure 4.4 for the intensity of words illustrate, there was a significant interaction between accent and appearance for both intensity of vowels (F (1, 41) = 10.927; p < .05; η^2_p = .210) and the intensity of words (F (1, 41) = 10.549; p < .05; η^2_p = .205): specifically, intensity in speech to foreign-accented interlocutors was not affected by appearance (if anything, in speech to foreign-accented interlocutors, intensity was marginally higher for those with native than foreign appearance); while intensity in speech to native-accented interlocutors was significantly decreased for native appearance and largely increased for foreign appearance.

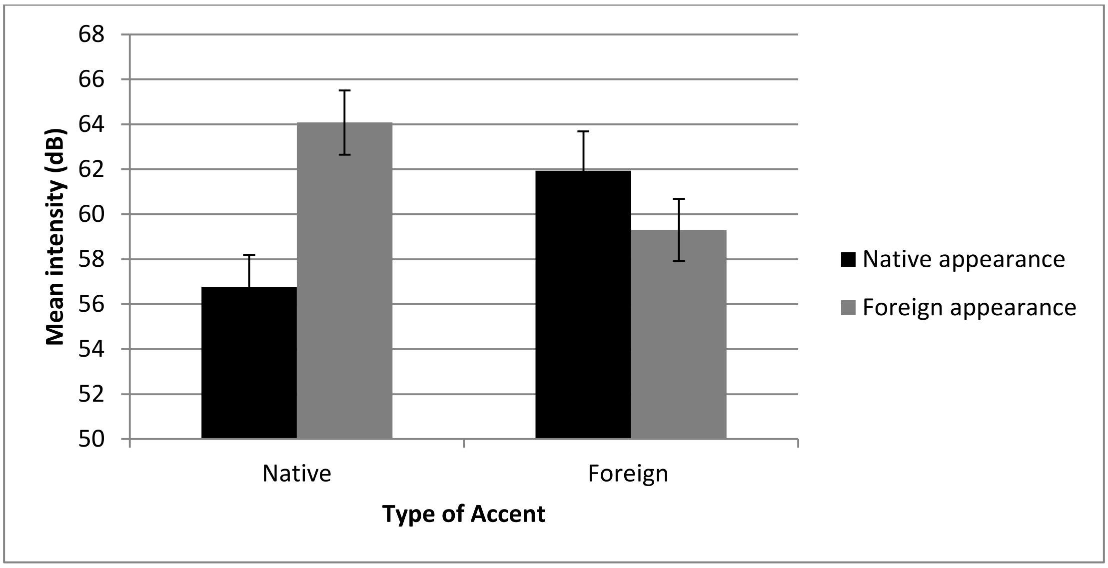


Figure 4.3: Mean intensity (in dB) as a function of accent (native sounding; foreign sounding) in interaction with appearance (native looking; foreign looking) for. Error bars show + /-1 standard error from the mean.

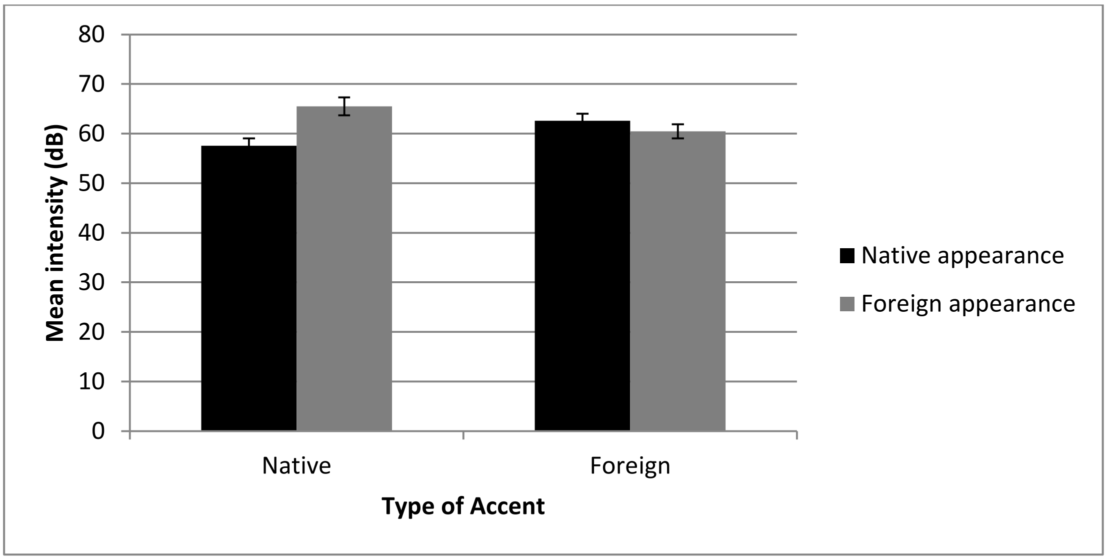


Figure 4.4: Mean intensity (in dB) for target words as a function of accent (native sounding; foreign sounding) in interaction with Appearance (native looking; foreign looking). Error bars show + /-1 standard error from the mean.

These findings have not been observed in previous literature and seem to suggest that interlocutors' foreign appearance might contribute to increased intensity in speech when their accent is native than foreign. This can be considered to not support the hypothesis according to which sound intensity was predicted to be exaggerated in speech to foreign sounding interlocutors irrespective of appearance**.**

4.4.1.4 Length of target vowels and target words

There was no significant main effect of appearance for target vowel length. However, there was a significant main effect of accent (Figure 4.5). Specifically, vowel length was significantly larger for the foreign sounding condition than the native sounding condition (F (1, 41) = 17.331; $p < .05$; $\eta^2_p = .297$). Considering foreigners as language learners, this result can be viewed similar to the finding of longer vowel length in speech to other language learners such as children (Garnica, 1977; Liu et al., 2009) and infants (Fernald et al., 1989; Kuhl et al., 1997) and computers (Burnham et al., 2010). This finding therefore supports the hypothesis that vowel length will be longer in speech to foreign-sounding interlocutors than native-sounding interlocutors.

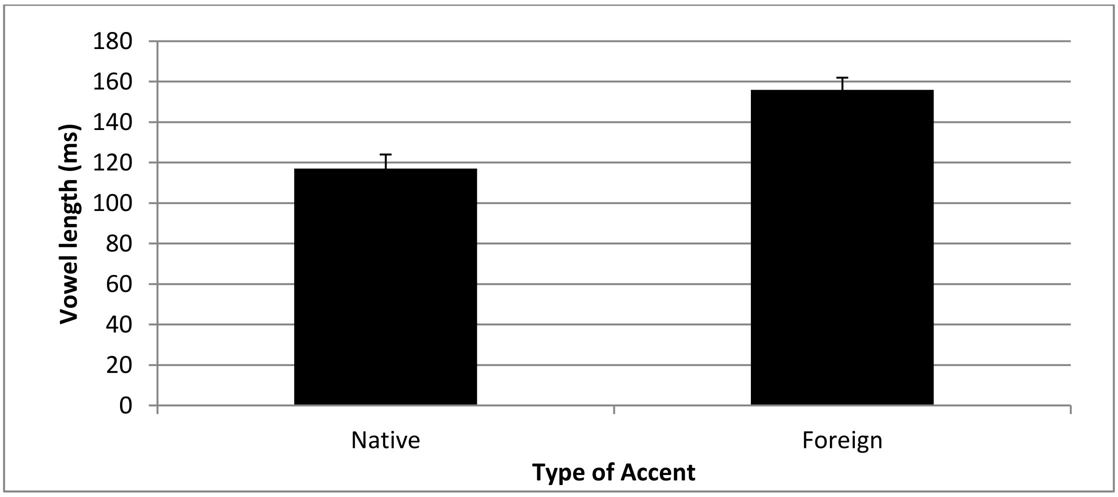


Figure 4.5: Mean vowel length (in ms) of target vowels as a function of accent (native sounding; foreign sounding). Error bars show + /-1 standard error from the mean.

There was no significant main effect of appearance for target word length. However, there was a significant main effect of accent (Figure 4.6 below). Specifically, word length was significantly larger for the foreign sounding condition than the native sounding condition (F (1, 41) = 21.204; $p < .05$; $\eta^2 p = .341$).

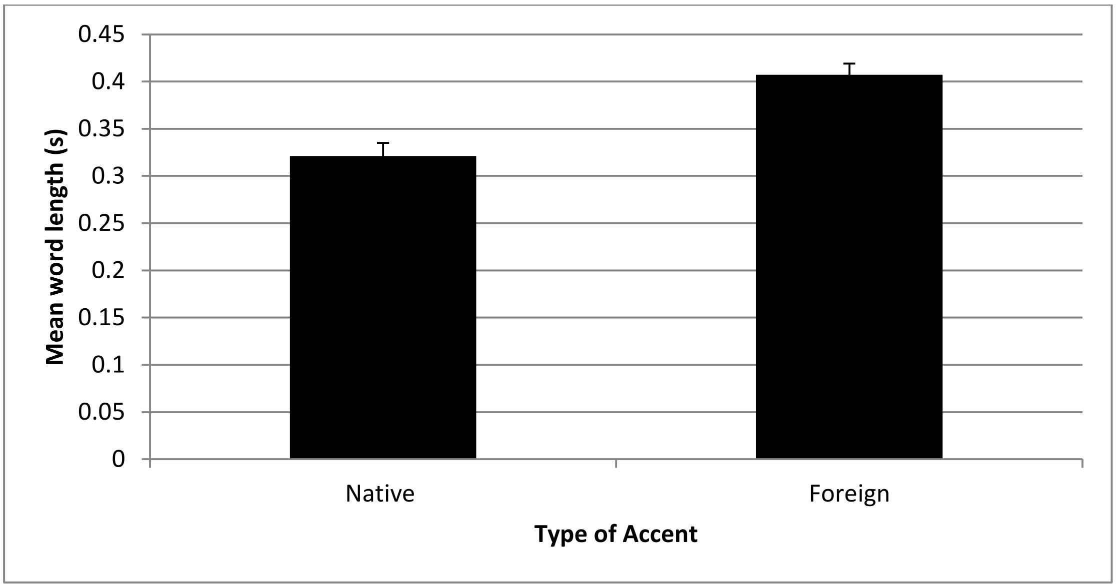


Figure 4.6: Mean word length (in s) of the target words as a function of accent (native sounding; foreign sounding). Error bars show + /-1 standard error from the mean.

This finding is similar to previous studies (Bond & Moore, 1994), which showed words were elongated in clear speech. This finding supports the hypothesis that word length would be longer in speech to foreign-sounding interlocutors than native-sounding interlocutors.

4.4.2 Affective measures of speech and speech rate

Tables 4.2 and 4.3 provide an overview of the results of speech rate and affect from the MANOVA. While speech rate and affect are presented in the first column, the next two columns include the main effects of accent and appearance. The final column presents the interaction between accent and appearance.

Table 4.2 Summary table of results from MANOVA for speech rate. Results are indicated as significant (p< .05) or as not significant (ns).

Speech rate	accent	appearance	accent **x** appearance
Number of words	**ns**	**ns**	**ns**
Number of syllables	**ns**	**ns**	**ns**
Number of pauses	**ns**	**ns**	**ns**
Length of pauses	**ns**	**ns**	**ns**

Table 4.3 Summary table of results from MANOVA for vocal affect. Results are indicated as significant (p< .05) or as not significant (ns).

Vocal affect	accent	appearance	accent **x** appearance
Positive affect	**ns**	**ns**	**ns**
Negative affect	**ns**	**ns**	**ns**
Encouragement of attention	**ns**	**ns**	**ns**

4.4.2.1 Speech rate

A MANOVA showed that speech rate including number of words, number of syllables, number and length of pauses surprisingly did not significantly differ across appearance and accent. This non-significant finding is different from previous research in which clear speech produced upon instruction to speak as if speaking to non-native or hearing-impaired listeners was accompanied by reduced speech rate

(Bradlow et al., 2003). This finding therefore does not support the hypothesis that speech rate to foreign sounding interlocutors would be slow compared to that to native sounding interlocutors.

4.4.2.2 Speech ratings

Mean ratings of positive affect and negative affect across appearance for foreign and native sounding conditions are shown in Figure 4.7, with encouragement of attention in Figure 4.8. A MANOVA showed that there was no significant main effect of accent or appearance for positive affect, negative affect and encouragement of attention. Although there was a significant interaction between accent and appearance for positive affect (F (1, 41) = 6.972; p < .05; η^2p = .145) and encouragement of attention (F (1, 41) = 11.675; p < .05; η^2p = .222), pairwise comparisons did not reveal any differences in positive affect or encouragement of attention across recipient conditions. This finding is different from prior research that showed more negative affect in FDS than ADS (Uther et al., 2007) and therefore does not support the hypothesis that speech to foreign-sounding interlocutors will express more negative emotion than to speech to native sounding interlocutors.

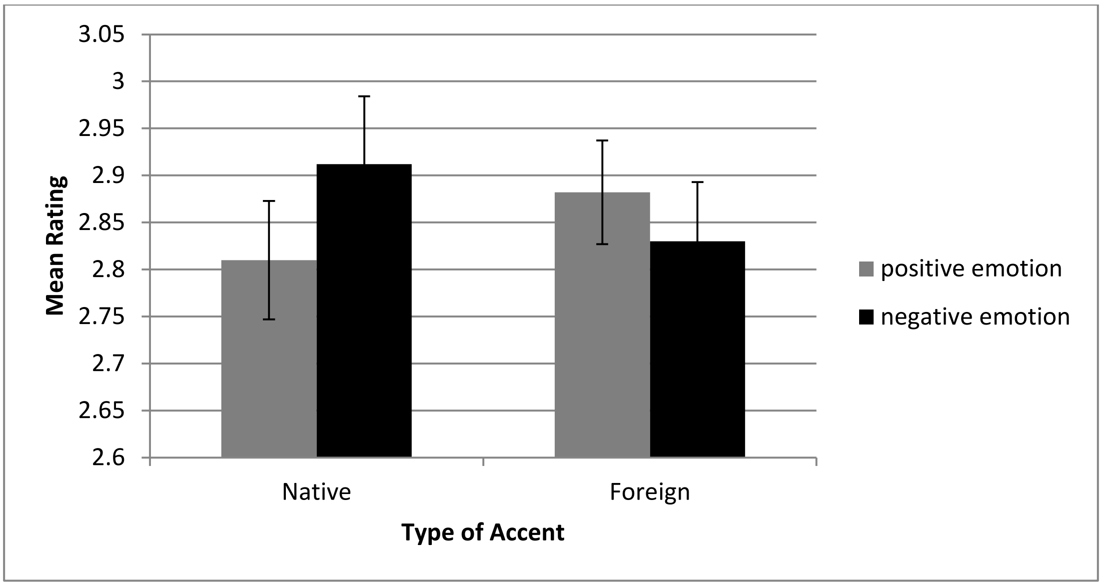


Figure 4.7: Mean ratings of positive affect and negative affect across appearance for speech type (native sounding; foreign sounding) on a scale from 1 (= very positive; very negative) to 4 (=not positive at all; not negative at all). Error bars show + /-1 standard error from the mean.

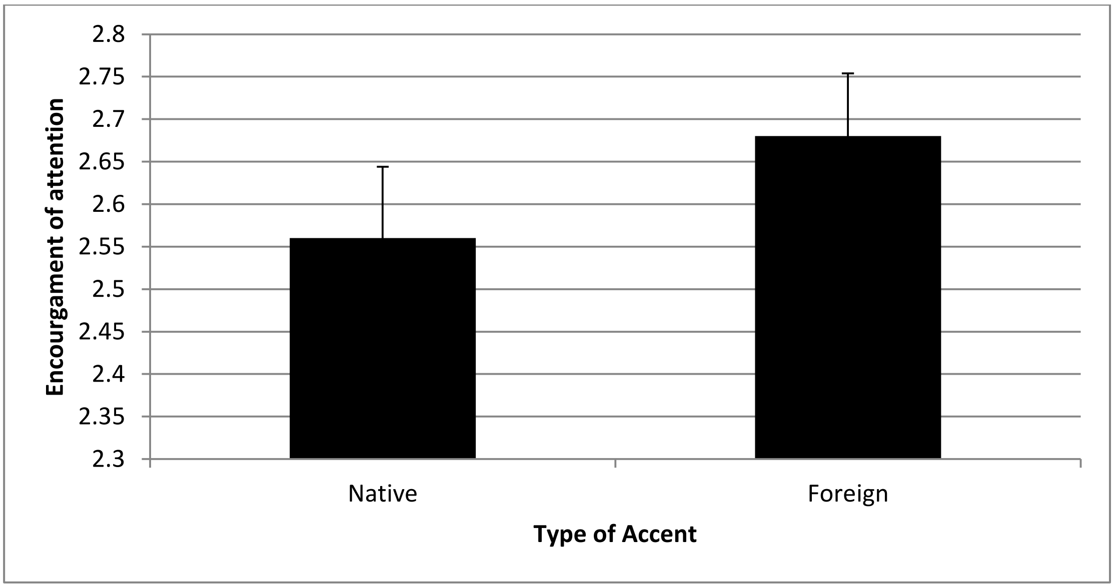


Figure 4.8: Mean ratings of encouragement of attention across appearance for speech type (native sounding; foreign sounding) on a scale from 1(= high encouragement of attention; very clear; very natural) to 4 (=low encouragement of attention; not clear at all; not natural) at all). Error bars show standard error. Error bars show + /-1 standard error from the mean.

4.4.3 Measures of (sub) conscious prejudice

To find out if changes observed in the acoustic aspects of speech were due to subconscious or conscious prejudice, both measures of explicit attitudes (RA, MRS) and implicit attitudes (IAT) were recorded. Table 4.4 provides an overview of the results from Pearson’s correlation for the implicit and explicit attitude measures. These measures are presented in the first column, and in the first row alongside vowel space.

Table 4.4 Summary table of results from Pearson's correlation for implicit and explicit attitude measures. Results are indicated as significant (p< .05) or not significant (ns). The direction of the significant correlation is indicated by 'positive' or 'negative'.

Implicit & Explicit attitude measures	vowel space	IAT	RA(own group1)	RA(White European)	RA(own group2)	RA(South east Asian)	MRS (White European)	MRS (Southeast Asian)
IAT	ns	/	ns	ns	ns	ns	ns	ns
RA(owngroup1)	ns	ns	/	Negative ($p < .05$)	Positive ($p < .05$)	Negative ($p < .05$)	ns	ns
RA(White European)	ns	ns	Negative ($p < .05$)	/	Negative ($p < .05$)	Positive ($p < .05$)	ns	ns
RA(owngroup2)	ns	ns	Positive ($p < .05$)	Negative ($p < .05$)	/	Negative ($p < .05$)	ns	ns
RA(Southeast Asian)	ns	ns	Negative ($p < .05$)	Positive ($p < .05$)	Negative ($p < .05$)	/	ns	ns
MRS(White European)	ns	ns	ns	ns	ns	ns	/	Positive ($p < .05$)
MRS(Southeast Asian)	ns	ns	ns	ns	ns	ns	Positive ($p < .05$)	/

4.4.3.1 Correlation of vowel space with the measurement of implicit attitude 'Implicit Association Test' (IAT)

A Pearson's correlation showed that vowel space did not correlate with IAT.

4.4.3.2 Correlation of vowel space with the measurements of explicit attitude 'Resource Allocation' (RA) task and 'Modern Racism Scale' (MRS)

A Pearson's correlation showed that vowel space correlated with neither RA nor MRS. When speakers in the RA task were required to allocate money to their own group or to an outgroup (White European group) and make the same decision with a second outgroup (Southeast Asian group), the speakers' hypothetical contributions was found to correlate positively with their own contributions (r = .856, p (two-tailed) < .05) and to negatively correlate with their contribution to the first outgroup (White European group) (r = -1, p (two-tailed) < .05) and to the second outgroup (Southeast Asian group) (r = -.856, p (two-tailed) < .05). Speakers' performances on the MRS scales on both White European and Southeast Asian groups were found to correlate positively (r = .748, p (two-tailed) < .05), indicating that native speakers' explicit attitudes towards Southeast Asian groups is comparable to that towards White European groups. The findings of implicit and explicit measures therefore do no supports the hypothesis that the acoustic change of vowel space expansion in native speakers' speech to foreigners might be motivated by native speakers' implicit and explicit attitudes towards foreigners.

4.5 Discussion

This experiment aimed to determine whether foreign appearance or foreign accent drives hyperarticulation in FDS. The goal of this experiment was to find out under what circumstances modifications in vowel space, pitch, intensity, vocal affect, speech rate, and word-and vowel length occur, if at all.

It was found that vowel space was expanded in the presence of a foreign-accented interlocutor. The finding of larger vowel space in speech to foreign sounding interlocutors extends prior studies that observed larger vowel formant space in speech to foreigners who both looked and sounded foreign (Scarborough et al., 2007; Uther et al., 2007).

The evidence of speech to foreign sounding listeners that includes expanded vowel space as a feature of clear speech supports previous research that showed vowel space expansion in clear speech (Johnson et al., 1993). This finding can be considered to uphold the H&H theory (Lindblom, 1990), according to which speech production is a listener-oriented modification because speakers adapt their speech using information of their environmental listening circumstances and of their listener’s background knowledge to provide the listener with sufficient information to understand speech. Therefore considering speech as a result of listener-dependent modification, it can be proposed that native speakers in speech to foreign-sounding interlocutors are more intelligible than those in speech to native-accented interlocutors (Bond & Moore, 1994; Bradlow et al., 1996; Byrd, 1994; Hazan & Markham, 2004).

This result may, to an extent, also be explained using communication accommodation theory (CAT) since the vowel space expansion in native speakers’ speech to foreign-sounding interlocutors can be viewed as native speakers’ attempt to linguistically converge with the foreign-accented interlocutor (Giles, 2001). The result also seems to be theoretically consistent with the notion of audience design according to which speakers are able to design their speech style for a diverse range of listeners (Bell, 2001). However, because the accommodation theory mainly focused on paralinguistic aspects of speech such as utterance length, pitch or speech rate, and not specifically on segmental features such as formant frequencies (Bell, 2006), the H&H theory can be considered to better account for this specific variation in speech than accommodation theory.

Similarly, because audience design only emphasised the practical character of language style in presentation and identity creation (Bell, 2006), the H&H theory can better explain the observation of increased vowel space expansion in speech to foreign-accented interlocutors than audience design. Specifically, because the interaction task was a problem-solving task that encouraged native speakers to be especially instructive to ensure the efficient solution of the task, which was the shared goal between native speakers and interlocutors, the task did not create a situation in which speakers and interlocutors were required to emphasise or attenuate their social identity. Similarly, the task did not create a situation that invoked in interlocutors the need for social inclusiveness. Thus, the resulting outcome on vowel space is at the linguistic level theoretically best explained by the notion of H&H.

Nevertheless, it can be argued that because of the shared goal, and because of the fact that except from the ethnical dimension, both speakers and non-native interlocutors shared social identities at other levels such as being students at the same university and being in the same age group, it can be considered that social convergence might have occurred in interaction between native and non-native communicators (Zuengler, 1991). The CAT claims that if native and non-native speakers have social characteristics in common, non-native speakers can be considered to become similar to native speakers in their use of language (Giles & Johnson, 1987; Young, 1988).

According to Young, the general extent of social convergence determines variation in speech in interaction between native speakers and non-native interlocutors (1988). Consequently, it was argued that in casual situations, non-native interlocutors would converge linguistically with native speakers when they have social aspects in common such as age, academic background, profession, gender, ethnicity and region of origin (Young, 1988).

Research, for instance, in which Chinese interlocutors were interviewed by a Chinese speaker in one situation, and by a native English speaker in another, showed that high-proficient non-native interlocutors' extent of social convergence with native speakers corresponded positively to their extent of linguistic convergence (Young, 1988). This indicates that a set of characteristics and not solely ethnicity can cause speech variation in interaction between native speakers and non-native interlocutors, and that communicators use this set to evaluate how comparable they are (Young, 1988). This evidence therefore clearly endorses the element of CAT that deals with similarity and attraction (Zuengler, 1991).

However, these observed variations in speech were produced by non-native interlocutors. By contrast, Experiment 1 did not measure variations in non-native interlocutors' speech. It can be speculated that if, for example, non-native interlocutors showed the inclination to use a more deliberate and considered speech style themselves in interaction with native speakers, native speakers' hyperarticulated speech style may reflect non-native interlocutors' style of speech to indicate native speakers' psychological convergence towards them. Thus, although sociolinguistic factors of convergence and divergence might maybe play a role in communication between native speakers and non-native interlocutors, and since the present experiment did not directly test those sociolinguistic factors, the consequent psychological convergence between native speakers and non-native interlocutors as result of non-native interlocutors' presumed linguistic convergence can only be assumed at this stage.

Therefore further experiments need to be carried out to find out how non-native interlocutors' speech actually varies in interaction with native speakers, and how it varies in response to native speakers' choice of speech strategies. Previous research showed how high-proficient Japanese learners of English did vary their speech to low-proficient Japanese learners of English but in a different manner than did native English speakers in speech to low-proficient Japanese learners of English (Tani, 2011). For example, while native speakers' speech to low-proficient interlocutors was characterized by lower speech rate and lower number of statements compared to their speech to high-proficient interlocutors, high-proficient speakers' speech to low-proficient interlocutors consisted of higher speaking rate while including a less complex and shorter sentence structure (Tani, 2011). This indicated that both native speakers and also high-proficient learners of English adapt their speech to the level of English proficiency of non-native interlocutors, with accommodation taking place only in case understanding was hindered, and when non-native speakers are themselves able to accommodate (Tani, 2011). Additional research can also explore to what extent non-native interlocutors' linguistic convergence occurs as result of social convergence, and to what extent psychological convergence occurs as result of linguistic convergence. Thus, it can be said that compared to sociolinguistic and psycholinguistic theories, the H&H theory provides the simplest explanation of the data regarding formant frequencies.

The findings also revealed that mean word length and vowel length were significantly larger in speech to foreign sounding interlocutors than in speech to native-accented interlocutors. The finding on word length is consistent with prior research that showed larger word length in clear speech (Bond & Moore, 1994). The finding of long vowel length extends previous studies that showed that vowel length is elongated in speech to foreigners (Ashby, 2004; Sankowska et al., 2011; Scarborough et al., 2007). This finding on vowel length can be likely explained by the fact that the vowels /a:/, /i:/ and /u:/ are tense vowels and they are compared to lax vowels longer in duration in the standard accent of Standard English (Received Pronunciation) in the UK. This finding is also in line with previous research that showed that long/tense vowels are lengthened in clear speech, and are longer in duration than short/lax vowels (Uchanski, 1988, 1992). The data thus suggest that in addition to exaggeration in vowel space, exaggeration in word and vowel length is elicited in speech to foreign sounding listeners compared to speech to native-accented interlocutors.

Moreover, the finding that mean intensity was higher in speech to foreign looking interlocutors with native accent than foreign accent does not support the hypothesis that the acoustic aspect of intensity will be exaggerated in speech to foreign-sounding interlocutors irrespective of their appearance. Moreover, it is not in line with prior evidence, according to which FDS is considered to be louder than ADS when foreigners are foreign in appearance and accent (Lipski, 2005). Nonetheless, consistent with previous research, this result appears to indicate that intensity is not a characteristic of speech that is necessary to be modified for non-native speakers to acoustically perceive speech better (Sommers et al., 1994). Similar to intensity, the finding on speech rate was not consistent with what was hypothesised before because speech rate did not differ across interlocutors who varied in their extent of looking and sounding foreign. Contrary to prior research (Bradlow et al., 2003; Smiljanic & Bradlow, 2005), this result therefore implies that clear speech, as produced in interaction with foreign looking and foreign sounding listeners, does not have to involve exaggerated speech rate.

This result on speech rate shows that long vowel length can occur in the absence of slower speech rate. This seems to indicate that a longer duration at phoneme level does not necessarily have to produce a large effect on speech rate unlike acoustic features at suprasegmental level that are associated with speech rate such as the occurrence of pauses and duration of pauses. Moreover, it has been shown that the extent to which vowel length alters with changes in speech rate is not stable, and also be affected by other speech aspects such as the syntactic formation and length of an utterance, thereby indicating the complexity of the effect of speech rate on vowel length (Miller, 1981).

It has previously been argued that clear and conversational speech styles do not have to differ in speech rate as both speaking styles can be produced at equal speech rate (Zwicky, 1972). Moreover, it has also been shown that speech rate does not have to be modified in order to generate clear speech (Krause & Braida, 2002). Accordingly, it was found that at normal speech rate, clear speech was more intelligible than conversational speech by fourteen percentage points and that at slow speech rate, clear speech was more intelligible than conversational speech by twelve percentage points (Krause & Braida, 2002). Nonetheless, one has to note that the speakers involved in producing clear speech in that study were experienced in public speaking and received additional training in clear speech production (Krause & Braida, 2002). Thus, it might be that clear speech might not be possible to be generated at normal speech rate by speakers with no experience in public speaking. Thus, because of no significant difference in speech rate between speech to interlocutors who are native sounding and foreign sounding and because of expanded vowel space in speech to foreign sounding interlocutors in the present experiment, the present result seems to support the view that clear speech can be produced effectively at the same speech rate as conversational speech.

Moreover, consistent with what was previously hypothesised, no changes in the pitch of vowels were revealed across accent and appearance conditions. This is as expected as previous research showed that native listeners did not modify their pitch in speech to neither foreign interlocutors nor native interlocutors (Uther et al., 2007). This result is also comparable to the finding that imaginary interactions between native speakers and foreigners did not differ in pitch (Knoll et al., 2009).

In addition, regarding affect, native speakers did not differ in their expression of positive and negative affect in their speech across interlocutors. This is inconsistent with prior findings according to which speech to foreign interlocutors expressed more negative emotion than speech to native interlocutors, and speech to native interlocutors conveyed higher positive affect than speech to foreigners (Uther et al., 2007). Nonetheless, the present results are consistent with the observation of no differences in pitch between speech to foreign and native interlocutors. The results might be explained by the fact that speakers for this experiment were recruited from a multicultural student population at Brunel University and therefore were probably used to daily contact with foreigners. These speakers were different from the speakers in Uther et al.'s study (2007) who were from the Hampshire region, which is likely to contain a predominance of White British residents (Hantsweb, 2001). Thus, a possible factor that can be considered to affect whether native speakers' emotion in speech to a type of foreign interlocutors can be linked to their appearance, depends on whether native speakers reside in an environment in which that type of foreigners are prevalent.

Finally, implicit and explicit measures of prejudicial biases did not significantly correlate with vowel space expansion. Thus, there seems to be no link between speakers' implicit and explicit attitude towards foreigners and speakers' way of modifying vowel formants towards foreign sounding interlocutors. This finding therefore is not consistent with what was previously hypothesised according to which changes in vowel space in native speakers' speech to foreigners might be motivated by native speakers' implicit and explicit attitudes towards foreigners. Based on prior research (McConnell & Leibold, 2001) it can be assumed that the observation of negative implicit attitudes in native speakers towards foreigners would have been manifested in speech in form of verbal mistakes and hesitations. Prior research, for example, showed that white speakers, who demonstrated more negative implicit attitudes to people of color, were perceived to show more negativity in communication with a black investigator than a white investigator (McConnell & Leibold, 2001). They also revealed comparatively more negative explicit attitudes towards people of colour (McConnell & Leibold, 2001). Thus, this implies that if native speakers had exhibited negative implicit attitudes towards foreigners in experiment 1, these attitudes would have been caused by foreigners' foreign appearance.

Nonetheless, recently the use of the IAT to measure implicit prejudice (Greenwald et al., 1998) has been questioned because recent IAT studies have suggested that the IAT might actually measure the strength of one's membership with the in-and out-group and not one's racial bias (Blair et al., 2010; Popa-Roch & Delmas, 2010). In addition, although it was demonstrated that individuals with a strong IAT effect tend to behave in an explicitly prejudiced manner, the size of the correlation was small (Greenwald et al., 2009; Nosek, Greenwald & Banaji, 2007). Thus, it seems to be that the IAT appears not to measure implicit attitudes per se and therefore the finding of an absent correlation of vowel space with IAT will not be interpreted further.

In conclusion, Chapter 4 has addressed the research question whether appearance or accent motivates acoustic hyperarticulation in speech: this chapter has shown that based on theory (H&H theory by Lindblom (1990)) and previous evidence (Burnham et al., 2003; Sankowska et al., 2011; Uther et al., 2007), native speakers acoustically exaggerate speech to foreign-accented listeners who require linguistic clarifications by including vowel hyperarticulation, increased word length and vowel length compared to speech to native-accented listeners (Lindblom, 1990; Scarborough et al., 2007). These modifications do not seem to be motivated by native speakers' implicit and explicit attitudes towards foreigners if they come from an environment that is characterised by regular contact with foreigners.

Chapter 5

Comparison of vowel space in infant-directed speech and read speech

5.1 Acoustic modifications of read speech and infant-directed speech (IDS)

IDS contains phonetically exaggerated properties, such as exaggerated vowel space and increased vowel length, which have been correlated with speech intelligibility in IDS (Liu et al., 2003). Termed a 'hyperarticulated' speech style, exaggerated vowel space has been reported to occur to a larger extent in speech to younger infants than older ones (Liu, Tsao, & Kuhl, 2009). Recently, it has been suggested that infants' word recognition might be enhanced by vowel space exaggeration (Song et al., 2010) (see 3.2.1 in chapter 3 for more information). The present study aimed to more directly test whether the acoustical modifications used in IDS are uniquely didactic and distinguishable from other forms of clear (but not necessarily hyperarticulated) speech such as read speech, which has been used in previous research to produce clear speech as well (Harnsberger, Wright, & Pisoni, 2008; Kain et al., 2008; Picheny et al., 1985).

Reading tasks have been used in experimental studies for various purposes such as to produce a hyperarticulatory speech style (Hazan & Baker, 2010; Kain et al., 2008). The resulting outcome of cautiously regulated read speech produced under ideal conditions has been shown to include the acoustic features of clear speech such as higher mean pitch, larger pitch range, decreased speech rate (Bradlow, Kraus, & Hayes, 2003; Hazan & Baker, 2010), articulatory stretched vowel space (Picheny et al., 1985) and elongated and recurring pauses (Picheny et al., 1985; Liu & Zeng, 2006). Similarly, in the presence of noise, it was found that read speech revealed higher vowel formant frequency for *F*1, decreased speech rate, and also higher rms energy to enhance speech clarity and counteract the masking by noise (Lu & Cooke, 2008).

Although previous research has compared read speech with naturally elicited speech, it has focused on the comparison of acoustic-phonetic features of casual and clear speech produced in two task types: a spot-the-difference task and reading task (Hazan & Baker, 2010; Hazan & Baker, 2011). Hazan and Baker (2010) used a 'Spot the difference' task (termed a Diapix task), to elicit speech in optimal and adverse listening conditions. Within that study, each speaker was asked to read speech in a clear and a conversational manner. It was found that pitch median, pitch range and reduced speech rate were all larger for read speech than in the natural speech under clear and conversational speech styles (Hazan & Baker, 2010). These findings seem to suggest that read speech expressed in a clear speaking style represents an overemphasised type of clear speech compared to natural speech that is articulated in a clear manner (Hazan & Baker, 2010). Although these observations appear to be similar to those found previously in IDS, that study only examined ADS and not IDS. Nonetheless, this study shows that other forms of speech such as read speech can be, similar to speech to infants, articulated in an exaggerated style of speech as well.

Similarly, a recent study (Hazan & Baker, 2011) focussed on comparing talkers' acoustic-phonetic adjustments made communicatively with interlocutors in a challenging and ideal listening condition with those made in read speech produced in a clear and conversational speech style respectively. The study involved 40 speakers completing a Diapix task in pairs. Adverse listening conditions were created using either a noise vocoder or simultaneous babble noise (Hazan & Baker, 2011). The read speech in the conversational style was produced upon the instruction to read as if speaking to a friend while the read speech in the clear speech style was produced following the instruction to read as if speaking to a hearing-impaired listener (Hazan & Baker, 2011). Comparable to the previous study (Hazan & Baker, 2010), it was found that median pitch in read speech was larger than in the communicative Diapix task conditions. Median pitch was larger in the clear speaking style than in the conversational speaking style (Hazan & Baker, 2011). Moreover, the rise in pitch range was larger between the casual and clear speaking style in read speech than in the communicative conditions and it was more noticeable in read speech than in the communicative speech. This result suggests that speakers produce larger pitch when reading than when communicating with others (Hazan & Baker, 2011).

In addition, it was found that mean word duration was longer in clear read speech than when making speech clearer to another speaker (Hazan & Baker, 2011). Also, $F1$ range was more different between clear read speech and conversational read speech than for speech with communicative intent (Hazan & Baker, 2011). Similarly, the difference in $F2$ range was larger between clear read speech and conversational read speech than between the communicative challenging and ideal conditions. These results therefore suggest that vowel range was more affected by speaking style in read speech than in speech with communicative intent (Hazan & Baker, 2011). Thus, it seems that instructing speakers to read more clearly, produces consistently clear read speech. The authors argued in line with Lindblom's H&H theory (1990) that compared to clear read speech communication between two speakers might not be reliably clear because the communicative efficiency can vary with the presence and absence of misunderstandings (Hazan & Baker, 2011).

Similar to the previous study (Hazan & Baker, 2010), that study by Hazan and Baker (2011), however, focussed on ADS and not IDS. Moreover, contrary to prior research, it included formant frequency range instead of formant midpoints (e.g. Picheny et al., 1986; Moon & Lindblom, 1994; Uther et al., 2007; Wassink et al., 2006). Therefore the present experiment will address the research question of how naturally produced IDS differs acoustically from clear read speech as another type of clear speech. Specifically, it will investigate whether the acoustic characteristic of vowel space in IDS is similar to that produced in read speech. It will measure the midpoint of the first and second formant, and compare conditions within speakers. It thus compares vowel space of clear read speech and of spontaneous clear speech that is produced in naturalistic interactions with a communicative purpose for the same set of speakers. Based on theory (Kuhl's NLM theory (Kuhl et al., 2008) (see 2.6 in chapter 2 for more information); H&H theory by Lindblom (1990)) and prior research, it is predicted that compared to read speech IDS will have acoustic exaggerations in vowel space, vowel length, word length, pitch and pitch range (Burnham et al., 2002; Fernald & Kuhl, 1987; Kuhl et al., 1997; Uther et al., 2007). IDS will include higher pitch and a broader pitch range than read speech, which has been shown to have higher pitch range than interaction between adults (Hazan & Baker, 2011). Differences in mean intensity between IDS and read speech are not expected (Fernald & Kuhl, 1987; Hazan & Baker, 2011; Lu & Cooke, 2008).

5.2 Methods

5.2.1 Participants

Fourteen mothers (20-45 years) were recruited through Brunel University. All mothers and their infants (12-20 months) were White British and from the West London area. All mothers were native speakers with a Southern British English accent. There were no reported language or speech disorders in the participants.

5.2.2 Design

This study used a within-subjects design in which the target vowels /a:/, /i:/ and /e/ were produced in both the mother-infant interaction condition and the reading condition using target cues. The dependent variable was vowel triangle area (indexed by mean vowel space plotted using $F1$ and $F2$ values of each of the three corner vowels, /a:/, /i:/, /e/), in the target words in which one of the three target vowels was present. Other dependent variables were mean $F0$, mean pitch range, mean intensity, mean vowel and word length. Because of the need of a longer sample to detect changes in $F0$ range, the $F0$ range of words and not vowels will be analysed.

5.2.3 Materials

For the purpose of eliciting the target vowels /a:/, /i:/ and /e/, the words 'car', 'green' and 'red' were chosen as specific target words, which contain each of the target vowels respectively. These target vowels were elicited from the mothers in the interaction condition by using 19 picture cards (see Appendix M) that depicted pictures of everyday objects with their names on them (e.g. cat, shoes, door). A collection of 27 sentences (see Appendix N) with syntactically simple organization was designed. It included the target words 'car', 'green' and 'red' for the reading condition. The sentences carried meaningful weight, e.g. '*green* tea can improve your health/ stop signs are *red*’. A digital voice recorder Edirol R-09HR by Roland (sampling rate: 44,100 Hz) was used to record speech production in both conditions. Each interaction was recorded as a mono 16-bit file in wav format.

5.2.4 Procedure

In each half an hour audio-recorded interaction, mothers were instructed to show their infants the 19 picture cards with everyday objects on them such as ‘shoes’ and ‘dress’ and to try to engage their infants in a dialogue. After this part, mothers were asked to read out aloud in a clear voice the short list of sentences and then finally to complete a short online questionnaire on their demographic and linguistic backgrounds (see Appendix S). All interactions were recorded in a quiet room in participants’ homes to ensure a naturalistic communication between mothers and infants.

5.2.5 Data analysis

This study employed a repeated measures ANOVA for the analysis of the acoustic measures (mean vowel triangle area, mean *F*0, mean intensity, mean vowel and word length), with the reading and mother-infant interaction conditions and the three target vowels /a:/, /i:/ and /e/ representing the within-subjects factors (2x3 within-subjects design).

Initially ten target words (“car”, “blue”, “green”, “red”, “shoes”, “door”, “ball”, “box”, “rocks”, “cup”) were recorded from mothers during their interaction with their infants (see Appendix M), which included one of seven target vowels (/a:/, /u:/,/i:/, /e/, /ɒ/, /ɔ:/, /ʌ/). The target words “car”, “green” ,“red”, “blue”, “shoes” and “shop” with the vowels /a:/, /i:/, /e/, /u:/ and /ɒ/ respectively were chosen due to sufficient tokens of each of them produced by each participant in each condition.

5.3 Results

A 2x3 repeated-measures Analysis of Variance (ANOVA) was used to analyse vowel space, pitch, pitch range, intensity, vowel length and word length.

5.3.1 Acoustic measures

5.3.1.1 Formant analysis

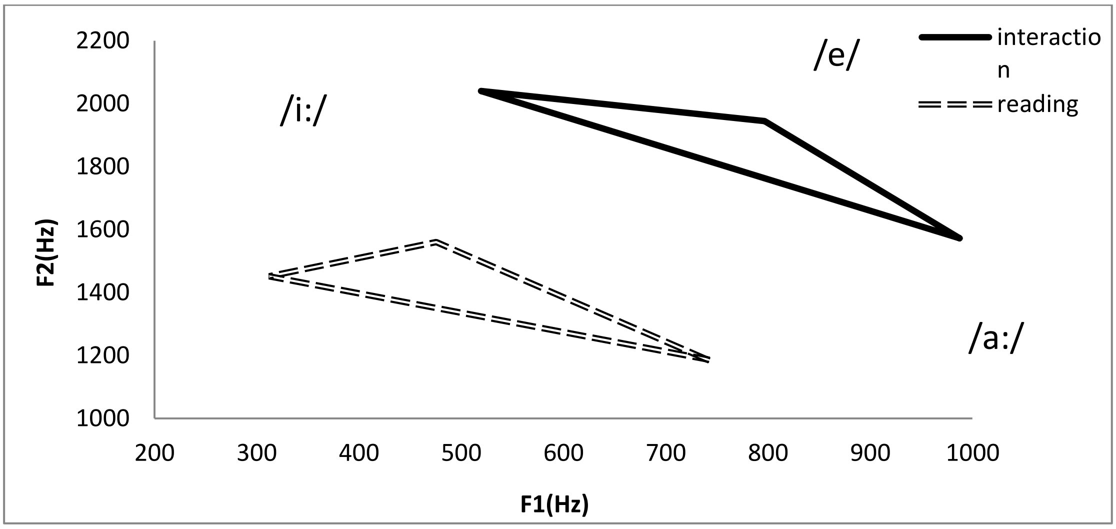


Figure 5.1: Exaggeration of target vowels in the speech type condition of mother-infant interaction and absence of hyperarticulated vowels in the reading condition.

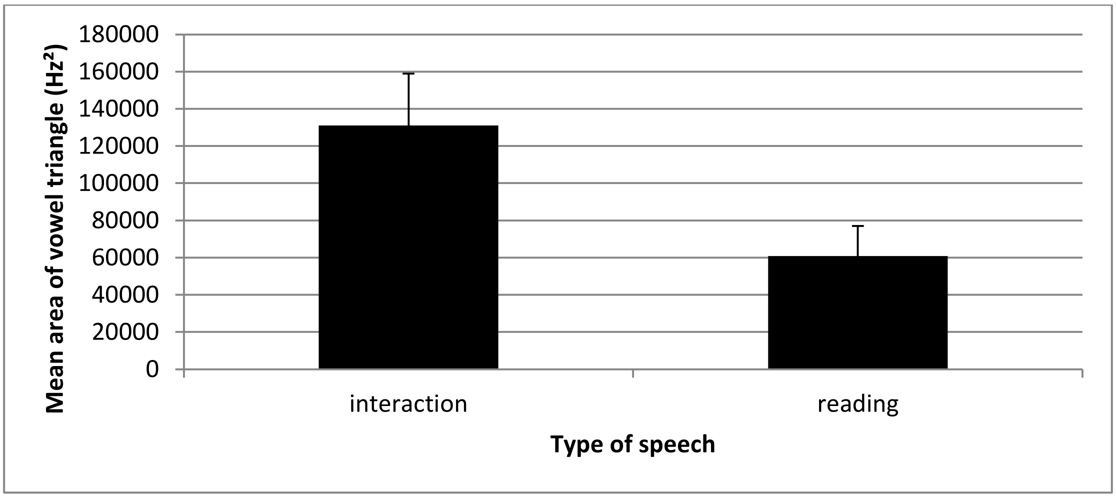


Figure 5.2: Mean area calculated from the vowel triangle in F2/F1 space in the mother-infant interaction condition and the reading condition. Error bars show + /-1 standard error from the mean

Significantly larger vowel space was revealed for the infant-directed speech condition than the read speech condition (F (1, 10) = 85.612.630; p < .05; η^2_p = .895) (Figures 5.1 and 5.2). Thus, this result shows that mothers' articulatory vowel space in speech to infants is significantly more expanded than their vowel space expressed when reading sentences. This finding thus confirms the hypothesis of expanded vowel space in IDS compared to read speech. Supporting the didactic role of IDS, this finding is consistent with prior research that showed that IDS includes expanded vowel space in mother-infant interactions across different languages (Kuhl et al., 1997). This result also seems to indicate that the didactic needs in natural IDS might be larger than in a reading context.

5.3.1.2 Fundamental frequency (F0) – mean and range

5.3.1.2.1 Mean F0 of vowels

The results showed that there was a significant main effect of condition: unsurprisingly, the mean $F0$ was significantly larger for the infant-directed condition than the reading condition (F (1, 10) = 125.630; p < .05; η^2_p = .926) (Figure 5.3).

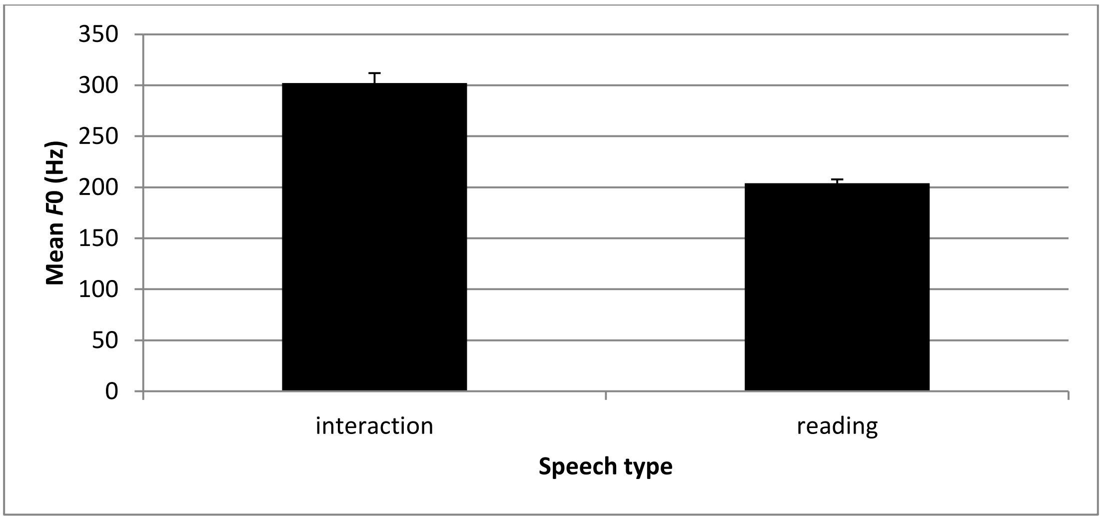


Figure 5.3: Mean pitch (in Hz) as a function of speech type (interaction; reading) across vowels. Error bars show + /-1 standard error from the mean.

5.3.1.2.2 Mean F0 range of words

There was a significant main effect of speech condition with mean *F*0 range being significantly larger in the infant-directed condition than the reading condition (F (1, 10) = 44.971; p < .05; $\eta^2 p$ = .818) (see Figure 5.4). These findings of higher mean pitch of vowels and wider pitch range for words support the hypothesis that pitch range would be greater in IDS than read speech, and they agree with studies that showed broader range in *F*0 and more increased *F*0 in IDS in tonal and nontonal languages (Fernald et al., 1989; Grieser & Kuhl, 1988; Xu, Burnham, & Kitamura, 2007).

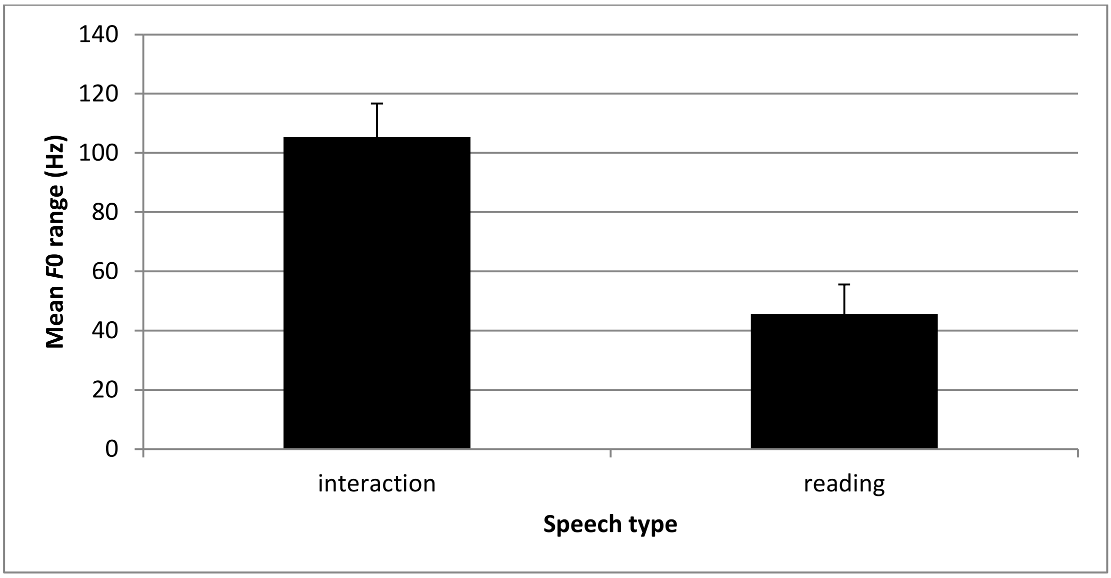


Figure 5.4: Mean pitch range (in Hz) as a function of speech type (interaction; reading) across vowels. Error bars show + /-1 standard error from the mean.

5.3.1.3 Mean intensity of vowels and words

There was no significant main effect of speech condition for mean intensity of vowels or of words. These findings therefore support the hypothesis according to which differences in mean intensity between IDS and read speech would not be expected.

5.3.2 Length of utterances

5.3.2.1 Target vowel length

There was a significant main effect for speech condition (F (1, 10) = 43.919; p < .05; η^2_p = .815). This is shown in Figure 5.5, which reveals larger vowel length for the interaction condition than the reading condition across all vowels. There was no main effect of vowels for target vowel length (F (2, 20) = .435; p > .05; η^2_p = .042).

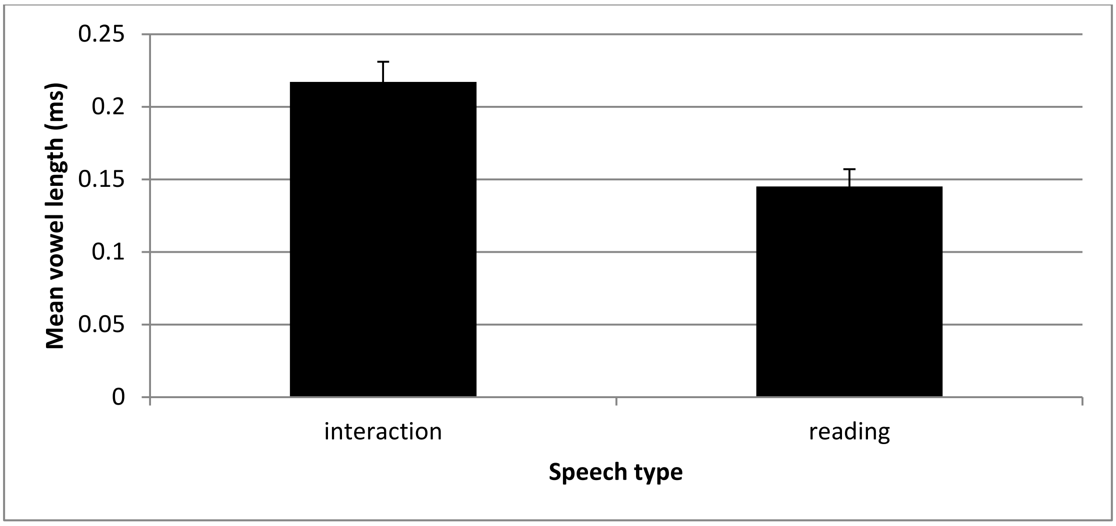


Figure 5.5: Mean vowel length (in ms) of target vowels as a function of speech type (interaction; reading). Error bars show + /-1 standard error from the mean.

5.3.2.2 Target word length

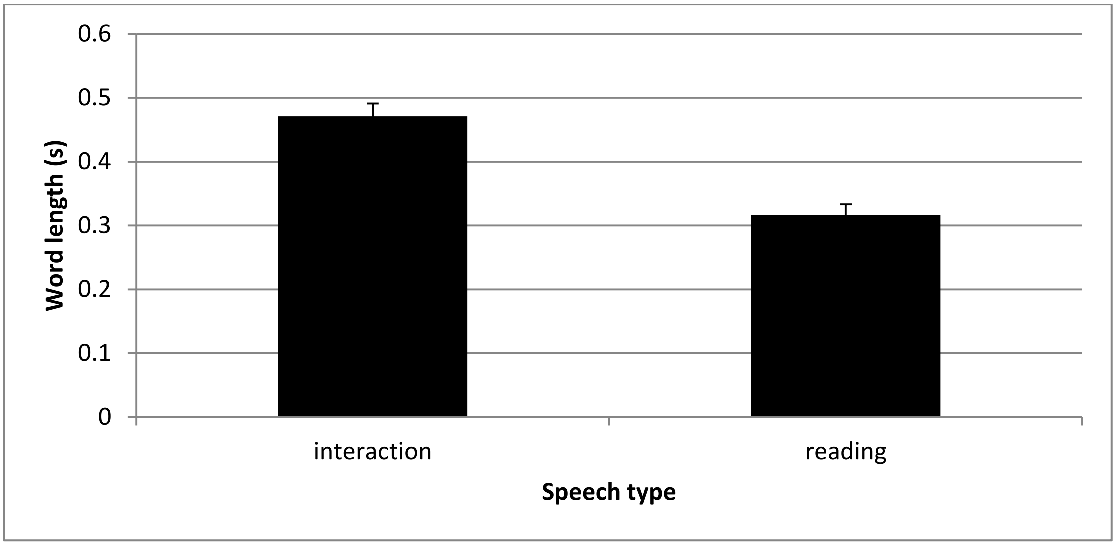


Figure 5.6: Mean word length (in s) of target words ('car', 'green', 'red') as a function of speech type (interaction; reading). Error bars show + /-1 standard error from the mean.

There was a significant main effect for speech condition (F (1, 10) = 75.389; p< .05; η^2_p = .883). This is shown in Figure 5.6, which reveals larger word length for the interaction condition than the reading condition across all target words. There was no main effect of target words for target word length (F (2, 20) = 0.035; p > .05; η^2_p = .004).

5.4 Discussion

This experiment aimed to answer the research question of how infant-directed speech (IDS) differs acoustically from read speech. It therefore compared the acoustic-phonetic features of clear read speech with those of speech to infants. The aim of this experiment was to determine if IDS and clear read speech differed acoustic-phonetically with regards to vowel space, mean pitch, mean intensity, mean word-and vowel length.

First, this experiment showed that vowel space in speech to infants was significantly more expanded than vowel space in read speech. Although this study focused on the elicitation of clear speech only, this result is to some degree similar to a previous study in which women were observed to show the inclination to produce more extended vowel range in casual speech than men (Hazan & Baker, 2010).

This result supports previous studies that showed that in IDS acoustic cues of phonetic units are exaggerated whereby differences between phonetic categories are overstressed (Bernstein–Ratner, 1984; Kuhl et al., 1997; Burnham et al., 2002). The fact that vowel space was reduced in read speech compared to IDS suggests that hyperarticulation in IDS may be uniquely didactic (as compared to other forms of clear speech) and might therefore contribute to infants' development of vowel categories (Andruski et al., 1999; Kuhl et al., 1997). As such, it could be suggested that IDS requires more acoustic exaggeration than read speech due to the natural linguistic needs of the listener as a language learner that is the infant (Kuhl et al., 1997). This result thus shows that the acoustic modifications in IDS are distinctively informative (Kuhl et al., 1997; Uther et al., 2007). This result agrees with the H&H notion of hyperarticulation, in which speech is acoustically exaggerated (Lindblom, 1990) (see section 3.1.2 for more information on hypo-and hyper speech).

This result on vowel space indicates that IDS seems clearer than clear read speech, and this might be possibly explained by the age range of infants who took part in the experiment. The age range was between 12-20 months and was based on previous IDS research such as by Song et al (2010), in which infants were on average 19 months old. It is known that by 12 months, infants have become language-specific learners of the language of their ambient environment (Werker & Tees, 1984), have gone through a notably sensitive stage for phonetic learning (Kuhl, Tsao & Liu, 2003), and have developed phonetic abilities to discover phonotactic patterns in speech production that are essential to identify words (Saffran et al., 1996). Infants' improved phonetic capacities may support the acquisition of words earlier in development (Tsao et al., 2003). By then, mothers might have become used to talking to their infants in a manner that helps infants' development of phonetic categories.

An additional reason for the difference in vowel space between IDS and clear read speech can be attributed to the nature of the interaction task, which was instructive in nature: in the interaction task mothers were asked to actively teach their infants the words from the picture cards. It can be argued that because mothers might have tried to aim to facilitate infants' development of vowel categories, the speech style that mothers chose to use in the interaction task may have been very instructive in nature. Moreover, speech to infants was produced in the participants' homes to ensure an interaction that is as naturalistic as possible between mothers and their infants. After talking with their infants, mothers were asked to read out a short list of sentences, which means that contrary to the study by Hazan and Baker (2010), for instance, the reading task in this experiment was conducted in an environment that participants are familiar with rather than in a laboratory setting. The fact that during the reading task, infants were not present in the room in which mothers completed the reading task might also have contributed to the difference in vowel space between IDS and clear read speech.

The observed lack of a tendency towards vowel space expansion in read speech that in size is similar to or larger than that in IDS is contrary to previous studies (Picheny et al., 1985; Hazan & Baker, 2010) and research in which vowel range was more affected by speaking style in read speech than in speech with communicative intent (Hazan & Baker, 2011). Thus, although read speech expressed in a clear speaking style was previously observed to represent an overemphasised type of clear speech compared to natural clear ADS (Hazan & Baker, 2010), IDS seems to be a listener-oriented speech style as it is targeted at a first language learner that is the infant.

This result is in line with Kuhl's Native Language Magnet (NLM) Theory (enhanced), which proposes that the overemphasis of relevant phonetic differences in IDS compared to in ADS is a contributor of change that in infants' perception leads forward the transition from a universal phonetic speech perception to a speech perception that is particular to a native language (Kuhl et al., 2008). Thus, expanded vowel space in IDS serves to overemphasise important phonetic contrasts through which the speech perceptual abilities of infants in their native language might be supported (Kuhl 2000b; 2004; Kuhl et al., 2008; Liu et al., 2003). In addition to previous studies in which expanded vowel space was observed in Mandarin (Liu et al., 2003), American English, Russian and Swedish (Kuhl et al., 1997), Australian English (Burnham et al., 2002), British Southern English (Uther et al., 2007), the present experiment can be considered to demonstrate larger vowel space in British Southeast English (West London region), suggesting that IDS is a uniquely didactically-oriented speech style (Kuhl et al., 1997; Uther et al., 2007). Thus, this finding has shown additional evidence for the linguistic contribution of expanded vowel space in IDS and therefore supports the previous hypothesis of exaggeration in vowel space in IDS compared to read speech.

This result therefore seems to suggest that the occurrence of vowel hyperarticulation in IDS seems to be facilitated by the use of a didactic task that involves teaching words with corner vowels in British English to infants after the age of 12 months. This view appears to be supported by recent investigations on IDS that reported improved consonant contrasts in speech to infants aged beyond 13 months of age (Cristia, 2010), that did not report vowel hyperarticulation in speech to infants that were younger than 12 months old (Cristia & Seidl, 2013), that showed no improved contrast for central vowels in IDS (McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013), and that did not observe vowel hyperarticulation in other languages such as Norwegian (Englund & Behne, 2006). This might indicate that vowel hyperarticulation in IDS might be useful for older infants who are active word learners (Bernstein-Ranter, 1984; Cristia, 2010), and that vowel hyperarticulation in IDS may be language-specific and applies to point vowels but not central vowels (Cristia & Seidl, 2013).

On the other hand, one could argue that if the purpose of IDS is to convey high positive affect, changes to the first two formant frequencies in IDS such as a rise in formant frequencies might be a side effect of positive affect in IDS, such as smiling (Benders, 2013). Smiling has previously been argued to acoustically result in an increase of formant frequencies (Ohala, 1980, 1984). A recent study reported the lack of vowel hyperarticulation in IDS, and positive affect being conveyed through a higher second formant in speech to 11 months old infants than to 15 months old infants (Benders, 2013). This implies that if positive affect is emphasised in speech to younger infants, IDS might not be necessarily motivated by didactic intent.

Thus, one would consider such non-verbal expression of high positive affect not to occur after around the age of 12 months since the effect of smiling on the second formant of corner vowels such as /i/ can be detrimental for regular articulation (Benders, 2013). Therefore, it seems that if the expression of increased positive emotion is the purpose of IDS, such as in speech to infants aged less than 12 months old, an increase in the formant frequencies such as F2 and F3 has been suggested to occur at the cost of enhanced contrast between vowels. However, not sufficient data are available to support this possibility. Moreover, further research is required that can replicate Bender's experiment (2013) in another language, and that uses the same methodologies as in Bender's experiment because the present experiment 2 is methodologically not identical to Bender's experiment (2013). Further research can explore whether in speech to both younger and older infants, the systematic manipulation of smiling affects formant frequencies.

It has also been argued that vowel hyperarticulation might be a by-product of slow speech rate in IDS (McMurray et al., 2013). It has recently been observed, for instance, that alterations in VOT of plosives are a side-effect of altered speech rate (McMurray et al., 2013). However, since VOT is a timing cue that helps listeners differentiate between aspirated and unaspirated stop consonants, it is probable that temporal measures such as VOT are most affected by speaking rate, while vowel formants represent a phonetic measure of the clarity of vowel articulation that are affected by arrangements within the vocal tract. In contrast to vowel length, however, vowel formants may not be influenced by speech rate.

Second, as previously hypothesised, mean pitch and mean pitch range were exaggerated in IDS than in clear read speech at both vowel and word level: all target vowels and words were articulated with higher pitch and pitch range. The finding of larger pitch and pitch range in IDS has been well documented (Grieser & Kuhl, 1988; Stern et al., 1983), both in tonal and non-tonal languages (Fernald et al., 1989; Kitamura et al., 2001; Xu & Burnham, 2010).

It has been previously observed that in IDS mean pitch and pitch range are increased in order to maintain infants' attention (Kitamura & Burnham, 2002; Stern et al., 1982). Thus, although it was reported that speakers produced larger pitch and higher pitch range when reading than when communicating with others (Hazan & Baker, 2011), the fact that the present experiment dealt with speech to first-language learners in form of infants whose attention needs be drawn and sustained through high pitch explains why pitch and pitch range were larger in IDS than in read speech, and why in previous research that involved adult native speakers (Hazan & Baker, 2010, 2011) higher pitch was reported in read speech than conversational speech (Cooper & Aslin, 1997; Kuhl et al., 1997; Lam & Kitamura, 2012; Uther et al., 2007; Werker & Mcleod, 1989). Thus, the present experiment demonstrated that in British Southeast English mean pitch is larger in IDS than in read speech at both vowel and word level.

Third, as previously hypothesised, read speech did not differ from IDS in mean intensity. Moreover, the finding that all target vowels and all target words were longer in IDS than read speech is as previously hypothesised and also agrees with prior studies that reported expanded vowel space and long vowel duration (Cristia & Seidl, 2103), and showed decreased speaking rate, giving rise to elongated vowels in IDS (Kuhl et al., 1997; Stern et al., 1983). It was previously shown that clear read speech had longer mean word duration than speech to adults in which speech was degraded by noise (Hazan & Baker, 2010, 2011).

It seems to follow logically that in speech to a first-language learner, words are even more elongated than in clear read speech that does not involve any language learners such as infants or non-native speakers. This finding therefore shows that even if speakers might reduce their speech rate in reading compared to when they communicate with an adult speaker (Hazan & Baker, 2010, 2011), they decrease their pace of speech more in interaction with language learners such as infants than when reading clearly. Thus, it can be said that the results of expanded vowel space, increased pitch, pitch range, target vowels and words in IDS compared to clear read speech is consistent with the H&H theory according to which speech production is listener-oriented and speakers adapt their speech depending on the requirements of the communicative situation (Lindblom, 1990).

One has to note, however, the fact that the way clear read speech was elicited in this experiment is not exactly comparable to the manner in which clear read speech was elicited in previous research (Hazan & Baker, 2011): clear read speech in the present experiment was elicited by instructing mothers to read out sentences as clearly as possible while in Hazan and Baker's study (2011) clear read speech was produced upon the instruction to read as if speaking to a hearing-impaired listener (Hazan & Baker, 2011). Similar to the difference that speaking by pretending to speak to another interlocutor is not exactly comparable to speech produced in a genuine interaction with real interlocutors (Charles-Luce, 1997), this difference between the instruction to read out sentences clearly in the present experiment and the instruction to read as if speaking to a hearing-impaired listener in previous research might have contributed to differences in the findings concerning vowel space, pitch, pitch range, vowel length and word duration. Thus, because there is no general agreement on how clear read speech should be produced, clear read speech in the present experiment is not exactly comparable to clear read speech in prior read speech research.

In conclusion, Chapter 5 addressed the research question how IDS acoustically differs from other types of clear speech, specifically clear read speech: the experiment in this chapter extends prior research (Burnham et al., 2002; Fernald & Kuhl, 1987; Kuhl et al., 1997; Uther et al., 2007) by showing that IDS is a uniquely distinct style of speech compared to read speech due to acoustic exaggerations in vowel space, pitch, pitch range, vowel length and word length in IDS. These results therefore support the idea of hyperspeech by the H&H theory (Lindblom, 1990), and the notion of NLM according to which one of the causes towards infants' development of a language-specific speech perception is the acoustic exaggeration of acoustic cues of phonetic units.

Chapter 6

Native listeners' intelligibility of vowel hyperarticulatory speech to adults and infants

6.1 Expanded vowel space and intelligibility

Stretched vowel space is one acoustic-phonetic feature that has been reliably revealed to naturally be part of clear speech (Chen, 1980; Krause & Braida, 2003; Krause & Braida, 2004; Picheny et al., 1986; Uther et al., 2007). Previous research has shown that enlarged vowel space correlates with speech intelligibility (Bond & Moore, 1994; Liu et al., 2003; Monsen, 1976). It was also found that those speakers, who can naturally articulate phonetic contrasts accurately on a segmental level and employ larger vowel space, are more intelligible than those speakers with smaller vowel spaces (Bradlow et al., 1996; Byrd, 1994; Hazan & Markham, 2004)

Nevertheless, it can be said that despite the large number of studies that have investigated the relationship between vowel space and intelligibility, to this date, only an association between vowels space expansion and speech intelligibility has been revealed. Nonetheless, because correlation does not necessarily imply causation, to this date, no study has directly tested whether speech with expanded vowel space, produced under naturalistic circumstances and not merely by instructing speakers to speak as if they are talking to foreign listeners or hearing-impaired listeners are more intelligible than speech samples with non-expanded vowel space in a controlled setting.

In order to move beyond the correlational data linking intelligibility and vowel space, the present chapter will investigate whether expanded vowel space could be manipulated to produce enhanced intelligibility of clear speech. Experiment 3 of this thesis, presented in 6.5of this chapter, will address the research question what effect vowel hyperarticulation has on clarity by focusing on the evaluation of the intelligibility of clear speech produced in speech to foreign sounding and native sounding interlocutors. Experiment 4 of this thesis, presented in 6.6 of this chapter, will address the same question by attending to the assessment of the intelligibility of IDS compared to that of read speech. Due to the use of the same methods employed for both experiments, the chapter will look at the methods of both experiments together in section 6.4, and deal with the results of each experiment separately.

6.2 Intelligibility of speech to foreign-accented interlocutors versus native-accented interlocutors

Previous clear speech studies did not specifically assess the efficacy of articulatory stretched vowel space on speech intelligibility using, for instance, perceptual clarity rating tasks and natural speech materials that were recorded from real-life communicative interactions. Most clear speech investigations have focused on the link between speaker variability and the variability in intelligibility advantage (e.g. Ferguson, 2004; Ferguson & Kewley-Port, 2007; Gagne et al., 1994). In a more recent study, however, stretched vowel space was found to relate to intelligibility although the relationship between expanded vowel space and intelligibility was found to be equivocal due some inconsistency in the production of expanded vowel space among speakers (Ferguson & Kewley-Port, 2007) (for more detailed explanations please refer to 3.1.3 of chapter 1). Thus, the purpose of Experiment 3 presented in 6.5 was to ascertain whether speech with expanded vowel space, as elicited in speech to foreign sounding adults, leads listeners to perceive enhanced intelligibility. Dependent measures of interest included an orthographic transcription task as intelligibility measure with a confidence rating of that transcription using a Likert Scale, a goodness rating task and a clarity rating task in Likert scale format as intelligibility measures. Samples used in the three tasks of Experiment 3 of this thesis consisted of word samples collected from Experiment 1 of this thesis presented in chapter 4.

Based on previous research (Burnham et al., 2003; Uther et al., 2007) and theory (H&H theory by Lindblom (1990)), it is hypothesised that native listeners will identify more words correctly that were extracted from native speakers' speech toward foreign sounding interlocutors than those from native speakers' speech toward native sounding interlocutors. This is because speech to foreign-sounding interlocutors might be spoken more clearly than speech to native-sounding interlocutors due to the linguistic needs of foreign-sounding interlocutors (Burnham et al., 2002; Munro & Derwing, 1999; Uther et al., 2007). Consequently, native listeners will feel more confident in the accuracy of their transcription of native speakers' speech toward foreign sounding interlocutors than that of native speakers' speech toward native sounding interlocutors.

In addition, it is hypothesised that native listeners will rate the words from native speakers' speech to foreign sounding interlocutors as more typical of the words in English that listeners consider them to be than words produced in speech to native sounding interlocutors. This is because these words include phonetically better prototypes of a particular vowel category than the words from native speakers' speech to native sounding interlocutors (Burnham et al., 2002; Iverson & Kuhl, 1995; Kuhl, 1991). It was previously shown that speech to foreign sounding interlocutors included vowel hyperarticulation compared to speech to native sounding interlocutors (Uther et al., 2007). Thus, it is hypothesised that words to foreign sounding interlocutors will be rated as clearer than words in speech to native sounding interlocutors. The following section will look at the role of expanded vowel space in enhanced intelligibility in IDS for Experiment 4 of this thesis.

6.3 Expanded vowel space in IDS and intelligibility

Previous research established the occurrence of stretched vowel space in IDS (Kuhl et al., 1997; Uther et al., 2007) and reported an association between the qualitative speech effort by mothers and infants' performance on phonetic speech perception (Liu et al., 2003). One way to find out if expanded vowel space actually helps infants to perceive words more clearly is the use of speech containing maternal vowel hyperarticulation in speech perception tasks. Thus, the purpose of Experiment 4 was to ascertain whether speech with expanded vowel space, as elicited in IDS, leads listeners to perceive enhanced intelligibility. Dependent measures included the same tasks as in Experiment 3 (see 6.2). Samples included word samples from IDS and mothers' read speech (see chapter 5).

It is hypothesised that because IDS is clearer than read speech, native listeners will identify more words correctly from IDS than mothers' read speech (Kuhl et al., 1997). Native listeners will thus feel more confident in the accuracy of their transcription of IDS than of read speech. Additionally, words in IDS will be rated as more typical of the words in English that listeners consider them to be than words in read speech. Moreover, native listeners will rate words from IDS clearer than from read speech.

6.4 Methods for Experiment 3 and Experiment 4

6.4.1 Participants

6.4.1.1 Participants used to collect speech samples

For Experiment 3, adult speech samples were taken from Experiment 1 in Chapter 4 (see section 4.4.1). For Experiment 4, infant speech samples were taken from Experiment 2 in Chapter 5 (section 5.4.1).

6.4.1.2 Listeners

The listeners were 21 native speakers of English from the Southeast London area. All stated to have normal hearing and basic orthographic skills.

6.4.2 Speech Materials

6.4.2.1 Stimuli for Experiment 3

Stimuli used in Experiment 3 in section 6.5 of this thesis included target words that were produced in Experiment 1 in chapter 5. Twenty-eight tokens from the foreign sounding experimental condition and 22 tokens from the native sounding experimental condition across the three vowels /a:/, /i:/ and /u:/ were taken randomly of Experiment 1 for Experiment 3 of this thesis. Tokens from these vowels were used since previous analysis showed these stimuli to significantly differ in terms of perceptual vowel space between the native sounding and the foreign sounding conditions. In addition, 25 distractor tokens from the foreign sounding experimental condition and 15 distractor tokens from the native sounding experimental condition were randomly taken from across the three vowels /i/, /e/ and /ɒ/. These distractors did not reveal statistically significant expanded vowel space across the four experimental conditions.

6.4.2.2 Stimuli for Experiment 4

Stimuli used in Experiment 4 in section 6.6 of this thesis included target words that were produced in Experiment 2 in chapter 5.

Twenty-four tokens from the mother-infant condition and 17 tokens from the reading condition across the three vowels /a:/, /i:/ and /e/ were taken randomly of Experiment 2 for Experiment 4 of this thesis. Tokens from these vowels were used since previous analysis showed these stimuli to significantly differ in terms of perceptual vowel space between the native sounding and the foreign sounding conditions. In addition, 18 distractor tokens from the mother-infant condition and 15 distractor tokens from the reading condition were randomly taken from the three distractor words "blue", "shoes" and "shop". These distractors did not reveal statistically significant expanded vowel space across the two experimental conditions.

6.4.2.3 Production of speech stimuli

Stimuli used in Experiment 3 were produced in Experiment 1 in Chapter 4 (see section 4.4.4). Stimuli used in Experiment 3 were produced in Experiment 2 in Chapter 5 (see section 5.4.4).

6.4.3 Perceptual evaluation tasks

Listeners completed three sessions. The first session included transcription and confidence rating tasks, the second session included the goodness rating task and the third session included a clarity rating task. The ratings consisted of participants pressing one of six buttons, which included the numbers 1, 2, 3, 4, 5 and 6 on a Likert scale format (1 = very confident/typical/clear; 6 = not confident/typical/clear at all).

Each session was presented through a separate e-prime programme. Stimuli were presented in a random order. All three sessions lasted approximately 35 minutes in total.

6.4.3.1 Orthographic transcription and confidence rating

In the first session, participants were asked to listen to each word stimulus with care and then on the keyboard to accurately type out what they heard in the space provided on the screen. This orthographic transcription task was a measure of speech intelligibility. Participants then indicated on a scale from 1 to 6 how confident they were in the accuracy of their transcription (1 = very confident; 6 = not confident at all). A new word stimulus was presented 500 milliseconds after listeners had rated their confidence in their transcription. The presentation of the next word was indicated by an arrow that appeared for 200 milliseconds.

6.4.3.2 Typicality rating task

In the second session, participants viewed the words on the screen for 300 milliseconds and after another 300 milliseconds, they listened carefully to the viewed words. They then rated the words for typicality on a scale from 1 to 6, where 1 equals very typical and 6 equals not typical at all. A new word stimulus was presented 500 milliseconds after listeners had completed rating the goodness of the previous one. The presentation of the next word was indicated by an arrow that appeared for 200 milliseconds.

6.4.3.3 Clarity rating task

In the third session, participants first viewed the words on the screen for 300 seconds and after another 300 milliseconds, they listened carefully to the viewed words. They then rated the words for clarity on a scale from 1 to 6 (1 = very clear; 6 = not clear at all). A new word stimulus was presented 500 milliseconds after listeners had completed rating the goodness of the previous one. The presentation of the next word was indicated by an arrow that appeared for 200 milliseconds.

6.4.4 Data analysis

Paired samples *t*-test was employed in Experiment 3 and Experiment 4 of this thesis to analyse listeners' evaluation of the intelligibility of speech to native sounding interlocutors versus speech to foreign sounding interlocutors, as well as their evaluation of the intelligibility of IDS versus read speech. The evaluation was assessed using the orthographic transcription task, the confidence rating, the typicality and clarity rating tasks that were completed by the same 21 listeners. Thus, the paired samples *t*-test was used because the four tasks represented four different measures that are not related.

6.5 Results

6.5.1 Experiment 3

6.5.1.1 Transcription

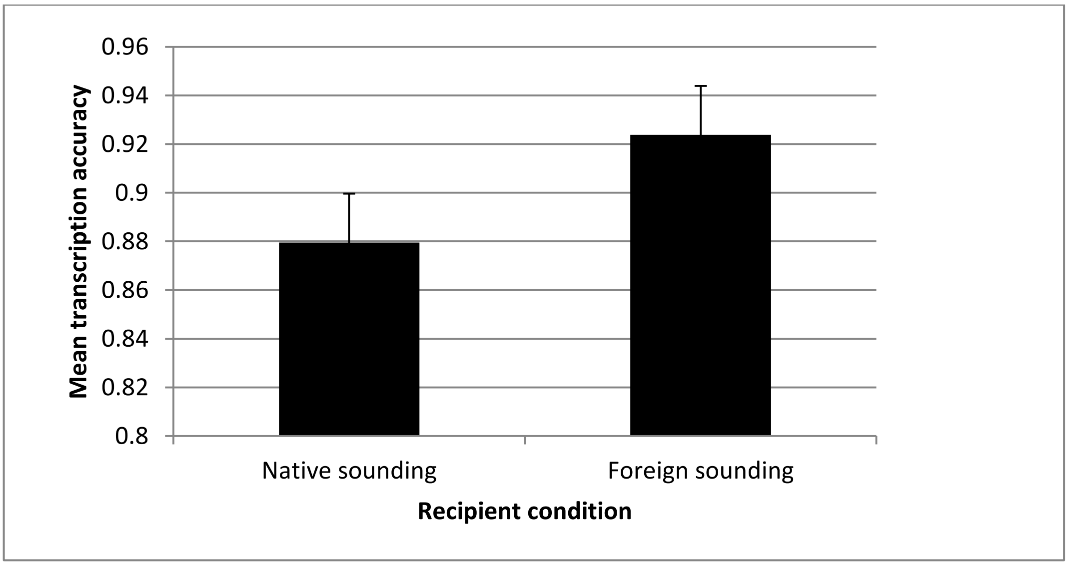


Figure 6.1: Transcription accuracy of native speakers for the native sounding and the foreign sounding recipient conditions. Error bars show + / -1 standard error from the mean.

As shown in Figure 6.1, the mean transcription accuracy was higher for speech samples directed to foreign sounding interlocutors than native sounding interlocutors. This difference was statistically significant (t (20) = 2.426, p < .05, r = .520). This supports the hypothesis that listeners would show higher transcription accuracy for speech directed to foreign than native sounding interlocutors.

6.5.1.2 Confidence

The mean confidence rating was not significantly different between the foreign speaker directed speech and the native-speaker directed speech. This finding therefore does not support the hypothesis according to which listeners would be more confident in their transcription accuracy of speech to foreign sounding interlocutors than native sounding interlocutors.

6.5.1.3 Typicality

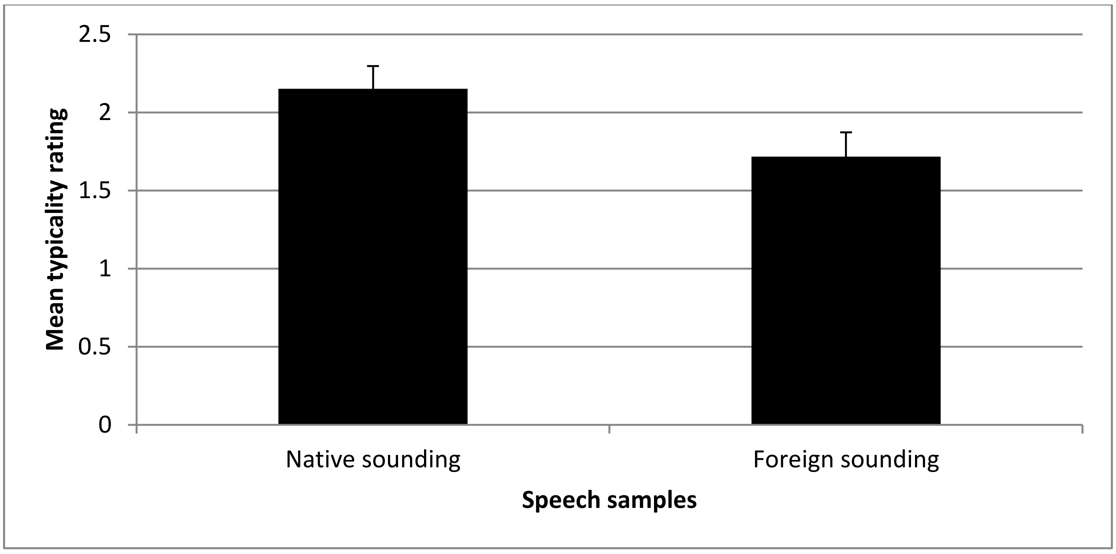


Figure 6.2: Mean typicality rating of native speakers for native and foreign samples on a scale from 1 (very typical) to 6 (not typical at all). Error bars show + /-1 standard error from the mean.

As shown in Figure 6.2, above, native speakers rated speech samples directed to foreign sounding interlocutors as more typical of representing the word in the English language than the speech samples directed to native sounding interlocutors. This difference was statistically significant (t (20) = -5.036, p < .05, r = .722). This supports the hypothesis that native listeners would rate words to foreign sounding interlocutors as more typical of the words in the English language than the words to native sounding interlocutors.

6.5.1.4 Clarity

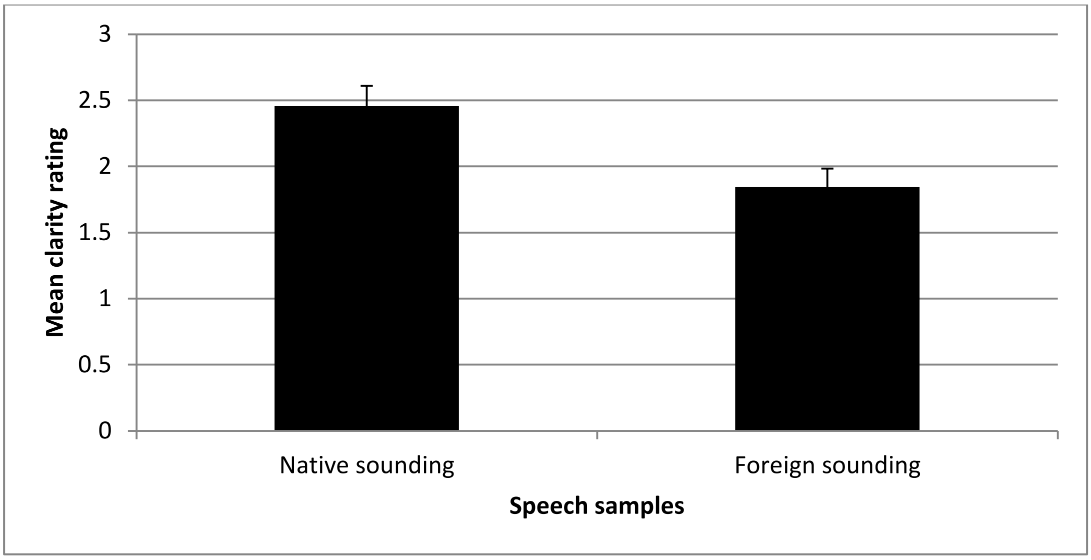


Figure 6.3: Mean clarity rating of native speakers for native and foreign samples on a scale from 1 (very clear) to 6 (not clear at all). Error bars show + /-1 standard error from the mean.

As shown in Figure 6.5.3, on a scale from 1 (very clear) to 6 (not clear at all), native speakers rated speech samples directed to foreign sounding interlocutors as clearer than the speech samples directed to native sounding interlocutors. This difference was statistically significant (t (20) = -6.066, p < .05, r = .764). This finding supports the hypothesis that native listeners would rate words to foreign sounding interlocutors as clearer than the words to native sounding interlocutors.

6.6 Results

6.6.1 Experiment 4

6.6.1.1 Transcription

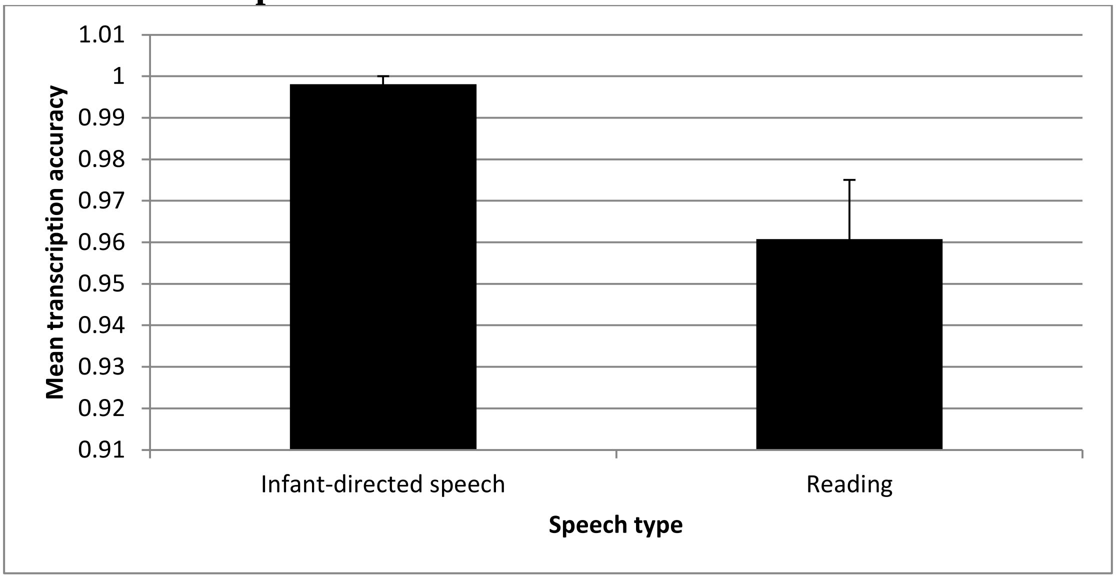


Figure 6.4: Transcription accuracy of native speakers for infant-directed speech and read speech samples. Error bars show + /-1 standard error from the mean.

As shown in Figure 6.4, the mean transcription accuracy was higher for speech samples directed to infants than for samples of read speech. This difference was statistically significant (t (20) = 2.693, p < .05, r = .275) and is as previously hypothesised.

6.6.1.2 Confidence

As shown in Figure 6.5, the mean confidence rating shows that on a scale from 1 (very confident) to 6 (not confident at all), native speakers were more confident in their transcriptions of IDS than of read speech samples.

This difference was statistically significant (t (20) = 6.279, p < .05, r = .354). This finding supports the hypothesis that native listeners would be more confident in their accuracy of transcribing words in IDS than in read speech.

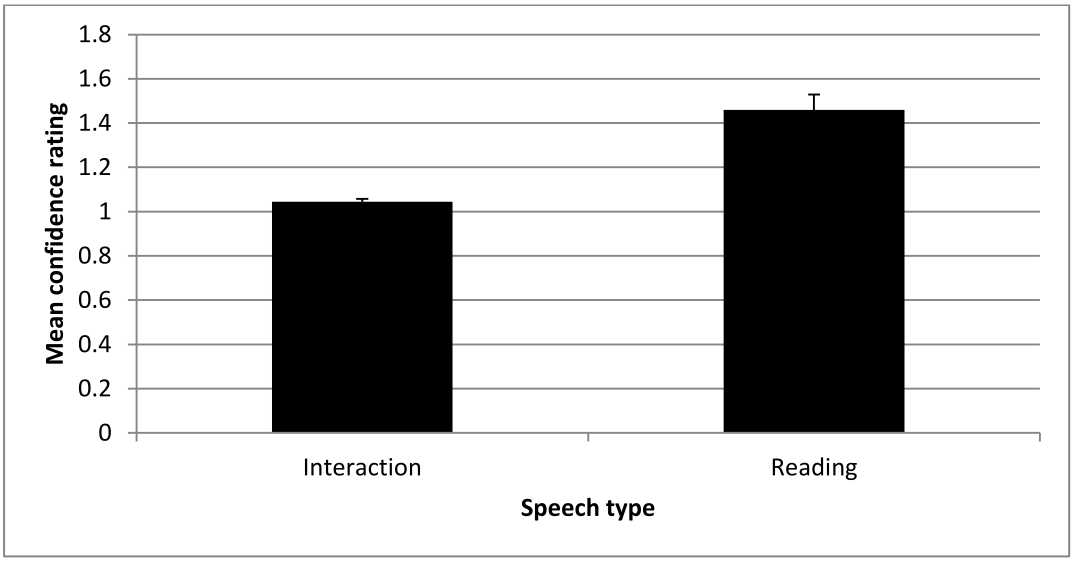


Figure 6.5: Mean confidence ratings of native speakers for infant-directed speech and read speech samples on a scale from 1 (very confident) to 6 (not confident at all). Error bars show + /-1 standard error from the mean.

6.6.1.3 Typicality

Mean typicality ratings by native speakers were not significantly different between IDS and read speech samples. This result does not support the hypothesis that native listeners would rate words in IDS as more typical of the words in the English language than in read speech.

6.6.1.4 Clarity

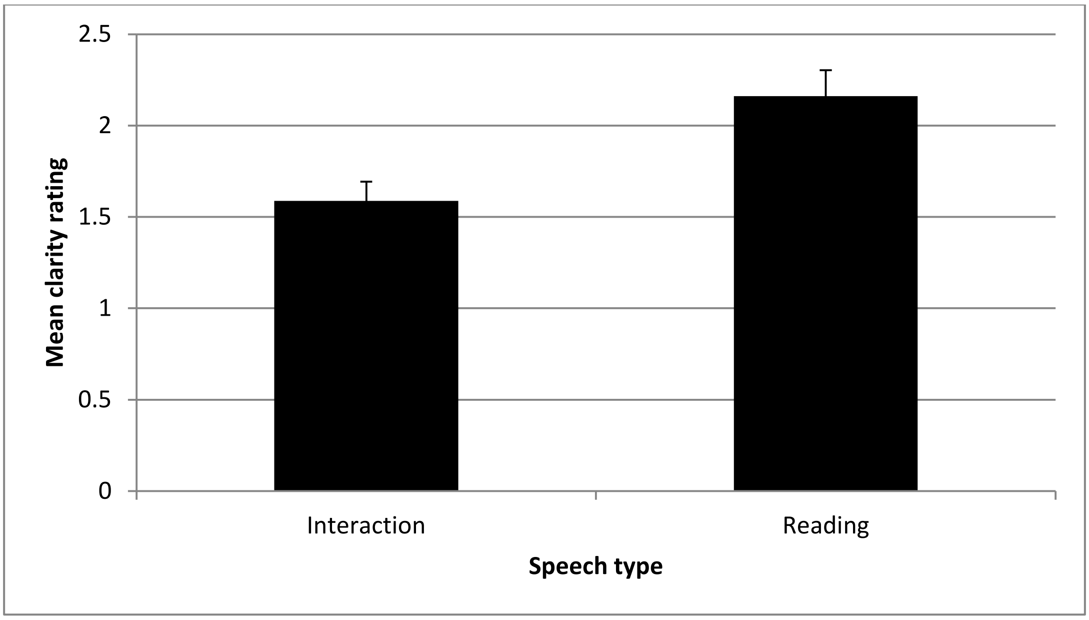


Figure 6.6: Mean clarity ratings of native speakers for infant-directed speech and read speech samples on a scale from 1 (very clear) to 6 (not clear at all). Error bars show + /-1 standard error from the mean.

As shown in Figure 6.6, on a scale from 1 (very clear) to 6 (not clear at all), native speakers rated infant directed speech samples as clearer than read speech samples. This difference was statistically significant (t (20) = -6.393, p < .05, r = .775). This supports the hypothesis that native listeners would rate words in IDS as clearer than the words in mothers' read speech.

6.6.2 Multiple regression

A multiple regression was conducted on the clarity rating of the adult samples and infant samples to distinguish between the relative contribution of vowel space expansion, vowel length and word length to the outcome in the clarity ratings in 6.5.1.4 and 6.6.1.4 respectively. Based on previous research on what might contribute to vowel/word intelligibility (Ferguson & Kewley-Port, 2007; Smiljanić & Bradlow, 2009; Uchanski, 2005), predictors were entered hierarchically in order of their importance in predicting the outcome that is first F1, F2, before vowel length and then word length.

6.6.2.1 Clarity rating with speech samples directed to foreign-sounding interlocutors

When only F1 and F2 are used as predictor, this is the simple correlation between F1 and F2 with clarity rating (.438). F1 and F2 can account for 19.2% of the variation in the clarity rating as the value of R squared is .192 (model 1). When vowel length is included (model 2), this value increases to .231 or 23.1%. When word length is added as a predictor (model 3) 23.2% of the variance in clarity ratings can be accounted for. Thus, vowel and word length account for an additional 4% of the variance in clarity ratings. The F-ratio is 2.963 for the first model, which is close to significance (p=.070). For the second and third model, the F-ratio decreases gradually, and is not significant either. Therefore the initial model is not significantly good at predicting the outcome in the clarity rating.

6.6.2.2 Clarity rating with speech samples directed to infants

When F1 and F2 are used as predictor, this is the simple correlation between F1 and F2 with clarity rating (.087). F1 and F2 can account for 0.8% of the variation in the clarity rating (model 1).When vowel length is included (model 2) the value of R increases to .012 or 1.2%. When word length is added as a predictor (model 3) 6.5% of the variance in clarity ratings can be accounted for. Thus, word length accounts for an additional 5.3% of the variance in clarity ratings. The F-ratio is 0.080 for the first model, which is not significant. For the second and third model, the F-ratio improves gradually, but is not significant either. Therefore none of the three models are significantly good at predicting the outcome in the clarity rating.

6.7 Discussion

6.7.1 Intelligibility of speech to foreign-accented interlocutors versus native-accented interlocutors

Experiment 3 of this thesis aimed to investigate what effect vowel hyperarticulation in speech to interlocutors with foreign accents has on intelligibility. Experiment 3 therefore compared whether speech to foreign sounding or native sounding interlocutors will lead native listeners to experience perceptually better intelligibility of speech.

First, the first task of Experiment 3 was a transcription task. Similar to prior studies (Bradlow & Bent, 2002; Bradlow & Alexander, 2007; Smiljanić & Bradlow, 2011; Munro & Derwing, 1999; Lane, 1963), this transcription task did not accept any word candidates that were recognised half-way by native speakers but only transcribed words that were identified correctly. This was done so as not to be unclear about whether near hits resulted because of typing errors or because listeners actually did not recognise the presented word. It was shown that native listeners' mean transcription accuracy was higher for speech to foreign sounding interlocutors than native sounding interlocutors. This confirms the hypothesis that native listeners would transcribe more words more accurately that had been articulated in speech to foreign sounding interlocutors than to native sounding interlocutors.

This finding might be explained by the observation in Experiment 1 in Chapter 4 that speech to foreign sounding interlocutors involved expanded vowel space compared to speech to native sounding interlocutors. Thus, the present finding on transcription implies that the acoustic alteration of vowel space in speech to foreign sounding interlocutors might have contributed to native listeners' higher transcription accuracy compared to the lower transcription accuracy of speech spoken to native sounding interlocutors. This result therefore seems to suggest in line with the revised form of the cohort model that expanded vowel space might have facilitated native listeners' bottom-up processing in spoken word recognition (Marslen-Wilson, 1990).

According to the revised form of the cohort model, spoken word recognition is affected by bottom-up processes that are determined by the heard stimulus (Marslen-Wilson, 1990). Therefore, when a word is presented aurally, words that conform to the series of sounds heard become activated (Marslen-Wilson, 1990). It can be considered that this activation might have been facilitated by the fact that stimuli represented words from everyday life.

This result of higher transcription accuracy of speech spoken to foreign sounding interlocutors also upholds previous research according to which foreigners have special linguistic needs, to whom speech needs to be acoustically modified correspondingly (Burnham et al., 2003; Uther et al., 2007). In addition, this finding extends previous evidence of expanded vowel space contributing to vowel intelligibility (Ferguson & Kewley-Port, 2007).

By contrast, the confidence rating data (which showed no difference between speech to native vs. foreign speakers) did not support the hypothesis that listeners would be more confident in their transcription of foreign-directed speech. This might be attributed to the nature of the rating scale used for the confidence rating task, which in contrast to the majority of speech rating studies (e.g. Iverson & Kuhl, 1995; Smiljanić & Bradlow, 2011), used a scale from 1 to 6, with 1 being least and 6 being most confident in one's transcription accuracy. This pattern of the present rating scale, which was also used in Munro & Derwing's study (1999), is opposite to what is actually used in speech rating tasks.

Nonetheless, the clarity rating data (which showed speech to foreign speakers to be clearer than speech to native speakers) supports the above finding of listeners' higher transcription accuracy of words in speech to foreign speakers than native speakers. This supports the hypothesis that speech to foreign speaking interlocutors would be perceived as clearer than to native speaking interlocutors. This seems to suggest that the effect of vowel hyperarticulation on intelligibility appears to be that in FDS it significantly enhances the clarity of speech.

Previously, appropriate spoken word identification and therefore speech intelligibility has been commonly found to be affected by how accurate vowels are articulated (Liu et al., 2005; Weismer, Laures, Jeng, Kent, & Kent, 2001). Accordingly, it was reported that the association between the vowel working space area and the extent to which speech is intelligible is essential when exploring how listeners' recognition of spoken speech can be influenced by speakers' size of their vowel working space area (Liu et al., 2005). The result of the clarity rating task thus appears to underline the importance of this association. This result therefore seems to extend previous studies that correlated vowel space expansion with speech intelligibility (Bond & Moore, 1994; Bradlow et al., 1996; Byrd, 1994; Hazan & Markham, 2004; Johnson et al., 1993; Picheny et al., 1986).

This finding also appears to be in line with the H&H theory according to which hyperarticulated speech allows more accurate phonetic units to be more easily perceived as being acoustically distinct (Lindblom, 1990). This result therefore suggests that vowel hyperarticulation in speech to foreign speakers might be a useful didactic device in order to increase their intelligibility of spoken speech.

However, since in Experiment 1 in Chapter 4 of this thesis, longer vowel duration and longer word length were reported in speech to foreign sounding interlocutors than native sounding interlocutors, a multiple regression was conducted to distinguish between the relative contributions of the different acoustic measures to the outcome of the clarity rating. Since none of the models was successful in predicting the higher clarity ratings of speech to foreign-sounding interlocutors compared to native-sounding interlocutors, it may be that this outcome might be due to a small number of cases in the data (N=28), which is a little more than the required minimum of N=15 for a multiple regression analysis.

This result suggests that a more effective approach to differentiate between the contributions of vowel space expansion, long vowel and word length to enhancing clarity in speech would involve a separate perceptual experiment. Specifically, in the first experimental condition native listeners would rate words for clarity that are accompanied by vowel space expansion only. In the second experimental condition, native listeners would rate the same words, which now are characterised by enhanced vowel length only but not by vowel space expansion. In a third condition, these words would be presented with enhanced word length but no vowel space expansion and long vowel length. A baseline condition can involve the rating of the same words but with none of the three acoustic features.

The finding of higher typicality ratings for speech to foreign speakers than native speakers seems to suggest that words in speech to foreign sounding interlocutors included phonetically better prototypes of particular vowel categories than the words in speech to native sounding interlocutors. This might be because of the fact that acoustic features in ADS are normally underspecified compared to FDS and is of hypoarticulated nature (e.g. ‘and’ becomes ‘nd’ in casual speech) (Burnham et al., 2002; Uther et al., 2007). It seems therefore understandable that words spoken to foreign sounding speakers were perceived as phonetically more typical than words spoken in conversation between two native speakers. This finding therefore supports the hypothesis that speech to foreign sounding interlocutors would be more typical than speech to native sounding interlocutors because it has good exemplars of phonetic categories. This finding therefore does not contradict the finding that speech to foreign sounding interlocutors is clearer than that to native sounding interlocutors.

Therefore, the findings of Experiment 3 seem to confirm the role of speech to foreign sounding interlocutors to be of didactic advantage. It can therefore be proposed that speech to foreign sounding listeners might help second language (L2) learners to gain sufficient phonemic discrimination skills to fully generate and comprehend L2. Although the findings seem to indicate the usefulness of clear speech with expanded vowel space, whichmight onlyapply to words and if spoken to normal hearing listeners in quiet, it cannot be said with absolute certainty that expanded vowel space contributed to the higher clarity rating. Therefore additional research would be required to tease out the relative contributions of the different acoustic aspects to clarity The findings of Experiment 3 therefore suggest that vowel space expansion as it is used together with other acoustic-phonetic features in speech to foreign sounding listeners might be useful to be employed in linguistic training programs to facilitate foreign sounding listeners’ intelligibility of the target language.

6.7.2 Intelligibility of speech to infants versus read speech

Experiment 4 of this thesis aimed to investigate what effect vowel hyperarticulation in speech to infants has on intelligibility. Experiment 4 therefore investigated whether speech to infants (IDS) or clear read speech as an alternative type of clear speech would lead native listeners to experience perceptually better intelligibility of speech.

First, the first task of Experiment 4 was a transcription task. It was shown that native listeners transcribed mothers' speech to infants more accurately than read speech. This finding confirms the hypothesis that native listeners would transcribe more words more accurately in IDS than in read speech. Native listeners also showed more confidence in their accuracy of transcribing mothers' speech to their infants than of mothers' read speech. This supports the hypothesis that listeners would be more confident in transcribing IDS than read speech. These transcription and confidence rating data might be explained by the observation that in Experiment 2 in Chapter 5 IDS revealed significantly expanded vowel space compared to mothers' read speech. It can therefore be argued that the acoustic-phonetic alteration of vowel space might have contributed to native listeners' higher accuracy and confidence in transcribing speech to infants (Kuhl et al., 1997; Uther et al., 2007).

Experiment 4 showed that neither mothers' read speech nor their IDS were rated to be significantly more typical of speech in English. This does not support the hypothesis that words in IDS would be regarded by listeners to be more typical of the words in English that they consider them to be. However, this result might be explained by the fact that IDS is of higher pitch compared to ADS (Burnham et al., 2002; Lam & Kitamura, 2012; Uther et al., 2007). Therefore words in IDS might not sound as appropriately representative of the words in English than when expressed in ADS.

This finding that words in IDS were not rated as more typical of the words in English that listeners consider them to be is therefore not in conflict with the finding in Experiment 3 that words in FDS were rated as being significantly more typical of the words in English compared to words in ADS. This is because similar as words in ADS, words in FDS are low in pitch but include phonetically better prototypes of vowel categories.

The finding that native listeners rated IDS as being clearer compared to read speech supports the above results of listeners' higher transcription accuracy and higher confidence in IDS than read speech. This finding is not in conflict with the absent finding of significantly higher typicality rating for IDS compared to clear read speech, because it suggests that the clarity of words in IDS is not dependent on the quality of pitch, which however, seems to affect the extent to which the words in IDS are rated to be typical of words in English.

This finding on clarity in IDS therefore supports the hypothesis that IDS would receive higher clarity ratings than read speech. It therefore appears that vowel hyperarticulation in IDS might have significantly enhanced the clarity of speech. This finding seems to provide direct evidence connecting vowel hyperarticulation to increased intelligibility of speech. This finding therefore appears to extend previous studies (Liu et al., 2003; Song et al., 2010) that showed that the extent of expanded vowel space correlated with infants' phonetic discrimination abilities (Liu et al., 2003) and that infants' word recognition was supported by extended vowel space to a certain extent (Song et al., 2010). Thus, Experiment 4 seems to support the idea proposed by other researchers that vowel hyperarticulation in IDS is didactic in nature (Kuhl et al., 1997; Uther et al., 2007).

The finding of higher clarity ratings for IDS samples than mothers' read speech is in line with Kuhl's perceptual Native Language Magnet Theory (enhanced) theory. The theory suggests that expanded vowel space in IDS is used to overstress crucial phonetic contrasts through which infants' speech perception skills can be assisted in the acquisition of their native language (Kuhl, 2001, 2004; Liu et al., 2003). The findings of Experiment 4 therefore suggest that expanded vowel space as it occurs in IDS in combination with other acoustic-phonetic characteristics, such as high pitch, might be useful to be employed in linguistic training programs to facilitate infants' development of phonetic discrimination abilities.

However, because Experiment 4 of this thesis seems to only establish a link between expanded vowel space and increased intelligibility at word level, this finding might not apply to native listeners' intelligibility of sentences. Moreover, the fact that native listeners did not rate mothers' speech to infants more typical than mothers' read speech implies that speech produced in mother-infant interaction might not reflect the manner in which words are normally generated in daily life. This might be because mothers' speech to infants is usually of higher pitch and also includes overemphasised intonation contours (Fernald & Simon, 1984; Grieser & Kuhl, 1988; Stern et al., 1983). On the other hand, speech to adults in daily life normally does not include higher fundamental frequency or overemphasised intonations (Burnham et al., 2002; Uther et al., 2007). Moreover, exaggerations in pitch and also slower speaking rate are utilised in IDS to sustain infants' attention in communication (Kitamura & Burnham, 2003; Lam & Kitamura, 2012) while these acoustic-phonetic changes are not used to maintain adults' attention in speech to normal-hearing adults (Lam & Kitamura, 2012; Uther et al., 2007).

In addition, Experiment 4 of this thesis seems to suggest that expanded vowel space might support infants' perceptual word recognition abilities. However, it is important to acknowledge the fact that Experiment 4 did not directly measure the effect of expanded vowel space on infants' perceptual skills. Also, due to the limiting nature of the speech stimuli from Experiment 2, Experiment 4 consisted of rating tasks only and did not involve any phonetic discrimination tasks.

Moreover, since in Experiment 2 in Chapter 5 of this thesis, longer vowel duration and longer word length were also reported in speech to infants compared to read speech, a multiple regression was conducted to distinguish between the relative contributions of the different acoustic measures to the outcome of the clarity rating. Since none of the models was successful in predicting higher clarity ratings of speech to infants, it may be that this outcome might be due to a small number of cases in the data (N=24), which is slightly more than the required minimum of N=15.

It might therefore be useful to differentiate between the contributions of expanded vowel space, long vowel and word length to improving clarity by conducting an additional perceptual experiment: specifically, in one experimental condition native listeners would rate words for clarity that include enhanced vowel formants only, while in a second condition the same words would be presented to native listeners, which, however, are now characterised by enhanced vowel length only and not by vowel space expansion. In a third experimental condition, these words would be presented with enhanced word length only and no vowel space expansion and long vowel length. A baseline condition can involve the rating of the same words with neither vowel space expansion nor long vowel or word length. Therefore, it can be suggested that overall the findings of Experiment 4 seem to confirm that the role of IDS provides listeners with a didactic advantage.

In conclusion, Chapter 6 addressed the research question what effect vowel hyperarticulation in speech to infants and foreign speakers has on intelligibility. It evaluated the intelligibility of clear speech produced in speech to foreign sounding and native sounding speakers in Experiment 3, and it assessed the intelligibility of IDS and clear read speech in Experiment 4. Both experiments suggested that vowel hyperarticulation, as it occurs together with other acoustic-phonetic features in speech, seems to enhance intelligibility of both IDS and of speech to foreign speakers, thereby providing evidence for vowel hyperarticulation as a didactic instrument.

Chapter 7

Native and non-native listeners' speech comprehension performance under adverse listening conditions

7.1 Speech perception of native and non-native listeners in noise

In noiseless environments, speakers of a second language (L2) perform like native speakers in speech perception tasks (e.g. Nábělek & Donahue, 1984). However, when interrupted by background noise, their speech perception in L2 is more affected than in their first language (L1) (Florentine, 1985a, b; Garcia Lecumberri & Cooke, 2006; Mayo, Florentine, & Buus, 1997; Takata & Nábělek, 1990). This effect has been suggested to be associated with listeners' age (Bergman, 1980), the time period of L2 study (Florentine, 1985a, b) and the environmental situation under which listening occurs (Takata & Nábělek, 1990).

In the presence of noise, non-native speakers' performance on L2 speech perception tasks has been shown to depend on the age at which they acquire L2 (Florentine, 1985b; Mayo et al., 1997). For example, research by Florentine (1985b) revealed that exposure to L2 from infancy onwards helped two L2 listeners to perform as well as L1 speakers on speech perception tasks in the presence of increasing noise. By contrast, L2 listeners who had been exposed to L2 only after puberty did not perform at the same level as L1 listeners of American English even after massive exposure (Florentine, 1985b). Moreover, L2 listeners did not make use of any contextual cues, which contrasts to the effects seen in L1 listeners (Florentine, 1985b). These data are interpreted as indicating a sensitive period after which learning a second language negatively affects L2 listeners' perception of L2 in noise (Florentine, 1985b). It was shown that in speech perception tasks with noise, early learners of L2 performed better and benefitted more from sentence-level contextual information compared to late but very proficient L2 learners. However, early L2 learners' ability to perceive L2 in noise has been suggested to be decreased and not be like that of native listeners' due to intervention by L1 experience (Mayo et al., 1997).

It has therefore been argued that early L2 learners' better performance might be due to the age at which L2 was acquired and not the average time length of L2 exposure (Mayo et al., 1997). Consequently, if L2 study is not started in early childhood, L2 listeners will have difficulty perceiving L2 in noise even with extensive exposure. This has been illustrated by the observation that early L2 learners showed higher levels of tolerating noise than late L2 learners (Mayo et al., 1997). However, L1 English listeners had higher noise-tolerance levels than early L2 English learners (Mayo et al., 1997). L1 listeners have thus been claimed to be able to recover quickly from noise-induced disturbance because of their linguistic knowledge of established L1 categories (Bradlow & Alexander, 2007). On the contrary, late L2 listeners are not able to recover their speech perception that is disrupted by noise as quickly as L1 listeners because their lacking linguistic knowledge of L2 causes their recovery from noise to be too slow (Bradlow & Alexander, 2007).

As result of late L2 listeners' limited exposure to L2, it is argued that late L2 listeners do not respond to clear speech as well as early L2 learners or L1 listeners (Bradlow & Bent, 2002; Smiljanić & Bradlow, 2011). Specifically, clear speech is considered to have signal enhancements such as slow speech rate and broad pitch range that all listeners are regarded to be able to access (Hazan & Simpson, 2000). However, clear speech also includes subtle enhancements that are specific to the target language and that are considered to improve the acoustic distance among phonologically different contrasts in the target language (Bradlow & Bent, 2002; Smiljanić & Bradlow, 2011). It has been proposed that these enhancements are only accessible and beneficial to native listeners and early L2 learners because compared to late L2 learners they are very experienced in the sound structure of L2 and in processing L2 (Bradlow & Bent, 2002; Smiljanić & Bradlow, 2011). It is therefore, for example, considered that only L1 listeners and early L2 learners, who are familiar with the difference in duration between short and long vowels in the English language, will be able to show sensitivity to and thus benefit from an exaggerated dissimilarity between these vowels in clear speech (Bradlow & Bent, 2002).

As a consequence, it has been argued that late L2 listeners might benefit only very little from clear speech compared to conversational speech under degraded situations. Evidence for this comes from Bradlow and Bent (2002) who aimed to find out if low-proficient L2 listeners can benefit from clear speech produced by L1 English speakers under different noise conditions. In that study, in which slow speech rate, broad pitch range and larger sound pressure levels were considered aspects that lead to the improved signal of clear speech, late L2 listeners showed a smaller benefit from clear speech compared to L1 listeners. This outcome has been suggested to be caused by late L2 listeners' limited experience with the L2 sound structure. The authors therefore argued that the nature of clear speech is not oriented towards L2 listeners but towards L1 listeners (Bradlow & Bent, 2002). However, one has to note that clear speech in their study was produced by instructing L1 speakers to read sentences as if talking to hearing-impaired or foreign listeners. Clear speech in that study was therefore not elicited in a natural interaction with a real interlocutor, and it was not specifically aimed at foreigners. Accordingly, whether clear speech is oriented towards a specific target audience depends on whether the kind of clear speech is directed at that audience as well. Naturally produced clear speech directed at foreigners can thus be regarded to be oriented to L2 listeners.

There is an abundance of literature on the effects of clear speech and intelligibility (for detailed information see section 6.1 of chapter 6) (Bradlow & Bent, 2002; Bradlow et al., 2003). However, there are gaps in terms of research that indicates which clear speech properties are beneficial for L1 and L2 listeners' speech comprehension under noisy conditions. Previous studies mainly looked at the relationship between intelligibility and several clear speech properties at vowel level, and highlighted the role of expanded vowel space in enhancing vowel intelligibility (Ferguson & Kewely-Port, 2002, 2007). Intelligibility of speech does not equal speech comprehensibility (Hustad & Beukelman, 2002): comprehension assesses a listener's ability to construe the meaning of an acoustic signal in order to be able to answer questions about its contents while intelligibility indicates the extent to which a listener can precisely retrieve the acoustic signal (Hustad, 2008). It is notable that previous comprehensibility studies that were administered to ask listeners for sentence-level information (Hustad & Beukelman, 2002) or narrative-level information (Hustad, 2008) were presented to native listeners. No research has been done on which clear speech properties (e.g. expanded vowel space) produce a clear speech benefit at word level for L2 learners.

Experiment 5 of this thesis therefore aimed to address the research question whether expanded vowel space improves clarity for listeners, thereby testing the hypothesis that expanded vowel space in speech will be more comprehensible for both L1 British

English speakers' and early and late L2 British English learners in quiet and in noise conditions.

7.2 Methods

7.2.1 Participants

7.2.1.1 Participants used to collect speech samples

Stimuli for Experiment 5 were obtained using the samples gathered in Experiment 1 (section 4.4.1 in Chapter 4).

7.2.1.2 Listeners

The listeners consisted of three groups: 16 monolingual speakers of British English (between 18-45 years) from the Southeast London area and 16 native speakers of Mandarin Chinese (between 18-45 years) who learned English before the age of twelve years and 16 native speakers of Mandarin Chinese (between 18-45 years) who learned English after the age of twelve years. The average age of twelve was chosen based on prior research (Flege, 1995; Flege & MacKay, 2004). Non-native listeners were recruited from the Brunel Language Centre. All listeners were enrolled at Brunel University and had no speech or hearing impairments at the time of testing. Participants were paid £10 for taking part.

7.2.2 Speech stimuli

Stimuli used in Experiment 5 included target words that were produced in Experiment 1 presented in Chapter 4. Initially 150 target words belonging to one of thirteen target vowels were recorded from native speakers during the completion of the Diapix task (see Appendix I). The vowels /a:/, /i:/, /u:/, /i/, /e/ and /ɒ/ were chosen from the target words "car", "beach", "blue", "pink", "red" and "shop" as they contained a minimum of one sample. Five instances from each of the three vowels /a:/, /i:/ and /u:/ were taken randomly of each of the four experimental conditions of Experiment 1 for Experiment 5. In addition, instances from the three vowels /i/, /e/ and /ɒ/ were taken randomly of each of the four experimental conditions as distractors.

After the recordings were generated, target words were extracted from the digital speech files. Word-length target files were equated for root-mean-square amplitude before being mixed with white noise as background noise (similar as in Billings et al., 2009) (generated in MATLAB) at +16dB, +12dB and +8dB SNRs. The noise created for each target word had the same total duration as the speech signal. White noise was employed because this type of energetic masking was found to influence native and non-native listeners to the same degree for everyday words and syntactically and semantically simple speech material (Cutler et al., 2004; Garcia Lecumberri, Cooke, & Cutler, 2010). This type of noise is not specific to speech and thus represents environmental degradation of speech. Based on previous research, the SNR at +8dB SNR was chosen as medium noise, and the SNR at 16dB SNR was selected as very low noise, with +12dB SNR chosen as a middle noise level (Bradlow et al., 2003; Cutler et al., 2008).

7.2.3 Procedure

7.2.3.1 Speech production

For details on how speech samples were obtained for experiment 5, the reader is referred to Experiment 1 presented in section 4.4.4 of Chapter 4.

7.2.3.2 Speech perception task

Listeners completed the speech comprehension task in front of a computer monitor in an experimental cubicle. Listeners heard the audio stimuli using headphones with a comfortable listening level set prior to the start of the task. Stimuli were presented in a random order. The session lasted approximately 40 minutes.

In this task, participants were asked to listen to each word stimulus with care and then to indicate on a scale from 1-6 whether the heard stimulus was easy to understand (1 = not easy to understand at all; 6 = very easy to understand). There was a 500 milliseconds delay in presenting the subsequent stimuli after the participant indicated their response. The presentation of the next word was signalled by an arrow that appeared for 200 milliseconds.

Before the experimental session, a practice session with 16 trials was implemented in which four non-experimental practice words were presented at one of the four SNR levels so that listeners got accustomed to the nature of the task and to the stimuli with noise. None of the experimental target words were used for this practice session.

During the experimental session, each word stimulus was presented three times and listeners could take as long as necessary to give a response. To minimise learning effects over the time-length of the study, the order of presentation of the words stimuli was randomised. Signal-to-noise ratio (quiet vs. +8dB SNR vs. +12dB SNR vs. +16dB SNR), and speaking style (casual speech to native interlocutors versus clear speech to non-native interlocutors) represented the within-subjects factors while the independent variable of listeners (native English listeners vs. early native Mandarin learners of English vs. late native Mandarin learners of English) was varied between participants.

7.2.4 Data analysis

The design of this study was a 3x2x4x3 mixed design, with the vowels (/a:/, /i:/, /u:/), the recipient condition (native sounding and foreign sounding), and the noise levels (quiet vs. +8dB SNR vs. +12dB SNR vs. +16dB SNR) representing the within-subjects factors and the three listening groups (native listeners, early non-native listeners, and later non-native listeners) representing the between-subjects factor. Therefore a mixed ANOVA for the analysis of listeners' performance in the spoken comprehensibility task was used.

7.3 Results

A regression analysis showed that all listeners across all three listening groups rated all 360 speech samples consistently in reaction time and in their response choice.

7.3.1 Mean rating

A mixed ANOVA across all three listening groups showed that speech to foreign sounding interlocutors was easier to understand than speech to native sounding interlocutors (F (1, 45) = 205.002; $p < .05$, $\eta^2_p = .820$) (Figure 7.1). This supports the hypothesis that stimuli with expanded vowel space will improve listeners' comprehension.

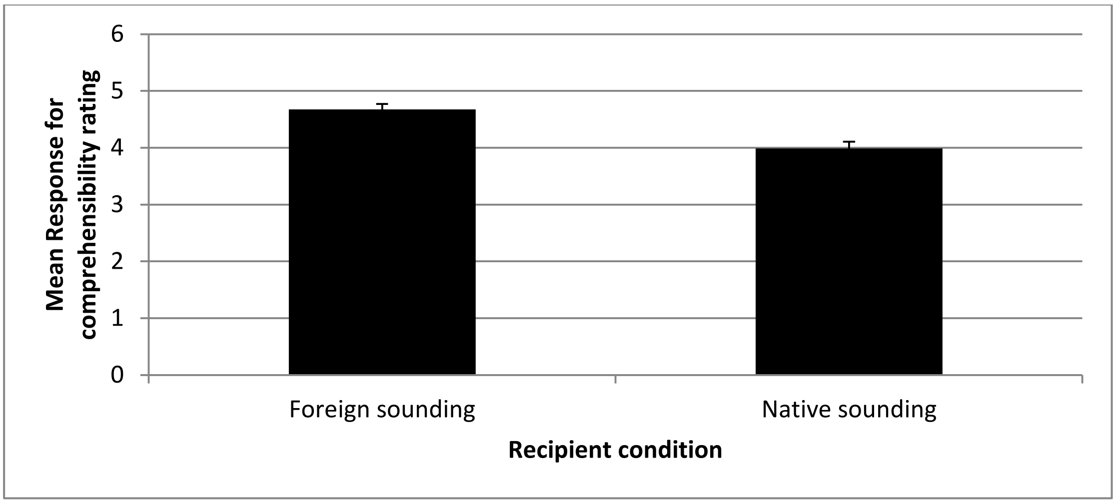


Figure 7.1: Mean response for comprehensibility ratings as a function of recipient condition (foreign sounding; native sounding). Error bars show +/ -1 standard error from the mean.

The assumption of sphericity was violated for noise, which according to Multivariate Tests was a significant main effect (F (3, 43) = 43.325; p < .05, η^2_p = .751) (Figure 7.2). Pairwise comparisons showed that while speech at quiet was easier to understand than at +16dB SNR (t (47) = 10.545, p < .0083, r = .797), at +12dB SNR (t (47) = 10.677, p < .0083, r = .734), and at +8dB SNR (t (47) = 11.427, p < .0083, r = .678), speech at +16dB SNR was easier to understand than at +12dB SNR (t (47) = 7.019, p < .0083, r = .981) and +8dB SNR (t (47) = 9.886, p < .0083, r = .958), while speech at +12dB SNR was easier to understand than at +8dB SNR (t (47) = 8.287, p < .0083, r = .981). These results show that in noise, speech at quiet or lower noise levels are easier to understand than speech at higher noise levels.

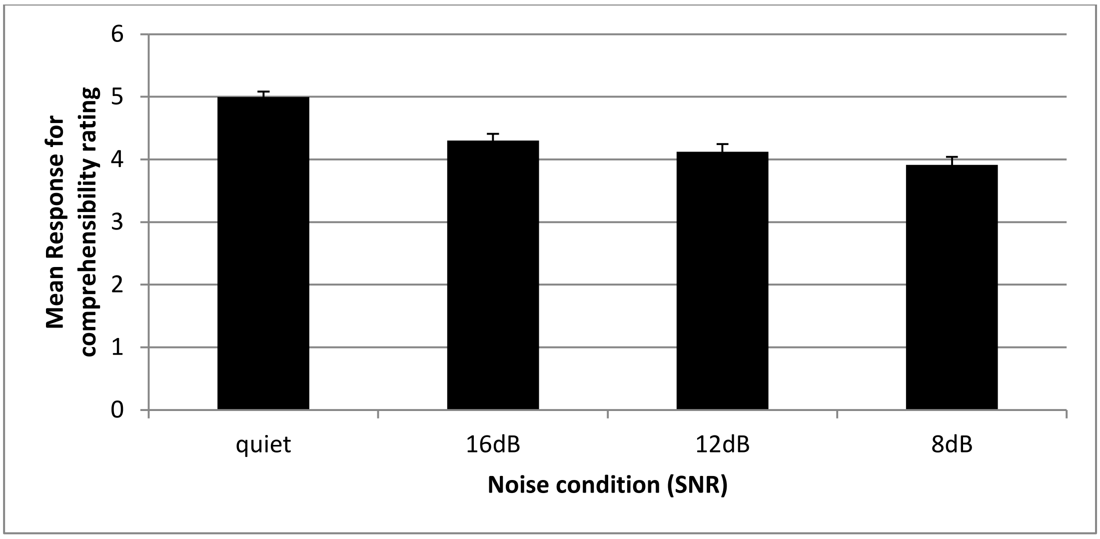


Figure 7.2: Mean response as a function of SNR (quiet, +16dB, +12dB and +8dB). Error bars show +/ - 1 standard errors from the mean.

Moreover, there was a significant interaction between recipient condition and noise: listeners' rating of speech to foreign sounding interlocutors was not affected by the level of noise as much as their rating of speech to native sounding interlocutors (F (3, 43) = 8.693; $p < .05$, $\eta^2_p = .378$) (Figure 7.3). Thus, this result shows that not only at quiet but also in the presence of noise stimuli with expanded vowel space were rated more comprehensible across listener groups than stimuli without expanded vowel space. This indicates a role of expanded vowel space in improving comprehension of speech that is presented in background noise.

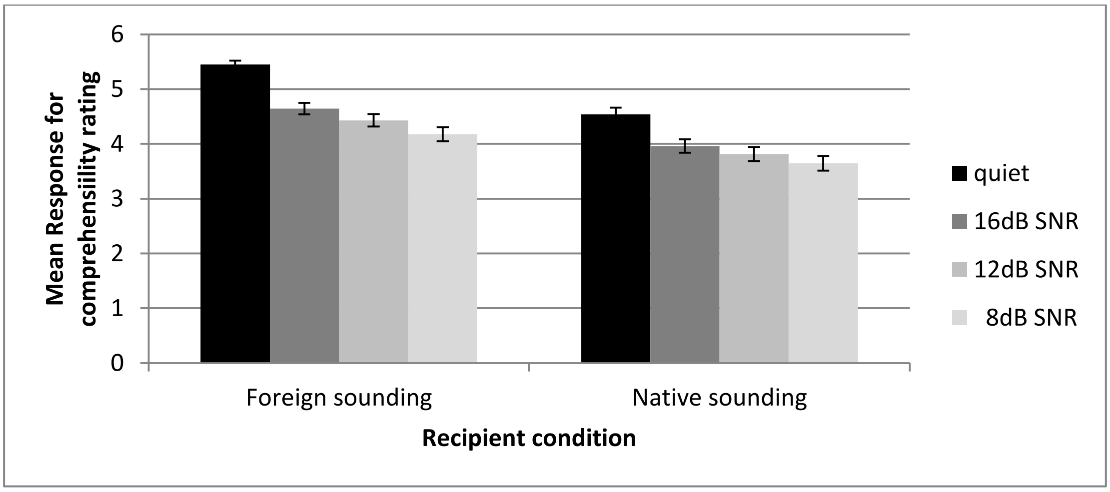


Figure 7.3: Mean response for comprehensibility ratings of speech to different recipient conditions (foreign-sounding; native-sounding) at different SNRs (quiet, +16dB, +12dB and the +8dB). Error bars show +/ -1 standard errors from the mean.

7.4 Discussion

Experiment 5 aimed to answer the research question what effect vowel space expansion has on L1 and L2 listeners' comprehensibility of speech. First, the investigation of mean rating revealed across listener groups that speech directed at foreign sounding interlocutors was easier to understand than speech directed at native sounding interlocutors. It was also observed that speech at quiet and low noise levels were easier to understand than speech at high noise levels. These findings are consistent with the previous hypothesis that speech with expanded vowel space will improve listeners' comprehension.

Moreover, it was found that at both quiet and in the presence of background noise speech with expanded vowel space was easier to understand than speech without stretched vowel space. There were no differences between listener groups in performance at quiet or at the different noise levels. These observations do not support previous research that proposed that at quiet early L2 learners would show a speech comprehension benefit from expanded vowel space that is comparable to that of L1 English listeners and that is larger than that of late L2 English learners (Bradlow & Bent, 2002; Smiljanić & Bradlow, 2011). Similarly, these observations do not support suggestions by previous research that in noise early L2 English learners will, compared to late L2 English learners, find stimuli with expanded vowel space more comprehensible but less than L1 English speakers (Florentine, 1985a, b; Garcia Lecumberri & Cooke, 2006; Mayo et al, 1997; Rogers et al., 2006; Takata &

Nábělek, 1990). Thus, it seems that despite of their varying proficiency levels in L2, both early and late L2 listeners appear to have equally benefitted from the stretched vowel space that was embedded in natural speech to foreign sounding interlocutors. These observations seem to suggest that at quiet and in noise vowel hyperarticulation actually assists listening comprehensibility for both L1 listeners, and early and late L2 learners of English.

This finding therefore appears to suggest that L2 listeners' recognition of words in English can be supported through vowel hyperarticulation. This result supports previous findings according to which vowel hyperarticulation was proposed to likely lead to increased comprehensibility of speech (Ferguson &Kewley-Port, 2007). The result also seems to confirm that vowel hyperarticulation, if elicited in a communicative setting, can lead to improved speech comprehensibility of words (Ferguson & Kewley-Port, 2007). The present experiment can therefore be seen as an extension to studies that showed that clear vowel hyperarticulated speech can lead to higher speech intelligibility (Bond & Moore, 1994; Bradlow et al., 1996; Byrd, 1994; Johnson et al, 1993; Hazan & Markham, 2004). The observation that speech to foreign sounding interlocutors was easier to understand than speech to native sounding interlocutors at quiet and at different noise levels support the H&H theory according to which adults modify their speech to maximise discriminability in order to provide the listener with sufficient information to make speech comprehension possible (Lindblom, 1992).

However, Experiment 5 does not uphold prior research that reported early L2 listeners have higher noise-tolerance levels than late L2 listeners (Mayo et al., 1997). This can therefore be considered to disagree with the previous finding that when their speech perception in L2 was interrupted by noise, late L2 listeners' perception of speech in L2 was more affected than their speech perception in L1 (Florentine, 1985a, b; Garcia Lecumberri & Cooke, 2006; Mayo et al., 1997; Takata & Nábělek, 1990). This lacking finding of a higher speech comprehension benefit for early L2 learners than late L2 listeners might have been due to the limited nature of the task of the present experiment as it employed a listening comprehension task (Munro & Derwing, 1999). The absent finding of higher noise-tolerance levels for early L2 listeners might therefore be accounted for by the limited speech material available and the simplicity of target words used.

Another reason for the absent finding of a higher speech comprehension benefit for early L2 learners than late L2 listeners might be due to the confound of the length of experience using the language between the early and late L2 learner groups in the L2 country. Consequently, even if early and late L2 learners might have started L2 acquisition at a different age, the difference in the length of their exposure to L2 might have contributed to this result. However, it can be argued that this confound is inevitable because even if early and late l2 learners are matched for the length of experience using L2 and differ in age of L2 acquisition, early L2 learners might have been exposed to more L2 when watching news or television programs in L2 in their native country compared to late L2 learners, and vice versa. Nonetheless, this result can be used by future experiments investigating the effect of age of L2 acquisition on L2 learners' performance on L2 comprehension tasks to look at additional factors that might lead to a difference in their performance between early and late L2 learners such as L2 learners' reported percentage use of L1 and the number of speakers they interact with in L1 on a regular basis (Flege & MacKay, 2004).

Similarly, the data did not show that early L2 learners' comprehensibility was lower than that of L1 English speakers. This is inconsistent with previous research in which L1 listeners were reported to recover more quickly than early non-native L2 listeners from adverse listening conditions due to their established L1 categories (Mayo et al., 1997). In addition, the data cannot be suggested to support previous research that showed that L1 listeners experienced a perceptually higher benefit than late L2 learners because of late L2 learners' insufficient experience in the L2 sound structure (Bergman, 1990). Thus, it cannot be argued that late L2 learners' L1 might have affected their performance in the listening comprehensibility rating in that they may have perceptually assimilated incoming L2 phonemes to L1 categories (Best & Tyler, 2007). Nonetheless, it has been suggested that late L2 learning does not prevent the perception of L2 vowels that functionally is similar to native like perception of L2 vowels (Flege & MacKay, 2004).

Finally, since long stimuli length was reported by native speakers in speech to foreign sounding interlocutors compared to native sounding interlocutors in Experiment 1 (Chapter 4), it can be argued that long vowel length or long word length might have contributed to the increased comprehension of speech to foreign sounding interlocutors than native sounding interlocutors. However, to tease out the effect of vowel length, another study would be necessary, in which the effect of long vowel duration on native listeners' comprehension is tested without vowel space expansion. Regarding word length, one has to note that word duration in the target words tested includes both vowel and consonants. Since Experiment 1 did not control for the type of consonant that come before and after the vowels in the words, it is not known to what extent at word level consonants might have played a role in long word length. This implies that depending on what level of speech is analysed (vowel, word, sentence, paragraph level), the contribution to comprehension of specific acoustic features could change (Boothroyd & Nittrouer, 1988).

Therefore, the findings of Experiment 5 seem to confirm the role of speech to foreign sounding interlocutors to be of didactic benefit. The findings of Experiment 5 therefore suggest that vowel space expansion as it is used together with other acoustic-phonetic features in speech to foreign sounding listeners might be useful to be employed in linguistic training programs to facilitate foreign sounding listeners' comprehension of the target language.

In conclusion, Chapter 7 addressed the research question what effect vowel space expansion has on L1 and L2 listeners' comprehensibility of speech. Across all listener groups (early L2 learners of English, late L2 learners of English and L1 English speakers) speech to foreign sounding interlocutors was easier to understand than speech to native sounding interlocutors at both quiet and all noise levels. It therefore seems that vowel hyperarticulation used together with other acoustic-phonetic features in speech to foreign sounding listeners has an enhancing effect on the comprehensibility in FDS. This chapter thus provided evidence for vowel hyperarticulation being a linguistic instrument that might be used for didactic purposes.

Chapter 8

General discussion and conclusion

Numerous studies have demonstrated that clear speech is more intelligible than conversational or casual speech for hearing-impaired individuals in quiet environments (Picheny et al., 1985), for young normal-hearing individuals in noisy environments (Bradlow & Bent, 2002; Krause & Braida, 2002), in reverberating settings (Payton et al., 1994), in noisy environments with reverberation (Payton et al., 1994) and for elderly normal-hearing individuals in noisy settings (Helfer, 1998). Compared to conversational speech, clear speech has also been reported to be more intelligible for users of cochlear implants (Liu et al., 2004), and auditory neuropathy patients whose transmission of sound information stream is interrupted (Zeng & Liu, 2006). These studies show that for individuals with and without hearing-impairments, clear speech seems to give listeners an advantage with regards to intelligibility (Uchanski, 2005).

The present thesis sought to focus on the possible role in communication of one acoustic modification of clear speech: expanded vowel space. This thesis investigated the speech production in English by British English native speakers in interaction with interlocutors with varying levels of looking and sounding foreign. The thesis assessed speech perception by both non-native learners of L2 English and British English native speakers, emphasising the effects of proficiency and environmental noise. More specifically, the thesis addressed the extent and ease of identifying naturally elicited vowel hyperarticulated speech by native listeners compared to conversational speech. It also investigated the extent to which this naturally elicited clear speech can be understood at word level by British English native speakers (L1 listeners) and Asian learners with English as their second language (L2 listeners). Possible effects of non-native speakers' varying proficiency in English on understanding speech was tested taking into consideration early non-native learners' extensive experience with the L2 sound structure compared to late non-native learners' low proficiency in L2. In addition, this thesis investigated whether naturally elicited vowel hyperarticulated speech provides an equally perceptual benefit to native and non-native listeners when presented under varying levels of background noise. The thesis also discussed the role of vowel hyperarticulation in interaction with foreign-accented listeners in light of the influential framework on speech production and perception that is the H&H theory, as well as leading learning models of L2 acquisition.

8.1 Speech production by British native speakers in interaction with interlocutors with varying level of looking and sounding foreign

Speech is modified according to whom the speaker is conversing with (Burnham et al., 2002, 2010). Adult-directed speech (ADS) is acoustically different from speech directed to foreigners (FDS) and infants (IDS) since vowels were reported to include expanded vowel space in FDS and IDS but not in ADS (Kuhl et al., 1997; Uther et al., 2007). The speech modification of hyperarticulating vowels has therefore been suggested to be of didactic use since vowels were exaggerated for infants but not for pets (Burnham et al., 2002; Kuhl et al., 1997; Uther et al., 2007; Xu et al., 2004). As past research on different speech registers used with foreigners has focused on those with both a foreign appearance and foreign accent (Littleford et al., 2005; Uther et al., 2007), it is not known whether native speakers' speech at native looking and foreign sounding individuals, on the one hand, is different from their speech at foreign looking and native sounding individuals, on the other hand.

Experiment 1 in Chapter 4 of this thesis extended the study by Uther et al (2007) by dissociating foreigners' accent from their physical appearance and by finding out whether foreign interlocutors' physical appearance or accent most influence speech registers. This was made possible by using four kinds of interlocutors: those who both look and sound native, those who appear foreign but linguistically sound like native speakers, those who appear native but sound foreign and those who both look and sound foreign. It was clear that in speech to foreign-accented foreigners, vowels were acoustically hyperarticulated irrespective of whether their appearance was native or foreign. It was observed that vowel length and word length were significantly larger for foreign sounding interlocutors than native sounding interlocutors. These observations are in line with Lindblom's H&H theory (1990), according to which speakers segmentally adapt speech output depending on listeners' needs within a communicative context.

Although it seems that the findings from Experiment 1 seem to also broadly agree with the CAT notion of linguistic convergence that native speakers show speech convergence in interaction with non-native speakers, and also with the audience design view, according to which speakers change their speaking style to different audiences respectively, it can be said that because the research question in Experiment 1 was not investigated from a sociopsychological or sociolinguistic standpoint, the findings are on a linguistic level better accounted for by the H&H theory.

Experiment 1 suggested there were no differences in the manner in which native British English speakers expressed pitch at vowel level towards foreign and native sounding interlocutors regardless of appearance. This is consistent with prior research in which no differences in pitch were observed in adult native speakers' communication with foreigners or native listeners (Uther et al., 2007), and native speakers' imaginary interactions with foreigners (Knoll et al., 2009). The first experiment also suggested that at vowel and word level, native speakers increase intensity if the interlocutor looks foreign irrespective of accent. This indicates, in line with past research (Sommers et al., 1994), that intensity is not an essential speech characteristic that is required to be modified for non-native listeners' speech perception.

Considering speech to native-sounding interlocutors as conversational speech, and speech to foreign-sounding interlocutors as clear speech, the finding on the occurrence of expanded vowel space is comparable to studies that showed that in clear speech, $F1$ and $F2$ came close to target values and were grouped together more closely than in conversational speech, indicating that tightly grouped single vowel categories might thereby be less likely confused across categories (Chen, 1980). Consequently, compared to conversational speech clear speech displayed a larger vowel formant space and diverged less from the anticipated formant targets (Johnson, Flemming, & Wright, 1993; Lindblom, 1990; Moon & Lindblom, 1994). Thus, overall it appears that in line with the H&H theory, FDS is an adaptive and instructive tool, and that vowel hyperarticulation is a didactic device that is elicited in speech to foreign-accented interlocutors regardless of appearance.

Broadly speaking, the findings from Experiment 1 can also be considered to agree with the CAT perspective and with audience design view, according to which speech to non-native interlocutors native speakers show convergence in speech, and speakers change their speaking style to different audiences respectively. However, because the research question in Experiment 1 was not investigated from a socio-psychological or socio-linguistic standpoint, the findings are on a linguistic level better accounted for by the H&H theory.

8.2 Intelligibility and comprehension of speech with and without expanded vowel space

Considering Ferguson and Kewley-Port (2002)'s suggestion that the stationary steady-state portion of the vowel assists intelligibility at vowel level, Experiment 3 in Chapter 6 of this thesis aimed at exploring the possible contribution of vowel hyperarticulation to intelligibility at word level. Naturally elicited speech towards foreign-accented interlocutors and native-accented interlocutors that had been collected from Experiment 1 were used in the perceptual rating tasks in Experiment 3. These tasks included an orthographic transcription task with a confidence rating, as well as a typicality rating and clarity rating.

It was clear that mean transcription accuracy was higher for words from speech to foreign-accented interlocutors than for words from speech to native-accented interlocutors although listeners did not show higher confidence in their transcription of speech to foreign-accented than of speech to native-accented interlocutors. Words from speech to foreign-accented interlocutors were rated as more typical of speech in English and as clearer than words from speech to native-accented interlocutors. The result of the confidence rating indicated a mismatch between native listeners' perception of their performance (confidence rating) and their actual performance on the intelligibility of stimuli with and without vowel space expansion (transcription, typicality, clarity task).

In addition, the clarity rating result seems to suggest that vowel hyperarticulation causally contributes to improved intelligibility of speech at word level for native normal-hearing listeners. This result appears to expand previous studies that correlated increased vowel space with speech intelligibility (Bond & Moore, 1994; Bradlow et al., 1996; Hazan & Markham, 2004; Picheny et al., 1986). This finding also seems to be in line with the H&H theory according to which hyperarticulated speech allows more accurate phonetic units to be more easily perceived as being acoustically distinct (Lindblom, 1990)and according to which adults modify their speech to generate a speech signal that maximises discriminability and therefore intelligibility (Ferguson, 1977; Kuhl et al., 1997; Lindblom, 1990; Liu et al., 2003; Song et al., 2010; Uther et al., 2007).

However, because an additional statistical analysis did not show expanded vowel space to be the leading acoustic characteristic underlying the higher clarity ratings of speech directed to foreign-sounding interlocutors compared to native-sounding interlocutors, it cannot be concluded with certainty that vowel space expansion is the sole cause for the higher clarity ratings. It may be that other acoustic features might also play a role, which would need to be explored in an additional experiment. Nevertheless, the finding of Experiment 3 can be considered to suggest that expanded vowel space is involved in improving the intelligibility of speech to foreign-sounding interlocutors.

Experiment 5 in Chapter 7 of this thesis expanded on the results of Experiment 3 by investigating whether expanded vowel formant space in clear speech compared to conversational speech directly improves both native listeners' and non-native listeners' comprehension of spoken speech. The listeners included native speakers as L1 listeners, early non-native learners as early L2 listeners, and late non-native learners as late L2 listeners. Naturally elicited speech towards foreign-accented interlocutors and native-accented interlocutors collected from Experiment 1 were presented at quiet and at varying noise levels.

It was clear that across all three listener groups, speech to foreign-accented listeners was easier to understand than to native-accented listeners. Moreover, Experiment 5 showed that at both quiet and background noise speech with expanded vowel space was easier to understand than speech without expanded vowel space. Native and non-native listeners did not differ in their speech comprehension performance. This result seems to demonstrate for the first time the benefit of articulatory expanded vowel space on listeners' comprehension of spoken words, which supports the didactic role of this speech modification for language learners of English as L2 (Kuhl et al., 1997; Uther et al., 2007). This finding also appears to imply that speech to foreign-accented listeners was modified to a certain extent through vowel hyperarticulation, through which comprehensibility of speech to foreign-accented listeners might have increased compared to that of speech to native-accented listeners. Although this experiment did not look at the perceptual effectiveness of other acoustic features that might operate together with vowel space expansion such as stimuli length to contribute to increased comprehensibility, the finding that speech with expanded vowel space improves comprehensibility in speech is in line with the H&H theory (Ferguson, 1977; Kuhl et al., 1997; Liu et al., 2003; Song et al., 2010; Uther et al., 2007).

The finding of no differences in performance between listener groups is inconsistent with prior studies that proposed that due to differences in the levels of exposure to the L2 sound structure, early L2 listeners will experience a larger speech comprehension benefit than late L2 listeners from speech with expanded vowel space, and a lower benefit than L1 listeners (Bradlow & Bent, 2002; Garcia Lecumberri & Cooke, 2006; Mayo et al., 1997; Smiljanić & Bradlow, 2011; Takata & Nábělek, 1990). A reason for the lacking higher speech comprehension benefit for early L2 learners than late L2 listeners may be due to the confound of the length of experience using the language between the early and late L2 learner groups in the L2 country.

Nonetheless, the observations from Experiment 5 are in line with research that suggested that early learning of L2 does not mean that it will lead to native like L2 vowel perception. At the same time, late L2 learning does not prevent the perception of L2 vowels that functionally is similar to native like perception of L2 vowels (Flege & MacKay, 2004). Thus, the findings from Experiment 5 on how non-native English speakers with varying proficiencies process speech with clear speech features such as expanded vowel space compared to native speakers can be most suitably accounted for by Flege’s Speech Learning Model (SLM) (Flege, 2002, 2003a; Flege & Eefting, 1986; Flege & Hillenbrand, 1987; Flege & Mackay, 2004).

8.3 Speech production by mothers as British native speakers in read speech and in interaction with infants

It has been argued that IDS has features that may benefit the listener, particularly in language development (Kuhl et al., 1997). Generally, IDS has been proposed to serve three main purposes, which are of affective, attentional and didactic nature (Cooper & Aslin, 1994; Cooper, Abraham, Berman, & Staska, 1997; Fernald & Kuhl, 1987; Grieser & Kuhl, 1988; Uther et al., 2007; Werker & McLeod, 1989). Accordingly, IDS might draw and sustain infants' attention (Cooper & Aslin, 1994; Cooper et al., 1997; Grieser & Kuhl, 1988; Werker & McLeod, 1989); it might express positive vocal affect (Werker & Leopold, 1989; Uther et al., 2007; Cooper et al., 1997; Fernald & Kuhl, 1987); and, it may support infants' language learning process (Burnham et al., 2002; Kuhl et al., 1997; Uther et al., 2007). Experiment 2 in Chapter 5 of this thesis focused on the didactic aspect by attending to the acoustic properties of IDS that have been proposed to contribute to speech intelligibility (Kai et al., 2008; Kuhl, 2004; Liu et al., 2003).

Specifically, it compared the acoustic-phonetic features in IDS with those of clear read speech. Eleven White British mothers communicated with their infants for about ten minutes by showing them picture cards with everyday objects on them and then read out a brief list of sentences in a clear speaking voice. It was found that vowels were hyperarticulated in IDS compared to read speech. This finding is in line with prior studies (Kuhl 2000b; 2004; Kuhl et al., 2008; Liu et al., 2003). Experiment 2 also showed that IDS had higher pitch and a broader pitch range compared to read speech, supporting previous findings (Fernald & Kuhl, 1987; Burnham et al., 2002). In terms of mean intensity, there was no difference between IDS and read speech. There were also longer vowel length and longer word length in IDS compared to in read speech. Thus, by demonstrating exaggerations in vowel space and stimuli length that are present in IDS and absent in clear read speech, Experiment 2 of this thesis extended prior studies (Burnham, et al., 2002; Uther et al., 2007) that had indicated that vowel hyperarticulation in IDS might be didactic in nature. It also supports the H&H theory by confirming IDS as a listener-oriented speech style, and it supports the Kuhl's NLM Theory (enhanced) (Kuhl et al., 2008), which suggests vowel space expansion in IDS has a didactic utility (Kuhl 2000b; 2004; Kuhl et al., 2008).

These findings can be considered to be ascribed to the very instructive nature of the interaction task, to the absence of infants during the reading task, and the age range of infants which was between 12-20 months, by which mothers might have become used to talking to their infants in a manner that helps infants' development of phonetic categories.

8.4 Speech intelligibility of IDS versus read speech for British native speakers

Prior studies suggested hyperarticulated speech to enable listeners to more easily perceive phonetic units as being acoustically distinct (Burnham et al., 2002; Lindblom, 1990; Uther et al., 2007). Experiment 4 in Chapter 7 of this thesis therefore investigated whether IDS or read speech would lead native listeners to experience perceptually better intelligibility of speech. Tasks consisted of an orthographic transcription task with a confidence rating, a typicality rating task, and a clarity rating task.

It was clear that more words were correctly identified in IDS than in mothers' read speech (Kuhl et al., 1997). Native listeners also felt more confident in the accuracy of their transcription of IDS compared to that of read speech, though they did not rate IDS as more typical of the words in English that they considered them to be compared to read speech. Thus, these results indicate that the hyperarticulated nature of IDS might have helped listeners become confident in their transcription of IDS, as previously hypothesised. The observation that listeners did not rate words in IDS as typical of the English words that they consider them to be might be explained by the fact that IDS was of high pitch and large pitch range. These acoustic characteristics are, however, not usually part of speech between native speakers. In support of prior theory (Kuhl's NML-enhanced (Kuhl et al., 2008), and studies (Kuhl et al., 1997; Uther et al., 2007), words from IDS were rated clearer than words from mother's read speech.

Thus, by indicating that hyperarticulated speech in IDS enabled listeners to perceive phonetic units more easily as acoustically distinct than clear read speech, these findings are in line with previous research (Burnham et al., 2002; Uther et al. 2007). The finding of IDS samples being clearer than mothers' read speech is also in line with Kuhl's perceptual Native Language Magnet Theory (NLM) (enhanced) theory, according to which overemphasised crucial phonetic differences in IDS contribute to improved L1 phonetic perception (Kuhl et al., 2008). Thus, while prior evidence for vowel hyperarticulation as a didactic feature was largely circumstantial or not clear-cut (Liu et al., 2003; Song et al., 2010), Experiment 4 of this thesis seems to demonstrate that vowel hyperarticulation in IDS has a didactic function compared to other types of clear speech that have no communicative intent such as read speech. Nonetheless, because an additional statistical analysis did not show expanded vowel space to be the leading acoustic characteristic underlying the higher clarity ratings of infant-directed speech compared to read speech, one cannot be completely sure that vowel space expansion solely led to to the higher clarity ratings of speech to infants. Because other acoustic features might also have played a role, this would need to be explored in an additional experiment. Nevertheless, the finding of Experiment 4 can be considered to suggest that expanded vowel space is involved in improving the intelligibility of speech to infants.

A summary of the originality, hypotheses, main findings, supported models and implications for all experiments is presented in Tables 8.1-8.5.

Table 8.1 Summary of Experiment 1 in Chapter 4.

Originality
• Differentiation between foreigners' appearance and accent in FDS (different from Burnham et al., 2002; Uther et al., 2007). • Use of natural interactions (similar to Leubecker & Bohannon III, 1982; Uther et al., 2007).
Hypotheses
• There would be hyperarticulation in speech to foreign-accented interlocutors compared to speech to native speakers, and no effect of appearance. • There would be no difference in pitch to interlocutors regardless of accent and appearance. • There would be higher intensity in speech to foreign-accented interlocutors regardless of appearance compared to native-accented interlocutors.
Main findings
• Vowel space was significantly larger in speech to foreign-accented interlocutors than native-accented interlocutors, with no effect of appearance. • There was no difference in pitch to interlocutors across different accent and appearance conditions. • There was increased intensity in speech to foreign looking interlocutors.

Models supported
• H&H theory (Lindblom, 1990).
Implications for interaction with foreigners
• Vowel hyperarticulation is a didactic device that is elicited in speech to foreign-accented interlocutors regardless of their appearance.

Table 8.2 Summary of Experiment 2 in Chapter 5.

Originality
• Comparison of acoustical modifications in IDS (with communicative intent) with those in read speech (without communicative intent) as another type of clear speech (different from Burnham et al., 2002; Uther et al., 2007; similar to Hazan & Baker, 2011 who however focused on ADS). • Use of natural interactions (similar to Burnham et al., 2002; Hazan & Baker, 2011; Uther et al., 2007).
Hypotheses
• IDS would be hyperarticulated compared to read speech due to lack of communicative intent in read speech. • There would be higher pitch and pitch range in IDS than in read speech because in read speech there is no interlocutor whose attention has to be maintained. • There would be no differences in intensity between IDS and read speech.
Main findings
• Vowel space was significantly larger in IDS than read speech. • Mean pitch and pitch range were larger in IDS than read speech. • There were no differences in intensity between IDS and read speech.
Models supported
• H&H theory (Lindblom, 1990); Kuhl's Native Language Magnet Theory (enhanced) (Kuhl et al., 2008).
Implications for interaction with infants
• Vowel hyperarticulation demonstrated in IDS of British Southeast English (West London region) is a uniquely didactically-oriented speech style compared to read speech.

Table 8.3 Summary of Experiment 3 in Chapter 6.

Originality

• Use of naturally elicited clear speech towards foreign-accented and native-accented interlocutors in perceptual rating tasks for native listeners.
Hypotheses
• Native listeners would transcribe words from speech to foreign-accented interlocutors more correctly than from speech to native-accented interlocutors. • Native listeners would be more confident in their transcription of words from speech to foreign-accented interlocutors than native-accented interlocutors. • Native listeners would rate words from speech to foreign-accented interlocutors as more typical of speech in English and as clearer than words from speech to native-accented interlocutors.
Main findings
• Mean transcription accuracy was higher for words from speech to foreign-accented than native-accented interlocutors. • Confidence in transcription between words from speech to foreign-accented and native-accented interlocutors did not differ. • Words from speech to foreign-accented interlocutors were rated clearer and as more typical of speech in English than those from speech to native-accented interlocutors.
Models supported
• H&H theory (Lindblom, 1990).
Implications for foreigners as second-language learners of English compared to native listeners
• Vowel hyperarticulation as elicited in speech to foreign sounding interlocutors causally contributes to improved intelligibility of words for native normal-hearing listeners.

Table 8.4 Summary of Experiment 4 in Chapter 6.

Originality
• Use of naturally elicited clear speech towards infants in perceptual rating tasks.
Hypotheses
• Native listeners would transcribe words from IDS more correctly than those from read speech. • Native listeners would be more confident in the transcription of words from IDS than from read speech. • Native listeners would rate words from IDS as more typical of speech in English and as clearer than words from read speech.

Main findings
• Mean transcription accuracy and confidence in transcription were higher for words from IDS than read speech. • Confidence in transcription was higher for words from IDS than read speech. Words from IDS were not rated as more typical of speech in English than words from read speech. Words from IDS were rated clearer than words from read speech. Words from IDS were rated clearer than words from read speech.
Models supported
• Kuhl's Native Language Magnet Theory (enhanced) (Kuhl et al., 2008).
Implications for foreigners as second-language learners of English compared to native listeners
• Vowel hyperarticulation as elicited in IDS is involved in contributing to improved intelligibility of words for young native normal-hearing listeners.

Table 8.5 Summary of Experiment 5 in Chapter 7.

Originality
• Use of naturally elicited clear speech towards foreign-accented interlocutors in speech comprehension tasks at quiet and adverse listening conditions for native L1 listeners and non-native L2 listeners with varying proficiency of English.
Hypotheses
• In quiet, L1 listeners, early and late L2 listeners would find stimuli with vowel space expansion easier to understand than stimuli without vowel space expansion. • In noise, L1 listeners, early and late L2 listeners would find stimuli with vowel space expansion more comprehensible than stimuli without vowel space expansion.
Main findings
• Speech to foreign sounding interlocutors was easier to understand than speech to native sounding interlocutors across listener groups. • Speech at quiet and low noise levels was easier to understand than speech at high noise levels across listener groups • At both quiet and noise speech with stretched vowel space was easier to understand than speech without stretched vowel space across listener groups.

• There were no differences between listener groups in rating speech at quiet or at noise.
Models supported
• H&H theory (Lindblom, 1990); Flege's Speech Learning Model (SLM) (Flege & MacKay, 2004).
Implications for foreigners as second-language learners of English compared to native listeners
Vowel hyperarticulation • a listener-oriented speech modification. • improves L1 and L2 listeners' speech perception in quiet and noise.

8.5 What drives hyperarticulation in speech?

The Experiments 1 and 2 discussed in this thesis seem to suggest in line with the H&H theory that in interactions between native speakers and language learners, speech is hyperarticulated at the segmental level. The first experiment, for instance, demonstrated that in interactions between native speakers and foreign-accented non-native listeners who solve a problem-solving task together, speech includes vowel hyperarticulation. Thus, it appears that at the beginning of an interaction between a native speaker and foreign-sounding interlocutor, the communication will linguistically be focused on the assessment of the non-native interloctor's linguistic competence so that native speakers can evaluate to what extent they can talk to non-native interlocutors in a manner that both ensures the comprehension of their speech and matches non-native interlocutors' level of linguistic ability. This seems to be especially the case if both communicators have not met before and do not know each other on a personal level.

When such an interaction occurs in context of a problem-solving task that does not require the communicators to accentuate or attenuate their social identity, and that does not highlight social context as triggers for variation in speech, it seems that the results from Experiments 1 and 2 are more effectively explained by the H&H theory than by sociolinguistic and psycholinguistic theories such as CAT and audience design. This is because the H&H theory is a phonetic theory that can directly explain hyperarticulated speech with a didactic purpose aimed at language learners. As a result, in line with the requirements to solve the task, non-native listeners would be concerned with finding the differences between the pictures, and therefore be focused on understanding the native speaker's speech to the best of their ability.

One could also argue that sociolinguistic factors of convergence might explain native speakers' observed hyperarticulation towards non-native interlocutors because in addition to sharing social identities (e.g. being students and in the same age group), they both shared the goal of finding the differences between the pictures together (Zuengler, 1991). Accordingly, if native and non-native speakers have social characteristics in common, non-native speakers are regarded to become similar to native speakers in their use of language (Giles & Johnson, 1987; Young, 1988). As a result one would expect to observe linguistic convergence between native speakers and non-native interlocutors or the use of downward convergence in form of informal language with each other (Giles & Smith, 1979).

Interlocutors would perceive each other as more attractive, and would feel more of a sense of agreement when they become progressively similar in communicative aspects such as speech rate, which can consequently lead to psychological convergence (Buller, LePoire, Aune, & Eloy, 1992). That native speakers' adaptations in convergence can range from very small to nearly identical to that of non-native interlocutors' speech, was supported by evidence that foreigner-talk can include interlocutors' errors in speech (Snow et al., 1981).

Thus, since sociolinguistic ideas of accommodation may play a role in communication between native speakers and non-native interlocutors, and because Experiment 1 did not directly test these ideas, further experiments are required to test these sociolinguistic notions. Qualitative methods such as thematic analysis would help dissect native speakers' variations in speech in terms of their social meanings by developing theory-driven codes that can be used as labels that describe different sections of their speech such as native speakers' intergroup and interpersonal connections (Bell & Johnson 1997; Boyatzis, 1998). Thus, it can be argued that the H&H theory presently provides the most appropriate explanation of the data discussed in Experiments 1 and 2.

Experiments 3-5 suggest that vowel space expansion as it occured together with other acoustic-phonetic features in naturally elicited speech to language learners seems to be important in contributing to improved speech intelligibility and comprehension for both native and non-native listeners. In experiments 3 and 5, for instance, speech to foreign sounding listeners was observed to be hyperarticulated, which according to the H&H theory maximizes clarity in speech (Lindblom, 1990). Since sociolinguistic and psycholinguistic theories mainly focus on variations in speakers' speech style that are elicited in and interpreted using social contexts, it is evident that the H&H theory seems to theoretically justify most appropriately the outcomes of improved speech clarity and comprehension in Experiments 3-5.

Thus, it can be argued that the use of segmental variation in speech and its effect on perception is naturally better explained by phonetic theories such as the H&H theory because the thesis explores the linguistic benefit of vowel hyperarticulation in speech to language learners while the sociolinguistic and psycholinguistic theories do not address speech variation at the segmental level but explore the communicators' relationship with each other more closely at the social and psychological level in terms of the speech produced.

8.6 Limitations and future research

It can be said that according to Experiments 3, 4 and 5, vowel hyperarticulation seems to be a beneficial feature at the segmental level to increase the intelligibility and understanding of words. One limitation of the present research in this thesis is that it only focused on the possible linguistic role of expanded vowel space in improving speech perception. Although long vowel duration was found in clear speech production in Experiments 1 and 2, it is not clear to what extent vowel duration contributed to intelligibility. A future experiment can therefore investigate this issue by separating long vowel duration from expanded vowel space in speech to explore its effect on listeners' speech comprehension.

Similarly, although increased monosyllabic word duration was found in clear speech production in Experiments 1 and 2, it is not clear to what extent long word duration contributed to intelligibility at word level. Prior research only indicated long word duration to be positively related to intelligibility at word level (Hazan & Markham, 2004). However, since the target words used in Experiments 3-5 included both vowels and consonants, and because they did not control for the consonants coming before and after the vowels, it is not known to what extent consonants might have played a role in causing long word length. Therefore a future experiment can investigate the role of long word length in improving listeners' speech comprehension by controlling for consonants and by separating long word duration from expanded vowel space in speech comprehension tasks.

Nonetheless, one has to consider the possibility that acoustic features that can be used to increase intelligibility of isolated words such as vowel hyperarticulation might not be useful to improve intelligibility of the same words at sentence level. Moreover, expanded vowel space might not be sufficient on its own to be useful to native listeners and non-native listeners to increase their understanding at sentence level. This is because other speech aspects such as speech rate might affect intelligibility at sentence level, which, however, cannot be compared to speech rate that originates from the duration of single words. Nevertheless, previous research has not found speech rate to correlate with intelligibility at sentence level (Bradlow et al., 1996).
One also has to note that the importance of some acoustic characteristics can vary depending on what recognition level (e.g. syllable, word, sentences) intelligibility is investigated. In addition, one needs to consider the possibility that differences in intelligibility at word and sentence level can be affected by attributes of different types of speakers such as speakers with dysarthria (Dongili, 1994; Liu et al., 2005; Yorkston & Beukelman, 1978) or late learners of English (Bradlow & Smiljanic, 2011).

Another limitation is that the present thesis only looked at the acoustic characteristics that might be useful for the recognition of words for native and non-native listeners. However, it is clear that the importance of acoustic features that are essential for word intelligibility can differ across listeners such as hearing-impaired adults or elderly adults with hearing impairments (Cervera, Soler, Dasi, & Ruiz, 2009; Gordon-Salant, 1986). Similarly, a further limitation is that IDS samples in Experiment 4 were not tested with infants as listeners but with native listeners. Therefore future research can look at how expanded vowel space in naturally elicited IDS contributes to word intelligibility for infants. The main limitations of the research in this thesis can therefore be considered to include the limited nature of stimuli, the recruitment of native speakers and native and non-native listeners, and the focus on one clear speech modification at the segmental level.

In conclusion, the research presented in this thesis has demonstrated the hyperarticulatory nature of speech directed at foreigners and infants. It has highlighted the role of vowel hyperarticulation as a listener-oriented speech style that is used in speech to foreign-accented speakers irrespective of their appearance. It has shown that, in infant-directed speech, vowel hyperarticulation is uniquely didactic compared to other forms of clear speech such as clear read speech. Moreover, the research has shown that vowel hyperarticulation, as it occurs with other acoustic-phonetic features in speech to foreign-accented speakers and infants, improves native listeners' intelligibility of spoken speech and enhances the listening comprehension of native listeners and non-native listeners with varying proficiency in English under ideal and adverse listening conditions.

Future research should investigate the role of other acoustic features of clear speech such as stimuli length to find out to what extent they contribute to increased intelligibility and comprehension together with and separately from vowel hyperarticulation at different levels of speech (at the sentence, paragraph, and narrative level). It should investigate their utility in enhancing speech understanding for different types of listeners such as non-native learners of English, infants and hearing-impaired listeners. The resulting findings would have implications for facilitating non-native learners' acquisition of English in linguistic training programs and for enhancing pedagogic strategies that support the language acquisition of infants as first language learners.

References

Alexander, R. M. (1988). *Elastic mechanisms in animal movement*. Cambridge: Cambridge U.P.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Allport, G. W. (1954). *The nature of prejudice*. Cambridge: M.A.: Addison-Wesley.

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants ' and parents' utterances in longitudinal recordings. *Journal of Acoustical Society of America, 119*, 1636-1647.

Andruski, J. E., & Kuhl, P. K. (1996). Acoustic structure of vowels in mothers' speech to infants and adults. *Proceedings of the 1996 International Conference on Spoken Language Processing, ICSLP. Part 1 (of 4), Philadelphia, PA, USA, 3* 1545-1548.

Andruski J.E., Kuhl P.K. & Hayashi, A. (1999). The acoustics of vowels in Japanese women's speech to infants and adults. *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, California, USA. 3, 2177-2181.

Anisfield, M., Bogo, N., & Lambert, W. (1962). Evaluational reactions to accented English speech. *Journal of Abnormal Social Psychology, 69*, 89-97.

Ashby, S. (2004). *Is infant-directed speech hyperspeech? An acoustical analysis of speech to infants and other accommodative speech styles.* PhD Thesis, University College Dublin, Ireland.

Baker, R., & Hazan, V. (2009). Acoustic-phonetic characteristics of naturally elicited clear speech in British English. *157th Meeting of the Acoustical Society of America*, Portland, Oregon.

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1; 2. *Journal of Child Language, 32*, 159–173.

Benade, A. H. (1976). *Fundamentals of musical acoustics*. New York: Oxford University Press.

Bergman, M. (Ed.). (1980). *Aging and the perception of speech*. Baltimore: University Park Press.

Bernstein-Ratner, N. (1984). Patterns of vowel modification contained by mother-child speech. *Journal of Child Language*, *11,* 557-578.

Best, C. T. (1989). Phonologic and phonetic factors in the influence of the language environment on speech perception. *International Conference on Event Perception and Action* Miami University, Miami, OH.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window om early phonological deveopment. In B. Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year*. Dordrecht, The Netherlands: Kluwer Academic.

Best, C. T. (1994b). Learning to perceive the sound pattern of English. In C. Rovee-Collier, & L. Lipsitt (Eds.), *Advances in infancy research* (pp. 217-304). Hillsdale, NJ.: Ablex Publishers.

Best, C. T. (1994a). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In H. Nusbaum (Ed.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception In W. Strange (Ed.), *Speech perception and linguistics experience: Issues in crosslanguage research* (pp. 171–204). York, Baltimore:

Best, C. T. (1999). Development of language-specific influences on speech percepyion and production in pre-verbal infancy. *International Conference of Phonetic Sciences*, San Francisco. 1261-1264.

Best, C. T., & Avery, R. A. (1999). Left-hemispehere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science, 10*, 65--69.

Best, C. T., Faber, A., & Levitt, A. G. (1996). Perceptual assimilation of non-native vowel contrasts to the American English vowel system. *The Journal of the Acoustical Society of America, 99*, 2602.

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech, 46*, 183-216.

Best, C. T., McRoberts, G. W., & Goodell, E. (2000). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *Journal of the Acoustical Society of America, 109*, 775-794.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). American listeners' perception of non-native consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America*, *109*, 775-794.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two non-native consonant contrasts. *Infant Behaviour and Development*, *18*, 339-350.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganisation for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Biersack, S., Kempe, V., & Knapton, L. (2005). Fine-tuning speech registers: A Comparison of the prosadic features of child- directed and Foreigner- directed speech. *Proceedings of the 9Th European Conference on Speech Communication and Technology,* Lisbon.

Billing, C.J., Tremblay, K.L., Stecker, G.C., Tolin, W.M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, *254*, 15-24.

Blair, I. V., Judd, C. M., Havranek, E. P., & Steiner, J. F. (2010). Using community data to test the discriminant validity of ethnic/racial group IATs. Zeitschrift fu¨r Psychologie/Journal of Psychology, 218, 36–43.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41.

Bond, Z. S., Moore, T. J., & Gable, B. (1989). Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *Journal of the Acoustical Society of America*, *85*, 907–912.

Bond, Z. S., Moore, T. J., 1994. A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, *14*, 325–337.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America*, *84*, 101–114.

Bowen, C. Speech-language-therapy dot com Caroline Bowen PhD Speech-Language Pathologist. Retrieved January, 2014, from h http://www.speech-languagetherapy.com/index.php?option=com_content&view=article&id=29:admin&catid=11:admin&Itemid=117.

Boyatzis, R.E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, London, & New Delhi: SAGE Publications.

Bradlow, A. R., & Alexander, J. (2007). Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of Acoustical Society of America*, *112*, 2339–49.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of Acoustical Society of America, 112*, 272–284.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech. I. global and fine-rained acoustic-phonetic talker characteristics. *Speech Communication*, *20*, 255–272.

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research, 46,* 80–97.

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 760-773.

Brink, J., Wright, R., Pisoni, D.B. (1998). Eliciting speech reduction in the laboratory: Assessment of a new experimental method. In: Research on Spoken Language Processing Progress Report No. 22 Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 396–420.

Brown, R. (1995). *Prejudice : Its social psychology*. Oxford: Blackwell.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155-180.

Burnham, D., Joeffry, S., & Rice, L. (2010). Computer-and human-directed speech before and after correction. *Thirteenth Australasian International Conference on Speech Science and Technology*, Melbourne, Australia, 13-17.

Burnham, D., Joeffry, S., & Rice, L. (2010). Computer-and human-directed speech before and after correction. *Thirteenth Australasian International Conference on Speech Science and Technology, Melbourne, Australia*, 13-17.

Byrd, D. (1994). Relations of sex and dialect to reduction *Speech Communication, 15*, 39–54.

Campbell, G. A. (1910). Telephonic intelligibility. *Philosophical Magazine, 19*, 152–159.

Campell, C., Schlue, K., & Vander Brook, S. (1977). Discourse and second languageacquisition of yes-no questions. *Proceedings of the los angeles second language research forum. Los Angeles: U. C. L. A.*

Cargile, A. C., & Giles, H. (1997). Understanding language attitudes: Exploring listener affect and identity. *Language & Communication, 17*, 195-217.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of languagespecific phoneme representations in the infant brain. *Nature Neuroscience, 1*, 351-353.

Carlson, R., Granström, B., & Fant, G. (1970). Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report, 2-3*, 19-35.

Charles-Luce, J. (1997). Cognitive factors involved in preserving a phonemic contrast *Language & Speech, 40*, 229-48.

Chen, F. (1980). *Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level* (Unpublished Master's thesis). Massachusetts Institute of Technology, Cambridge.

Chomsky, N. (1981). *Rules and representations*. New York: Columbia University Press.

Cervera, T., Soler, M., Dasi, C., & Ruiz, J. (2009). Speech recognition and working memory capacity in young-elderly listeners: Effects of hearing sensitivity. *Canadian Journal of Experimental Psychology, 63,* 216-226.

Cole, R. A., & Rudnicky, A. I. (1983). What's new in speech perception? the research and ideas of william chandler bagley. *Psychological Review, 90,* 94–101.

Conboy, B. T., Sommerville, J., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology, 44,* 1505-1512.

Cooke, M., & Garcia-Lecumberri, M. L. (2012). The intelligibility of lombard speech for non-native listeners. *Journal of Acoustical Society of America, 132,* 1120-1129.

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers *Journal of Acoustical Society of America, 128,* 2059–2069.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584–1595.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infant preference for motherese. *Infant Behaviour Development, 20*, 477–488.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Intelligibility of average talkers in typical listening environments. *Journal of the Acoustical Society of America, 81*, 1598-1608.

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality & Social Psychology Bulletin, 30*, 1332–1346.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: consistency, stability, and convergent validity. *American Psychological Society, 12*, 163-170.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context *Journal of Acoustical Society of America, 124*, 1264-1268.

Cutler, A., Norris, D., & Sebastián-Gallés, N. (2004). Phonemic repertoir and similarity within the vocabulary *Proceedings of the 3rd International Conference on Speech Prosody , 1* 65-68.

Dees, T. M., Bradlow, A. R., Dhar, S., & Wong, P. C. M. (2007). Effects of noise on lexical tone perception by native and non-native listeners. *International Congress of Phonetic Sciences*, 817-820.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*, 462-496.

de Boer, B. & Kuhl, P. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters on-Line, 4*, 129–134.

DePaolis, R. (2005). The influence of production on perception: Output as input. *Supplememt to the Proceedings of the 29th Boston University Conference on Language Development.*

Diamond, A., Werker, J. F., & Lalonde, C. (1994). Toward understanding commonalities in the development of object search, detour navigation, categorization, and speech perception In G. Dawson, & K. W. Fischer (Eds.), *Human behavior and the developing brain* (pp. 380-426). New York: Guilford Press.

Diehl, R., & Lindblom, B. (2004). Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness. In S. Greenberg, W. A. Ainsworth, A. N. Popper & R. R. Fay (Eds.), *Speech processing in the auditory system* (pp. 101-162). New York: Springer-Verlag.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male female differences in the phonetic realization of vowel categories. *Journal of Phonetics, 24*, 187-208.

Dilley, L. C., & Brown, M. (2001). Distinct relative F0 levels elicit categorical effects in F0 maximum and minimum alignment. *Journal of the Acoustical Society of America, 115,* 2398- 2398.

Dongilli, P. (1994). Semantic context and speech intelligibility. In J. Till & K. Yorkston & D. Beukelman (Eds): *Motor Speech Disorders: Advance in Assessment and Treatment* (pp 175–191). Baltimore, Brookes.

Dovidio, J.L., Evans, N., & Tyler, R.B. (1986). Racial stereotypes: the contents of their cognitive representations. J*ournal of Experimental Social Psychology, 22,* 22-37.

Dovidio, J. F., Kawakami, K. & Beach, K. R. (2001). Implicit and explicit attitudes examination of the relationship between measures of intergroup bias. In R.Brown & S. Gaertner (Eds.), *Blackwell Handbook of Social Psychology: Intergroup Processes*. Oxford: Blackwell.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82,* 62-68.

Draegert, G. L. (1951). Relationships between voice variables and speech intelligibility in high level noise. *Speech Monograph, 18*, 272–278.

Dreher, J., & O'Neill, J., (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *The Journal of the Acoustical Society of America, 29,* 1320–1323.

Englund K. (2005). “Voice onset time in infant directed speech over the first six months,” *First Language, 25*, 219–234.

Fant, G. (1960). *Acoustic theory of speech production*. 'sGravenhage, NL: Mouton.

Fayer, J. M., & Krasininski, E. (1987). Native and nonative judgements of inetlligibility and irritation.*37*, 313-326.

Ferguson, C. A. (1975). Towards a characterization of english foreigner talk. *Anthropological Linguistics, 17,* 1-14.

Ferguson, C. A. (1977). Baby talk as a simplified register. In C.E. Snow & C. A. Ferguson (Eds). *Talking to Children* (pp.209-35). Cambridge: Cambridge University Press.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of Acoustical Society of America*, *116*, 2365–2373.

Ferguson, S. H., Poore, M. A., & Shrivastav, R. (2010). Acoustic correlates of reported clear speech strategies. *Journal of the Academy of Rehabilitative Audiology, 43,* 45-64.

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *Journal of Speech, Language, and Hearing Research, 55*, 779–790.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*, 259–271.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language and Hearing Research*, *50*, 1241-1255.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*, 103-114.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, R. Golinkoff, M. Greenberg, C. Harding & J. Bohannon (Eds.), *The origins and growth of communication* (pp. 5-29). Norwood, NJ: Ablex.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181–195.

Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *Annals of Child Development*, *8*, 43-80.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behaviour Development*, *10*, 279-293.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*, 104-113.

Fernald, A., Taeschner, T., Hirsh-Pasek, K., & Jusczyk, P. W. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477-501.

Fiedler, K., Messner, C. & Bluemke, M. (2006). Unresolved problems with the "I", the "A" and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology*, *17*, 74-147.

Flege, J. (1992a). Speech learning in a second language In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and application.* (pp. 565-604). Timonium, MD: York Press.

Flege, J. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 233-272). York: Timonium, MD.

Flege, J. E. (1987a). A critical period for learning to pronounce foreign languages? *Applied Linguistics, 8*, 162-77.

Flege, J. E. (1988b). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders. A review* (pp. 224-401). Norwood, NJ.: Ablex.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. *An integrated view of language development: Papers in honor of henning wode* (pp. 217-244). Trier, Germany: Wissenschaftlicher Verlag Trier.

Flege, J. E. (2003a). Assessing constraints on second-language segmental production and perception. *Phonetics and phonology in language comprehension and* *production: Differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop-consonants. *Phonetica, 43*, 155-71.

Flege, J. E., & Hillenbrand, J. (1987). Differential use of closure voicing and release burst as cue to top voicing by native speakers of french and english. *Journal of Phonetics*, *15,* 203-8.

Flege, J. E., MacKay, I., & Meador, D. (1999a). Native italian speakers' production and perception of English vowels *The Journal of the Acoustical Society of America*, *106,* 2973-2987.

Flege, J. E., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1-34.

Flege, J. E., Munro, M., & MacKay, I. (1995a). Factors affecting degree of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97,* 3125-34.

Flege, J. E., Munro, M., & MacKay, I. (1995b). Effects of age of second-language learning on the production of english consonants. *Speech Communication, 16,* 1-26.

Flege, J. E., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40,* 467-491.

Fletcher, H. (1929). *Speech and hearing*. New York: Van Nostrand.

Florentine, M. (1985a). Non-native listeners' perception of american-english in noise. *Proceedings of Inter-Noise '85,* 1021-1024.

Florentine, M. (1985b). Speech perception in noise by fluent, non-native listeners. *Proceedings of the Acoustical Society of Japan,* Japan. H-85-16.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception &Psychophysics, 48*, 559-570.

Freed, B. F. (1978). *Foreign talk: A study of speech adjustment made by native speakers of English in conversation with non-native speakers* (Unpublished PhD). University of Pennsylvania, USA.

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19*, 90–119.

Fry, D. B. (1979). *The physics of speech*. Cambridge: Cambridge University Press.

Gagne, J. P., & Willie, K. (1989). Relative effectiveness of three repair strategies on the visual identification of misperceived words. *Ear and Hearing, 10,* 368-374.

Gagne, J. P., Masterson, V., Munhall, K. G., Bilida, N., & Querengesser, C. (1994). Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *Journal of the Academy of Rehabilitative Audiology, 27*, 135–158.

Garcia Lecumberri, M. L., & Cooke, M. P. (2006). Effect of masker type on native and non-native consonant perception in noise *Journal of Acoustical Society of America*, *119*, 2445–2454.

Garcia Lecumberri, M., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review *Speech Communication*, *52*, 864-886.

Garnica, O. (1977). On some prosodic and paralinguistic features of speech to young children. *Talking to children: Language input and acquisition* (pp. 63-88). Cambridge: Cambridge University Press.

Gaskell, W., Cambell, C., & Vander Brook, S. (1977). Some aspects of foreigner talk. In C. A. Henning (Ed.), *Proceedings of the Los Angeles Second Language research forum*. Los Angeles: U. C. L. A.

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology*, *49*, 171-180.

Gelfand, S. A., Piper, N., & Silman, S. (1985a). Consonant recognition in quiet as a function of aging among normal hearing subjects. *Journal of Acoustical Society of America*, *78*, 1198-1206.

Giolas, T. (1965).Comparative intelligibility scores of sentence lists and continuous discourse. The *Journal of Auditory Research, 6,* 31-38.

Giolas, T., & Epstein, A. (1963).Comparative intelligibility of word lists and continuous discourse. *Journal of speech and hearing research, 6,* 349-358.

Gordon-Salant, S. (1986). Recognition of natural and time/ intensity altered CVs by young and elderly subjects with normal hearing. *Journal of Acoustical Society of America*, *80*, 1599-1607.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds L and R. *Neuropsychologia*, *9*, 317–323.

Granit, R. (1977). *The purposive brain*. Cambridge, M.A.: MIT Press.

Greenwald, A. G., Mc Ghee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cogition: The Implicit association task. *Journal of Personality and Social Psychology*, *74*, 11464- 1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta–analysis of predictive validity. Journal of Personality and Social Psychology, 1, 17–41.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24*, 14-20.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*, 577-588.

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107,* 2711-2724.

Hanley, T. D., & Steer, M. D. (1949). Effect of level of distracting noise upon speaking rate, duration and intensity. *Journal of Speech and Hearing Disorders, 14*, 363–368.

Hantsweb, Hampshire Count Council. Hampshire Facts and Figures. 2001 Census: Ethnic Group. Retrieved July 2014, from http://www3.hants.gov.uk/planning/factsandfigures/population-statistics/census_pages/census_information/pop_ethnicity.htm.

Harnsberger, J. D., Wright, R., & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication, 50,* 323-336.

Hatch, E., Shapira, R., & Gough, J. (1975). 'Foreigner-talk' discourse. *Working Papers in English as a second language*. Los Angeles: U.C. L. A.

Hawkins, S. & Midgley, J. (2005). Formant frequencies of RP monophtongs in four age groups of speakers. Journal of the International Phonetic Association, 35, 183-199.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America, 130,* 2139–2152.

Hazan, V., & Baker, R. (2010). Does reading clearly produce the same acoustic-phonetic modifications as spontaneous speech in a clear speaking style? *Proceedings of DiSS-LPSS Joint Workshop,* Tokyo. 7-10.

Hazan, V., Markham, D., (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the American Academy of Audiology*, *116*, 3108–3118.

Hazan, V., Simpson, A., 1998. The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Speech Communication*, *24*, 211–226.

Helfer, K. S. (1998). Auditory and auditory-visual recognition of clear and conversational speech by older adults. *Journal of the American Academy of Audiology, 9*, 234-42.

Higginbotham, D.J., Drazek, A.L., Kowarsky, K. , Scally, & C.,Segal, E.(1994). Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. *Augmentative and Alternative Communication, 10,* 191-202.

Higgins, C. &, M. (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology*, *10*, 271-277.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hvd/ utterances: Effects of formant contour *Journal of the Acoustical Society of America*, *406*, 3509–3523.

Hillenbrand, J. M., Clark, M., & Houde, R. (2000). Some effects of duration on vowel recognition. Journal of the Acoustical Society of America, 108, 3013-3022.

Hillenbrand, J., & Gayvert, R. (1993a). Identifications of steady-state vowels synthesied from the peterson and barney measurements. Journal of the Acoustical Society of America, 94, 668-674.

Hillenbrand, J., & Gayvert, R. (1993b). Vowel classification based on fundamental frequency and formant frequencies. Journal of Speech & Hearing Research, 36, 694–700.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels The Journal of the Acoustical Society of America, 97, 3099–3111.

Hirsh-Pasek, K., & Treiman, R. (1982). Doggerel: Motherese in a new context. *Journal of Child Language*, *9*, 229-237.

Hoemke, K. A., & Diel, R. L. (1994). Perception of vowel height: The role of F1-F0 distance. *Journal of the Acoustical Society of America, 96*, 661-674.

Hosoda, M., Stone-Romeko, E. F., & Walter, J. N. (2007). Listeners' cognitive and affective reactions to english speakers with standard American English and Asian accents. *Perceptual and Motor Skills, 104*, 307 326.

Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. *Nature*, *292*, 239–240.

Howard, D. M., & Angus, J. A. S. (2009). Acoustics and psychoacoustics. Oxford: Focal Press.

Hubley, P., & Trevarthen, C. (1979). Sharing a task in infancy. In I. Uzgiris (Ed.), *Social interaction During infancy: New directions for Child Development* (pp. 57-80). San Francisco: Jossey-Bass.

Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., & Strauser, L. E. (2007). Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effects of presentation rate. *Journal of Speech, Language, and Hearing Research*, *50,* 283–303.

Hunter, G., & Kebede, H. (2012). Formant frequencies of British english vowels produced by native speakers of Farsi. *Proceedings of the Acoustics 2012 Conference, Nantes, France,* 23-27.

Hustad, K. C. (2007). Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phoniatrica Et Logopaedica, 59*, 306-317.

Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language & Hearing Research, 51*, 562-573.

Hustad, K.C., & Beukelman, D.R. (2002). Listener comprehension of severely dysarthric speech: effects of linguistic cues and stimulus cohesion. *Journal of Speech, Language, and Hearing Research, 45,* 545–558.

Hustad, K. C., & Cahill, M. A. (2003). Effects of presenta- tion mode and repeated familiarization on intellligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12,* 1-11.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *Neuroreport, 17*, 957-962.

Ingram, D. (1995). The cultural basis of prosodic modifications to infants and children: a response to Fernald's universalist theory. *Journal of Child Language*, *22*, 223–233.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching english vertical bar r vertical bar-vertical bar l vertical bar to japanese adults. *Journal of the Acoustical Society of America, 118*, 3267–3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of Acoustic Society of America, 97*, 553–562.

Iverson, P., & Kuhl, P. K. (1996). *Journal of Acoustical Society of America, 99*, 1130-1140.

Jenkins, J., Strange, W., & Edman, T. (1983). Identification of vowels in "vowelless" syllables. *Perception & Psychophysics*, *34*, 441-50.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69*, 505–528.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y. & Jusczyk, A M. (1993). Infants' Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language, 32,* 402–420.

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 17-64). Timonium, MD: York Press.

Kain, A., Amano-Kusumoto, A., & Hosom, J. P. (2008). Hybridizing conversational and clear speech to determine the degree of contribution of acoustic features to intelligibility. *Journal of the Acoustical Society of America, 124,* 2308-2319.

Kangatharan, J. (2009). A Study of the Effects of Verbal Communication on Interaction Regardless of Appearance and Language Proficiency (Unpublished undergraduate dissertation) Brunel University, Uxbridge.

Kangatharan, J., Uther, M., & Gobet, F. (2012). The effect of physical appearance and accent on the elicitation of vowel hyperarticulation by British English native speakers in speech to foreigners. *Poster Presented at the 164th Meeting of the Acoustical Society of America*, Kansas, Missouri, USA.

Karpinski, A. (2004). Measuring self-esteem using the implicit association test: The role of the other. *Personality and Social Psychology Bulletin, 30,* 22–34.

Karpinski, A., & Hilton, J.L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology*, *81*, 774– 788.

Kelly, C., Dancer, J., & Bradley, R. (1986). Correlations of SPINE test scores to judges' ratings of speech intelligibility in hearing impaired children. *Volta Review, 88*, 145-150.

Kemper, S., Ferrell, P., Harden, T., Finter-Urczyk, A., & Billington, C. (1998). Use of elderspeak by young and older adults to impaired and unimpaired listeners. *Aging, Neuropsychology and Cognition, 5*, 43-55.

Kim, H., Diehl, M.M, Panneton, R., Moon, C. (2006). Hyperarticulation in Mothers' Speech to Babies and Puppies, X*Vth Biennial International Conference on Infant Studies, Kyoto, Japan.*

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

Kitamura, C., Thanavishuth, C., & Burnham, D. (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 Behaviour and Development*, *24*, 372-392.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). 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*, 372-392.

Knoll, M. A. & Scharrer, L. (2007). Acoustic and affective comparisons of natural and imaginary infant-, foreigner- and adult-directed speech. *Proceedings of the 8th Interspeech conference.* Antwerp, Belgium.

Knoll, M. A., Uther, M., MacLeod, N., O'Neill, M., & Walsh, S. A. (2006). Emotional linguistic or cute? the function of pitch countour in infant- and foreigner-directed speech. *Proceedings of the 3rd International Conference on Speech Prosody*, Germany.

Knoll, M., Scharrer, L., & Costall, A. (2009). Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech. *Speech Communication*, *51*, 296-305.

Koopmans-Van Beinum, F. (1980). *Vowel contrast reduction: An acoustic and perceptual study of dutch vowels in various speech conditions*. Amsterdam: Academische Presse.

Koster, C. (1987). *Word recognition in foreign and native language*. Dordrecht, The Netherlands: Foris.

Krause, J. C., & Braida, L. D. (1995). The effects of speaking rate on the intelligibility of speech for various speaking modes *The Journal of the Acoustical Society of America, 98*, S2982.

Krause, J. C. (2001). *Properties of naturally produced clear speech at normal rates and implications for intelligibility enhancement* (Unpublished Doctoral Dissertation).

Krause, J. C., & Braida, L. D. (2003). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America, 115*, 362-78.

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America, 112*, 2165-72.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates *Journal of the Acoustical Society of America, 115*, 362–78.

Kuhl, P. K. (1991a). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93-107.

Kuhl, P. K. (1991b). Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech In S. E. Brauth, W. S. Hall & R. J. Dooling (Eds.), *Plasticity of development* (pp. 73-106). Cambridge, MA: MIT Press.

Kuhl, P. K. (1993a). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics, 21*, 125-39.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*, 812-822.

Kuhl, P. K. (1998). Effects of language experience on speech perception *The Journal of the Acoustical Society of America, 103*, 29-31.

Kuhl, P. K. (2000a). A new view of language acquisition. *Proceedings of the National Academy of Science, 97*, 11850-11857.

Kuhl, P. K. (2000b). Language, mind, and brain: Experience alters perception In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 99–115). Cambridge, MA: MIT Press.

Kuhl, P. K., & Iverson, P. (1995). Linguistics experience and the perceptual magnet effect In W. Strange (Ed.), *Speech perception and linguistics experience: Issues in cross-language research* (pp. 121–154). York, Baltimore: York Press.

Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684–686.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B, 363*, 979-1000.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. C. (2005b). Early speech perception and later langauge development: Implications for the 'critical period'. *Language Learning and Development, 1*, 237-264.

Kuhl, P. K., Kirtani, S., Deguchi, T., Hayashi, A., Stevens, E.B., Dugger, C.D., & Iverson, P. (1997). Effects of language experience on speech perception: American and Japanese infants' perception of /ra/ and /la/. *Journal of the Acoustical Society of America, 102*, 684-686.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*, F13–F21.

Kuhl, P. K., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, 100*, 9096-9101.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Language experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606-60.

Lacerda, F., Sundberg, U., Andersson, C., & Rex, A. (1995). Young infant's perception of segmental and suprasegmental information: Preliminary results. *Proceedings of the International Conference of Phonetic Science*, *3*, 408–411.

Ladefoged, P. (1996). *Elements of acoustic phonetics*. University of Chicago: London.

Ladefoged, P. (2005). *Vowels and consonants: An introduction to the sounds of languages*. Oxford: Blackwell.

Ladefoged, P. Vowel chart. Retrieved June, 2013, from http://www.phonetics.ucla.edu/course/chapter1/vowels.html.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development, 18*, 459-475.

Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Developmental Science*, *15*, 212-221.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35*, 451-453.

Lane, H., & Tranel, B. (1971). Lombard sign and role of hearing in speech. *Journal of Speech & Hearing Research, 14*, 677–709.

Lashley, K. S. (1930). Basic neural mechanisms in behavior. *Psychological Review, 37*, 1-24.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), Speech production and speech modeling (pp. 403-39). Amsterdam, The Netherlands: Kluwer Academic.

Lindblom B. (1992). Phonological units as adaptive emergents of lexical development. In Phonological Development: Models, Research, Implications, ed. C.A. Ferguson, L. Menn, C. Stoel Gammon, pp. 131–63. Timonium, MD: York.

Lindblom B. (1996). Role of articulation in speech perception: Clues form production. *The Journal of the Acoustical Society of America, 99*, 1683-1692.

Lindblom, B., Brownlee, S., Davis, B., & Moon, S.-J. (1992). Speech transforms. *Speech Communication, 11*, 357–368.

Lipski, J. M (2005). “Me want cookie”: foreigner talk as monster talk. Invited lecture. ShippensburgUniversity.

Littleford, L. N., Wright, M. O., & Sayoc-Parial, M. (2005). White students' intergroup anxiety during same-race and interracial interactions: A multimethod approach. basic and applied. *Social Psychology, 27*, 85-94.

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6,* F1- F10.

Liu, H. M., Tsao, F-M. &Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligiblity in Mandarin-speaking young adults with cerebral palsy. *Acoustical Society of America, 117*, 3879-3889.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in mandarin infant-directed speech. *Developmental Psychology, 43,* 912–917.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study. *Journal of Child Language, 36*, 1-10.

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. G. (2004). Clear speech perception in acoustic and electric hearing. *Journal of the Acoustical Society of America, 116*, 2374-2383.

Liu. S. & Zeng, F.G. (2006). Temporal properties in clear speech perception. *Journal of the Acoustical Society of America, 120,* 424-32.

Lively, S. E. (1993). An examination of the perceptual magnet effect. *The Journal of the Acoustical Society of America, 93*, 2423.

Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de l'oreille, du Larynx, du Nez et du Pharynx, 37*, 101-119.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874-886.

Lu, Y., & Cooke, M. (2008). "Speech production modifications produced by competing talkers, babble and stationary noise," *Journal of the Acoustical Society of America, 124,* 3261–3275.

MacKain, K. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language, 9,* 527-542.

MacKain, K., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2,* 369-390.

MacKay, I., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58,* 103-125.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistics and computational perspectives*. Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language comprehension. *Cognition, 6,* 1-71.

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*, 465-494.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognitive Psychology*, *82*, B101-B111.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise *Journal of Speech, Language & Hearing Research, 40*, 686–93.

Meador, D., Flege, J., & Mackay, I. (2000). Factors affecting the recognition of words in second language *Bilingualism: Language and Cognition, 3*, 55-67.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale In J. F. Dovidio & S. L Gaertner (Eds), *Prejudice, Discrimination, and Racism*. Orlando, FL: Academic Press.

McConnell, A. R. & Leibold, J. M. (2001). Relations aong the implicit association test, discriminatory behaviour, and explicit measures of racial attitudes. Journal of Experimental Social Psychology, 37, 435-442.

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*, 483–510.

Meisel, J. (1977). Ausländerdeutsch und Deutsch ausländischer Arbeiter. Zur moglichen Entstehung eines Pidgin in der BRD. *Zeitschrift Für Linguistik and Literaturwissenschaft, 5,* 18.

Miller, J. L. (1981). Effect of speaking rate on segmental distinctions. In P.D. Eimas & J. L. Miller (Eds.), *Plasticity of development* (pp. 39-70). Hillsdale, New Jersey: Lawrence Erlbaum.

Miyawaki, K., Strange, W., Verbrugge, R., & Liberman, A. M. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331–340.

Moats, L. (2006). How spelling supports reading. *American Educator, 29*, 42-43.

Monsen, R. B. (1976). Normal and reduced phonological space: The productions of English vowels by deaf adolescents", *Journal of Phonetics, 4,* 189-198.

Moon, S. (1995). Formant patterns of clear speech and their implications for phonetic invariance. *Linguistics in the Monrning Calm, 3*, 411-22.

Moon, S., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in english stressed vowels. *Journal of the Acoustical Society of America, 96*, 40-55.

Moore, B. (1997). *An introduction to the psychology of hearing*. London: Academic Press.

Morgan, J. L. (1990). Input, innateness, and induction in language acquisition. *Developmental Psychobiology, 23*, 661-678.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of segmental and suprasegmental information in prelingual speech segmentation. *Child Development, 66*, 911-936.

Mullen, B., Brown, R., & Smith, C. (1992). In-group bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology, 22*, 103–122.

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

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Language and Speech, 38*, 289-306.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 49*, 285-310.

Munro, M., Flege, J. E., & MacKay, I. (1996). Effects of age of second-language learning on the production of english vowels. *Applied Psycholinguistics, 17*, 313-334.

Nábělek, A. K., & Donahue, A. M. (1984). *Journal of Acoustical Society of America, 75*, 632–634.

Nave, R. Forming the vowel sounds. Retrieved June, 2013, from http://hyperphysics.phy-astr.gsu.edu/hbase/music/vowel.html

Neel, A. (1998). *Factors influencing vowel identification in elderly hearing-impaired listeners.* (Doctoral dissertation, Indiana University). Retrieved from http://phdtree.org/pdf/25129651-factors-influencing-vowel-identification-in-elderly-hearing-impaired-listeners/.

Nejime, Y., & Moore, B. C. (1998). Evaluation of the effect of speech rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss. *Journal of the Acoustical Society of America*, *103*, 572-6.

Nelson, N. W. (1976). Comprehension of spoken language by normal children as a function of speaking rate, sentence difficulty, and listener age and sex. *Child Development*, *47*, 299–303.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception *Journal of the Acoustical Society of America, 85,* 2088-2113.

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America, 80*, 1297-1308.

Newport, E. L., & Aslin, R. N. (2004). Learning at a ditance I. statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127-162.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow, & C. A. Ferguson (Eds.), *Talking to children: Language input and acquision* (pp. 109-149). Cambridge: C.U.P.

Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society, 112,* 1627–41.

Pisoni, D. B. (1996). Word identification in noise. *Language and Cognitive Processes, 11*, 681-687.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653–667.

O'Shaughnessy, D. (2000). *Speech communications: Human and machine* (2nd ed.). New York: IEEE Press.

Oviatt, S. (1998). The CHAM model of hyperarticulate adaptation during human-computer error resolution. *Proceedings of the International Conference on Spoken Lan*guage Processing, Sydney, Australia.

Oviatt, S., & Coulson, R. (2003). Predicting children's hyperarticulate speech during human-computer error resolution. *Journal of the Acoustical Society of America, 113*, 2296.

Oviatt, S., Levow, G., Moreton, E., & MacEachern, M. (1998). Modeling global and focal hyperarticulation during human–computer error resolution. *Journal of the Acoustical Society of America, 104*, 3080-3098.

Oviatt, S., MacEachern, M., & Levow, G. (1998). Predicting hyperarticulate speech during human–computer error resolution. *Speech Communication, 24*, 87-110.

Papousek, M., & Hwang, S. F. C. (1991). Tone and intonation in mandarin baby talk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, *12*, 481- 504.

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America, 95*, 1581–1592.

Pepperberg, I. M. (2002). Allospecific referential speech acquisition in grey parrots (psittacus erithacus): Evidence for multiple levels of avian vocal imitation. In K. Dautenhahn, & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 109–131). Cambridge, MA: MIT Press.

Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society, 112*, 1627–41.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels *Journal of the Acoustical Society of America, 24*, 175-184.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28*, 96–103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech & Hearing Research, 29*, 434-46.

Piske, T., Flege, 1., MacKay, I., & Meador, D. (2002). The production of English vowels by fluent early and late billnguals. *Phonetica, 59*, 49-71.

Pisoni, D. B. (1996). Word identification in noise. *Language and Cognitive Processes, 11*, 681-687.

Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C., & Yuchtman, M.,. (1985). Some acoustic-phonetic correlates of speech produced in noise. *ICASS*P, Tampa, FL, 1581-1584.

Plichta, B. (2010). Akustyk. Retrieved July/3rd, 2013, from http://bartus.org/akustyk/tutorials/formant_analysis_video.html.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961-2977.

Polka, L. (1992). Characterising the influence of native experience on adult speech perception. *Perception & Psychophysics, 52*, 37-52.

Polka, L., & Bohn, O. (1996). A cross-language comparison of vowel perception in English learning and German-learning infants. *The Journal of the Acoustical Society of America, 100*, 577-592.

Polka, L., Colantonio, C., & Sundara, M. (2001). A crosslanguage comparison of /d/–/ð/ perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America, 109*, 2190–2201.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421–435.

Popa-Roch, M., & Delmas, F. (in press). Prejudice Implicit Association Test effects. Zeitschrift fu ̈r Psychology/Journal of Psychology, 218, 44-50.

Raphael, L. J., Borden, G. J., & Harris, K. S. (Eds.). (2007). *Speech science primer*. Baltimore, MD: Lippincott, Williams & Wilkins.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14,* 287-290.

Richeson, J. A., & Trawaltar, S. (2005), Why do interracial interactions impair executive function? A Resource Depletion Account. *Journal of Personality and Social Psychology, 88*, 934- 947.

Richeson, J. A., Trawaltar, S., & Shelton, J. N. (2005). African Americans’ implicit racial attitudes and the depletion of executive function after interracial interactions. *Social Cognition*, *23*, 336-352.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native- and non-native speech contrasts in seven- and eleven-month-old American infants. *Developmental Science, 8,* 162-172.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7*, 532-47.

Rostolland, D. (1982a). Acoustic features of shouted voice. *Acustica, 50*, 118-125.

Rostolland, D. (1982b). Phonetic structure of shouted voice. *Acustica, 51*, 80-89.

Rostolland, D., & Parant, C. (1974). Physical analysis of shouted voice. *Eighth International Congress on Acoustics*, *London, Great Britain.*

Rossing, T., Moore, R., & Wheeler, P. (2002). *The science of sound*. San Francisco, CA: Addison Wesley.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology*: *General*, *133*, 139–165.

Sankowska, J., Garcia Lecumberri, M. L., Cooke, M. (2011). Vowel shortening in altered speech modes elicited via an interactive task. *International Congress of Phonetic Sciences (ICPhS), Hong Kong.*

Sankowska, J., Garcia Lecumberri, M. L., Cooke, M. (2011). Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech. *Poznan Studies in Contemporary Linguistics, 47*, 109-119.

Sato, C. J. (1991). Sociolinguistic variation and language attitudes in Hawaii. In J. Cheshire (Ed.), *English around the world: Sociolinguistic perspectives* (pp. 647-663). Cambridge: Cambridge University Press.

Sawusch, J. R. (2005). Acoustic analysis and synthesis of speech. In D. B. Pisoni, & R. Remez (Eds.), *The handbook of speech perception* (pp. 7-27). Oxford, UK: Blackwell.

Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., & Dmitrieva, O. (2007). An Acoustic study of real and imagined foreigner- directed speech. *17th International Congress of Phonetic Sciences,* Germany.

Schaeffler, F., Kempe, V., Biersack, S. (2006). Comparing Vocal Parameters in Spontaneous and Posed Child-Directed Speech. In: *Proceedings of the 3rd Speech Prosody Congress*, Dresden, Germany.

Scherer, K. R. (1981a). Speech and emotional states. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 189-220). New York: Grune and Stratton.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143-165.

Schum, D. J. (1996). Intelligibility of clear and conversational speech of young and elderly talkers. *Journal of the American Academy of Audiology*, *7*, 212-18.

Schum, D. J. (1997). Beyond hearing aids: Clear speech training as an intervention strategy. *The Hearing Journal, 50*, 36-40.

Sebastian, R., Ryan, E., Keogh, T., & Schmidt, A. (1980). The effects of negative affect arousal on reactions to speaker In H. Giles, W. Robinson & P. M. Smith (Eds.), *Language: Social psychologica1 perspective* (pp. 203-208). Oxford: Pergamon.

Sikveland, R. O. (2006). How do we speak to foreigners? - Phonetic analyses of speech communication between L1 and L2 speakers of Norwegian. *Working Papers, 52*, 109-112.

Smiljanić, R., & Bradlow, A. (2005). Production and perception of clear speech in Croatian and English. *Journal of Acoustical Society of America, 118*, 1677-88.

Smiljanić, R., & Bradlow, A. R. (2007). Clear speech intelligibility: Listener and talker effects *Proceedings of the XVIth International Congress of Phonetic Sciences* Germany.

Smiljanić, R., & Bradlow, A. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness *Journal of Acoustical Society of America, 130,* 4020–4032.

Snow, C. E., van Eeden, R., & Muysken, P. (1981). The interactional origins of foreigner talk: Municipal employees and foreign workers. *International Journal of the Sociology of Language, 81*, 81-91.

Sommers, M. S., & Barcroft, J. (2006). Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification. *Journal of the Acoustical Society of America, 119*, 2406-2416.

Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II: The effects of age and hearing impairment. *Journal of the Acoustical Society of America*, *101*, 2278-2288.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America, 96*, 1314–1324.

Son Hing, L. S., Chung- Yan, G. A., Hamilton, L. K., & Zanna, M. P. (2008). A two-dimensional model that employs explicit and implicit attitude to characterize prejudice. *Journal of Personality and Social Psychology, 94*, 971- 987.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America, 128*, 389–400.

Sound and Perception. Retrieved July 16, 2012, from http://flylib.com/books/en/1.500.1.12/1/#_/term_.

Stent, A., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication, 50*, 163-178.

Stephens, R., & Bate, A. E. (1966). *Acoustics and vibrational physics*. London: Edward Arnold.

Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants Developmental Psychology, 8, 727-735.

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language, 10*, 1-15.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.

Storti, C. (1999). *Figuring foreigners out: A practical guide*. London: Intercultural Press.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85*, 2135–2153.

Strange, W. (1986). Speech input and the development of speech perception. In J. F. Kavanagh (Ed.), *Otitis media and child development* (pp. 12-26). Parkton, MD: York.

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In E. Tohkura, E. Vatiskiotis-Bateson & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic strcuture*. Tokyo, Japan: Ohmsha.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. *Language evolution: The states of the art* (pp. 235-254). Oxford, UK: Oxford University Press.

Subtelny, J. (1977). Assessment of speech with implications for training. In F. Bess (Ed.), *Childhood deafness* (pp. 183-194). New York: Grune&Stratton.

Summers, W., & Leek, M. (1992). The role of spectral and temporal cues in vowel identification by listeners with impaired hearing Journal of Speech and Hearing Research, 35, 1189–1199. *Language, 10*, 1-15.

Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American, 236*, 88-91.

Sussman, J. E., & Lauckner-Morano, V. J. (1995). Further tests of the "percptual magnet effect" in the perception of [i]: Identification and change-no-change discrimination. *The Journal of the Acoustical Society of America, 97*, 539-552.

Tajfel, H., Billig, M. G., Bundy, R. P., Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149-177.

Takagi, N. (1993). *Perception of American English /r/ and /1/ by adult Japanese learners of English: A unified view.* (Unpublished Ph.D dissertation). University of California-Irvine.

Takagi, N., & Mann, V. A. (1995). Signal detection modeling of Japanese listeners' /r/-/l/ labeling behaviour in a one-interval identification task. *Journal of the Acoustical Society of America, 97*, 539-552.

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by japanese and american listeners *Journal of Acoustical Society of America*, *88*, 663–666.

Tani, M. (2011). Accommodation Theory in Interlanguage Talk: Convergence across different levels of English (Unpublished thesis). The Faculty of Foreign Studies. Tokyo University of Foreign Studies.

Taylor, C. (2007). The physics of sound. In P. Kruth, & H. Stobart (Eds.), *Sound* (pp. 34-64). Cambridge: Cambridge University Press.

ThinkQuest. (http://library.thinkquest.org/19537/). The soundry. Retrieved June, 2011, from http://library.thinkquest.org/19537/.

Tomasello, M. (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant- directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review, 9*, 335-340.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*, 466–472.

Tsao, F., Liu, H., Kuhl, P. K., & Tseng, C. (2000). Perceptual discrimination of a Mandarin fricative-affricate contrast by English-learning and Mandarin-learning infants. Poster presented at the meeting of the International Society on Infant Studies, Brighton, U.K.

Tsao, F., Liu, H., & Kuhl, P. K. (2006). Perception of native and non-native affricate–fricative contrasts: Cross-language tests on adults and infants. *Journal of Acoustical Society of America, 120*, 2285-2294.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., & et al. (1994). Discrimination of English/r/-/l/and/w-y/by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. *Third International Conference on Spoken Language Processing*, Yokohama, Japan.

Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38*, 1001-1013.

Uchanski, R. M. (1988). *Spectral and temporal contributions to speech clarity for hearing impaired listener.* PhD Thesis, Massachusetts Institute of Technology, USA.

Uchanski, R. M., Choi, S., Braida, L., & Durlach, N. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research, 39,* 494–509.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49,* 2-7.

Vallabha, G. K., & McClelland, J. L. (2007). Success and failure of new speech category learning in adulthood: Consequences of learned hebbian attractors in topographic maps. *Cognitive, Affective & Behavioral Neuroscience, 7,* 53-73.

Vance, T. (1987). *An introduction to Japanese phonology*. Albany, NY.: State University of New York Press.

Van Dijk, T.A. (1987). Communicating racism: Ethnic prejudice in thought and talk. London: Sage.

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America, 84,* 917-928.

Wallace, S. The Basics of Hyperarticulated Speech. Retrieved December 2, 2010, from http://blogs.iit.edu/csl_ug_summer_research/category/sean/.

Walker, R. & Jussim, L. (2002). Do people lie to appear unprejudiced. *The Rutgers Scholar, 4*, 1-19.

Walley, A., & Flege, J. E. (2000). Effects of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics, 27*, 307-332.

Warren-Leubecker, A., & Bohannon, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development, 55*, 1379-1385.

Wassink, A. B., Wright, R. and Franklin, A. (2007). Intraspeaker variability in vowel production: an investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics,* 35, 353-379.

Weiher, E. (1975). *Lautwahrmnehmung und lauproduktion in englischunterricht für Deutsche* [Working Papers, 3]. University of Kiel: Institute of Phonetics.

Weinberger, S. (1990). Minimal segments in second language phonology. *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech,* Amsterdam: University of Amsterdam, Netherlands. 263-311.

Weinreich, U. (1957). *Languages in contact, findings and problems*. The Hague: Mouton.

Weismer, G., Laures, J. S., Jeng, J., Kent, R. D., & Kent, J. F. (2001). The effect of speaking rate manipulations on speech timing, formant frequencies, and perceptual judgments in dysarthric speakers with amyotrophiclateral sclerosis. *Folia Phoniatrica Et Logopaedica, 52*, 201-219.

Weitz, S. (1972). Attitude, voice and behavior: A repressed affect model of interracial interactions. *Journal of Personality and Social Psychology*, *24*, 14-21.

Werker, J. F. (1994). Cross-language speech perception: Developmental change does not involve loss. In *Speech Perception and Word Recognition*, J. Goodman & H. Nusbaum (Eds.) MIT, Cambridge.

Werker, J. F., Gilbert, J. V. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development,* 52, 349-355.

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

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of Acoustical Society of America, 75*, 1866-1878.

Werker, J. F., & Lalonde, C. E. (1988). *Developmental Psychology, 24*, 672-683.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception &Psychophysics, 37*, 35-44.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and implications* (pp. 285–311). Timonium, MD: York Press.

Werker, J. F., & Polka, L. (1993). Developmental changes in speech perception: New challenges and new directions. *Journal of Phonetics, 21*, 83-101.

Wode, H. (1978). The beginning of non-school room L2 phonolgoical acquisition. *International Review of Applied Linguistics, 16*, 109-25.

Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2004). Polly want a cracker? On talking to babies, puppies and parrots. *International Conference on Infant Studies, Chicago, USA.*

Xu, N., Burnham, D., & Kitamura, C. (2007). Vowels and tones in infant directed speech: Hyperarticulation for both, but different developmental patterns. *8th Annual Conference of the International Speech Communication Association, Interspeech 2007*, Antwerp, Belgium.

Xu, N., & Burnham, D. (2010). Tone hyperarticulation and intonation in cantonese infant directed speech. *Proceedings of Speech Prosody,* Chicago. 1-4.

Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog- and infant- directed speech. *Anthrozoos, 26*(3), 373-380.

Yamada, R., & Tohkura, Y. (1995). Age and acquisition of second language speech sounds: Perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience, issues in cross-language research* (pp. 305-320). Timonium, MD: York Press.

Yorkston, K., & Beukelman, D. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders, 13*, 15-31.

Yorkston, K., & Beukelman, D. R. (1981b). Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *Journal of Speech and Hearing Disorders, 46*, 296-301.

Yorkston, K., & Beukelman, D. R. (1983). The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In W. Berry (Ed.), *Clinical dysarthria* (pp. 155-164). San Diego: College Hill.

Yorkston, K., & Beukelman, D. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11*, 499-512.

Yrttiaho, S., May, P. J. C., Tiitinen, H., & Alku, P. (2011). Cortical encoding of aperiodic and periodic speech sounds: Evidence for distinct neural populations. *NeuroImage, 55*, 1252-1259.

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognitive Development, 6*, 179–208.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *Neuroimage, 26*, 703-720.

Zeng, F. G., & Liu, S. (2006). Speech perception in individuals with auditory neuropathy. *Journal of Speech, Language, and Hearing Research, 49*, 367-380.

Zwicky, A. M. (1972). On causal speech *Papers from the Eighth Regional Meeting.* Chicago Linguistic Society, Chicago. 607–615.

Appendices

A Ethics of Experiment 1

B Ethics of Experiment 2

C Ethics of Experiment 3 and 4

D Ethics of Experiment 5

E Informed Consent and Debrief for Experiment 1

F Informed Consent and Debrief for Experiment 2

G Informed Consent and Debrief for Experiment 3 and 4

H Informed Consent and Debrief for Experiment 1

I Diapix Task for Experiment 1

J Affect Rating for Experiment 1

K Resource Allocation (RA) and Modern Racism Scale (MRS) for Experiment 1

L Demographic and Linguistic Questionnaire

M Picture cards for Experiment 2

N Reading task for Experiment 2

Appendix A

Ethics of Experiment 1

Appendix B

Ethics of Experiment 2

Appendix C

Ethics of Experiment 3 and 4

Appendix D

Ethics of Experiment 5

Appendix E

Informed Consents and Debriefs for Experiment 1

Appendix F

Informed Consents and Debriefs for Experiment 2

Appendix G

Informed Consents and Debriefs for Experiment 3 and 4

Appendix H

Informed Consents and Debriefs for Experiment 4

Appendix I

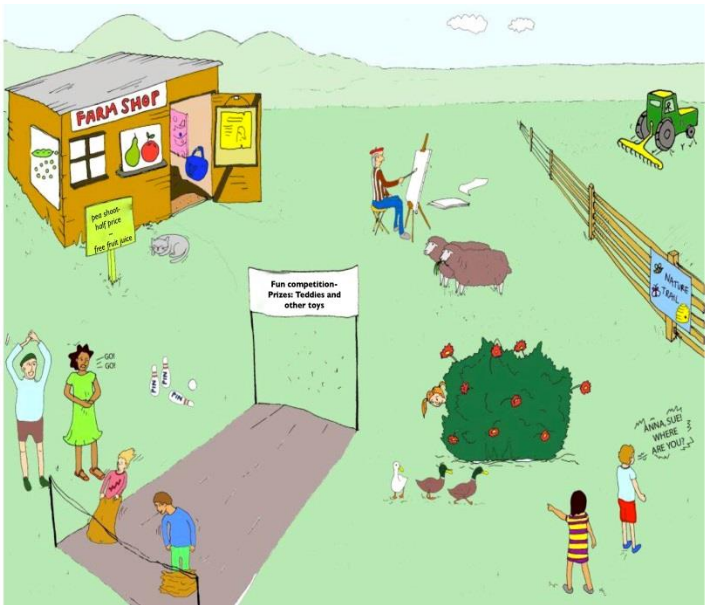
Diapix Task for Experiment 1

‘Beach’ picture (A)

‘Beach’ picture (B)

‘Farm’ picture (A)

‘Farm’ picture (B)



'Street' picture (A)

'Street' picture (B)

Appendix J

Affect Rating for Experiment 1

Appendix K

Resource Allocation (RA) and Modern Racism Scale (MRS) for Experiment 1

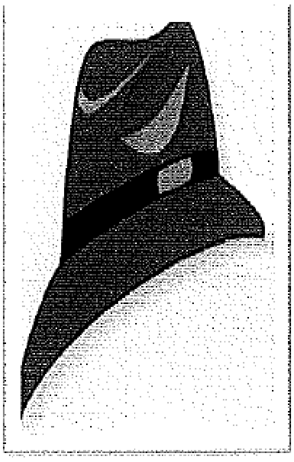
Appendix L

Demographic and Linguistic Questionnaire

Appendix M

Picture cards for Experiment 2

THREE BEES

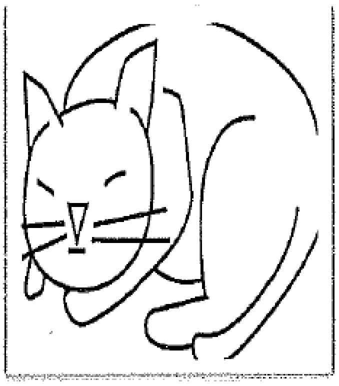


GREEN HAT

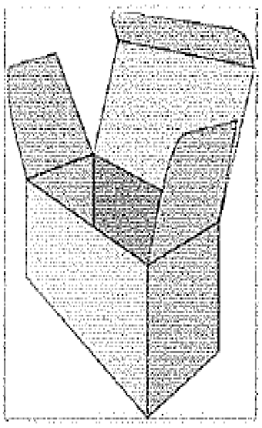
KITE

BLUE SHORTS

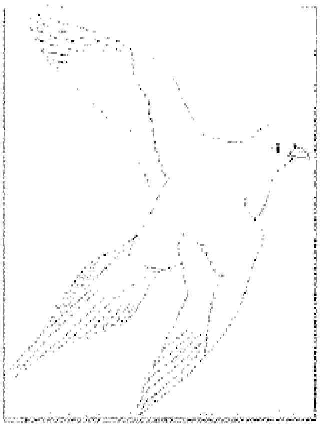
BLUE DRESS



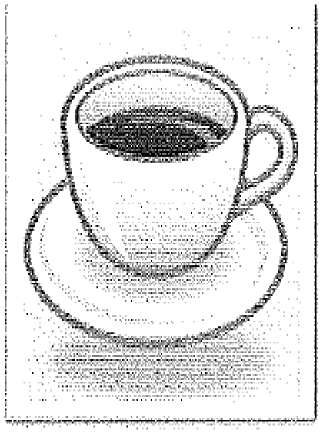
CAT



BOX

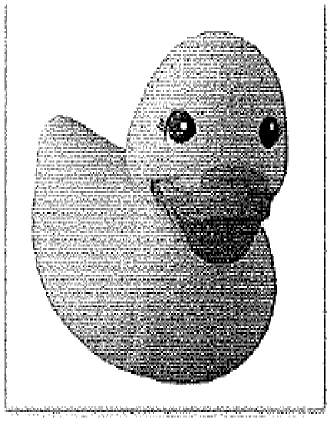


WHITE DOVE



CUP

BLACK TIE



RUBBER DUCK

Appendix N

Reading task for Experiment 2

Clearly read out the following lists of sentences

1. List

(a) Sharks have big and sharp teeth.
(b) Teddy bears are popular among little children
(c) Sheep are protected by shepherds.
(d) Fruit Juice is made out of fresh fruits.

2. List

(a) Cars can cause a traffic jam during rush hour.
(b) Pencils are sharpened with a sharpener.
(c) Many trees can be found in a park.
(d) There are different types of shoes.

3. List

(a) Gardens are full of flowers.
(b) Eggs are produced by hens.
(c) Bees produce honey.
(d) Stool is a chair without arm or back rests.

4. List

(a) Bananas can differ in size and color.
(b) Stop signs are red.
(c) There are different types of teas.
(d) Blueberries can reduce the risks of certain cancers.

5. List

(a) There are different types of baskets.
(b) New York taxis are yellow.
(c) Green tea can improve your health.
(d) Fruits and vegetables should be eaten every day.

6. List

(a) A black ties is part of a suit.
(b) A cat is sleeping on the floor.
(c) A box can be filled with books.
(d) A green hat can be worn by anybody.

7. List

(a) The door to the kitchen is open.
(b) There are rocks on the mountain.
(c) A cup of tea can be enjoyed in the afternoon.
(d) Dogs like to play with s ball.

8. List

(a) A white dove is building a nest.
(b) Children can play with a kite.
(c) A rubber duck can be found in the bathroom.