GROWING UP IN AN IMMIGRANT COMMUNITY: THE PHONEMIC DEVELOPMENT OF SEQUENTIAL BILINGUAL CHILDREN

Kathleen Marie McCarthy

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Division of Psychology and Language Sciences
Department of Developmental Science
UCL

Declaration

I, Kathleen Marie McCarthy, 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.

ACKNOWLEDGMENTS

Firstly, I would like to thank my supervisors Merle Mahon and Bronwen Evans. They have given me a great amount of support throughout my PhD. I want to thank Merle for always encouraging my ideas and for giving me freedom in my work. I want to thank Bronwen for teaching me so much about conducting thorough phonetics research. Both Merle and Bronwen have always been there for me when needed. I am extremely grateful to have had such a brilliant supervision.

Secondly, I would like to thank all of the families, children and schools for taking part in this research. I would like to thank the children for being so much fun to work with, and for sitting patiently through my experiments. I would like to thank the families who welcomed me into their homes and community. I cannot count the number of tasty Bangladeshi curries and sweets that I have eaten over the course of my PhD. I would like to thank the nurseries and schools for helping organise everything.

There are a number of people from Chandler House who I would like to thank. Paul Iverson, Katrin Skoruppa and Nathan Klinedinst have given me much support and invaluable advice throughout my PhD. Stuart Rosen, Debi Vickers and Andrew Faulkner have given me technical and experimental advice, they have been a great source of knowledge. Gordon Craig has taught me most of what I know about statistics. Mike Coleman, Steven Nevard and David Cushing have provided me with technical support. I would have been lost without them. I would also like to thank my fellow PhD students and researchers, in particular, Felicity Deamer, Melanie Pinet, Lucy Hughes, Anita Wagner, Fleur Corbett, Paraic Scanlon, Yasna Pereira, Sonia Granlund, Yasuaki Shinohara, Sam Green, Kathrine Mair and Tanya Denmark.

My family and friends have been an invaluable source of support. Thank you for putting up with me over the last 3 years. In particular, I must thank Amy Jackson for being a great friend. Not only has she sat through numerous practise presentations, she has also given me a tremendous amount of moral support throughout my PhD. Lastly, I would like to thank Anjo Perplies for his encouragement, support and most importantly for giving me so much happiness outside of research.

For my mum

ABSTRACT

The majority of bilingual speech research has focused on simultaneous bilinguals. Yet, in immigrant communities, children are often initially exposed to their heritage language (L1) before becoming gradually immersed in the host country's language (L2) when they start full-time education. This is typically referred to as sequential bilingualism. These children are often exposed to differing amounts of the L1 and L2, as well as accented variants. To date, little is known about the developmental trajectories of such children. This thesis investigates the influence of this highly variable language environment on the acquisition of L2 phonemes. Specifically this thesis focuses on Sylheti-English speaking children from the London-Bengali community.

To provide a baseline of the children's speech environment, Study 1 investigated the speech production of Sylheti (L1) and English (L2) by adult speakers from the London-Bengali community. The results show differences in production of both the L1 and L2 depending on the speaker's language background. Studies 2 and 3 tracked the acquisition of English vowel and plosive contrasts, both perception and production, by Sylheti-English bilingual children and their monolingual peers. Using a longitudinal design, children were tested at two time points: after seven months of English language experience in nursery (Time 1) and approximately one year later, when the children were in the first year of Primary school (Time 2). At Time 1 the bilingual children displayed difficulties with phonemic contrasts that do not exist in Sylheti. However, by Time 2, the bilingual children had rapidly changed to match that of their monolingual peers. Studies 4 and 5 explored the influence of language exposure and caregiver speech on the bilingual children's English phoneme acquisition. The results suggest that sequential bilingual children are particularly sensitive to the amount of language exposure to each language as well as fine-grained phonetic differences in caregiver speech.

Table of Contents

List of figures

List of Tables

1. Literature review

This aim of this chapter is to outline the developmental trends in phonemic development, from infancy to adulthood. The first section will review the research on monolingual development, as well as the current theoretical frameworks that have been proposed to explain the developmental speech perception and production patterns. The following sections will review the research on bilingual and second-language (L2) phonemic acquisition, as well as L2 speech acquisition theories. This section will primarily focus on an immigrant community setting, and in turn discuss the possible influences of sociocultural and input factors on L2 phoneme acquisition. The final section will focus on the current thesis, including a brief history of the London-Bengali community, the phonetics of Sylheti and English, as well as the research objectives for this thesis.

1.1. The development of phonemic perception and production: monolinguals

1.1.1. Perception

Infants are born with the perceptual abilities that allow them to discriminate between many phonemic contrasts, e.g., /p/-/b/ within the first few weeks of life (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The ability to discriminate between acoustically similar speech sounds, was initially demonstrated in infants using the High Amplitude Sucking (HAS) method by Eimas and colleagues (1971). In this study, infants aged 1-4 months learning English demonstrated the ability to discriminate the voicing contrast /ba/-/pa/. The stimuli used consisted of a voicing continuum ranging from -20ms to 80ms voice onset time (VOT) in 20ms steps. The infants displayed an increase in sucking rate when the stimuli changed from 20ms (/ba/) to 40ms (/pa/), the adult phoneme boundary, indicating that they had discriminated this difference in voicing, and thus the differing phonemes. Similar early patterns of perception have subsequently been demonstrated for many phonemic contrasts. For example, Kikuyu-learning and Japanese-learning infants can discriminate English contrasts that do not exist in their native language (Streeter, 1976; Tsushima et al., 1994).

Yet, over the course of the first year of life, the ability to discern non-native speech sounds decreases (e.g., Kuhl et al., 2006; Werker & Tees, 1984). The seminal studies by Werker and Tees (1983, 1984) were the first to illustrate this developmental pattern. They tested 6-12 month-old infants on native and non-native phonemic contrasts using the conditioned headturn procedure. English-learning infants were tested on the Hindi dental and retroflex stop contrast /t̪/-/ʈ/, and the Salish Nthlakampx velar and uvular ejective stop contrast /k'/-/q'/, both of which do not exist in English. The findings showed that at 6-8 months of age the infants were able to discriminate these contrasts, whereas English-speaking adults could not. However, the 10-12 month-old infants showed a decline in their ability to discriminate these contrasts, whereas 11-12 month-old Hindi- and Salish-speaking were able to discriminate them. These findings suggest that infants become attuned to their native language speech sounds within the first year of life (see Figure 1.1 for a timeline displaying this developmental trajectory). This well-established pattern of decline has been replicated for many other consonant contrasts (e.g., /ɬ/-/ɮ/, Best & McRoberts, 2003) as well as vowel contrasts (e.g., /U/-/Y/, Polka & Werker, 1994). For vowels, the findings suggest that these native language effects are present slightly earlier than for consonants, at around 4 – 6 months of age (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Polka & Werker, 1994).

In addition to displaying a decline in non-native discrimination abilities, more recent cross-linguistic research has shown that infants increase their sensitivity to speech sounds in their ambient language (e.g., Kuhl et al., 2006; Tsao, Lui, & Kuhl, 2006; Narayan, Werker & Beddor, 2010). Using the conditioned headturn procedure, Kuhl et al. (2006) tested 6-12 month-old American and Japanese infants on English /r/-/l/, an English phonemic contrast that does not exist in Japanese, and has consistently been shown to be difficult for Japanese adult speakers (Hattori & Iverson, 2009). The study showed an improvement in discrimination of the /r/-/l/ contrast between 6-12 months old for the English-learning infants, suggesting a refinement in perceptual abilities for their native-English contrast. The Japanese-learning infants, in contrast, displayed a decline in perception for this non-native contrast. Similar patterns of improvement have been found for other phonemic contrasts (e.g., Eilers, Wilson, & Moore, 1979; Polka, Colantonio & Sundara, 2001; Sato, Sogabe & Mazuka, 2010), with some developing slightly later than others.

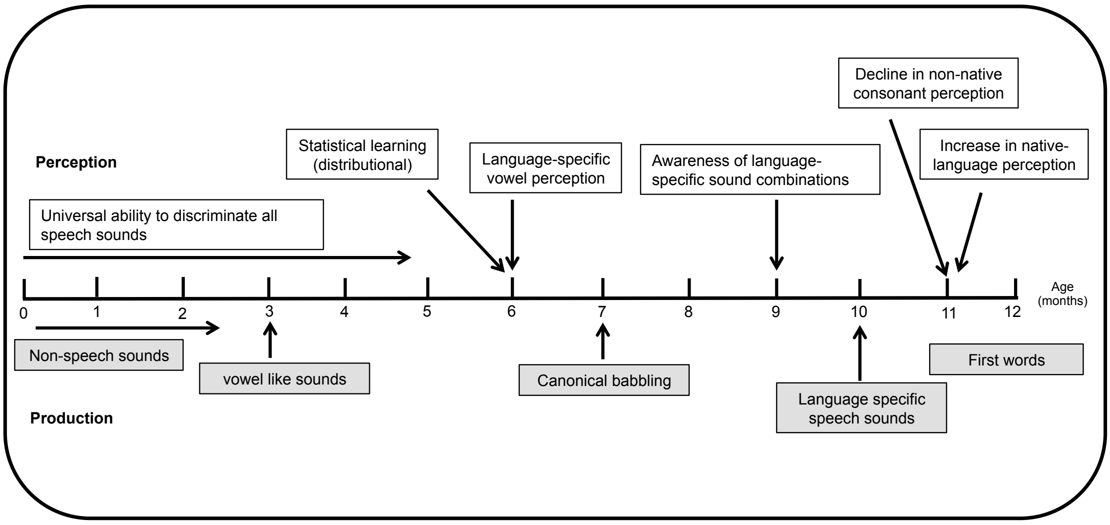


Figure 1.1. Timeline of typically developing infants' speech perception and production in the first year of life. Adapted from Kuhl, 2004; Kuhl et al., 2008

To account for the developmental pattern observed in infant speech perception, Aslin and Pisoni (1980) outlined four possible developmental trajectories: *maintenance, facilitation, induction, and loss*. Initial selectionist theories of speech perception development primarily focused on *maintenance* and *loss*. That is, the seemingly complete *loss* in the ability to discriminate non-native contrasts that were found in the initial infant experiments (e.g., Eimas et al., 1971), and the influence of language experience that has been shown to *maintain* the ability to discriminate native contrasts (e.g., Werker & Tees, 1984). Two prominent examples of such theories are Liberman's motor theory (Liberman, Cooper, Shankweiler, Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) and Eimas' phonetic feature detection account (Eimas, 1975). The core of both theories suggests that infants are born with an innate phonetic-specific mechanism that allows them to distinguish between speech contrasts. Specifically, Liberman's motor theory suggested that infants are born with an innate ability to detect all phonetically relevant gestures, e.g., lip rounding, tongue-backing, jaw raising (Liberman & Mattingly, 1985), that allow infants to detect the phonetic distinctions. In contrast, Eimas' phonetic feature detectors account (1975) posits that infants are born with acoustic-phonetic detectors that are sensitive to the phonetic distinctions between speech sounds. For both theories, it is argued that these innate mechanisms i.e., acoustic feature or gestural detectors, allow the infant to maintain language-specific categories, and that with native language experience they lose the ability to discriminate non-native contrasts.

Such selectionist accounts were hard to uphold in the light of the research findings that have been presented over the last four decades. Firstly, the discovery of categorical perceptual abilities in non-humans (Kuhl & Miller 1975, 1978; Kuhl & Padden, 1983) suggested that the initial categorical perception findings in infants are in part likely to be due to more general auditory mechanisms rather than a linguistic-specific phoneme detector (see e.g., Eimas, 1975). Further, as discussed, infant and child research shows that the perceptual abilities in infants are not simply maintained or lost with native language experience. Instead, research has demonstrated that from 6-months-old to well beyond the first year of life, infants and children refine their native language perceptual abilities, thus demonstrating an improvement in their perceptual ability to discriminate and categorize native phonemic contrasts (e.g., Hazan & Barrett, 2000; Nittrouer & Miller, 1997; Sundara, Polka & Genesee, 2006). It thus seems that language experience is likely to facilitate phonemic learning in infants and children rather than simply to maintain it.

With regards to loss, both adult and infant research has shown that language experience does not result in complete loss. For example, some infant studies have shown no decline in non-native contrasts (Best, McRoberts & Sithole, 1988; Polka & Bohn, 1996; Polka et al., 2001; Rivera-Gaxiola, Silva-Pereyra & Kuhl, 2005). For example, Best et al. (1988) showed that both 14 month-old English-learning infants and adults were able to discriminate the non-native isiZulu phonemic contrast [ǀ]-[ǁ]. Thus, the pattern of decline does not seem to be as universal to all non-native phonemic contrasts. Rather, some non-native contrasts remain perceptually discriminable. Likewise, some native contrasts, that are thought to be less phonetically salient, have been shown to be particularly difficult for infants and thus develop slightly later (e.g., negative voice onset time, see Burnham, 1986). Furthermore, adult research has shown that the perception of non-native contrasts can be improved with training (Hazan, Sennema, Iba & Faulkner, 2005; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999) indicating that adults still retain a certain level of plasticity for speech acquisition.

Such findings have given rise to theoretical frameworks that account for these developmental patterns. It has been suggested that infants acquire their phonemic categories through distributional-learning (Maye, Werker & Gerken, 2002; Maye, Weiss & Aslin, 2008), where infants track the specific features of their ambient input in order to establish phonemic categories. A number of theoretical models have been put forward to explain the developmental patterns observed in infant speech perception. Here, three primary models will be discussed: Kuhl's Native Language Magnet theory (NLM-e, Kuhl, 1994, 2004; Kuhl et al., 2008), Best's Perceptual Assimilation Model (PAM, Best, 1994, 1995; Best & McRoberts, 2003), as well as a more recent model, PRIMIR (Curtin, Byers-Heinlein, & Werker 2011; Werker & Curtin, 2005). It is important to note that this is not an exhaustive review of existing speech perception development models (see also, Attunement Theory, Aslin & Pisoni, 1980; Burnham, 1986; Boersma, Escudero, & Hayes, 2003; Jusczyk, 1992, 1997; Distributional Learning, Anderson, Morgan & White, 2003; Maye, Werker & Gerken, 2002), rather the aim here is to discuss the primary current theories concerned with phonemic development during the first year of life.

Native Language Magnet Theory (NLM-e)

Kuhl's NLM suggests that infants become attuned to the acoustic phonetic details of speech in their ambient environment. In Kuhl's revised version of the NLM theory, NLM-*expanded* (NLM, Kuhl, 1993; NLM-e, Kuhl et al., 2008), Kuhl describes four primary phases of phonemic development that the infant progresses through in the first year of life. During the first phase, infants' perceptual abilities reflect domain-general auditory abilities that allow them to discriminate the acoustic cues of many native and non-native phonemic contrasts. In phase two, native language experience increases infants' sensitivity to the distributional patterns in their native language speech, referred to by Kuhl as native language neural commitment (Kuhl, 2004). Infants become attuned to the sounds of their ambient language and in turn their sensitivity to non-native phonemes is reduced. During this phase, Kuhl et al. (2008) note the importance of social interaction as a key facilitative tool in the processes of perceptual refinement (see e.g., Kuhl, Tsao, & Liu, 2003; Kuhl, 2007; Maye, Werker & Gerken, 2002). Further, NLM-e posits that the stored perceptual information allows the infants to map between their phonetic representations and developing vocalizations, creating a perception and production link. By phase three, language specific phonetic categories begin to develop, referred to as prototypes. These prototypes subsequently warp the infants' perceptual space. It is argued that these language specific skills in turn support later language learning e.g., phonotactic patterns and word learning (Tsao, Liu, & Kuhl, 2004). According to NLM-e, in the final phase of development the infants' phoneme representations are relatively stable, yet they still allow for new phonemic category learning (see e.g., Kuhl et al, 2003).

Perceptual Assimilation Model (PAM)

In contrast to the NLM-e, Best's Perceptual Assimilation model (PAM, Best, 1993, 1994, 1995; Best & McRoberts, 2003), although originally a model proposed for adult perception, recent applications of PAM to infant research posit that infants perceive phonemes in terms of the way their 'dynamic articulatory gestures shape the speech signal' (Best & McRoberts, 2003 pp. 3; Tyler, Best, Goldstein, Antoniou & Krebs-Lazendic, 2008). Based on a direct realist approach (e.g., Browmen & Goldstein, 1989; Goldstein & Fowler, 2003; Gibson & Gibson, 1955), PAM suggests that the decline in non-native perception at around 10-12 months arises because the infants' have started to develop native-language articulatory-phonetic patterns. Thus, PAM proposes that the discriminability of non-native sounds depends on how infants assimilates based on their developing native language. In this way, PAM can account for the findings that have shown that older infants do not always perceive non-native contrast in the same way, i.e., the developmental pattern seems to differ across contrasts (e.g., Best, McRoberts & Sithole, 1988; Polka et al., 2001).

Unlike adults who have developed a native-language phonology, infants are still developing. Thus it is not entirely clear if the assimilation patterns observed in adults apply to young infants. However, it is thought that PAM could provide some insight into how native phonemic categories begin to shape infants' perception. The most recent developmental PAM model (PAM/AO, Best & McRoberts, 2003) incorporates the Articulatory Organ Hypothesis (Goldstein & Fowler, 2003; Studdert-Kennedy & Goldstein, 2003). This hypothesis holds that infants perceive contrasts in terms of the primary articulatory organ that produced it e.g., lips. PAM/AO suggests that infants will display a decline in perception for non-native contrasts that are produced with the same primary articulator, 'within-organ' contrast e.g., lips, /p/-/b/, than for contrasts that are produced with different primary articulators, 'between-organ' contrast e.g., lips vs. tongue, /b/-/d/. For example, Werker and colleagues (1981) demonstrated that English-learning infants display a decline in non-native contrasts that are produced with the same primary articulators e.g., Hindi dental-retroflex contrast (tongue). However, no decline in discrimination has been observed in non-native contrasts that are produced with different primary articulators e.g., dental vs. lateral isiZulu clicks (Best & McRoberts, 2003; Best et al., 1988; see also Tyler et al., 2007).

Processing Rich Information from Multidimensional Representations (PRIMIR)

More recently, Werker and Curtin (2005; Curtin et al., 2011) have proposed the PRIMIR model of speech perception development and word learning. In contrast to the NLM-e and PAM that hold that the main source of information for the perceptual learning process is phonetic i.e., acoustic or gestural phonology respectively, PRIMIR posits that the infant exploits all sources of information, both phonetic and indexical e.g., visual information, voice quality. PRIMIR claims that the infant brings three perceptual biases to process of learning: a preference for infant directed speech, a preference for syllable form and point vowels, and an ability to process rhythm. According to PRIMIR, these initial biases, combined with the infants' developmental level and the task demands, act as filters in the learning process. The PRIMIR framework includes representational spaces and learning mechanisms that drive the process. Specifically, the General Perceptual space stores the language-specific phonetic and indexical information. The Word Form space stores the word form sound sequences, which initially are not associated to meaning. As the word forms increase and the infants vocabulary expands, the Phoneme space emerges, allowing the child to distinguish the phonemic features that are needed to recognize words. As vocabulary grows the phonemic representations become more refined. Importantly, the representational spaces do not develop in a hierarchical fashion; rather they interact throughout development, causing the organization of the system to change as the child phonemic knowledge expands.

To date, there has been less focus on speech perception abilities after the first year of life. However, research suggests that phonemic perception continues to develop up until early adolescence (e.g., Hazan & Barrett, 2000; Nittrouer & Miller, 1997; Mayo & Turk, 2004; Polka et al., 2001; Sundara, Polka, & Genesee 2006; Tsukada, Birdsong, Bialystok, Mack, Flege, 2005). However, as noted by Sundara et al. (2006), understanding the exact developmental trajectory from infancy to childhood is often limited by differences in the tasks employed with each age group e.g., infant headturn procedure vs. identification tasks. Of the few cross-sectional studies that have used the same paradigm across a range of age groups a clear developmental trend can be observed. For example, Sundara et al. (2008) tested monolingual English 10-12 month-olds, 4-year-olds and adults on the English phonemic contrast /d/-/ð/ using an adapted conditioned headturn procedure, where the adults and older children were required to press a button when they detected a change in the stimuli. The results showed that the 4-year-olds were significantly better than the 10-12 month-olds at discriminating the /d/-/ð/ contrast but less consistent than the adults. This suggests that infants and children continue to refine their perceptual skills beyond the first year of life.

Similar developmental trends have been observed for non-native contrasts. Using the adapted conditioned headturn procedure, Werker & Tees (1983) tested 4-, 8-, and 10-year-old children on two Hindi contrasts, the dental-retroflex stops /t^h/-/d^h/ and voiced-voiceless aspirated dental stop /d̪^{h}/-/t̪^{h}/, as well as the English /ba/-/da/ contrast. The results showed no difference between the age groups for all three contrasts. All children were worse at discriminating between the Hindi contrasts than their native English contrast. For the English contrast, no difference was observed between the children, but when compared to 6-8 month-old English-learning infants and English-speaking adults (Werker, Gilbert, Humphrey, & Tees, 1981) the 4-year-olds were significantly worse at discriminating the Hindi contrasts than the infants, but not worse than the adults.

In addition to discrimination tasks, other studies have utilized identification paradigms to investigate children's ability to categorize phonemes. Such tasks require children to go beyond simply identifying a difference between two speech sounds, but instead they test if children perceive two speech sounds as phonemically contrastive e.g., **p**ea vs. **b**ee. Indeed, such experiments are particularly important for understanding the developmental process beyond the first year of life. Such experiments shed light on how children refine these perceptual abilities with age, as well as children's processing abilities for certain acoustic cues. Thus a more fine-grained analysis of speech perception development is presented. However, in comparison to infant research, studies that have investigated speech perception development beyond the first year of life are limited. From the findings to date, it has been hypothesized that during phoneme categorization, children pay attention to different acoustic cues than do adult listeners (Ohde & German, 2011; Ohde & Haley, 1997 Nittrouer, 2005), often referred to as 'perceptual cue weighting' (Nittrouer, 1996;

Nittrouer & Miller, 1997). Nittrouer's Developmental Shift Weighting hypothesis (DWS, 1996), suggests that children are initially more sensitive to more global measures such as formant transitions, as they are considered to be perceptually more salient than static cues such as duration. Evidence for this hypothesis comes from cue weighting studies that have shown that children are more sensitive to formant transition cues in voicing and vowel identification than (e.g., Howell, Rosen, Lang & Sackin, 1992; Ohde & Haley, 1997; Sussman, 1989). Nittrouer (1996) suggests that this pattern of perception is determined by a child's growing lexicon (see also Jusczyk & Derrah, 1987). That is, as a child's lexicon grows they are required to have a more refined representation of speech sounds, resulting in the transition from a more global perception to a specific adult-like pattern. Likewise, Burnham's robust-fragile hypothesis (1986) postulates that salience may influence speech perception development. Specifically, Burnham (1986) suggests that infants' and children's categorization is determined by the acoustic salience of the contrast i.e., robust contrasts are of high word frequency and thus easier to discriminate.

To assess phoneme categorization/identification, tasks generally require the listener to associate a stimulus, either naturally produced or along synthesized continuum, with a particular label or picture. From this, the child's accuracy of labeling a particular phoneme category can be assessed, and in the case of synthesized continua, where certain acoustic cues are modified, the consistency in labeling a specific phonemic category based on the specific acoustic cue e.g., voice onset time, can be assessed. In such paradigms a phoneme categorization slope can be extracted from the child's responses at each point on the continuum. Thus, the gradient of this slope reflects the children's consistency of labeling the continuum i.e., a steeper slope reflects more refined phonemic categories.

Overall, studies using this method have shown that children have significantly shallower identification slopes than that of adults, and in many cases a developmental trend between younger and older children can be observed (e.g., Hazan & Barrett, 2000; Mayo & Turk, 2001, 2004; Nittrouer, 1995, 2005; Ohde & Haley, 1997; Ohde, Haley, & McMahon, 1996). For example, using a forced-choice identification task, Hazan and Barrett (2000) investigated the phonemic categorization abilities of 6-, 12-year-olds and adults on four phonemic contrasts differing in voicing i.e., /g/-/k/, /s/-/z/ and place of articulation i.e., /d/-/g/, /s/-/ʃ/. In addition to investigating a general developmental trend between the age groups, this study tested Burnham's fragile-robust hypothesis (1986) i.e., robust plosive contrasts vs. fragile fricative contrasts. The results showed a significant increase in categorization between the age groups. The 12-year-olds had significantly steeper categorization slopes than did the 6-year-olds, but shallower than the adults. Interestingly, in-line with Burnham's fragile-robust hypothesis, children were more consistent at labeling the robust plosive contrasts than they were the fragile fricative contrasts. Overall, these findings suggest phonemic development continues well beyond the first year of life, and possibly up until adolescence. Similar developmental patterns between children and adults have been found for vowels (e.g., Ohde & German, 2011), as well as other consonants (e.g., fricatives, Nittrouer, 2002; Mayo & Turk, 2004).

1.1.2. Production

As displayed in Figure 1.1 (pp. 3), during the first year of life, infants progress through a series of developmental stages in their speech production. Research suggests that this somewhat reflects their speech perception abilities, and that infants' vocalizations become language specific at a similar time to their perception (see e.g., Kuhl et al., 2008). Of the early developmental stages, the canonical babbling period, around 10-months-old, has been found to be particularly related to later speech development (de Boysson-Bardies, 1993; de Boysson-Bardies, Hallé, Sagart & Durand, 1989; de Boysson-Bardies et al., 1992). In general, research has shown that infants' vocalizations during this stage of development are somewhat 'universal' (Kern & Davis, 2009; Locke, 1983). Plosives, particularly labial and coronal e.g. /ba/, /ga/, are more common than nasals and laterals (Locke, 1983; Roug, Landburg & Lundburg, 1989; Vihman, Ferguson & Elbert, 1986; but see Lee, Davis & MacNeilage 2008 for a different pattern). For vowels, mid, low-front and central vowels are most common (Kent & Bauer, 1985; Lieberman, 1980). Unlike perception, research suggests that the early stages of production development observed in infants' production are constrained by motoric/gestural development (see e.g., Frame-content theory, Davis & MacNeilage, 1995). Thus, much of this variability and restriction to specific sounds articulation patterns is thought to be due to the anatomical development of the vocal tract, often termed 'gestural reorganization' (MacNeilage & Davis, 1993, 2000).

During the later stages of the babbling period infants' vocalizations have been found to reflect that of the ambient language (de Boysson-Bardies et al 1984, 1989; Lee, Davis & MacNeilage, 2008). For example, de Boysson-Bardies and colleagues

(1989, 1984) compared the vowel and consonantal sound distributions of French, English, Cantonese and Algerian-learning infants. An acoustic analysis revealed a significant difference for vowels and consonants across the four languages, where the infants' productions were significantly related to their language-specific inventories. In addition, more recent longitudinal research has demonstrated that the consistent vocalizations (referred to as 'vocal motor schemes' by Vihman, e.g., Vihman & Croft, 2007) are directly related to the sounds produced in their first words (Davis & MacNeilage, 1995, 2000; McCune & Vihman, 2001; MacNeilage, Davis, Kinney & Matyear, 2000). For example, Davis and colleagues in series of cross-linguistics studies showed that the sounds produced during the babbling predict sounds produced in the child's first words (Davis & MacNeilage, 1995; Davis, MacNeilage & Matyear 2000). In general it seems that, as for perception, infants' early productions are related to their ambient language, as well as later word productions (but see, Kern, Davis & Zink, 2009).

However, much of the research on production beyond the first year of life focuses on developmental norms, including the age of acquisition of sounds (e.g., Dodd, Holm, Hua & Crosbie, 2003; Smit, 1990), the specific acoustic correlates and how these change with age e.g., VOT patterns (Macken & Barton, 1980), and error patterns (e.g., Dodd, 1995; Bankson & Bernathal, 1998) that are present in children's whole word productions. As in babbling, much variability has been found in children's early productions, and this variability seems to reduce with age (Nittrouer, 1993; Palethorpe, Wales, Clark & Senserrick, 1996; Stoel-Gammon & Dunn, 1985; Vihman, 1996). In general, more research has been conducted on consonants than vowels. For consonants, bilabial sounds e.g., /b/, /p/, /m/, /n/, are acquired first, at around 2-3 years-old, whereas fricatives and affricates are the last sounds to be fully acquired, as late as 7 years old (see Table 1.1 for summary of the studies, Dodd et al., 2003; Poole, 1934; Prather, Hendrick & Kern, 1975; Smit, Hand, Freilinger, Bernthal & Bird 1990; Templin, 1957; Wellman, Case, Mengert & Bradbury, 1931). For vowels, Templin's study on the acquisition of vowels by English learning children suggests that vowels are acquired by 3 years of age. More specifically, Allen and Hawkins (1980) claimed

Table 1.1. Summary of English phoneme age of acquisition studies, adapted from Dodd et al., (2003). Note, *Smit 1990, f = female, m = male

	Welman (1931)	**Poole (1934)**	**Templin 1957**	**Prather (1975)**	**Smit (1990)***	**Dodd (2003)**
/m/	3	3;6	3	2	3	3;0 – 3;5
/n/	3	4;6	3	2	3;6f, 3m	3;0 – 3;5
/ŋ/	> 6	4;6	3	2	7-9	3;0 – 3;5
/p/	4	3;6	3	2	3	3;0 – 3;5
/b/	3	3;6	4	2;8	3	3;0 – 3;5
/t/	4	4;6	6	2;8	4;6f, 3;6m	3;0 – 3;5
/d/	4	4;6	4	2;4	3f, 3;6m	3;0 – 3;5
/k/	4	4;6	4	2;4	3;6	3;0 – 3;5
/g/	4	4;6	4	3	3;6f, 4m	3;0 – 3;5
/f/	3	5;6	4	2;4	3; 6 I; 5;6 F	3;0 – 3;5
/v/	5	6;6	3	> 4	5;6	3;0 – 3;5
/θ/	> 6	7;6	6	> 4	6f, 8m	> 7
/ð/	6	6;6	6	4	4;6f, 7m	> 7
/s/	5	7;6	7	3	7-9	3;0 – 3;5
/z/	5	7;6	4;6	> 4	7-9	3;0 – 3;5
/ʃ/	-	6;6	7	3;8	6f, 7m	5;0 – 5;5
/ʒ/	6	6;6	4;6	4	-	4;0 – 4;5
/tʃ/	5	-	4;6	3;8	6f, 7m	4;0 – 4;5
/dʒ/	6	-	7	> 4	6f, 7m	4;0 – 4;5
/l/	4	6;6	6	3;4	5f, 6m I; 6fm 7m F	3;6 – 3;11
/ɹ/	5	7;6	4	3;4	8	3;6 – 3;11
/w/	3	3;6	3	2;8	3	3;6 – 3;11
/j/	4	4;6	3;6	2;4	4f, 5m	3;6 – 3;11
/h/	3	3;6	3	2	-	3;0 – 3;5

that children accurately produce vowels in stressed syllables by 3-years-old, however vowels produced in an unstressed syllable develop later, between ages 4-5-years.

In general, acoustic-phonetic studies have demonstrated that children's speech production is higher pitched than that of adults, with higher frequency values and with more variability than that of adult speech (e.g., Nissen & Fox, 2005). Of the few acoustic analyses of the development of children's vowel production, children have been shown to produce speech using higher formant values and variability than adults (Eguchi & Hirsch, 1969; Gilbert, 1973; Lee, Potamianos & Narayanan, 1999; Lieberman, 1980; Palethrope, Wales, Clark & Senserrick, 1996). For example, in a large study of vowel production by 437 children aged 5 to 17 years old, Lee et al (1999) found that the formant frequency variability did not reflect adult values until around 14 years of age. For consonants, a similar pattern of development is observed; studies have shown significant differences between adult and child speech (Dalston, 1975; Nissen & Fox, 2005). For example, Nittrouer (1995) found that adults displayed a larger spectral mean in their production of the voiceless fricatives /s/ and /ʃ/ than did children aged 3 to 7 years old.

Of interest in this research is the production of plosives. From an acoustic perspective, the plosive contrast can be differentiated by the presence or absence of voicing e.g., /b/-/p/. In their pioneering cross-linguistic study of word-initial plosives, Lisker & Abramson (1964) defined VOT as "the time interval between the burst that marks the release of the stop closure and the onset of quasi-periodicity that reflects the laryngeal vibration" (p. 422). They demonstrated that VOT was a reliable cue with which to distinguish the voiced-voiceless plosive contrast in English, especially in word-initial position e.g., ***p**ea* - ***b**ee*. In turn, the voicing patterns of both adult and child speech production has been a major cross-linguistic research topic over the last few decades (see e.g., Davis, 1995; Kehoe, Lleó, & Rakow; Khattab, 2000). It is important to note that other acoustic cues such as F1 transition and frequency at voicing onset are also important in the perception of the voicing contrast (Stevens & Klatt, 1974). That is, plosives with a longer F1 transition and/or lower F1 onset frequency are classified as voiced, and vice-versa for voiceless plosives.

Macken & Barton (1979) outlined 3 stages of production development for English stops; 1) children initially produce all plosives in the short-lag region, 2) a distinction between voiced and voiceless plosives is made, though it may not be perceivable by adult listeners, and 3) at around 2-years-old, children produce target voiced short-lag and voiceless long-lag plosives, where the long-lag has been found to be more extreme than in adult production. In contrast, studies of languages that contain prevoicing (e.g. French & Spanish) suggest that this develops slightly later, and may not be acquired until 5 years old or later (Macken & Barton, 1979; Zlatin & Koenigsknecht, 1976). It has been suggested that this is because the articulatory demands of prevoiced production i.e., the coordination of laryngeal control with supralaryngeal articulatory gestures, make it difficult for children to master (Kewley-Port & Preston, 1974).

1.3. Learning two languages from birth: the case of simultaneous bilinguals

A bilingual child is faced with the challenge of acquiring the phonology of two languages. In turn, they are required to map the different and possibly overlapping distributions to later phonological development and word learning for two languages (see e.g., Fennell, Byers-Heinlein, & Werker, 2007; Werker, Byers-Heinlein & Fennell, 2009). To date, our knowledge of bilingual speech development is scarce, and is predominantly limited to cross-sectional research on infants and on adult studies. Thus, little is known about the exact developmental trajectories of bilinguals. In general, research on simultaneous bilingual adults has shown that they are able to discriminate contrasts from both languages, and produce distinct categories for each language (e.g., Sundara, Polka & Baum, 2006). However findings suggest that the patterns of perception and production of simultaneous bilinguals are not identical to their monolinguals counterparts for each language. For example, in an acoustic analysis of coronal stops produced by Canadian French, English and bilingual French-English speakers, Sundara et al (2006) found that the bilinguals produced the stops with language specific VOT values, however, these values were not exactly the same as those found in French and English monolingual speakers.

Research on simultaneous infant and child bilinguals suggest that they typically acquire the phonemes of both languages (e.g., Bosch & Sebastián-Gallés, 2003a, 2003b; Burns, Yoshida, Hill & Werker, 2007; Sundara, Polka & Molna, 2008). However, research to date has shown two main patterns of development. Some studies indicate that in the first year of life bilingual infants display similar patterns to their monolingual peers, in that they are able to discriminate contrasts that exist in both of the languages being learned (e.g. Sundara et al., 2008; Sundara & Scutellaro, 2011).

For example, Sundara et al. (2008) tested 6-8 and 10-12 month-old bilingual French-English infants and their monolingual English and French peers on their ability to discriminate between French /d/ i.e, dental, and English /d/ i.e., alveolar. The findings showed that 6-8 month monolinguals and bilinguals were able to discriminate [d̪] and [d]. However, by 10-12 months only the French-English bilinguals and English monolinguals (similar to that of monolingual English adults) were able to distinguish the contrast. Such patterns suggest that, like monolingual infants, a simultaneous bilingual child can successfully track the distributions of both languages in their ambient environment.

Other studies have shown that bilingual infants display different developmental trajectories (e.g., Bosch, & Sebastián-Gallés, 2003a, 2003b; Burns, Yoshida, Hill & Werker, 2007). For instance, Garcia-Sierra et al. (2011) investigated the discrimination of the Spanish and English /t/-/d/ contrast using an ERP paradigm. In contrast to previous studies, bilinguals in this experiment showed no neural discrimination of Spanish or English contrasts at 6-9 months of age, unlike their monolingual English peers, but by 10-12 months were able to discriminate both contrasts. Similar patterns of perception have been found for vowels. For example, in a series of experiments that investigated vowel perception by Spanish-Catalan bilingual infants (e.g., Bosch & Sebastián-Gallés, 2003a; Sebastián-Gallés & Bosch, 2009) a U-shaped pattern in development was observed: 4-month-old and 12-month-old bilinguals are able to discriminate acoustically similar sounds in their two languages e.g., /o/ and /u/ (i.e., Sebastián-Gallés & Bosch, 2009) and /e/ and /ɛ/ (i.e., Bosch & Sebastián-Gallés, 2003a, 2003b). In contrast, 8-month-olds were unable to discriminate the contrasts.

There have been a number of explanations postulated for the disparities found in bilingual infant data. One possibility is that a bilingual child may need more time to accumulate the amount of experience required to become attuned to the phonetic differences in the two languages being learned (e.g., Garcia-Sierra et al., 2011; Kuhl et al. 2008). Specifically, Werker and colleagues (2009) suggested that U-shaped pattern in perception may be due to a ‘temporary broadening of phonetic categories’ (p. 3652), that allows the bilingual infants to reorganize their perceptual space to account for both languages. Another possibility, is that the developmental differences in bilinguals at 6-9 months may be due, in part, to the phonetic similarities between the languages being learned, as well as the amount of variability (e.g., dialectal variants) infants are exposed to in each language. For instance, Sundara & Scutellaro (2011) have suggested that Spanish-English bilingual infants are aided in their acquisition of acoustically similar vowel contrasts that are phonemic in one language but not the other, because Spanish and English are rhythmically different (syllable-timed vs. stress-timed, respectively).

Recently, models of speech perception development have started to account for bilingual development. Based on the bilingual speech perception and word learning research from the last decade Curtin et al. (2011) presented the PRIMIR-*in focus* model of speech perception and word learning. Here, Curtin and colleagues extend the PRIMIR model (see section 1.1) to account for bilingual development. The PRIMIR-*in focus* posits that bilingual infants posses the same representation spaces, dynamic filters, and learning mechanisms as monolinguals. However, in addition, the model introduces a new learning mechanism that is needed by bilinguals to organise the information from both languages, namely the Comparison-Contrast mechanism. This mechanism combined with the multidimensional space described in PRIMIR (Werker & Curtin, 2005) allows the bilingual infant to organise the sounds of his/her language. In turn, as shown in research to date, by the end of the first year of life, bilingual infants are able to discriminate the sounds of his/her languages. In terms of the Word-Form space, PRIMIR-*in focus* holds that language dominance will likely play a key role for bilingual infants, where more word representations will develop in the infant's dominant language. According to PRIMIR-*in focus* the words of each language are stored in separate distinct clusters. Finally, PRIMIR holds that phonemes emerge in the Phoneme space once the child has enough 'word-meaning linkages' in their representations. In the case of monolinguals the growth in vocabulary assist with the development of distinct phonemic categories. Curtin et al. (2011), note that although bilinguals have a similar vocabulary size to that of their monolingual peers when the words from both languages are combined (see e.g., De Houwer, Bornstein & De Coster, 2006), bilingual children tend to have a smaller sized vocabulary in each language (see e.g., Bailystok, 2009). In turn, PRIMIR-*in focus* predicts that infants and children's phoneme categories will be in part determined by their vocabulary size in each language. Thus, phoneme acquisition may be slightly later for bilinguals. Evidence for this comes from studies have shown that bilingual toddlers are less sensitive to word mispronunciations i.e., phoneme changes, in their two languages than their monolingual peers (see e.g., Fennell, Byers-Heinlein & Werker 2007; Ramon-Casas, Swingley, Sebastián-Gallés & Bosch, 2009).

The findings of the development of bilingual production are also mixed, and often limited to small sample sizes. For example, a study of three English-Arabic bilingual children (aged 5, 7 & 10 years) demonstrated that all children developed the voicing contrast for English plosives, but not for Arabic (Khattab, 2000). Specifically, the youngest and oldest bilingual child did not use Arabic prevoicing for voiced stops. As suggest by Khattab, the amount of input may have played a role in these findings. That is, at the time of the recordings all children received an increase in Arabic input, the oldest child may have been past the critical age for the acquisition of prevoicing. Production research on younger bilingual children suggests that a bilingual child may also develop merged categories that incorporate both languages. For example, Johnson and Wilson (2002) found that a 2;10 year-old Japanese-English bilingual child produced both English and Japanese plosives in the English voicing region. However, 4;8 year-old Japanese-English bilinguals were able to separate these categories, albeit different from their Japanese or English monolinguals peers.

1.3. Learning a second language later in life

1.3.1. Age effects

It is well established that to successfully acquire an L2 sound system, earlier is better as far as the age of acquisition is concerned (e.g., Flege, Munro & MacKay, 1995; Flege, Schirru & MacKay, 2003; Munro, Flege & Mackay, 1996). Unlike simultaneous bilinguals who typically acquire the sounds of both of their languages in a native-like way (e.g., Baker & Trofimovich, 2005; Sundara, Polka, & Baum, 2006), late L2 learners (i.e., those who acquire their second language as adults), have difficulties acquiring the sounds of their L2 (e.g., Flege et al., 2003, MacKay, Flege, Piske, & Schirru, 2001). Within large multiethnic cities such as London there are substantial pockets of immigrant groups, made up of speakers who acquire their heritage language and the language of the host country at different ages and in different environments. Some speakers are born in the home country and arrive late. These speakers often arrive in their 20s or later, i.e., beyond the assumed critical period, when L1 acquisition is thought to be complete (e.g., Flege, Mackay & Meador, 1999). Initially they will have acquired and used their heritage language (L1) in their home country and they only start to acquire the host language (L2) on arrival, i.e., at a later stage in development. Others arrive in the host country at an early age i.e., Early

L2 learners. These speakers typically use only their heritage language up until they arrive in the host country, but acquire the host country language on arrival, early in their development. Still others are born in the host country. These speakers initially acquire their heritage language from their parents, but are then immersed in host country's language when they start preschool. They may also have some experience with English from birth, e.g., through older siblings, the media. These learners are typically referred to as sequential bilinguals.

As well as differing in their early experiences with the host language, these groups also differ in their usage of their heritage and host language. The dominant language in these communities is likely to be their heritage language rather than that of the host-country (see e.g., Fox, 2010; Rasinger, 2007). The language of the home and its use is supported through local shops, markets, as well as cultural and religious practices (e.g., Boishakhi Mela in Tower Hamlets London, to celebrate Bengali New Year). Patterns of usage however, differ between groups. Often those who are considered to be late learners will continue to use the heritage language as their dominant language at home, within the community, and in some cases at work (e.g., Li Wei, Milroy & Pong, 1992; Rasinger, 2007). In contrast, those who arrive early in life or are born in the host country such as the UK, attend English speaking preschool and full-time school from a young age, and in turn tend to develop ties outside of their heritage community (e.g., Sharma & Sankaran, 2011). Their usage of English is therefore more extensive; they use English with a wider variety of interlocutors. Indeed, although both these groups acquire their heritage language initially, they typically become dominant in their L2. This results in a highly complex linguistic community; speakers differ not only in their early experience with the host language, but also differ in their usage of their heritage and host languages

One could imagine that the differences in patterns of language acquisition result in differences in the perception and production of both the L1 and L2. To date, research has shown that for perception and production, early learners generally pattern more with native language learners more than with late learners (Flege, Mackay & Meador, 1999; Flege & Mackay, 2004; Højen & Flege, 2006; Iverson & Evans, 2007; Tsukada et al., 2005). For example, Tsukada et al. (2005) compared adult and child Korean learners of English in their perception and production of English vowels. All subjects had been resident in North America for 2-4.9 years. For perception, Korean children were able to discriminate English vowels more accurately than Korean adults, but less accurately than native English-speaking children. For production, Korean children were judged to be more native-like than Korean adults, and did not differ from the native English-speaking children. Likewise, studies that have compared adult L2 speakers who varied in their age of arrival have shown similar patterns of perception (see e.g., Flege et al., 1999)

Similar age effects have been found in production (e.g., Baker & Trofimovich., 2005; Flege et al., 1995 Guion, 2003; Sundara, et al., 2006). Guion (2003) found that late Quichua-Spanish L2 learners used their existing Quichua vowels when producing Spanish vowels. Interestingly, although the early L2 learners produced distinct vowels for both Spanish and Quichua, they produced them differently from their monolingual counterparts. Flege (1991) investigated the production of English /t/ by two groups Spanish learners of English; those who started acquire English before 6 years of age (early learners) and late learners who began learning English in adulthood, with a mean age of 19 years. In Spanish, plosives are produced using a shorter VOT, including prevoicing. The results showed that the late learners produced the English stops with shorter VOT values than did the monolingual English speakers, whereas no significant difference was found between the early L2 speakers and monolingual English controls.

Theories of second language acquisition have proposed that these age effects are due to differences in phonological organisation. Flege's Speech Learning Model (SLM; Flege, 1995, Flege et al., 2003) suggests that the difficulties arise because the L1 and L2 phonemes exist in a common "phonological space", rather than in two separate phonological systems, i.e., when the new L2 categories are similar to the existing L1 categories. The SLM claims that this is because L2 learners are likely to use their existing L1 categories when producing the sounds of their L2, thus resulting in accented speech. It is the perceived similarity or difference between the listeners' L1 and L2 categories that are thought to determine this effect. Therefore, L2 sounds are thought to be easier to acquire if they fall into an unoccupied space, i.e., are dissimilar to the sounds in the L1. In the case of a similar L2 sound, the listener is first required to detect the differences between the two categories. Importantly, according to the SLM, the ability to acquire L2 sounds remains intact throughout life (Flege & Mackay 2004) and age effects arise because L1 categories become more established with age (see also Iverson et al, 2003).

As discussed earlier in this literature review, Best's Perceptual Assimilation Model (PAM; Best, 1994; Best, McRoberts & Goodell, 2001) is concerned with how adult listeners perceive non-native phonemic contrasts. PAM suggests that L2 listeners perceptually assimilate non-native phonemes to the native phonemes that they perceive as most similar, and that these patterns of assimilation are driven by the relationship between the L1 and L2 phonemes. Best (1994; Best & McRoberts, 2003; Best & Tyler, 2007) outlines different types of assimilation patterns that can arise depending on the L1 phonemic categories of the L2 listener. Specifically, L2 discrimination is thought to be native-like in the case of Two-Category (TC) assimilation. Here, each of the L2 phonemes corresponds to an existing L1 contrast. For example, German listeners are relatively good at discriminating the English /i/-/ɪ/ contrast (Bohn & Flege, 1990; Flege et al, 1997). This is likely due to the similarities to the German vowel inventory, where German contains a similar vowel contrast. However, Spanish listeners find this English contrast particularly difficult (Escudero & Boersma, 2004). Such a pattern is an example of what Best describes as Single-Category (SC) assimilation, where both English /i/ and /ɪ/ assimilate equally well to the single Spanish /i/ phoneme. Category-Goodness (CG) assimilation occurs when both phonemes map onto a single native category, however one phoneme is slightly different. Here, discrimination is thought to be moderate to very good (see e.g., Best & Strange, 1992). In cases where the L2 contrast cannot be mapped onto any native contrast, Uncategorized-Uncategorized (UU) assimilation, the discrimination abilities are thought vary depending on the target contrast. When one phoneme of an L2 contrast is categorized to a native phoneme but the other is not (Uncategorized-Categorized, UC), perception is thought to be good (see e.g., Polka, 1992). Finally, in cases where both phonemes in the L2 contrast are not perceived as any L1 phoneme, Non-Assimilable (NA), perception is predicted to be very good, as both phonemes are perceived as non-speech sounds e.g., isiZulu clicks by English listeners (Best et al., 1988).

1.3.2. The role of social factors

Within immigrant communities the effects of language experience on the acquisition of both the host and heritage language cannot easily be isolated from the effects of sociocultural factors. The age at which the speakers start to acquire the host country language (i.e., L2) is likely to be influenced by their language experiences in the home country, such as knowledge of the L2 prior to arrival and the social networks they develop in the host country, as well as how the groups differ in their identity, e.g., whether or not they identify more with their host country (typically those born in UK and are early learners) or less (typically late learners).

Previous sociolinguistic studies have suggested that identity and social networks play a role in the use of specific phonetic variants in speech production (Fox, 2010; Samant, 2010). For example, Fox (2007; 2010) investigated the use of the PRICE and FACE vowels by Bangladeshi, mixed-race and white British adolescents in the East End of London. The study found that Bangladeshi boys had adopted variants that are not typically associated with the traditional East London variety of English (see e.g., Wells, 1982). Specifically, For PRICE, the Bangladeshi boys predominately used [æ] [aɪ] and [ɐɪ], and for the FACE vowel they used [e̝ɪ], [eɪ] and [ɛɪ]. In contrast, the white British boys and girls used the more traditional [ɑ̝ɪ], [ɑɪ], [ɑː] variants for the PRICE vowel, and, [æɪ], [aɪ] variants for the FACE vowel. A detailed analysis of the speakers' friendship groups and social practices revealed that such factors were strongly related to the use of the variants. For example, the white British girls who had little contact with the Bangladeshi boys did not use the non-traditional variants that were used by the Bangladeshi boys, whilst the white British boys who had more social interaction with the Bangladeshi boys had adopted these variants.

At a broader level, it might be the case the speakers who reside in large multicultural cities adopt a 'multiethnolect', that consists of phonetic features that reflect the ethnic make-up of their community e.g., Multicultural London English (Cheshire, Kerswill, Fox & Torgersen, 2011). Furthermore, previous research has shown that some speakers acquire a repertoire of phonetic variants from their heritage and host language, which they use to selectively adapt their accent according to their interlocutor (Sharma, 2011; Lambert, Alam, & Stuart-Smith, 2007).

Similar factors and language use patterns are likely to influence the production of the heritage language (e.g., Pallier et al., 2003). Although all speakers will acquire the heritage language as an L1, it is likely that the exposure and usage patterns will be very different (e.g., Flege et al., 1995; Flege et al, 1999). For example, Khattab's (2000) study of bilingual Arabic children discussed earlier provides some insight into how the heritage language may be established by speakers who are born in the host country. Khattab found that all children produced English stops with a VOT similar to their monolingual English peers. However for Arabic, only the 5-year-old produced voiced Arabic stops with the correct voicing lead. Khattab attributed this finding to an increase in Arabic input for that child. In contrast, children who grow up in a dense immigrant community such as the London Bengali community are likely to be exposed to the heritage language within and outside the home from a young age. It seems likely then that as a result of this relatively large amount of input, both early arrivals and those born in the host country will establish separate phonetic categories for both the heritage and host country language.

However, this picture is complicated by the possibility that the host language influences the production of the heritage language. For example, in a speech production study of the UK Bradford Panjabi community, Heselwood & McChrystal (1999) found that second-generation speakers (i.e., those born in the host country) produced voiceless and aspirated stops in a similar way to speakers who acquired Panjabi in Pakistan before coming to the UK (late arrivals). However, for the voiced Panjabi stops (which are realised with a prevoicing in Panjabi), the second-generation speakers used a voicing pattern that was similar to British English voiced stops, i.e. short lag. The authors suggest that such patterns in production may either be due to the influence of English categories on the production of Panjabi, or be the result of a broad linguistic change that is occurring in the Panjabi spoken in Bradford. Late arrivals also showed fewer transfer effects than second-generation speakers, an effect which could be linked to those born in the host country becoming dominant in their L2.

But what about speakers who are born in the home country and arrive during childhood (early learners)? Like second-generation speakers, early arrivals will acquire the host country language during childhood, mainly at school. Thus, we might expect them to behave much like those born in the host country. However, the early experience in the home country and the arrival in the host country with a family less established within the host community, might lead them to use their heritage language more. That is, they might have fewer contacts outside their heritage community, identify more with the heritage community, and in turn show production patterns similar to that of late arrivals. In sum, within immigrant communities it is likely that speakers will differ in their usage patterns and early experiences. As a result, the different speaker groups may display different patterns of perception and production in both their L1 and L2.

1.3.4. The role of input

In addition to the differences in language use patterns, speakers who reside in immigrant communities are also likely to differ in the amount and quality of language that they are exposed to (see e.g., De Houwer, 2009; Werker & Byers-Heinlein, 2008). As discussed in the previous section, speakers who arrive as adults are likely to have fewer contacts that are native speakers of the host country language than those who arrive in childhood, who will enrol in full-time education and have more native contacts than late arrivals (see e.g., Wei et al., 1992). Although the L2 literature acknowledges the possible influence of input on L2 speech learning (see e.g., Flege & Piske, 2009; Flege, Yeni-Komshian and Lui, 1999), few studies have attempted to investigate this relationship directly. For adults this is particularly difficult as it is often confounded by other factors such as age of arrival (AOA), language use and length of residence (LOR). The findings from Flege & Lui (2001) on late-learners provide some insight into the effect of input on L2 learning. In this study tests of L2 comprehension, grammar sensitivity and final phoneme identification were administered to a group of Chinese L2 Late learners. The participants were grouped according to LOR (mean short: 2;7 vs. long: 6;6 years) and occupation (student vs. non-student). When compared in terms of LOR no significant difference was found between the groups. However, when occupation was taken into account a significant difference between the student and non-student group was found. That is, the student long LOR had the highest L2 scores, in contrast to the non-student short LOR group. The authors attribute these findings to the input-rich L2 environment to which the student long group will likely have been exposed.

Likewise, research on childhood bilinguals has shown that the exact amount exposure in both languages is significantly related to language outcomes (De Houwer, 2009; Hoff, 2006, Hoff, Core, Place Rumiche, Senor & Parra, 2012; Place & Hoff, 2011). In previous studies, exposure has been indexed by the percentage of each language used with the child. This data is normally collected in the form of a detailed caregiver interview, regarding the proportion of the child's waking hours that the child is exposed to the differing languages (e.g., Parra, Hoff & Core, 2011), or more recently through detailed caregiver diary records of daily language exposure (Place & Hoff, 2011). For example, Place & Hoff (2011) used this method to record which language was used with the child during every 30 minutes of the child's waking hours over the course of a week. A subsequent analysis found that the diary data was significantly correlated with the parental reports. In general, parental language use has been shown to be a significant predictor of language development patterns in bilingual children. For example, De Houwer (2004) found that parental language use accounted for 84% of the variance in the bilingual children's language development patterns.

To date, the majority of this research is limited to vocabulary, grammar, and general language development (see e.g., Conboy & Mills, 2006; De Houwer, 2007, 2009, 2011; Place & Hoff, 2011; Pearson, Fernandez, Lewedeg & Oller, 1997), which has shown that the language development of bilingual children is directly linked to the amount of exposure in each language. For example, Hoff et al. (2012) tracked the vocabulary and grammar development of Spanish-English bilinguals aged 1;10, 2;1 and 2;10 using the English and Spanish version of the MacArthur-Bates Communicative Development Inventory (CDI, Fenson et al., 1993; El Inventario del Desarrollo de Habilidades Communicativas, IDHC, Jackson-Maldonado et al., 2003). They found that the proportion of English exposure was positively correlated with the English outcomes and negatively correlated with Spanish outcomes. That is, the bilingual children with the highest English exposure had English language outcomes closer to that of their monolingual peers, however they had the lowest scores on the Spanish outcome measures.

Fewer studies have looked at the effect of language exposure on speech perception and production. A recent study by Garcia-Sierra et al., (2011) investigated the influence of language exposure on simultaneous bilingual infants' phonetic discrimination abilities. Specifically, 6-12 month-old Spanish-English learning infants discrimination of Spanish and English /da/-/ta/ VOT contrast was assessed (where Spanish has a shorter VOT than English) using event-related potential (ERP) measures. Language exposure was measured using a bilingual questionnaire (Conboy & Mills, 2006). The questionnaire gathered information regarding the amount of interaction the infant had with each individual in the home as well as the languages used with the infant (i.e., English, Spanish or Both). Exposure was indexed using a rating-scale i.e., 1-10, low-high exposure. The results showed that high exposure in both Spanish and English was associated a significant change in amplitude in ERP responses (i.e., better discrimination) with age.

In addition to the amount of language exposure, a bilingual child is likely to hear both of the languages from native (e.g., teachers, peers) and non-native speakers (e.g., caregivers, uncles), see e.g., Fernald (2006). Depending on the background of their main caregivers they are likely to be exposed to accented variants in the host country's language and possibly in their L1 (e.g., from second-generation speakers). Research on the influence of accented input on phonemic development has only recently started to develop over the last decade, and is primarily focused on monolingual infants, and often investigating the child's ability to process dialectal variations (e.g., Best, Tyler, Gooding, Orlando & Quann, 2009; Cristià, 2011; Schmale, Cristià, Seidl, Johnson, 2010; Schmale & Seidl, 2009), with some work on young children (e.g., Nathan, Wells & Donlan, 1998; Nathan & Wells, 2001). For example, Cristià (2011) investigated 12-14 month-old English speaking infants' sensitivity to fine-grained subphonemic differences in caregivers production of /s/ using the Visual Habituation procedure. Cristià showed that the specific acoustic characteristics of the caregivers' /s/ productions were significantly related to the infants ability to discriminate the /s/-/ʃ/ phonemic contrasts. That is, infants of caregivers' who produced an acoustically more extreme /s/ i.e., making the /s/-/ʃ/ contrast more distinct, were significantly better at discriminating the /s/-/ʃ/ contrast. This suggests that infants are possibly sensitive to fine grained variations in caregivers' speech.

Studies have shown that infants can deal with highly variable speech (e.g., Kuhl et al., 2008) as well as accented speech from around 14 months (e.g., Schmale & Seidl, 2009). Such findings raise interesting questions regarding simultaneous and sequential bilinguals. A sequential bilingual who grows up in an immigrant community is likely to be exposed to accented L2 speech input, especially if his/her main caregivers are late-arrival L2 learners. Possible issues arise when this accented speech is linguistically meaningful i.e., affecting phonemic categories i.e., English /i/-/ɪ/ contrast, rather than allophonic e.g., clear vs. dark /l/. In such cases, it could be suggested that the infant, if only exposed to the L2 from his/her main caregiver, may not initially acquire the L2 contrast e.g., the caregiver production of *sheep* and *ship*: /sh**i**:p/ -> [sh**i**p] and /sh**ɪ**p/ -> [sh**i**p].

In sum, previous research suggests that sequential bilingual children who grow up in an immigrant community may, at least initially, display different patterns of development to their monolingual and simultaneous bilingual peers. It could be suggested that, as shown in monolingual research, these children will initially be sensitive to the phonemic contrasts in their ambient heritage-language (L1). In turn, when these children start to acquire the host country's language (L2), on entry to full-time education, they may display difficulties with phonemic contrasts that do not exist in their first language. With time, these children will accumulate experience in the L2. As shown in second language research, it is likely that these children, referred to as early-learners, will eventually acquire the phonemes of their L2. However, little is known about the developmental trajectory of such children. Further, as discussed in the previous sections, the language environment of these children is likely to be further complicated by differences in language exposure of the L1 and L2, as well as exposure to accented variants of the L2.

This thesis will focus on children who grow up in the London-Bengali community. As will be discussed in the sections to follow, the London-Bengali community is a particularly interesting community to study for several reasons. Not only does it represent one of the largest immigrant communities in the UK, it also has a high number of Bangladeshi-heritage children under the age of 10-years-old. The continuation of emigration to the UK means that these children often have first-generation parents (i.e., late learners). Further, it is common for extended family members to be living in the home e.g., grandparents, uncles and aunts from Bangladesh. In turn, this creates a rich heritage language environment for the children. From a phonetics perspective, the heritage language, Sylheti, provides an interesting comparison with English. In terms of vowels, Sylheti is thought to have fewer vowels than English, where specific English phonemic contrasts do not exist in Sylheti. In term of plosives, Sylheti utilizes different voicing patterns to that of English. The sections to follow will provide a brief history of the London-Bengali community, as well as a description of the phonetics of English and Sylheti.

1.4. This thesis: London-Bengali community

1.4.1. A brief history of the London-Bengali community

The majority of London-Bengalis originate from the rural district of Sylhet, in the northeast of Bangladesh, where the local vernacular is Sylheti, an Indo-Aryan language. The Bangladeshi community is the fourth most common ethnic group in London [Office of National Statistics (ONS), 2009]. The majority of London-Bengalis reside in the London boroughs of Newham, Tower Hamlets, and Camden. This thesis focuses on the latter two boroughs: Tower Hamlets, where 30% of the population are of Bangladeshi heritage (Tower Hamlets Council, 2012), and Camden, where people of Bangladeshi heritage are the largest minority ethnic group (Camden Key Facts, 2012). See Figure 1.2 for the location of the target boroughs.

Relevant for the research reported in this thesis is the high proportion of Bangladeshi-heritage children in both of the target boroughs. In Tower Hamlets, 64% of primary school aged children (under 10 years of age) are of Bangladeshi-heritage (Spring School Census, Tower Hamlets 2012). In Camden, Bangladeshi-origin children are the largest minority group accounting for more than 17% of primary school aged children (GLA, 2011). It is important to note that in contrast to Tower Hamlets, where the Bangladeshi community are dominant throughout, the Camden Bangladeshi community seem to be dominant in specific pockets of the borough, resulting in differences in proportion of Bangladeshi children in schools. This thesis focuses on the children who reside in Bangladeshi dominant areas, who attend schools where 50-82% of the children are of Bangladeshi heritage.

Figure 1.2. Map of London and Greater London illustrating the location of the London Bangladeshi community: the London boroughs of Camden, Tower Hamlets and Newham.

Tower Hamlets is often referred to as 'Bangla Town', and the close-knit nature of the community has been highlighted by many researchers (e.g. Fox, 2010; Husain, 1991; Kershen, 2000; Rasingner, 2007). Continuous migration since the 1950s has resulted in a community made up of first-, second- and subsequent generations. The post Second-World War labour shortage in Britain resulted in the government encouraging people from former British colonies to come to Britain as migrant workers. The upheaval in Bangladesh, due to the partition of India and East Pakistan in

1947 meant that Sylhet, on the border of this partition, no longer had access to the trade and work from the Indian ports of Kolkata and Bombay, in turn this encouraged the migration to the UK for economic reasons (Asgar, 1966). The East End of London, being the 'point of arrival' (Bermant, 1975) for many immigrants (e.g., Huguenots and the Irish, motivated by the surrounding East End docks), received an influx of Bangladeshi immigrants (Adams, 1987). As a result, from 1950 to 1980 the number of Bangladeshi immigrants in the UK grew by 30 percent, from 6,000 to more than 162,000 people (Eade, 1997). At first, the Bangladeshi community was male dominant, the 'myth of return' to families Bangladesh meant that the wives and children originally remained in Bangladesh (Anwar, 1979). It was not until the 1970s that families started to join the men resulting in a peak in immigration.

By the 1980s Bangladeshis were dominant in Tower Hamlets. In 1984, 46% of live births in Tower Hamlets were to Bangladeshi mothers (Foreman, 1989). The men continued to work in the community and had the most contact with non-Bangladeshis, whereas many of the women had little contact outside of home or family network (Eade, Vamplew & Ceri, 1996). In the years to follow, Tower Hamlets remained the center for the London Bangladeshi community, but also spread neighbouring boroughs of Camden and Newham. Although migration to the UK has reduced over the years, the practice of arranged marriages in the Bangladeshi community has resulted in continued migration to the UK. Thus new first-generation Bangladeshi born men and women arrive to join their spouse in the UK. As a result, still today many Bangladeshi children in UK are of first-generation mothers, who were born in Bangladesh.

1.4.2. The phonetics of Sylheti and English

To date, there is little published work on Sylheti, and no published work on the acoustic-phonetics of Sylheti. Whether Sylheti is a separate language from Standard Bengali (SB; the official language of Bangladesh) is a much-debated issue (Rasinger, 2007). Chalmers (1996) describes Sylheti as a distinct language that is 'mutually unintelligible to a Standard Bengali speaker' (p. 6), but anecdotal evidence from members of the London-Bengali community suggests that the differences are relatively small (Rasigner, 2007). In terms of the phonetics of the language, Chalmers (1996) describes Sylheti as being 'simpler than SB, with fewer vowels and consonants' (ibid p. 8). A review of the phonetic descriptions of Sylheti and Standard Bengali show discrepancies between the phonetic described inventories. For consonants, descriptions range from 19-30 consonants (see Table 1.2 for a comparison of English and Sylheti consonants). In contrast to English that only has a voiced-voiceless distinction, SB plosives and affricates contrast in voicing and aspiration e.g., /p/, /p^h/, though this contrast is thought to not be present in Sylheti (see below). Further, unlike English, SB contains tap, trill and retroflex sounds.

For vowels, a similar discrepancy is evident, in particular for diphthongs. Specifically, SB contains seven monophthongal vowels, where vowel length is not contrastive (Khan, 2010). Chalmers (1996) states that there is very little difference between the Standard Bengali and Sylheti vowel systems, and outlines 7 vowels. In contrast, Standard Southern British English (SSBE) contains 11 monophthongs. It is unclear whether Standard Bengali contains diphthongs. Inventories range from none (Maddieson, 1984) to 31 (Hai, 1960). Descriptions of SSBE outline 8 diphthongs (Ashby, 1995).

This thesis focuses on English and Sylheti plosives and monophthongal vowels which have been shown to differ in the two languages. For plosives, in contrast to SB where stops contrast in voicing and aspiration (Khan, 2010, Mikuteit & Reetz., 2007), Sylheti has been reported to not contain the aspiration contrast instead the aspirated pairs are substituted by fricatives (Chalmers, 1996; Grierson, 1903). Specifically, the voiceless bilabial stops, /p/, /p^h/, are replaced by [f], and the voiceless aspirated velar stop, /k^h/, is replaced by [x] (Chalmers, 1996). In contrast, English contains voiced-voiceless stop pairs: /p/-/b/, /t/-/d/, and /k/-/g/. For place of articulation, English has bilabial, alveolar and velar stops, whereas SB and Sylheti have been described as having bilabial, dental, post-alveolar and velar stops. Within the literature, there are discrepancies in terms of the exact position of the post-alveolars in SB. As highlighted by Khan (2010), they are inconsistently described, for example: “apico-alveolar” (Dasgupta, 2003, p. 359), “retroflex-alveolar” (Ray, Hai & Ray, 1966, p. 6), and “alveolar” (Tunga 1995, p. 139). Likewise, in the limited descriptions of Sylheti there are discrepancies: Chalmers (1996) reports that dental and retroflex /t/ and /d/ exist, whereas early reports of Sylheti suggest that the ‘distinction between cerebral and dental consonants has almost vanished’ (Grierson, 1903, p. 225) i.e., the distinction between dental and retroflex (‘cerebral’) stops.

In terms of acoustic-phonetic descriptions, there are few descriptions of Standard Bengali, and no acoustic descriptions of Sylheti. For plosive contrasts, this thesis focuses on voice onset time. VOT, originally described by Lisker and Abramson

Table 1.2. Sylheti/Standard Bengali and English consonant inventory.

	bilabial	labiodental	dental	Dental-alveolar	alveolar	Post-alveolar	retroflex	palatal	velar	glottal
Plosive	**p b** p^{h} $b^{ɦ}$			t d t^{h} $d^{ɦ}$	t d t^{h} $d^{ɦ}$	**t d**	ʈ ɖ	ɟ	**k g** k^{h} $g^{ɦ}$	
Nasal	**m**	ɱ	**n**				ɳ		**ŋ**	
Trill					r similar to English Wl after a stop					
Tap/flap							ɽ $ɽ^{h}$ (rare)			
Fricative		**f v*** Only Sylheti, (Chalmers 1996)		**θ ð**	**s z** Only Sylheti, (Chalmers 1996	ʃ **ʒ***		ç	x	**h**
Lateral fricative										
Approximant					**ɹ**		ɻ			
Lateral approximant					**l** (Sylheti/SB: never dark)					

Bold = English; *English only; shaded grey = discrepancies between inventories

(1964) in their pioneering cross-language study of voicing as "the time interval between the burst that marks the release of the stop closure and the onset of quasi-periodicity that reflects laryngeal vibration" (p. 422), was found to be an effective measure in separating phonetic categories. The few descriptions of Standard Bengali combined with the reports on similar Indo-Ayran languages (e.g., Hindi, Davis, 1994) suggest that, as with other Indo-Ayran languages e.g., Hindi, the voiced-voiceless distinction ranges from prevoiced (i.e. below 0ms VOT) and short lag region (i.e., 0-30ms). The voiced-aspirated stops e.g., /bʱ/ likely contain a breathy voiced portion, likewise the voiceless-aspirated stops have a longer VOT (Davis, 1994; Mikuteit & Reetz, 2007; Ladefoged & Maddieson, 1996). In contrast, SSBE English voiced stops have a short-lag, and voiceless stops have a long-lag, though British English voiced stops with prevoicing have been described (Docherty, 1992). Specifically, English voiced stops range from 0 to 40 milliseconds mean VOT, with /g/ containing the longest VOT and /b/ the shortest. Voiceless stops range from around 40 to 80 milliseconds mean VOT, with /k/ containing the longest VOT and /p/ the shortest (Docherty, 1992; Lisker & Abramson, 1964)

For monophthongal vowels, previous phonetic descriptions of SB and Sylheti (Chalmers, 1996; Grierson, 1903; Khan, 2010) suggest that Sylheti has 7 monophthongs: /i/, /e/, /ɛ/, /a/, /u/, /ɔ/, /o/, where vowel length is not contrastive (Khan, 2010). In Kolkata Bengali (India), a contrast between oral and nasal monophthongs has been described (Madieson, 1984; Masica, 1991). Khan (2009) however, notes that this contrast is not present in SB (Bangladesh) and attributes the nasality in Kolkata Bengali to the Eastern dialect influence e.g., Hindi. In contrast, Standard Southern British English (SSBE), has 11 monophthongs: /iː/, /ɪ/, /ɛ/, /æ/, /ʌ/, /ɜː/, /ɑː/, /ɒ/, /ɔ/, /uː/, /ʊ/ (Hawkins & Midgley, 2005; Wells, 1982). SSBE also contains the schwa vowel /ə/, which is only present in unstressed syllables. Thus, unlike Sylheti, SSBE has central vowels, as well as tense-lax pairs e.g., /i/-/ɪ/. See Figure 1.3 for a comparison of the English and Sylheti vowel space.

For vowels, there are no known acoustic descriptions for Sylheti and Standard Bengali. However, Alam, Habib & Khan (2008), describe the typical formant frequencies and duration values for Kolkata Bengali (Indian), based on the recording of speakers (exact number of participants not known) aged 25 to 29 years old. Table 1.3 displays the mean formant and duration values.

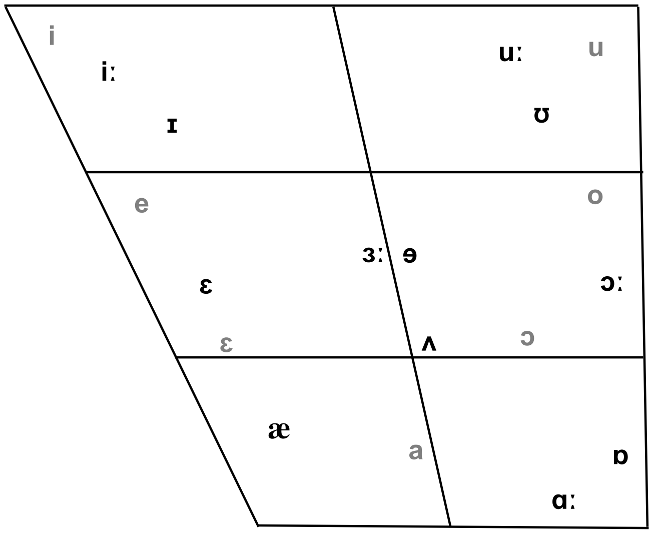


Figure 1.3. Vowel quadrilateral illustrating SSBE (black) and Bengali (grey) vowels. The Bengali vowels are based on phonetic descriptions of Standard Bengali (Khan, 2010)

For SSBE, Hawkins and Midgley (2005) report the typical formant values and duration values for four male SSBE speaking adults: 20-25, 35-40, 50-55 years old, and over 65 years old. Equivalent female formant data is available for ten SSBE female speakers from two age groups: 59-76 and 20-25 years old (Moreiras, 2006). See Table 1.4 for the mean duration and formant values from the 20-25 year old age group. The data shows that English vowels can be defined in terms of formant values, as well as duration for the tense-lax pairs e.g., /i/-/ɪ/.

Table 1.3. Formant (Hertz) and duration (milliseconds) values for Kolkata Bengali (Indian) vowels (based on Alam et al., 20

Vowel	**F1 (Hz)**	**F2 (Hz)**	**Duration (s)**
i	358	2318	0.68
e	501	1750	0.75
ɛ	693	1566	0.91
a	766	1451	0.98
ɔ	664	1192	0.75
o	445	1042	0.75
u	364	1023	0.68

Table 1.4. Formant (Hertz) and duration (milliseconds) values for SSBE speakers aged 20-25 years. The formant values are averaged over the male speakers from Hawkins & Midgley (2005) and female speakers (Moreiras, 2006). The duration measures are from Wells (1962).

Vowels	**F1 (Hz)**	**F2 (Hz)**	**Duration (s)**
iː	311	2636	0.29
ɪ	428	2382	0.14
ɛ	620	2089	0.17
æ	960	1640	0.21
ɑː	665	1142	0.33
ɒ	541	975	0.18
ɔː	425	711	0.33
uː	326	1873	0.29
ʊ	450	1556	0.14
ʌ	716	1355	0.15
ɜː	580	1574	0.31

1.5.3. Research objectives

This thesis investigates the acquisition of the English monophthongal vowels and plosives (voicing contrast) by Sylheti-English speaking sequential bilingual children. The children were all born in the UK, and raised within the London Bengali community, in the London boroughs of Camden and Tower Hamlets. These children were predominately exposed to Sylheti, and had less than 20% exposure to English before entering full-time education. The children were tested twice: Time 1 (Time 1) after an average of 7 months full time English exposure in nursery (54.7 months-old) and Time 2 (Time 2), approximately 12 months later when the children were in the first year of Primary school education (mean 64.7 months-old), thus having 11-12 months exposure to English.

Chapter 2 presents a detailed acoustic-phonetic analysis of Sylheti and English monophthongal vowels and plosives as produced by adults from the London-Bengali community. As there are currently no acoustic-phonetic descriptions of Sylheti, and in particular, no description of the Sylheti spoken by immigrants in London, the primary aim of Chapter 2 is to provide acoustic-phonetic description of Sylheti, as produced by monolingual Sylheti and London-Bengali speakers. The second aim of Chapter 2 is to provide a baseline for the speech that the children who grow up in the London Bengali community are likely to be exposed to, i.e., the variability in the adults' production of English and Sylheti. To do so, speakers differing in age of arrival to the UK and language use patterns were recorded. Based on previous L2 production research, it was hypothesized that that the speakers' production would vary as a function of age of arrival and language use. These findings provide a baseline for the longitudinal child experiments reported in Chapters 4-7.

Chapter 3 provides an introduction to the longitudinal child study (Chapters 4-7). The aim of Chapter 3 is to outline the overall experimental design of the experiments reported in Chapters 4 to 7, as well as define the key terminology. Chapters 4 and 5 report the main phoneme acquisition experiments: the acquisition of English monophthongal vowels (Chapter 4) and the acquisition of the English voicing contrast (Chapter 5). Two experiments were conducted for each chapter: perception and production. The target phonemic contrasts in the experiments were chosen based on previous research on English and Sylheti phonetics (as reported 1.5.2, this chapter) as well as the findings from the London-Bengali adult production study (Chapter 2).

Chapters 6 and 7 attempt to account for some of the variability in the bilingual children's outcomes reported in Chapter 4 and 5. Specifically, Chapter 6 investigates the influence of language exposure and the children's social networks on their acquisition of English vowels and plosives. To do so, detailed caregiver interviews regarding the children's social networks and language exposure patterns were conducted. Chapter 7 explores the relationship between caregiver speech and the bilingual children's phoneme acquisition. Specifically, a detailed acoustic analysis of the main caregivers' English vowel and plosive production was conducted. This data was subsequently explored in relation to the child's outcomes reported in Chapter 4 and 5. Finally, Chapter 8 provides a general discussion of the main findings of the experiments conducted in this thesis and the implications of the results on bilingual and second-language research and theories.

2. Detailing the phonetic input: the production of Sylheti and English vowels and plosives by adult speakers from the London-Bengali community

2.1. INTRODUCTION

This chapter reports an acoustic-phonetic study of Sylheti and English bilabial, alveolar/post-alveolar, velar plosives and monophthongal vowels as produced by adult speakers from the London-Bengali community. The aim of this chapter was twofold. First, to provide the first acoustic-phonetic description of Sylheti. To date, there are no existing detailed phonetic descriptions of Sylheti. Second, to investigate the difference in the production of Sylheti and English by speakers who reside in the London Bengali community, and who differ in terms of age and place of the heritage (L1) and host language (L2) acquisition. Thus, this study provides a baseline for the research in the chapters to follow. Not only does it outline the phonetics of Sylheti plosives and vowels, it also provides a picture of the phonetic variation that occurs in both the L1 and L2 in the target immigrant community, and thus highlights the possible variation in phonetic input that the children who grow up in the London-Bengali community are likely to be exposed to within the community.

2.2. MATERIALS AND METHODS

2.2.1 Participants

Thirty-seven adults (18 – 63 years old) from five speaker groups were recorded. Using a background questionnaire (see appendix 1a), the speakers were grouped according to place of birth and age of arrival to the UK. All speakers (apart from the controls) had been resident in the UK for an average of 25 years.

In addition, information on language use was collected: Each participant was presented with a list of interlocutors (e.g. when talking to grandparents, parents, children), and social situations (e.g. at work, at university, in the home), and for each situation the speaker had to state whether they used English only, Sylheti only, or whether they varied between Sylheti and English (i.e. used “both”, either Sylheti or English). Based on the answered items, a percentage of English, Sylheti and Both language use was calculated. For example, if a participant spoke to 4/8 of their interlocutors in only English, 1/8 in only Sylheti, and 3/8 in both Sylheti and English (i.e., mixed), their language use percentage would be as follows: 50% English only, 10% Sylheti only, and 40% both English and Sylheti. See Table 2.1 for speaker group details.

All participants were recruited from community and Woman's centres in the London Boroughs of Tower Hamlets and Camden. To access the community, the researcher attended parent groups, Bangladeshi events (e.g., Bengali New Year celebrations) and assisted with Bengali language classes for children from 6-months prior to the start of the research. During this time, the researcher also gave short talks to members of the community centres to explain the purpose of the adult (Chapter 2) and child (Chapters 3-7) research. The regular attendance at the Women's and Community Centres meant that the researcher became well known within the target community. This was particularly important for the longitudinal child study (Chapters 3-7), as it helped minimalize dropout rates.

Monolingual controls

Sylheti and Standard South British English (SSBE) controls were recruited as control subjects. The SSBE speakers were all born in the south of England, and resident in the London borough of Hackney or Tower Hamlets at the time of testing. As it was not feasible to record the Sylheti speaking controls in Bangladesh, all control subjects were recruited in the UK. To be included in the study as a Sylheti speaking control, subjects had to have arrived from Sylhet, Bangladesh to the UK within the last 3 months and have Sylheti (learned in Sylhet, Bangladesh) as their first language. Although the Sylheti controls had reported some English language use (mean: 14 % English only), all speakers' dominant language was Sylheti (mean: 69 % Sylheti). All of the Sylheti control speakers had not attended formal English language classes before arriving in the UK, and had had a maximum of 1 month of English classes since their arrival in the UK.

Late arrivals

The Sylheti speaking late arrivals (hereafter, "Late") were all born in Sylheti, Bangladesh and had arrived in the UK after the age of 16 (mean age of arrival: 21-years-old). All participants either worked in local Bangladeshi community businesses (e.g. restaurants, news agents), community centres, or were full-time mothers. Thus,

Table 2.1. Speaker group details.

Speaker group	No.	Mean age of arrival (years)	Mean length of residence in the UK	Mean % of Sylheti only language use	Mean % of English only language use	Mean % of English or Sylheti language use
Sylheti controls	5 (2 male, 3 female)	26	3 months	69	14	17
Late arrivals	11 (4 male, 7 female)	21	24 years	58	23	19
Early arrivals	9 (3 male, 6 female)	6	27 years	23	41	36
Second-generation	6 (3 male, 3 female)	Born in the UK	26 years	25	51	24
SSBE controls	6 (3 male, 3 female)	Born in the UK	28 years	0	100	0

these speakers had little social contact outside of the Bangladeshi community. Sylheti was the dominant language for the Late group (mean: 58% Sylheti only), and they used both Sylheti and English at work, with family and members of the community. In addition, the Late group reported that they used 23% English only, 19% English and/or Sylheti.

Early arrivals

The Sylheti speaking Early arrivals (hereafter, "Early") were all born in Sylhet, Bangladesh and had arrived in the UK before age 16 (mean age of arrival: 6 years old). All of the Early group speakers entered full-time UK education when they arrived in the UK. At the time of testing, they all worked in an English-speaking environment (either within or outside of the London-Bengali community). For this group, Sylheti was the main language used when talking to family members, and to members of the community. However, English was the overall dominant language used (mean: 41 % English only). Moreover, from the 3 different groups, the Early speakers reported the highest percentage of situations where they would use both Sylheti and English (36% of English or Sylheti).

Second-generation

The second-generation speakers (hereafter, Second-Gen) were all born in the UK, and had grown up in the London-Bengali community. These speakers had been exposed to English and Sylheti since birth, but Sylheti had been their dominant language up until they entered full-time education at 4 years of age. All of the Second-Gen speakers were either working or studying in an English-speaking environment, and also spoke English with the younger members of the community. Overall, English was their dominant language spoken by the Second-Gen speakers (mean: 51% English only). In contrast to the Early speakers, they reported fewer situations where they would use both languages (i.e. English or Sylheti).

2.2.2 Target sounds

Plosives

Bilabial, post-alveolar, and velar Sylheti /p/, /pʰ/, /b/, /bʱ/, /t/, /tʰ/, /d/, /dʱ/, /k/, /kʰ/, /g/, /gʱ/ and English bilabial, alveolar, and velar /p/, /b/, /t/, /d/, /k/, /g/ were elicited in word-initial stressed position. All words were piloted on 6 speakers (2 Sylheti control, 2 Late, and 2 Second-Gen speakers). Due to varying levels of proficiency between language groups (e.g., Second-Gen speakers knew fewer Sylheti words than did the Late and Early speaker groups), all words had to be selected to be imageable for all speakers. Consequently, target sounds could not be elicited in the same phonetic context. However, to maximise comparisons between and across languages, for voiceless plosives, /i/ did not follow the test consonant, where possible, as this has been shown to significantly affect VOT in English compared to the other English monophthongs (Nearey & Rochet, 1994). The only exception was 'teeth', to allow for cross-language comparison with Sylheti 'tiya'[1]. Similarly, for voiced plosives, /u/ did not follow the test consonant as this has been shown to have the greatest effect on VOT for English voiced plosives (Nearey & Rochet, 1994). The final set of target

[1] Piloting showed that the second-generation speakers did not know the alternative Sylheti words containing the /i/ vowel

words (see Appendix 1b) consist of 6 English and 12 Sylheti plosives (including aspirated pairs). Each subject named each picture twice.

Monophthongs

Where possible, all vowels were elicited in a CVC context in the same CVC frame. However as the target words had to account for varying language proficiency levels as well as be imageable, the Sylheti /i/ vowel was elicited in a VC context. Further, all vowels could not be selected in the same CVC frame, though where possible initial alveolars were avoided to maximise comparisons (see Hillenbrand et al., 2001). For Sylheti, 7 monophthongs as identified by Khan (2010) for Standard Bengali were elicited: ‘eet’ /i/, ‘desh’ /e/, ‘dheka’, /ɛ/ ‘pata’ /a/, ‘bhut’ /u/, ‘dosh’, /ɔ/, ‘choto’ /o/. For English, 11 monophthongs were elicited: bead /iː/, kid /ɪ/, bed /ɛ/, cat /æ/, cup /ʌ/, ‘bird’ /ɜː/, card /ɑː/, cot /ɒ/, court /ɔ/, boot /uː/, book /ʊ/ (see Appendix 1c for word list). Each subject named each picture twice.

2.2.3 Procedure

To avoid any orthographic influence on production or literacy difficulties, target sounds were elicited using pictures. The pictures were presented on a computer screen. For Sylheti, pictures were named in a dialect-specific sentence: *abar____kho* আবার ____ খও *(gloss: say ______ again)*. For English, pictures were named in the carrier sentence *say _______ again*; each target sound was recorded twice in a random order.

All recordings were made in a quiet room in the subjects’ home or workplace using a H2 Zoom recorder with a sampling rate of 44.1 kHz, 16-bit resolution. All sessions were conducted by the author who is a native speaker of English. All subjects were recruited from local Bengali community centres or Asian women projects within Tower Hamlets and Camden, where the author is well known. For all of the Sylheti recordings, subjects were instructed to produce the sentences as if they were speaking to a member of their family. The languages and repetitions were recorded separately. This resulted in 4 word lists, 2 Sylheti, and 2 English. The order and word lists of the recordings in the two languages were counterbalanced across subjects.

2.2.4. Acoustic analyses

For Sylheti, a total of 7 plosive tokens: (1 /bʱ/, 2 /d/, and 4 /t/) and 1 vowel (/ɛ/) for 18 Sylheti speakers, were missing from the analysis. This was due to subjects not knowing the target word. In addition, all Sylheti /p/, /pʰ/ and /kʰ/ tokens were produced as labiodental fricative [f] or velar fricative [x]. Consequently, /ɛ/, /p/, /pʰ/ and /kʰ/ were taken out of the final analyses.

Plosives

A total of 558 tokens for Sylheti and 384 tokens for English were analyzed. The acoustic measurements were made in Praat (Boersma and Weenink, 2012). The VOT measurements were obtained from the waveform and checked against the corresponding spectrogram. All of prevoiced tokens showed voicing throughout most of the hold phase, as seen in other Indo-Aryan languages (Heselwood & McChrystal, 1999; Davis, 1994). For these tokens the prevoicing was measured using the waveform and spectrogram (from the cessation of higher frequency spectral striations to the release of the plosive.) For the remaining tokens, the lag VOT was measured as the time between the release of the plosive closure and to the zero-crossing of the first periodic wave form from the following vowel. See Figure 2.1

Monophthongs

A total of 372 tokens for Sylheti and 704 tokens for English were analyzed. The first (F1) and second (F2) formant were measured using Praat (Boersma and Weenink, 2012). The stimuli were located manually, and then F1 and F2 were extracted using hand-corrected LPC analyses. Formant frequencies were measured from the midpoint of the steady state portion of the vowel. The steady-state portion of the vowel was defined as the part of the vowel that was closest to the midpoint. So that data from male and female speakers could be compared, Lobanov's z-score transformation (Lobanov, 1971) was used to normalize F1 and F2 values:

$$F\frac{Lobanov}{ti} = \frac{Fti - \mu ti}{\delta ti}$$

Where μti is the average formant frequency across the vowels for talker t and δti refers to the standard deviation for average μti.

The decision was made to use Lobanov z-score transformation over other vowel normalization methods (e.g., Nearey, 1978) as it has been shown to be the most effective method at preserving phonemic variation, reducing anatomical/physiological variation, and preserving sociolinguistic variation (Adank, Smits & van Hout, 2004). These factors are particularly important in the research reported in this thesis, as the aim is to investigate the phonemic and sociolinguistic variation within the community as well as account for variation that is a result of anatomical/physiological variation.

All duration measurements were taken from voicing onset to the voicing offset. Voicing onset was determined by the beginning of the F2 transitions, voicing offset was determined by the end of the F2 transitions, corresponding to the closure of the final consonant. Due to the differing word structure (i.e., CV & CVCV) in Sylheti, vowel duration measurements were not possible for Sylheti.

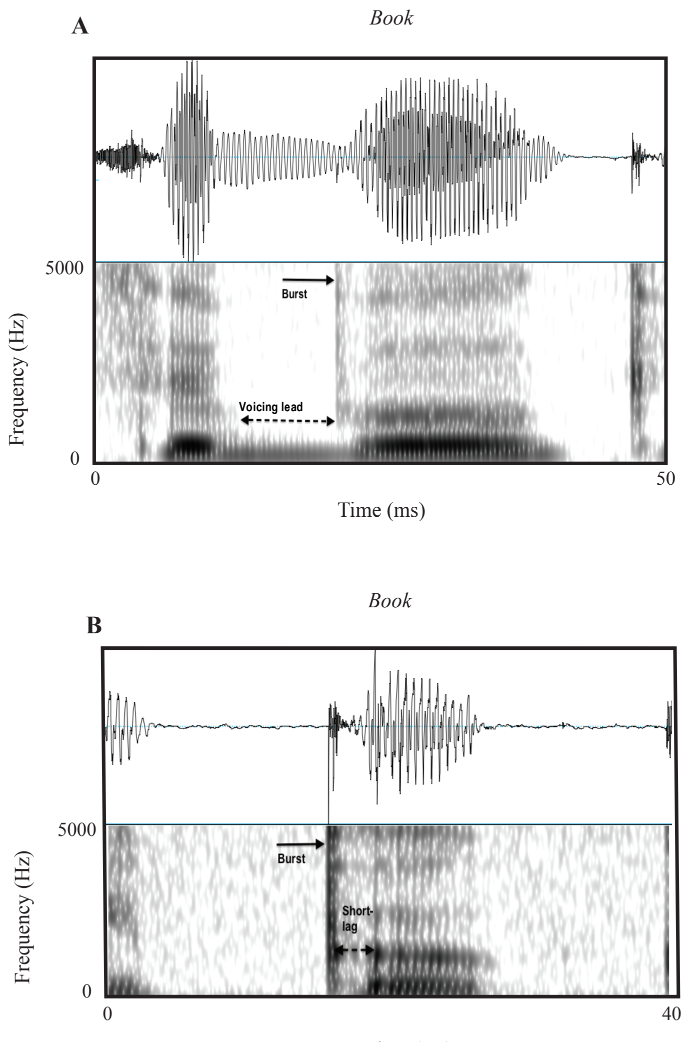


Figure 2.1. Waveform and spectrogram of book produced using voicing lead (A) and short-lag (B) VOT

2.3. RESULTS

Results from the two languages are presented in two parts: 1) Sylheti and 2) English. Each part is divided into two sections; a) Plosive Consonants and b) Monophthongs. For each language, data from the separate speaker groups [Sylheti Contols (for Sylheti only), Late, Early, Second-Generation & SSBE (for English only) speakers] will be compared. Finally, to investigate any potential L1-L2 influence, London-Bengali Sylheti speakers' (i.e., Late, Early & Second-Gen) production of plosive consonants in Sylheti and English will be compared.

To test for differences in production between the speaker groups, separate Generalized Linear Mixed Model (GLMM) analyses, using the identity link function, were conducted in SPSS 20. For all analyses, speaker group and plosive/vowel were treated as fixed factors. To account for individual differences, participant was treated was as a random factor in all analyses. Degrees of freedom were Satterthwaite corrected. Significant interactions were explored with post-hoc sequential Sidak tests; *p*-values were adjusted for multiple comparisons. There are several advantages of using GLMM analyses over typical General Linear Models (GLM) e.g., Analysis of Variance (ANOVA): 1) The model uses maximum likelihood methods (MLM) to estimate parameters, which provide better estimates of population characteristics than least square means, which are typically used in GLM; 2) The MLM allows for a robust estimation of missing data; 3) The model can account for random (e.g., participant) and fixed (e.g., voicing) effects. This means that levels that vary from experiment or condition (i.e., random effects) can be accounted for in the model.

2.3.1. Sylheti

Plosives

To reduce the number of multiple comparisons, plosives were divided into 3 groups, separated by place of articulation: (i) bilabial, (ii) post-alveolar, (iii) velar.

Bilabial: /b/, /bʱ/

Figure 2.2 shows differences between the groups for the bilabial voiced plosives: the Sylheti controls, Late and Early speakers produced prevoiced tokens. The Late, Early, and Second-Gen speakers produced /b/ & /bʱ/ with a similar VOT (Late: /b/ mean: -

59ms, /b^{ɦ}/ mean: -61ms; Early: /b/ mean: -16ms, /b^{ɦ}/ mean: -2ms; Second-Gen: /b/ mean: 11ms, /b^{ɦ}/ mean: 5ms) resulting in a loss of the contrast in terms of aspiration. Inspection of the confidence intervals suggests that Sylheti controls used a longer voicing lead for /b^{ɦ}/ (mean: -45ms) than for /b/ (mean: -81ms).

GLMM analyses revealed a main effect of plosive [$F(1,61)=15.73$, $p < .001$], speaker group [$F(3,33) = 34.07$, $p < .001$], and a significant interaction between plosive and speaker group [$F(3,61) = 5.49$, $p < .01$]. Post hoc sequential Sidak tests confirmed that the Sylheti controls were the only speaker group that distinguished between /b/ and /b^{ɦ}/ using significantly different VOT values ($p < .05$).

Post-alveolar: /t/, /t^{h}/, /d/, /d^{ɦ}/

Figure 2.2 shows that the Sylheti controls and Late speakers produced /t^{h}/ with a longer VOT value (Sylheti controls mean: 26ms, Late mean: 26ms) than they did for /t/ (Sylheti controls mean: 14ms, Late mean: 16ms). The Early and Second-Gen speakers produced /t/ and /t^{h}/ using a short-lag (Early mean: /t/ 23ms, /t^{h}/ 32ms; Second-Gen: /t/ 30ms, /t^{h}/ 35ms). For /d/ and /d^{ɦ}/, the Sylheti controls produced both voiced plosives with a voicing lead, using a longer lead for /d^{ɦ}/ (mean: -79ms) than for /d/ (mean: -40ms). The remaining speaker groups used voicing lead or short-lag VOT values for /d/ and /d^{ɦ}/ (/d/ mean: Late, -54ms; Early, 4ms; Second-Gen: 7ms; /d^{ɦ}/ mean: Late, 5ms; Early, 9ms; Second-Gen: 15ms).

GLMM analysis revealed a main effect of plosive [$F(3,92) = 239.80$; $p < .001$], speaker group [$F(3,35) = 44.98$; $p < .001$], and the interaction between plosive and speaker group [$F(9,92)= 46.94$; $p < .001$] were significant. Post-hoc sequential Sidak tests confirmed that the Sylheti controls and the Late speakers used prevoicing for /d^{ɦ}/ and /d/, whereas the Early and Second-Gen speakers used a short-lag ($p < .05$). The Sylheti controls and Late speakers were the only groups to distinguish between the unaspirated and aspirated pairs ($p < .05$).

Velar: /k/, /g/, /g^{ɦ}/

Figure 2.2 shows that all speakers produced /k/ short-lag (mean: 16ms). For /g/ and /g^{ɦ}/, the Sylheti controls and Late speakers used voicing lead (/g/ mean: Sylheti controls -71ms; Late -43ms; /g^{ɦ}/ mean: Sylheti controls 11ms; Late -16ms), whereas the Early and Second-Gen used a short-lag (Early mean: 3ms; Second-Gen mean: 14ms) similar

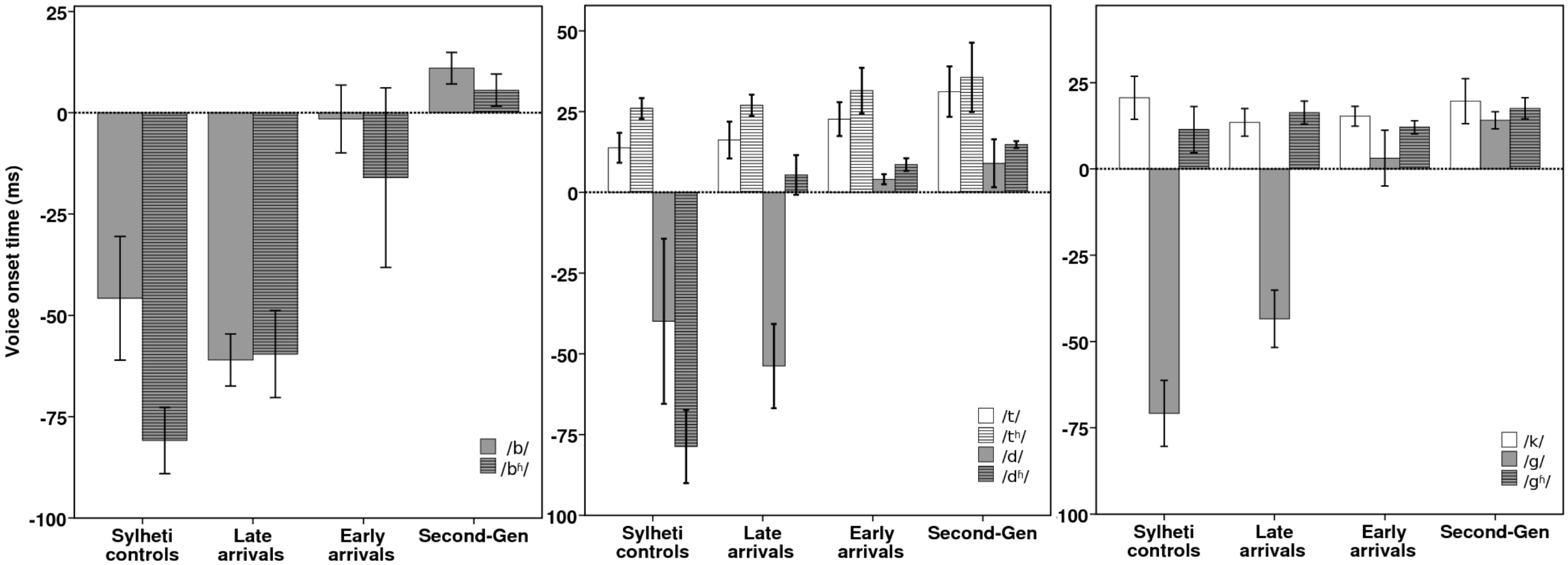


Figure 2.2. Bar charts representing the mean voice onset time in milliseconds for speakers' production of bilabial, post-alveolar and velar Sylheti plosives. The error bars indicate 95% confidence intervals.

to their /gʱ/ tokens (Early mean: 12ms; Second-Gen mean: 17ms). Inspection of the confidence intervals suggest that the Second-Gen speakers used a similar VOT for /k/, /g/ and /gʱ/.

GLMM analysis revealed a main effect of plosive [$F(2,107) = 179.88$; $p < .001$], speaker group [$F(3,42) = 36.26$; $p < .001$] and a significant interaction between plosive and speaker group [$F(6,107) = 42.45$; $p < .001$]. Post-hoc sequential Sidak tests confirmed that the Sylheti controls and Late speakers used significantly shorter (prevoiced) VOT for /g/ than did the other groups ($p < .05$). The Late speakers used VOT to distinguish between all velar plosives, apart from /gʱ/ - /k/ ($p > .05$). The Early speakers used a significantly different VOT for /k/-/g/ ($p < .05$), but not for /g/-/gʱ/ ($p > .05$). The VOT values for the Second-Gen speakers production of Sylheti velar plosives were not significantly different ($p > .05$).

Monophthongs

Figure 2.3 shows that all speakers production of Sylheti vowels reflected that of the Sylheti control speakers.

GLMM analysis revealed a main effect of vowel for F1 and F2 [F1, $F(5,104) = 456.80$; $p < .001$] and F2 [$F(5,54) = 943.28$; $p < .001$), and no main effect of speaker

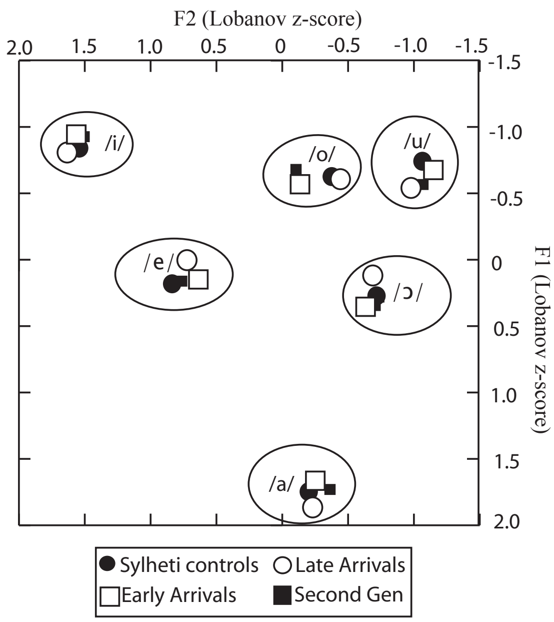


Figure 2.3. Mean z-score normalized formant frequencies values for Sylheti vowels for each speaker group. The x-axis displays decreasing F2 values from left to right. The y-axis displays increasing F1 values from top to bottom. Ovals are used to separate the vowel groups.

group for F1 and F2 ($p > .05$). There was an interaction between vowel and speaker group for F1 and F2 [F1, $F(15, 104) = 2.39, p < .05$; F2, $F(15, 118) = 2.36, p < .05$]. Post-hoc sequential Sidak tests revealed that the Second-Gen speakers used a lower F1 for /a/ than did the Late speakers, and the Early speakers used a higher F1 for /ɔ /than did the Late speakers ($p < .05$). For F2, the Early and Second-Gen speakers used a higher F2 for /o/ than did the Sylheti-controls and Late speakers.

Summary

For plosives, the analyses revealed that, as suggested by Chalmers (1996) and Grierson (1903), all of the Sylheti speaking groups used fricatives for voiceless plosives (i.e., /p/, /pʰ/ to /f/; /kʰ/ to /x/), resulting in the contrastive aspirated feature being lost for these sounds. The analyses revealed differences in production between the speaker groups. In general, the Late speakers displayed a similar voicing pattern to that of the Sylheti control speakers. In contrast, the Second-Gen speakers used a longer VOT for all plosives than did the other groups. The Early speakers demonstrated a VOT pattern that was midway between the Late speakers and Second-Gen speakers i.e., these speakers used longer VOTs than Late arrivals but shorter VOT values than the Second-Gen speakers. This pattern was particularly evident for the velar plosives, where the Late speakers used VOT to distinguish between voiced-voiceless pair (e.g. /k/-/g/) and the aspirated-unaspirated pairs (e.g. /g/-/gʱ/). The Early speakers used VOT to make the voiced-voiceless distinction, but not for the aspirated-unaspirated velar plosives. In contrast, the Second-Gen speakers did not use significantly different VOT values for the Sylheti velar plosives.

In general, the speaker groups did not differ in their production of Sylheti monophthongs. The main difference was for the /o/ vowel, where the Early and Second-Gen speakers used a higher F2 than did the Sylheti controls and the Late speakers.

2.3.2. English

To investigate differences in the production of English the different speaker groups were compared. Standard Southern British English (SSBE) controls were included in the analyses for a native British English target comparison.

Plosives

Bilabial: /p/, /b/

Figure 2.4 shows that the Late speakers produced /p/ and /b/ with a shorter VOT than did the other groups. The Late speakers produced /b/ with voicing lead (mean: -57ms). In contrast, the Early (/p/ mean: 35ms; /b/ mean: 10ms) and Second-Gen (/p/ mean: 28ms; /b/ mean: 13ms) speakers produced the bilabial plosives with a similar VOT pattern to the SSBE speakers (/p/ mean: 39; /b/ mean: 13ms). A gradual increase in VOT values is observed between Early, Second-Gen and SSBE speakers.

The GLMM analysis revealed a significant main effect of plosive [$F(1,76)=117.97$, $p < .001$], speaker group [$F(3,29) = 40.34$, $p < .001$] and a significant interaction between language group and plosive [$F(3,76) = 21.47$, $p < .001$]. Post hoc sequential Sidak tests revealed that the Late speakers used a significantly shorter VOT for /p/ than did the SSBE speakers ($p < .05$), but not the Early and Second-Gen speakers ($p > .05$). For /b/, the Late speakers used significantly shorter VOT values than did the Early, Second-Gen and SSBE speakers ($p < .05$).

Alveolar: /t/, /d/

Figure 2.4 shows that the SSBE controls produced voiced and voiceless alveolar plosives with a longer VOT than did the other groups. Specifically, the Late speakers used a short-lag VOT for /t/ (mean: 31ms) and voicing lead for /d/ (mean: -6ms), whereas Early (/t/ mean: 38ms; /d/ mean: 7ms) and Second-Gen speakers (/t/ mean: 55ms; /d/ mean: 12ms) produced these plosives with a similar VOT to the SSBE speakers (/t/ mean: 58ms; /d/ mean: 11ms).

GLMM analysis revealed no significant interaction between plosive and speaker group ($p > .05$). The main effect of plosive [$F(1,118) = 218.34$, $p > .001$] and speaker group [$F(3,118) = 16.70$, $p < .001$] were significant.

Velar: /k/, /g/

Figure 2.4 shows that Early and Second-Gen speakers produced /k/ (Early mean: 51ms; Second-Gen mean: 57ms) and /g/ (Early mean: 16ms ; Second-Gen mean: 21ms) with a similar VOT to the SSBE speakers (/k/ mean: 65ms; /g/ mean: 18ms). In

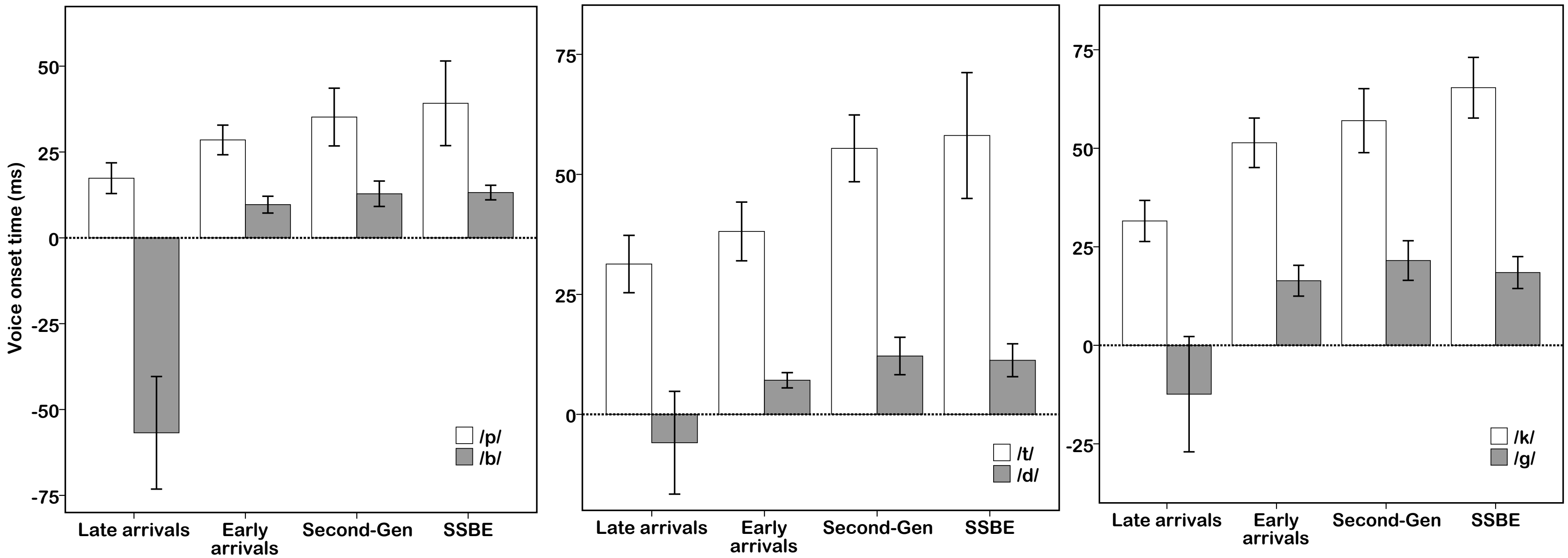


Figure 2.4. Bar charts representing the mean voice onset time in milliseconds for the speakers groups' production of bilabial, alveolar and velar English plosives. The error bars indicate 95% confidence intervals

contrast, the Late speakers produced /k/ (mean: 31ms) and /g/ (mean: -12ms) with a shorter VOT than did the other groups, using voicing lead for /g/.

GLMM analysis revealed no significant interaction between plosive and speaker group ($p > .05$). The main effect of plosive [$F(1,64) = 178.16$, $p < .001$] and speaker group [$F(3,24) = 15.72$, $p < .001$] were significant.

Monophthongs

Based on the comparison of the Sylheti and English vowel system (see section 1.4.2) and to avoid multiple statistical analyses, the English vowels were divided into 3 groups for the analyses: front: /iː/, /ɪ/, /ɛ/, æ; central: /ʌ/, /ɜː/; back: /uː/, /ʊ/, /ɔ/, /ɒ/, /ɑː/. To investigate the use of vowel duration, the production of the English tense-lax vowel pairs (i.e. /i/-/ɪ/, /ɑ/-/ɒ/, and /u/-/ʊ/) by the different speaker groups was compared.

Front vowels: /iː/, /ɪ/, /ɛ/, /æ/

Inspection of Figure 2.5 shows that the Late speakers produced /iː/ with a higher F1 and lower F2 than did the other groups, resulting in merged F1 and F2 values for /ɪ/ and /iː/. For /ɛ/ and /æ/ all groups used a different F1 and F2. In general, the SSBE speakers used a higher F1 and lower F2 than did the other speakers.

GLMM analysis revealed a significant main effect of vowel for F1 and F2 [F1, $F(3,72) = 561.34$, $p < .001$; F2 $F(3,130) = 90.21$, $p < .001$] and speaker group for F1 [F1, $F(3,72) = 4.62$, $p < .01$] but not for F2 ($p = .069$]. There was a significant interaction between vowel and language group for F1 and F2 [F1, $F(9,94)=8.659$, $p < .001$; F2 $F(9,130) = 2.33$, $p < .05$]. Post-hoc sequential Sidak tests confirmed that the Late speakers produced /iː/ with a higher F1 than did the SSBE, Second-Gen and Early speakers ($p < .05$). The differences between /iː/ and /ɪ/ were not significant for Late speakers for F1 ($p > .05$). For F2, all of the Sylheti speaking groups produced /iː/ with a lower F2 than did the SSBE speakers ($p > .05$).

Central vowels: /ɜ/, /ʌ/

Inspection of Figure 2.5 shows that the Early and the Second-Gen speakers produced /ɜ/ and /ʌ/ with a similar F1 and F2 to the SSBE speakers. In contrast, the Late speakers produced /ɜ/ and /ʌ/ with a higher F1 than did the SSBE speakers, resulting in a merged category for /ɜ/ and /ʌ/.

GLMM analysis revealed significant main effect of vowel for F1 and F2 [F1, $F(1,77)=74.116, p < .001$]; F2, $F(1,102) = 87.98, p < .001$], and a significant main effect of speaker group for F1 [$F(3,35) = 7.40, p < .01$], but not for F2 ($p > .05$).

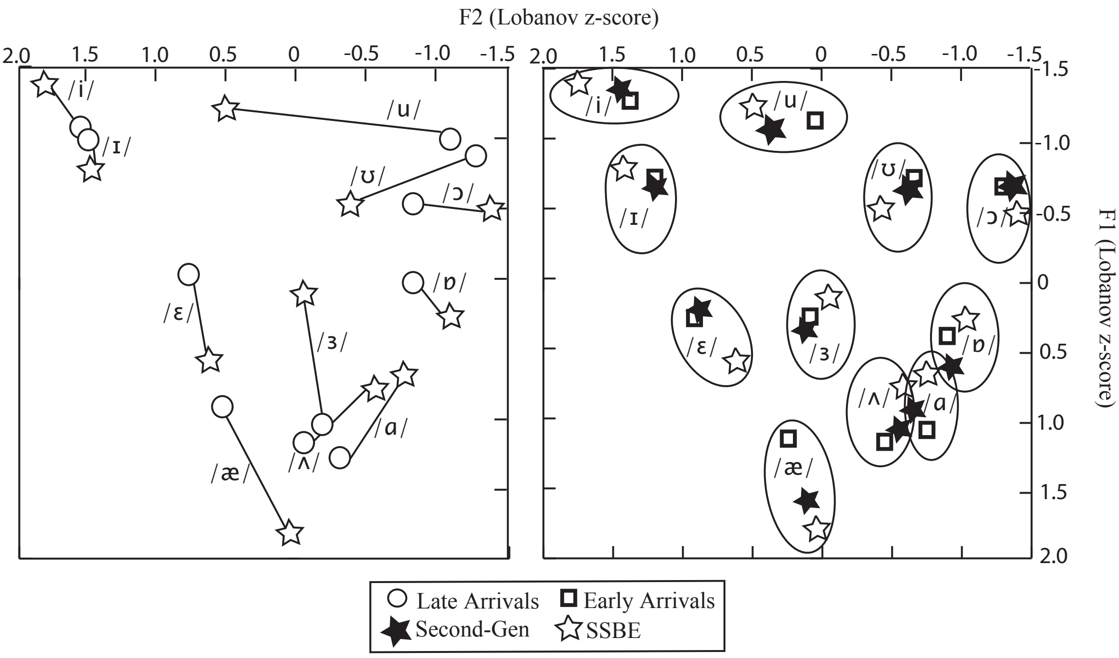


Figure 2.5. Mean z-score normalized formant frequency values for English vowels for each speaker group. The x-axis displays decreasing F1 values from left to right. The x-axis displays increasing F2 values from top to bottom. Circles are used in the right plot to separate the vowel groups. Lines are used in the left plot to illustrate the difference between the SSBE speakers (empty stars) and Late arrival speakers (circles).

There was also a significant interaction between vowel and language group for F1 and F2 [F1, $F(3,77) = 6.27, p < .001$; F2, $F(3,102) = 5.87, p < .01$]. Post-hoc sequential Sidak tests revealed that the Late speakers produced /ɜ/ with a significantly higher F1 and lower F2 and /ʌ/ with a higher F2 than did the other speaker groups ($p > .05$). Consequently, the F1 and F2 values for /ɜ/ and /ʌ/ were not significantly different for the Late speakers ($p < .05$).

Back vowels: /uː/, /ʊ/, /ɔː/, /ɒ/, /ɑː/

Inspection of Figure 2.5 shows that the Late speakers produced /ɑː/ with a higher F1 and F2, /ɔː/ with a higher F2, and /ʊ/ and /uː/ with a lower F2 than did the other speaker groups. This resulted in a merged category for /ʊ/ and /uː/.

GLMM analysis revealed a significant main effect of vowel for F1 and F2 [F1, $F(4,129) = 321.767$, $p < .001$; F2, $F(4,135)=82.765$, $p < .001$], and a main effect of language group for F2 [$F(3.47) = 7.448$, $p < .001$), but not for F1 ($p > .05$). There was a significant interaction between vowel and language group for [F1, $F(12,129)= 4.685$, $p < .001$; F2, $F(12,135) = 21.271$, $p < .001$]. Post-hoc sequential Sidak tests revealed that the Late speakers produced /ʊ/ with a lower F1 and F2, and /uː/ with a significantly higher F1 and lower F2 than did the Second-Gen and SSBE speakers ($p < .05$). This resulted in a merged category for /ʊ/ and /uː/ for the Late speakers ($p > .05$). Further, the Early speakers produced /u/ with a significantly lower F2 than did the Second-Gen and SSBE speakers ($p < .05$).

Duration

To compare speaker groups, vowel duration was normalized using z-scores. Inspection of Table 2.2 suggests that all speaker groups apart from the Late speakers use duration to mark the tense-lax distinction. Specifically the tense vowels, /iː/, /ɑː/ and /uː/, are produced with a longer duration than that of the lax counterpart vowels. The Late speakers used a shorter duration for /i/ and /u/ than did other speaker groups.

GLMM analysis revealed a main effect of vowel [$F(5, 342) = 161.17$, $p < .001$], speaker group [$F(3,342) = 4.18$, $p < .05$], and a significant interaction between speaker group and vowel [$F(15,342) = 6.134$, $p < .001$). Post-hoc sequential Sidak tests confirmed the Early and Second-Gen speakers used significantly longer vowel duration for the tense vowels than the lax vowels ($p < .05$). The Late speakers used a significantly longer duration for /iː/ than /ɪ/ and /ɑː/ than /a/ ($p < .05$), but not for /uː/ and /ʊ/ ($p > .05$)

Table 2.2. Speaker group mean vowel durations for tense-lax English vowels. Measurements are given in normalized z-score values. The shaded areas indicated the tense-lax pairs: /iː/-/ɪ/, /ɑː/-/ɒ/, and /uː/-/ʊ/.

Group	/iː/	/ɪ/	/ɑː/	/ɒ/	/uː/	/ʊ/
Late	-0.24	-0.76	0.86	-0.50	-0.23	-0.60
Early	0.70	-0.70	1.26	-0.70	0.07	-0.90
Second-Gen	0.88	-0.89	1.17	-0.58	0.32	-0.88
SSBE	0.91	-0.69	1.53	-9.70	0.27	-0.80

Summary

For plosives, the analyses revealed a similar pattern of production across place of articulation. Overall the SSBE speakers used a longer VOT than did the Sylheti speakers. Amongst the Sylheti speakers a gradual increase in VOT was observed between the Late to Second-Gen speakers. In general, the Early and Second-Gen speakers used VOT values that were closer to that of the SSBE speakers, whereas the Late speakers used short-lag and prevoiced tokens for English, using VOT values closer to that of their Sylheti categories. However, the Early and Second-Gen did not always differ from the Late speakers (i.e., for /p/), and for /t/ the Late and the Early speakers used shorter VOT values than did the Second-Gen and SSBE speakers.

The vowel analyses revealed differences between the speaker groups. The Late speakers produced the English vowels using vowels similar to those in the Sylheti vowel system. This resulted in the Late speakers merging vowels that do not have an exact phonetic counterpart in Sylheti e.g. /iː/-/ɪ/. However, the Late speakers used duration to distinguish between the English tense-lax pairs /iː/-/ɪ/ and /ɑː/-/ɒ/, but not for /uː/-/ʊ/. In contrast, the Early and Second-Gen speakers produced the English vowels using a monolingual-like English vowel space, as well as durational patterns similar to that of the SSBE controls. These speakers also demonstrate fronting of the /uː/ and /ʊ/ vowels, phenomenon known as GOOSE and FOOT fronting, respectively (Hawkins & Midgley, 2005; Wells, 1982).

2.3.3. Cross-Language Comparison: Sylheti vs. English plosive production

To investigate differences between the London-Bengali Sylheti speaking subjects' (Late, Early & Second-Gen) productions in Sylheti and English plosives i.e. to see if speakers use Sylheti-like variants when producing English plosives and vice versa, the speakers' production of Sylheti plosives was compared with their English plosives. Note that for all plosives a longer lag for voiceless than for voiced plosives is expected, and similarly a longer VOT for aspirated plosives than for unaspirated plosive pairs. Furthermore, a longer lag is expected for the speakers' English plosive production than in their productions of the equivalent Sylheti plosives. Due to the large and pervasive differences in the vowel inventories of Sylheti and English, a direct comparison between the speakers' Sylheti and English vowels was not made.

Bilabial plosives

Figure 2.6 shows that the Late speakers used voicing lead for all of the voiced bilabial plosives (Sylheti mean: -60ms; English mean: -57ms). In contrast, the Early speakers used prevoiced tokens for Syhleti /b/ and /bʱ/ (/b/ mean: -16; /bʱ/ mean: -2ms), and a short-lag for English /b/ (mean: 10ms). The Second-Gen speakers however, used a short-lag for voiced bilabial plosives (Sylheti /b/ mean: 11ms; Sylheti /bʱ/ mean: 5ms; English /b/ mean: 13ms). All speakers used a short-lag for English /p/.

GLMM analysis revealed a main effect of plosive [$F(3,64) = 144.002$, $p < .001$], speaker group [$F(2,49) = 85.931$, $p < .001$], and interaction between language group and plosive [$F(6,65) = 27.276$, $p < .001$] were all significant. Post-hoc sequential Sidak tests revealed that all speakers used a longer VOT for English /p/ than they did for all Syhleti voiced bilabial plosives ($p < .05$). For the voiced plosives, the Late speakers did not use VOT to distinguish between English /b/ and Sylheti /b/ and /bʱ/ ($p > .05$), using voicing lead for all voiced plosives. The Early speakers distinguished between Sylheti /b/ and /bʱ/, using voicing lead for /bʱ/ ($p < .05$). There was no significant difference Second-Gen speakers production of the English and Sylheti voiced bilabial plosives ($p > .05$).

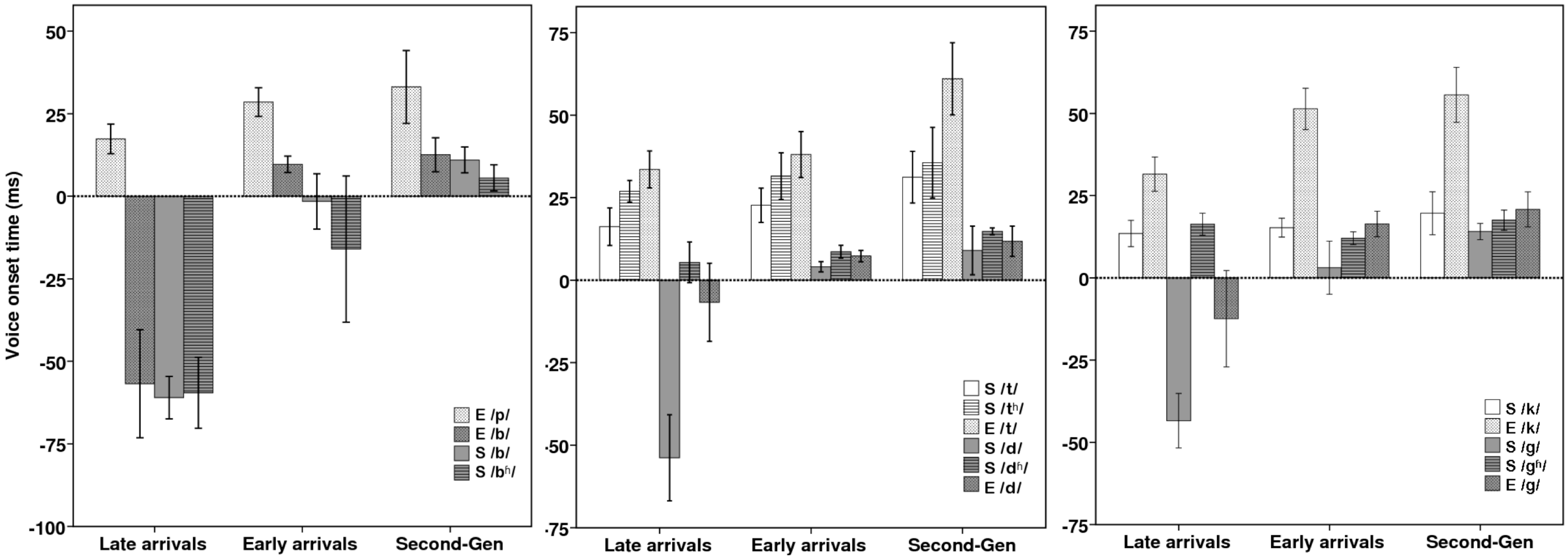


Figure 2.6. Bar charts representing the mean voice onset time in milliseconds for speakers groups' production of bilabial, alveolar/post-alveolar and velar Sylheti (S) and English (E) stop consonants. The error bars indicate 95% confidence intervals.

Alveolar/Post-alveolar

Figure 2.6 shows that all groups produced the English /t/ with a longer VOT (Late mean: 31ms; Early mean: 38ms; Second-Gen mean: 55ms) than they did Sylheti voiceless plosives (Late mean: /t/ 16ms, /t^{h}/ 26ms; Early mean: /t/ 23ms, /t^{h}/ 32ms; Second-Gen mean: /t/ 30ms, /t^{h}/ 35ms). For voiced velar plosives, the Early and Second-Gen speakers used a short-lag for Sylheti (Early mean: /d/ 4ms, /d^{ɦ}/ 9ms; Second-Gen mean: d/ 7ms, /d^{ɦ}/ 15ms) and English (Early mean 7ms; Second-Gen mean: 12ms). In contrast, the Late speakers used short and prevoiced tokens for Sylheti (mean: d/ -54ms, /d^{ɦ}/ 5ms) and English (mean: -6ms).

GLMM analysis revealed a main effect of plosive [$F(5,91) = 94.778$; $p < .001$], speaker group [$F(2,27) = 43.427$, $p < .001$], and interaction between language group and plosive [$F(10,92) = 8.782$, $p < .001$]. Post-hoc sequential Sidak tests revealed no significant difference between the VOT for Sylheti and English /t/ for all speakers, the Late speakers were the only group that distinguished between Sylheti /t/ and /t^{h}/ ($p < .05$). The Late speakers used a different VOT for all voiced plosives ($p < .05$), whereas the Early and Second-Gen speakers VOT for all voiced plosives was not significant ($p > .05$).

Velar plosives

Figure 2.6 shows that all groups used a longer VOT for English /k/ (Late mean: 31ms; Early: 51ms; Second-Gen: 57ms) than they did for Sylheti /k/ (Late mean: 13ms; Early: 15ms; Second-Gen: 20ms). For the voiced velar plosives, the Late speakers used voicing lead (Sylheti mean: /g/ -43, /g^{ɦ}/ -16ms; English mean: -12ms) whereas the Early (Sylheti mean: /g/ 3ms, /g^{ɦ}/ 12ms; English: /g/ 16ms) and Second-Gen (Sylheti mean: /g/ 14ms, /g^{ɦ}/ 17ms; English: /g/ 21ms) speakers used a short-lag. The Second-Gen speakers produced all voiced plosives using a similar VOT.

GLMM analysis revealed a main effect of plosive [$F(4,117) = 101.003$, $p < .001$], speaker group [$F(2,139) = 69.745$; $p < .001$], and an interaction between language group and plosive [$F(8,117) = 20.158$, $p < .001$]. Post-hoc sequential Sidak tests revealed that all speaker groups used a longer VOT for English /k/, than they did for all other velar plosives ($p < .05$). The Late speakers distinguished between all velar plosives, apart from Sylheti /g/ and /g^{ɦ}/ ($p < .05$). Early and Second-Gen speakers did not use significantly different VOT values for Sylheti and English voiced velar plosives ($p < .05$). However, the Early speakers used a significantly longer VOT for Sylheti /k/ than Sylheti /g/ ($p < .05$).

2.4. DISCUSSION

The aim of this chapter was to provide a baseline of Sylheti monophthongal vowels and plosives. The study also aimed to describe the variation in the production of Sylheti and English by speakers from the London Bengali community. To do this the study investigated the effect of age and place of acquisition on the production of both the heritage language (L1; Sylheti) and host country language (L2; English), by speakers who reside in the London Bengali community. Specifically, the study compared the production of stops and monophthongal vowels by speakers from the London-Bengali community. Speakers were grouped according to age of arrival; Late arrivals, Early arrivals and those born in the UK (Second-Gen).

Overall, the age and place of L2 acquisition influenced the production of both the L1 and the L2. Specifically, the speakers who had acquired their L2 late (i.e., after 20 years of age) used phonetic variants that reflected their L1. In general, speakers who had acquired their L2 during childhood used different phonetic variants in Sylheti and English, but these did not exactly match those of their monolingual counterparts. Despite similar experience with English in early childhood (e.g., attending school in the UK), the data also suggested that place of L1 acquisition (i.e., Early arrivals vs. Second-Gen) may affect the production of the L1 and L2.

The Late group had acquired the L2 during adulthood and had reported that Sylheti was their dominant language. In line with previous studies on L2 production in late learners (e.g., Baker & Trofimovich, 2005; Flege et al., 1999), the analyses found that the Late speakers produced native-like stops and vowels for Sylheti, and foreign-accented variants for English. Specifically, their production of English reflected that of their L1, i.e., they used shorter and prevoiced VOT values for stops, and collapsed some English vowel categories into single Sylheti-like categories e.g. English /i/-/ɪ/, both produced as /i/.

One possible explanation for this pattern of results is that Late speakers were using their native Sylheti categories for their English production (see e.g., Best, et al., 2001; Flege, 1995). As these speakers had acquired their L2 during adulthood, there was a decreased likelihood that they would have developed native-like L2 categories for sounds that do not have a phonetic counterpart in their L1. However, there was some evidence to suggest that they had begun to establish separate categories, at least for stops. A comparison between the Late speakers' production of Sylheti and English stops showed that although Late speakers used VOT patterns that closely reflected Sylheti, for some stops, in particular velar stops, they used significantly different VOT values in English and Sylheti. Such findings suggest that these speakers, who had begun to acquire their L2 at a late stage in their language development, were able to make some changes to their underlying phonetic categories (see also Iverson et al., 2003).

In contrast, the Second-Gen group had acquired monolingual-like English categories, but their production of Sylheti consonants differed from that of monolingual Sylheti and Late speakers. The Second-Gen speakers were born in the UK and were mainly exposed to their L1 until they entered school, around 4 years of age. These speakers reported that English was now their dominant language. Unlike the Late learners, the Second-Gen group used longer VOT values for Sylheti stops than that of the Sylheti controls and Late speakers. Furthermore, the Second-Gen group did not distinguish between the Sylheti unaspirated and aspirated pairs. The cross-language comparison for stops revealed that although the Second-Gen speakers used longer VOT values for English, this was only the case for voiceless velar stops and that overall, their Sylheti VOT values were similar to that of their English productions. This is somewhat surprising given that these Second-Gen speakers are likely to have received Sylheti input at home, and in the community from a variety of speakers (see e.g., Khattab, 2000).

One possible explanation for this is the influence of the shift to English as the dominant language. It is likely that English would have become dominant for these speakers once they started full-time school. During this stage in development (4-years-old) the speakers stop categories would have still been developing (e.g., Hazan & Barrett, 2000; McCarthy, Mahon, Rosen & Evans, under revision). This plasticity, combined with the dominant English use, may have resulted in the loss of voicing features that do not exist in English i.e., voicing lead. For vowels, the data are in line with previous studies that have shown that speakers, who acquire both languages during childhood, establish distinct vowel categories for the L1 and L2 (Baker & Trofimovich, 2005; Guion, 2003).

How might this discrepancy between stops and vowels be explained? The acquisition of the voicing contrast continues throughout childhood, up until early adolescence (Hazan & Barrett, 2000; Nittrouer & Miller, 1997). In contrast, vowels are thought to develop earlier, the perception of vowels has been shown to be language specific in infants as young as 8-months-old (Kuhl et al., 2006). Thus, one possibility is that the Second-gen speakers acquired robust Sylheti vowel categories early in life but that they did not acquire the Sylheti voicing contrast before they had started to acquire English and that this prevented them from acquiring voicing lead (Macken & Barton, 1980). Another possibility is that Sylheti has fewer vowel categories than English, and each has an existing counterpart in English (e.g., Best, 1995). The Second-Gen speakers may not have separate categories, but rather may use their dominant English categories when producing Sylheti vowels. Thus, difficulties did not arise for the Second-Gen speakers when speaking Sylheti because there is an exact phonetic counterpart in English (see e.g., Flege, 1995).

Unlike the Late learners, the Second-Gen speakers production of English was not significantly different from their monolingual SSBE counterparts. Such results are not surprising given that English was the speakers' dominant language, and they had started to acquire English during childhood (see also, Flege, Frieda & Nozawa, 1997). However, based on previous studies of the production of the host language in immigrant groups, and more broadly multicultural cities, a variety that reflects Multicultural London English may have been expected (MLE; see e.g., Cheshire, Fox, Kerswill & Torgersen, 2008; Cheshire et al., 2011; Fox, 2010; Torgersen, Kerswill & Fox, 2006). Although this was not a direct aim of the study, informal observational comparisons between the vowel data and that of the MLE data (Cheshire et al., 2011) suggest that some features of MLE are present in the data, namely, GOOSE-fronting (i.e., higher F2 for /u/) and FOOT-backing (i.e., lower F2 for /ʊ/). Likewise, informal observations of the English VOT data, suggest that the Second-Gen speakers used slightly shorter VOT values than did the SSBE speakers. Although this result did not reach significance, it is possible that this pattern reflects an on-going change in London English.

Alternatively, one could imagine that the Second-Gen speakers were influenced by their own L1 Sylheti stop categories, or were exposed to Sylheti-accented English variants within the community. Previous studies have shown that Early learners production is not identical to that of their monolingual peers (e.g. Baker &

Trofimovich, 2005), and the findings in this study suggest that even early learners may display differences in their production at a fine-grained acoustic level. Indeed, it is possible that speakers who acquire both languages from birth have a wider repertoire of phonetic variants, i.e., they may use MLE-like variants with certain interlocutors, in certain settings (see e.g., Sharma & Sankaran, 2011). The formal design of the study may have elicited speech that was closer to that of SSBE and likely does not reflect the full range of variants in their repertoire.

Although there was no significant difference between early and second-generation speakers' production of English and Sylheti stops and vowels, trends in the data suggest that early differences in language experience might have some effects on certain aspects of production. Early speakers arrived in the UK during childhood, and like the Second-Gen speakers, acquired English at school in the UK. However, the Early group were born in Bangladesh, and were monolingual Sylheti speakers up until they emigrated to the UK. English was the dominant language for both groups, but the Early group reported using either English or Sylheti i.e., both languages, in more situations than did the Second-Gen speakers, who mainly used English. The plosive data suggest that such differences in early language experience may have resulted in different production patterns. In general, the Early speakers used slightly shorter VOT values than did the Second-Gen speakers. This was particularly evident for Sylheti velar stops. Early speakers distinguished between Sylheti /k/-/g/, but Second-Gen speakers used similar VOT values for all Sylheti velar stops. Likewise, the Early speakers used different VOT values for English and Sylheti voiced and voiceless velar stops, but Second-Gen speakers only distinguished between the voiceless English and Sylheti velar stops.

It is possible that these small differences in production between Early and Second-Gen speakers may be the result of their different experiences with Sylheti. The Early speakers were born in Bangladesh and were solely exposed to Sylheti for the period before they arrived in the UK. In contrast, the Second-Gen speakers, likely had some exposure to English in London (e.g., through older second-generation family members) during infancy, even though they had been primarily exposed to Sylheti before attending English-speaking nursery. This early monolingual experience for the Early group may have allowed them to develop more robust Sylheti phonetic categories than the Second-Gen speakers. In addition, the Early group may have had less English input on arrival to the UK. The Early speaker would have arrived in the

UK with their monolingual Sylheti-speaking parents (Late speakers). Thus, one could imagine that as new immigrants to the London Bengali community, these Early learners may have had fewer native English-speaking contacts than the Second-Gen speakers whose parents had spent more time in the UK. (see e.g., Flege & Piske, 2006; Fox, 2010).

In sum, this chapter has shown differences in the production of Sylheti (L1) and English (L2) within the London-Bengali community. These differences reflect the language background of the speaker e.g., Late arrivals vs. Second-Gen speakers. For Sylheti, differences were observed for plosive production, but not for vowel production. For English, differences were observed between speaker groups, namely the Late arrivals production reflected their L1, whereas the Second-Gen speakers reflected their monolingual English peers. Such patterns of production pose interesting questions for the children who grow up in immigrant communities. These findings suggest that those born in London (i.e., Second-Gen) successfully acquire the speech sounds of the host country. However, such findings shed little light on how these sounds are initially acquired. These speakers were initially exposed to Sylheti up until they entered full-time English-speaking school at around four years of age, and could be considered as sequential bilinguals. Further, based on the findings from this study, it is likely that any English input that these speakers would have received from main caregivers would have been Sylheti-accented. To fully understand how and when these speakers acquired the English categories, investigation of the initial stages of acquisition are needed.

The chapters to follow will investigate the English phonemic acquisition of the Bangladeshi-heritage children who are born and grow up in the London-Bengali community. Chapter 3 will introduce the aims, research questions and outline of the longitudinal study presented in Chapter 4 – 7.

3. Introduction to chapters 4 - 7

This chapter provides an introduction to the longitudinal child study reported in Chapters 4 and 5. It will describe the aims of the child study, the main research questions, as well as the experimental design that was adopted. In addition, this chapter will provide definitions of some of the key terminology used in Chapters 4 and 5.

The overall aim of the studies reported in Chapters 4 and 5 was to track the acquisition of English phonemes by sequential bilinguals from the London-Bengali community. As discussed in Chapter 1, the majority of the children who grow up in the London-Bengali community are mainly exposed to the heritage language (L1), Sylheti, up until they enter full-time English speaking education at around 3.5 years-old. The tight-knit nature of the community combined with extended family living in the home provides the children with a rich Sylheti-speaking environment. In addition, as shown in Chapter 2, depending on the language background of the main caregiver, any English exposure from caregivers who acquired English during adulthood is likely to be Sylheti-accented. For many children from the London-Bengali community, the entry to nursery will be their first intensive exposure to English. From nursery age, these children will be in a predominately English-speaking environment five days a week at nursery/school. Such patterns of language exposure pose interesting questions from both a theoretical and educational perspective: Do sequential bilingual children initially establish phonemic categories that reflect their dominant Sylheti input? If this is the case, do they demonstrate difficulties with English phonemic contrasts that do not exist in Sylheti e.g., /iː/-/ɪ/? Furthermore, with English experience in nursery and school do sequential bilingual children acquire the English phonemic contrasts?

The target phonemic contrasts, English monophthongal vowels and plosives were based on the findings presented in Chapter 2. As shown in Chapter 2, Sylheti has a smaller vowel system than that of English and phonemic contrasts that are meaningful in English are not present in Sylheti, e.g., /i/-/ɪ/. For plosives, Sylheti uses prevoiced and short-lag voicing, in contrast to English that has a short-long lag distinction. Important for this study are the findings that show differences in Sylheti and English production from different speakers in the London-Bengali community. This means that the speech input that the children in this community are likely to be exposed to will depend on the parents' age of acquisition of English and Sylheti (i.e., Late arrivals vs. Early arrivals) as well as socio-cultural influences e.g., such as language use patterns, and length of residence.

The specific English monophthongal vowel and plosive contrasts chosen for investigation were based on two sources of information; 1) the comparisons between Sylheti and English i.e., the contrasts that exist in Sylheti, but not in English, and, 2) the contrasts that the adult Late arrival speakers (i.e., first-generation) from Chapter 2 found difficult. For example, the experiments in Chapter 2 revealed that Sylheti only contains the /i/ high front vowel, whereas English has the /iː/-/ɪ/ contrast.

3.1. Study design

To investigate the acquisition of English phonemic contrasts a longitudinal design was adopted. Such a design is advantageous over cross-sectional research in developmental research. As highlighted by Singer and Willet (2003), studying the same group of children over a period of time reduces cohort effects. That is, the control of confounding variables such as experience and cognitive factors is reduced in a cross-sectional design, where different sets of children of different ages are tested at one time point. This is particularly important when studying children from immigrant communities, where the children's life experiences are likely to vary e.g., place of birth, background of main caregivers and older siblings, time spent in the host country and home country. Testing the same group of children twice allows for some control of these factors, or at least, it allows for a detailed analysis of factors that may have influenced any changes in development.

The aim of the current longitudinal study was to track the acquisition of English phonemic contrasts by Sylheti-speaking sequential bilinguals from the London Bengali community. Specifically the aim was to investigate whether these children initially have difficulties with an English contrast that does not exist in their home language, Sylheti. If so, do these children acquire the English contrasts? Children were tested twice: during the period of initial exposure to English, and then again after a period of exposure to English. Therefore, all children were tested twice: during the first year of English-speaking preschool education, when the children had had an average of 7-months full-time English-speaking education with some Sylheti language support, and one year later when the children were in the first year of Primary education, reception class (see Figure 3.1). In addition, to gather information

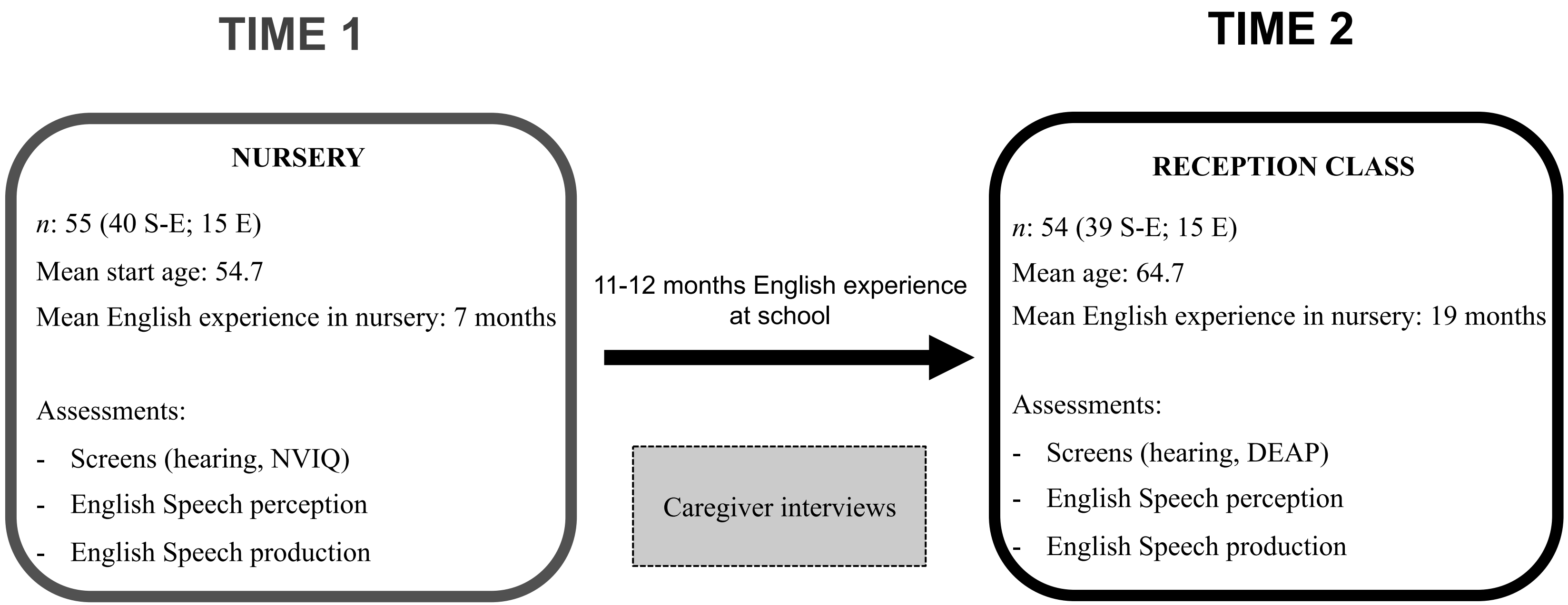


Figure 3.1. Diagram illustrating the longitudinal design of the child study reported in Chapters 4 to 7.

on the children's environmental factors e.g., patterns of language exposure, the main caregivers of the children were interviewed.

To ensure that the testing procedures were appropriate for the two age groups, all tests and procedures were piloted on 8 children [4 nursery age (2 Sylheti-English, 2 monolingual English), and 4 reception class age (3 Sylheti-English, 2 monolingual English)]

3.2. Terminology

The terms below will be defined as follows in current research:

Main caregiver/s: before the child enters full-time nursery education this term refers to the person who spends at least 70% of the child's wake time with him/her, as well as attending to the child's main needs: feeding, washing, playing. Once the child has entered full-time education, the main caregiver is defined as the person who spends the most time with child outside of nursery/school hours, as well as attending to the child's main needs.

Preschool education: refers to the education for children under 5 years old, also referred to as the Early Years Foundation stage.

Nursery education: refers to the point at which free preschool education is available in the UK (> 3 years) up until the first year of Primary School education (5 years of age). In the UK children aged 3-4 years old are entitled to 15 hours of free education.

Reception class: refers to first year of Primary school education. In the UK, children are required to start education by 5 years of age. Children normally start Reception class in a state school from the September after their 4th birthday.

4. The acquisition of English vowel contrasts

4.1. INTRODUCTION

This chapter reports the first part of the longitudinal study. The aim of this study was to track the acquisition of English vowel contrasts by Sylheti-English sequential bilinguals. In sum, Southern British English, the target second language in this research, has 11 monophthongal vowels: /iː, ɪ, ɛ, æ, ʌ, ɜː, ɑː, ɒ, ɔ, uː, ʊ/, including tense-lax pairs. In contrast, Sylheti has a smaller vowel inventory, containing 7 monophthongal vowels: /i, e, ɛ, a, u, ɔ, o/, where vowel length is not contrastive. The target English vowel contrasts used in the experiments in this chapter were chosen based on a comparison of Sylheti and English vowel space as well as the findings from the experiments reported in Chapter 2 (adult study). Not only does English have more vowels than Sylheti, English contains vowel contrasts that do not exists in Sylheti e.g., English /uː/-/ʊ/, whereas Sylheti only has the /u/ vowel (see Chapter 2 for more details). Chapter 2 showed that adult Sylheti speakers (especially late learners) have difficulties with such contrasts. Based on these findings, three English contrast were chosen: /iː/-/ɪ/, /ɑː/-/ʌ/ and /uː/-/ʊ/.

Two experiments were conducted to track the children's acquisition of the English vowel contrasts. Experiment 1 investigated the children's perception of naturally produced Standard Southern British English (SSBE) vowel contrasts. To avoid ceiling effects and to test how robust the children's English vowels categories were, the perception test was ran in quite and noise (2dB SNR), separately. To investigate the children's production of English vowels, Experiment 2 reports an acoustic-phonetic analysis of vowels produced in a picture-naming task.

4.2. Experiment 1: perception

4.2.1. MATERIALS AND METHODS

Participants

Fifty-five preschool children successfully completed the study: 40 Sylheti-English sequential bilinguals (mean start age: 52.7 months, range 46-57 months; 25 female, 15 male) and 15 monolingual English controls (mean start age: 54.2 months, range 47-57 months; 8 female, 15 male). The Sylheti-English children were recruited from preschool children's centres and primary school nurseries in the London Boroughs of Tower Hamlets and Camden, where children of Bangladeshi-origin were the majority (mean 70%, range 50% - 82%). To ensure that the monolingual English-speaking children had limited contact with Sylheti, they were recruited from school nurseries and children's centres in Camden and Hackney that had less than 20% children of Bangladeshi-origin.

To be included in the study, all children had to have: (1) normal hearing thresholds, (2) no documented history of chronic middle ear infections, and (3) no documented history of speech and language difficulties. All participants had to pass three screening tests: (1) an audiometric hearing screen of the frequencies 0.5, 1.0, 2.0 and 4.0 kHz presented at 25 dB HL, (2) a non-verbal IQ screen – within 1 standard deviation of the mean for the block design subtest of the BAS-II (British Ability Scales, 2nd edition, Elliot, 1996), and (3) a picture pointing and production screen of the target stimuli (children had to be able to identify and produce, without phonological errors, all of the target words used in the study). In addition, at Time 2, all children completed a phonology assessment screen from the DEAP (Diagnostic Evaluation of Articulation and Phonology, Holm, Dodd, & Hua, 2002). An additional 11 children participated in the study but were excluded from the analyses due to either failing the hearing screen ($n = 3$; 2 bilingual, 1 monolingual), failing the non-verbal IQ screen ($n = 3$; 2 bilingual, 1 monolingual), failing the phonology screen ($n = 2$; 1 bilingual, 1 monolingual), poor attention ($n = 1$; 1 bilingual), failing to complete the practice block ($n = 1$; 1 bilingual), or failing the target word screen ($n = 1$; 1 bilingual).

To be considered sequentially bilingual, a child needed to have a maximum of 20% exposure to English from the main caregivers from birth to entering pre-school education. Language exposure was measured using an adapted version of a language exposure questionnaire developed by McCarthy (2009), see Appendix 2a and Chapter 6 for more detail on how language exposure was scored. Based on previous bilingual questionnaires (Bosch & Sebastián-Gallés, 2001) the questionnaire required the main caregiver to provide an estimation (in hours, for every day of the week) of the exposure to English and Sylheti from the child's main caregivers (e.g. parents, grandparents) before the child entered nursery. All children, apart from two, were children of first-generation Bangladeshi-origin parents i.e., both parents were born in Sylhet, Bangladesh, and arrived in the UK either as children or adults. The remaining two children had one first-generation parent. All Sylheti-English bilinguals either resided in Tower Hamlets (*n* = 21) or Camden (*n* = 19). All children were tested twice: (1) in nursery, when the bilinguals had had an average of 7 months full-time English speaking preschool education (2) 11-12 months later, in the first year of Primary School education, reception class.

Socioeconomic status.

Demographic variables such as parents' level of education (in the UK or Bangladesh), place of language acquisition (English and Sylheti) and level of English (if attended classes in the UK) were collected during the interviews. Where possible the children were matched for socioeconomic status, based on maternal and paternal level of education (see e.g., Bradley & Corwyn, 2002). The level of education was categorized based on National Qualifications Framework (England, Wales & Northern Ireland, OFQUAL, 2004) or the Bangladeshi equivalent (based on UK NARIC: British Council, 2011), namely: Primary Education, General Certificate of Secondary Education (GCSE), Advanced Level General Certificate of Education (A-level)/College (Advanced Diploma or National Vocational Qualification level 5), Undergraduate University (Degree, Graduate Certificate or Diploma), and Postgraduate University degree (Masters or Doctorate).

Fifty of the fifty-five families responded to questions regarding the demographic information (46 fathers, 50 mothers). The majority of the parents either had an educational level of GCSE or A-levels/College (65%). Specifically, 9% had completed education up until Primary school, 30% up until GCSEs and 35% up until A-levels/college course. Nineteen percent had completed a Bachelors degree, and 7% a Masters degree.

Stimuli

The selection of the target words was constrained by several factors. All words had to be age-appropriate, and known by the bilingual children. Second, the stimuli were restricted in terms of phonetic environment. Nine real word minimal pairs containing the test English vowel contrasts, front: /iː/-/ɪ/ i.e., *sheep-ship*, open/open-mid back /ɑː/-/ʌ/ i.e., *cart-cut*, and back uː/-/ʊ/ i.e., *pool-pull* were selected. Each vowel contrast group consisted of test contrast and a control vowel i.e., *sheep-ship-**shape*** /eɪ/; *cart-cut-**cat*** /æ/; *pool-pull-**peel*** /iː/ (control vowel in bold). The control vowels were chosen based on two factors: 1) where possible, they were produced in the same vowel space region e.g., open-mid, yet they were distinct from the test vowels, and 2) based on the findings in Chapter 2 and previous descriptions of Standard Bengali (Khan, 2010). The Sylheti adult speakers from the experiments reported in Chapter 2 were able to produce the control vowels accurately i.e., /æ/ and /iː/, or similar sounds existed in reports of Standard Bengali i.e., diphthong /ei/.

To avoid lexical bias (see e.g., Walley & Flege, 1999; Thompson & Hazan, 2010) all of the target words were checked against age appropriate vocabulary lists i.e., Oxford Communicative Development Inventory Database (Hamilton, Plunkett & Schafer, 2000), MRC Lexical database (Wilson, 1988), Bristol Norms (Stadthagen-Gonzalez & Davis, 2006), as well as, classroom vocabulary list, class teachers, and parents. See Table 4.1 for the monolingual English age of acquisition and written word frequency (Thorndike-Lorge, 1944).

Table 4.1. Age of acquisition and written word frequency for the vowel identification target words.

Target word	***Age of acquisition (years)***	***Written word frequency***
Sheep	3.9	86
Ship	2.4	678
Shape	3.03	407
Cut	3.05	943
Cart	3.58	68
Cat	3.23	306
Pool	2.39	194
Pull	2.2	936
Peel	*Not available*	91

The stimuli were recorded by a Standard Southern British English (SSBE) female speaker. To elicit a child-directed speech style, the researcher instructed the speaker to name the words as if she was talking to a child. e.g., lively voice, slower speech rate. To facilitate this, the researcher demonstrated a child-directed speech style to the speaker. All recordings were made in a recording booth using a Rode NTime 1-A microphone and an Edirol UA25 audio interface with a sampling rate of 44.1 kHz, 16-

bit resolution. All recordings were made in a single session. Stimuli were edited with Cool Edit Pro 2.0 and individual stimulus files were created for each word. No silences were left at the beginning and end of the word. Spectral analysis of the formant frequencies at the central, steady-state portion of the vowels and vowel duration from voicing onset to the voicing offset were conducted (see Table 4.2). F1, F2, and duration values corresponded to those of standard descriptions of SSBE (Hawkings & Midgley, 2005; Moreiras, 2006).

Table 4.2. Vowel formant and duration measures for test stimuli.

Target word	*Test vowel*	*Duration (s)*	*F1 (Hz)*	*F2 (Hz)*
Sheep	/iː/	0.15	323	3097
Ship	/ɪ/	0.7	598	2615
Cart	/ɑː/	0.27	700	1236
Cut	/ʌ/	0.10	939	1374
Pool	/uː/	0.22	440	1418
Pull	/ʊ/	0.9	652	1102

For the noise condition, the digitized recordings were embedded in speech-shaped noise with a 0dB signal-to-noise ratio (SNR). The speech shaped noise was based on an approximation of the long-term average speech spectrum estimated from measurements for combined male and female voices (Table II of Byrne et al., 1994). The rms levels per 1/3 octave band were converted into spectrum level and then plotted on an octave scale. A three-line approximation was judged adequate for capturing the major part of the shape from about 60 Hz to 9 kHz. This consisted of a low-frequency portion rolling off below 120 Hz at 17.5 dB/octave, and a high-frequency portion rolling off at 7.2 dB/octave above 420 Hz, with a constant spectrum portion in-between. The final stimuli were peak normalized and the intensity of the sound file was scaled to 70 dB SPL.

The target pictures were colour photographs of the target object or action (see Appendix 2b). To ensure that the target pictures corresponded to the target words, the pictures were piloted on 10 nursery-aged children (5 monolingual-English, 5 Sylheti-

English) using a picture-pointing task. All children correctly matched the target word with the corresponding picture. Therefore all pictures were used in the main study.

Procedure

Children were tested in a quiet room in their nursery/school. The testing was conducted over three 15-minute sessions; 1) Screens, 2) a familiarisation phase and 3) a test phase. See Figure 4.1 for a flow diagram of the testing procedure. The tests were presented to the children in a computer game format on a laptop (see Figure 4.2). The stimuli were presented at 65 dB over Sennheiser HD 25-1 II headphones.

A three-alternative forced choice (3AFC) procedure was used. Children were presented with three pictures in a jungle scene with a monkey or a tiger, and presented with the scenario that *Monkey/Tiger is learning to say new words, and because you already know these words you're the best person to help him*. Children were instructed to *listen to Monkey/Tiger and point to what he says*. In the noise condition, the children were instructed that *Monkey/Tiger is talking in the rain, so it's a little bit noisy in the rainforest and you have to listen extra hard to find out what he says*.

As piloting showed that the nursery aged children could not independently use a mouse, a false touch screen method was used. The researcher recorded the children's responses by clicking on the picture that s/he pointed to. The researcher could not hear the stimuli. At the end of each trial the child was presented with an on screen reward (e.g. a picture of a banana). Following the reward the set of pictures for the next trial would automatically appear. To play the target stimuli the researcher would select the on screen "GO" button.

During the familiarisation phase, the researcher introduced each picture and the corresponding target word to the child. Following this, the children had to correctly name all of the pictures using a picture-naming task, and identify them in a picture-pointing game. In this phase, the target vowel contrasts were not presented together. To familiarize the children with the test procedure, all children completed a practice block of 3 trials. As the aim was to familiarize the children with the procedure, the test vowel contrasts e.g., *sheep – ship*, were not presented together.

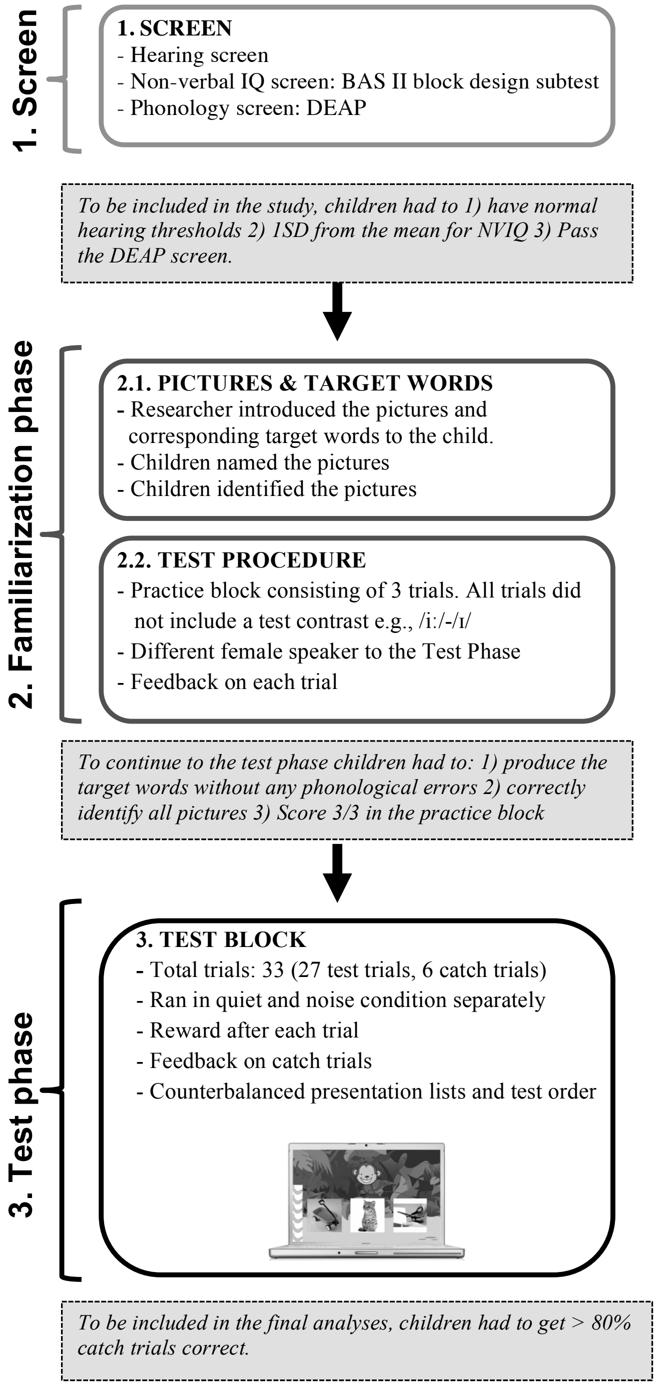


Figure 4.1. Flow diagram illustrating the test procedure for the vowel identification testing. Each phase was conducted in separate 15-minute blocks.

Further, the stimuli used in the practice block were recordings made by a different female SSBE speaker to that of the test phase. A single practice block trial included one control word: *shape, cat, or peel* and two test words from different test vowel groups e.g., *shape* (control word group), *cat* (open/open-mid back test group), *pool* (back test group). For the practice block, children received visual (tick or a cross) and auditory (*well done!* or a 'dong' sound) feedback after each trial, as well as an on-screen visual reward (a flower). Both the animal, and the display background in the practice block were different to the main test block. To continue to the testing phase, all children had to 1) produce the target words without any phonological errors, 2) correctly identify all of the target words in the picture pointing task, and 3) score 3/3 in the familiarisation block.

Figure 4.2. The 3-alternative-forced-choice test display, including the *cart-cut* contrast and on-screen banana rewards.

The test phase consisted of 33 trials: 27 test trials, and six catch trials. The quiet and noise conditions were presented in two separate tests. Test trials consisted of test vowel contrasts e.g., /iː/-/ɪ/, and the control vowel contrasts, e.g., /eɪ/. Catch trials did not include a test minimal pair group, instead they consisted of a control vowel, e.g. /eɪ/ and two test words from differing vowel groups e.g., /ɪ/ and /ʊ/. The aim of the catch trials was to track the children's attention and to motivate the child. These trials were interspersed every fifth trial. Children received an online reward e.g., a banana, after each trial. In addition, for the catch trials children received visual (tick or a cross) and auditory (*well done!* or a dong sound) feedback. To control for order effect, three counterbalanced presentation lists were developed. The presentation lists were divided equally amongst the children. For half of the children the quiet condition preceded the noise condition. For the remaining children, the order was reversed. After each test children were given a short break to put stickers in their scorecard booklet.

4.1.2. RESULTS

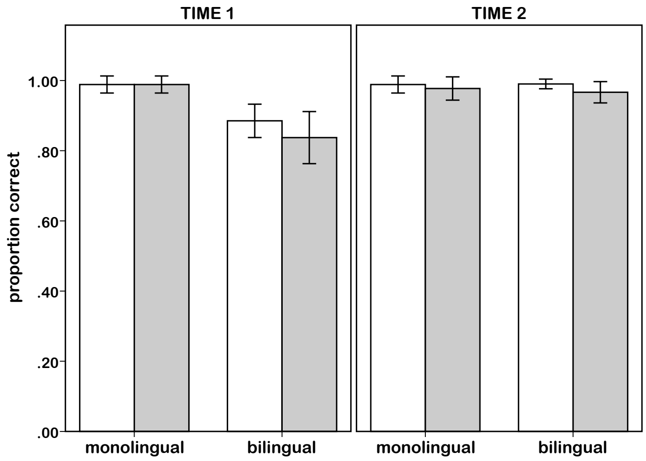
To investigate the influence of socioeconomic status (SES) on the target dependent variables (categorization slope and phoneme boundary), a preliminary analyses comparing the means of the different SES groups (indexed by maternal and paternal level of education) was ran. Separate one-way ANOVAs revealed that there was no significant difference in slope ($p > .05$) or phoneme boundary ($p > .05$) between SES groups. Thus, all of the participants were included in the final analyses.

To test for differences in perception, a Generalized Linear Mixed Model (GLMM) analysis, using the identity link function, was conducted in SPSS 20. Language group, time, vowel group and condition (i.e., quiet or noise) were treated as fixed factors. To account for individual differences between the children, participant was treated as a random factor. Significant interactions were explored with Sequential Sidak post-hoc tests. In addition, vowel confusions matrices for each separate vowel group will be presented. To be included in the final analysis children had to get 80% of the catch trials correct (as in Nittrouer, 2005). The final data set included results from 52 children (37 bilingual, 15 monolingual English).

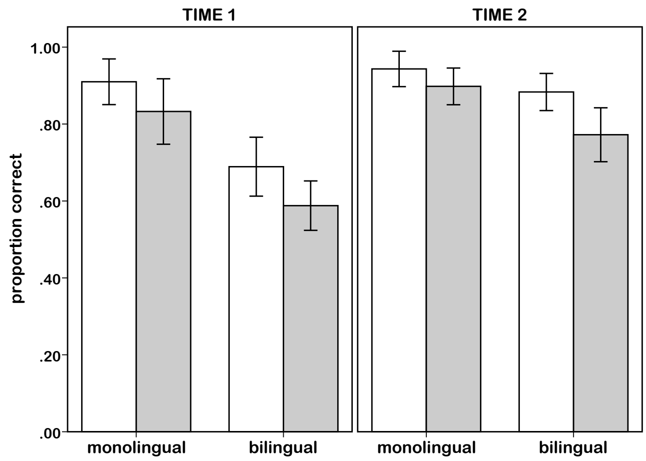
Figure 4.3 and Table 4.3 display the children's vowel identification (VID) scores for each vowel group. The GLMM analyses revealed a main effect of group, $F(1, 46) = 42.04$, $p < .001$, time, $F(1, 546) = 42.99$, $p < .001$, vowel group, $F(2, 538) = 113.09$, $p < .001$, condition, $F(1, 538) = 16.80$, $p < .001$, and a significant interaction between group and time, $F(1, 546) = 22.96$, $p < .001$, and group and vowel, $F(1, 538) = 14.38$, $p < .001$. Sequential Sidak post hoc analyses revealed that overall the children had significantly lower scores at Time 1 than they did at Time 2 ($p < .05$), where significantly lower scores were obtained in the noise condition than in the quiet condition ($p < .05$). Overall, the monolinguals had significantly higher VID scores than did the bilinguals ($p < .05$). All children obtained significantly lower VID scores for the /uː/-/ʊ/ contrast followed by the ɑː/-/ʌ/ and /iː/-/ɪ/ contrast respectively ($p < .05$). The bilinguals displayed a significant increase in VID scores from Time 1 to Time 2 ($p < .05$), however there was no significant difference between the monolingual VID scores at Time 1 and Time 2 ($p > .05$).

Inspection of the confusion matrices (Tables 4.4 and 4.5) shows that the bilinguals consistently confused the test contrasts, but not the control contrasts (all ≥ 96%). At Time 1, this is particularly evident for the /uː/-/ʊ/ contrast, followed by the

/iː/-/ɪ/



/ɑː/-/ʌ/



/uː/-/ʊ/

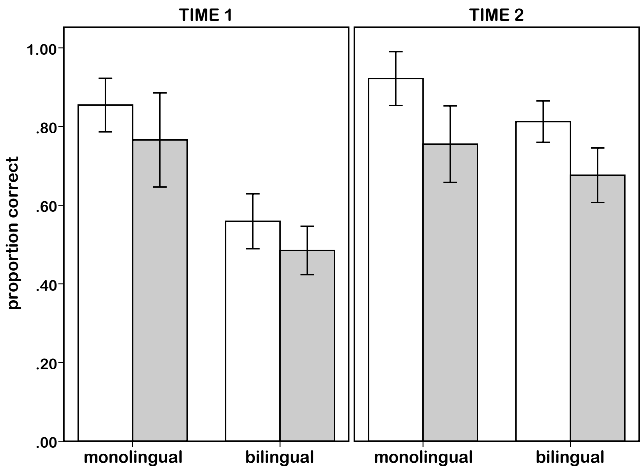


Figure 4.3. Bar charts representing children's mean VID score for the separate test vowel contrasts in quiet (white) and noise (grey), at Time 1 (nursery) and Time 2 (reception class). The error bars enclose 95% confidence intervals.

Table 4.3. Monolingual and bilingual children's VID scores in quiet (Q) and noise (N), at Time 1 and Time 2.

		Monolinguals			Bilinguals		
		Proportion correct					
		/iː/-/ɪ/	**/ɑː/-/ʌ/**	**/uː/-/ʊ/**	**/iː/-/ɪ/**	**/ɑː/-/ʌ/**	**/uː/-/ʊ/**
Time 1	*Mean Q*	.98	.90	.85	.88	.69	.58
	Median Q	1	1	.83	1	.67	.56
	SD Q	.04	.10	.12	.02	.22	.18
	Mean N	.97	.83	.76	.83	.59	.51
	Median N	1	.83	.83	1	.67	.56
	SD N	.04	.15	.21	.21	.18	.17
Time 2	*Mean Q*	.99	.94	.86	.99	.85	.81
	Median Q	1	1	.83	1	.83	.83
	SD Q	.04	.08	.12	.02	.20	.13
	Mean N	.98	.89	.78	.96	.77	.69
	Median N	1	.83	.83	1	.83	.67
	SD N	.06	.09	.17	.08	.20	.18

Table 4.4. Confusion matrices for the separate vowel groups in quiet and noise (shaded grey) as identified by the Sylheti-English bilingual children at Time 1 and Time 2. Percentage of responses are given, all values ≤ 1% are not shown.

	TIME 1				TIME 2		
	Response				***Response***		
Stimulus	/i/	/ɪ/	/eɪ/	***Stimulus***	/i/	/ɪ/	/eɪ/
/i/	**88**	10	2	/i/	**98**	2	
/ɪ/	13	**87**		/ɪ/		**100**	
/aɪ/			**100**	/aɪ/			**100**
/i/	**83**	15	2	/i/	**97**	3	
/ɪ/	15	**85**		/ɪ/	7	**92**	
/eɪ/		4	**96**	/eɪ/			**99**

	Response				***Response***		
Stimulus	/ɑ/	/ʌ/	/æ/	***Stimulus***	/ɑ/	/ʌ/	/æ/
/ɑ/	**68**	30	2	/ɑ/	**90**	9	
/ʌ/	25	**73**	2	/ʌ/	14	**86**	
/æ/			**99**	/æ/			**99**
/ɑ/	**58**	35	7	/ɑ/	**84**	15	
/ʌ/	38	**59**	3	/ʌ/	25	**74**	
/æ/		2	**97**	/æ/		2	**98**

	Response				***Response***		
Stimulus	/u/	/ʊ/	/ɪə/	***Stimulus***	/u/	/ʊ/	/ɪə/
/u/	**59**	31	10	/u/	**86**	14	
/ʊ/	35	**61**	4	/ʊ/	20	**78**	2
/ɪə/			**99**	/ɪə/			**100**
/u/	**54**	36	10	/u/	**76**	23	
/ʊ/	41	**50**	9	/ʊ/	41	**59**	
/i/	2	2	**96**	/i/			**100**

Table 4.5. Confusion matrices for the separate vowel groups in quiet and noise (shaded grey) as identified by the monolingual English children at Time 1 and Time 2. Percentage of responses are given, all values ≤ 1% are not shown.

TIME 1

	Response		
Stimulus	/iː/	/ɪ/	/eɪ/
/iː/	**100**		
/ɪ/	3	**97**	
/aɪ/			**100**
/iː/	**96**	2	2
/ɪ/	2	**98**	
/eɪ/			**100**

	Response		
Stimulus	/ɑː/	/ʌ/	/æ/
/ɑː/	**92**	7	
/ʌ/	9	**90**	
/æ/		2	**98**
/ɑː/	**88**	10	2
/ʌ/	15	**82**	3
/æ/		2	**98**

	Response		
Stimulus	/uː/	/ʊ/	/ɪə/
/uː/	**90**	10	
/ʊ/	21	**75**	4
/ɪə/			**100**
/uː/	**74**	24	2
/ʊ/	22	**78**	
/iː/	3		**97**

TIME 2

	Response		
Stimulus	/iː/	/ɪ/	/eɪ/
/iː/	**99**		
/ɪ/		**100**	
/aɪ/			**100**
/iː/	**100**		
/ɪ/	4	**96**	
/eɪ/	2	2	**96**

	Response		
Stimulus	/ɑː/	/ʌ/	/æ/
/ɑː/	**97**	3	
/ʌ/	6	**92**	2
/æ/			**100**
/ɑː/	**89**	7	4
/ʌ/	6	**91**	3
/æ/		2	**98**

	Response		
Stimulus	/uː/	/ʊ/	/ɪə/
/uː/	**83**	13	4
/ʊ/	7	**88**	5
/ɪə/			**100**
/uː/	**82**	14	4
/ʊ/	18	**76**	6
/iː/			**100**

/ɑː/-/ʌ/ contrast, and lastly the /iː/-/ɪ/ contrast. Specifically, the bilinguals identified /ʊ/ as /uː/ 34% in quiet and 41% in noise; and /uː/ as /ʊ/ 31% in quiet and 36% in noise. For /ɑː/-/ʌ/, the bilinguals identified /ɑː/ as /ʌ/ 30% in quiet and 35% in noise; and /ʌ/ as /ɑː/ 25% in quiet and 38 % in noise. For /iː/-/ɪ/, all identification scores were greater than 80% correct. At Time 2, most confusions were found for /ʊ/ in quiet (identified as /uː/ 20% of the time) and noise (identified 41% of the time as /uː/). In contrast, observation of Table 4.4 shows that the monolinguals correctly identified /iː/, /ɪ/, /ɑː/ and /ʌ/ more than 74% of the time at Time 1 and 2. Like the bilinguals though, the monolinguals were least accurate at identifying /uː/ and /ʊ/ at Time 1 in noise.

Summary of experiment 1

Overall the bilingual children displayed significantly lower VID scores than did the monolinguals at Time 1. All children displayed lower VID scores in the noise condition than in quiet. The bilingual children consistently confused the target test vowel contrast, and demonstrated accurate identification for the control contrasts. By Time 2, the bilingual children displayed a significant increase in VID scores from Time 1, however an overall significant difference between the groups remained. Closer inspection of the individual vowel groups showed a significant difference in identification accuracy across the vowel groups. Specifically, all children obtained the lowest VID scores for the /uː/-/ʊ/ contrast followed by the /ɑː/-/ʌ/ and /iː/-/ɪ/ contrast respectively.

4.3. Experiment 2: production

4.3.1. MATERIALS AND METHODS

Participants

The same children who participated in Experiment 1 also participated in Experiment 2 (40 Sylheti-English and 15 monolingual English). As in Experiment 1 all children were tested twice: in nursery (mean: 52.7 months) and one year later in the first year of Primary school education, reception class. Experiment 2 was conducted on a different day to Experiment 1. All of the 55 children were recorded at Time 1; at Time 2 one child was absent during the testing session.

Target sounds

The same English monophthong vowel contrast investigated in Experiment 1 were investigated here: front: /iː/-/ɪ/; open/open-mid back: /ɑː/-/ʌ/; back: /uː/-/ʊ/ were elicited in a CVC context in real English words. Each sound was elicited four times per child: two words per target sound, produced twice. The same target words used in Experiment 1 were used: *sh**ee**p* /ʃiːp/, *sh**i**p* /ʃɪp/, *c**u**t* /kʌt/, *c**ar**t* /kɑːt/, *p**u**ll* /pʊl/, *p**oo**l* /puːl/, plus six additional words in a *b*VC context: ***bea**d* /biːd/, ***bi**g* /bɪg/, ***bu**bble* /bʌbəl/, ***Bar**t* /bɑːt/, ***boo**k* /bʊk/, ***boo**t* /buːt/ were used.

As with Experiment 1, all of the words were checked against age appropriate vocabulary lists i.e., Oxford Communicative Development Inventory Database (Hamilton, Plunkett & Schafer, 2000), MRC Lexical database (Wilson, 1988), Bristol Norms (Stadthagen-Gonzalez & Davis, 2006), as well as, classroom vocabulary list, class teachers, and parents (see Table 4.6 and 4.1 for age of acquisition and written word frequency (Thorndike-Lorge, 1944). All words were elicited using pictures (see Appendix 2b and 2c). To ensure that the target pictures corresponded with the target word, all pictures were piloted on 10 nursery-aged children (5 monolingual-English and 5 Sylheti-English bilinguals) using a picture-pointing task. The pilot showed that all of the children correctly matched the target word with the corresponding picture.

Table 4.6. Age of acquisition and written word frequency for target /b/VC words.

Word	***Age of acquisition***	***Written word frequency***
Bead	*Not available*	*55*
Big	2.1	1773
Bubble	2.72	90
Bart*	*n/a*	*n/a*
Book	2.14	684
Boot	2.51	134

*Bart is the name of a well-known children's cartoon character.

Procedure

All recordings were made in a quiet room in the nursery/school using a H2 Zoom audio recorder with a sampling rate of 44.1 kHz, 16-bit resolution. The words were elicited in a randomised order using a picture naming technique.

Acoustic analysis

A total of 2376 vowel tokens were analyzed. Formant and duration measures were made in Praat (Boersma & Weenink, 2012). The stimuli were located manually, and then F1 and F2 were extracted using hand-corrected LPC analyses. Formant frequencies were measured from the midpoint of the steady state portion of the vowel. The steady-state portion of the vowel was defined as the part of the vowel that was closest to the midpoint and where the formant frequencies were most stable. So that data could be compared, Lobanov's z-score transformation (Lobanov, 1971) was used to normalize F1 and F2 values (see e.g., Palethorpe, Wales, Clark & Senserrick, 1996, for a discussion on this). All duration measurements were taken from voicing onset to the voicing offset. Voicing onset was determined by the beginning of the F2 transitions, voicing offset was determined by the end of the F2 transitions, corresponding to the closure of the final consonant.

4.3.2. RESULTS

To test for differences in vowel production, separate Generalized Linear Mixed Model (GLMM) analyses, using the identity link function, were conducted in SPSS 20. For all analyses, language group, time, and vowel were treated as fixed factors. To account for individual differences between the children, participant was treated as a random factor in all analyses. Significant interactions were explored with sequential Sidak post-hoc analysis. Separate analyses were conducted for F1, F2 and duration.

Preliminary analysis of the production data revealed a significant difference between the two phonetic environments, /p/V/l/ and /b/VC, for the back vowels /uː/-/ʊ/. Specifically, the /p/V/l/ (i.e., *pool, pull*) context resulted in a significantly lower F2 than did the /b/VC (i.e., *boot, book*) context. This is likely to be due to two factors: 1) GOOSE and FOOT fronting, a phenomenon observed in British English (Hawkins & Midgley, 2005; Torgersen et al., 2006; Wells, 1982), causing the back vowels to move forward in the vowel space (i.e., higher F2) in specific phonetic contexts i.e., for *book* and *boot,* and 2) The velarization of /l/ preventing fronting for *pull* and *pool* (see e.g., Torgersen, 2002).

A series of one-way ANOVAs confirmed a significant difference between *book* and *pool* for F1, $F(1, 207) = 83.39$, $p < .001$, and F2, $F(1, 207) = 242.47$, $p < .001$, and a significant difference between *boot* and *pull* for F1, $F(1, 207) = 161.39$, $p < .001$; and F2, $F(1, 207) = 131.48$, $p < .001$. For, /iː/-/ɪ/ and /ɑː/-/ʌ/, there was no significant difference in F1 and F2 for the differing phonetic environments for the remaining phonetic environments ($p > .05$). Thus, the two phonetic environments for the back vowels were analyzed separately and will be hereafter referred to as /uː/-/ʊ/ I for the /b/VC context and /uː/-/ʊ/ II for the /p/V/l/ context.

/iː/-/ɪ/

Figure 4.4 displays the monolingual and bilinguals normalized formant values for the target vowels at Time 1 and Time 2 (see Appendix 2d for more detail). Table 4.7 displays the children's vowel duration values. For /iː/-/ɪ/, the GLMM analyses revealed a significant main effect of vowel [F1, $F(1{,}829) = 609.14$, $p < .001$; F2, $F(1{,}803) = 504.51$, $p < .001$; duration, $F(1{,}776) = 493.09$, $p < .001$]. There were no other significant main effects or interactions ($p > .05$).

Sequential Sidak post hoc analyses revealed that /ɪ/ was produced with a significantly higher F1 and lower F2 than /iː/ ($p < .05$). No significant difference was found between Time 1 and Time 2 ($p > .05$). There was no significant difference between the bilinguals and monolinguals for F1, F2 and duration ($p > .05$). All children produced /iː/ with a longer vowel duration than /ɪ/, at Time 1 and Time 2 ($p < .05$).

/ɑː/-/ʌ/

For /ɑː/-/ʌ/, the GLMM analyses revealed a significant main effect of vowel [F1, $F(1{,}79) = 59.0$, $p = .001$; F2, $F(1, 808) = 166.27$, $p < .001$; duration, $F(1{,}768) = 976.39$, $p < .05$]. For F2, there was a significant main effect of time, $F(1{,}780) = 31.11$, $p < .001$, group, $F(1{,}57) = 19.22$, $p < .001$, and a significant interaction between group and vowel $F(1,811) = 30.20$, $p < .001$. There were no other significant main effects or interactions. Sequential Sidak post hoc analyses revealed that /ʌ/ was produced with a significantly higher F1 and F2 and shorter duration than /ɑː/. The bilingual children produced /ɑː/ with a significantly lower

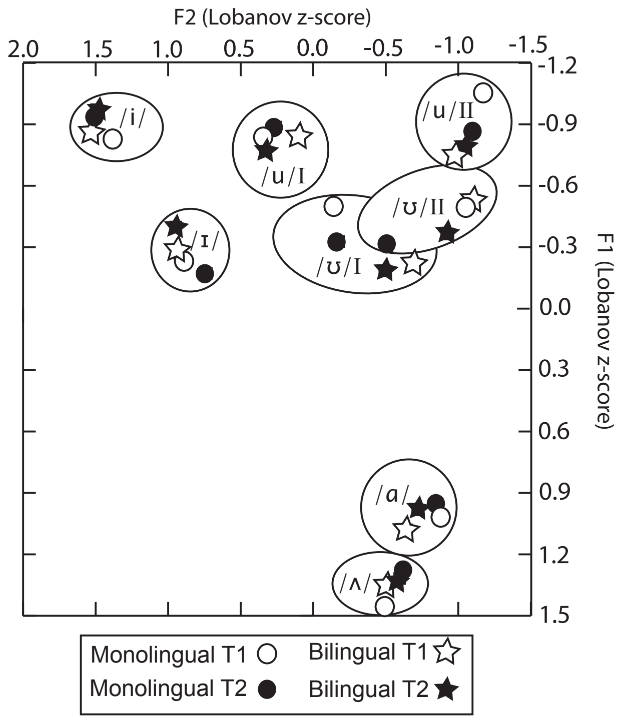


Figure 4.4. Mean z-score normalized formant frequency values for the bilingual and monolingual children's production of English vowels at Time 1 and Time 2. The x-axis displays decreasing F2 values from left to right, the y-axis displays increasing F1 values from top to bottom. Ovals are used to separate the test vowel groups.

F2 than did the monolingual children. Overall, the vowels were produced with a significantly higher F2 at Time 2.

/uː/-/ʊ/ I (boot – book)

For /uː/-/ʊ/ I, the GLMM analyses revealed a significant main effect of vowel [F1, $F(1,412) = 170.31$, $p < .001$; F2, $F(1, 172) = 105.753$, $p < .001$; duration, $F(1,343)= 167.77$, $p < .001$]. For F1 and F2, there was a significant main effect of group [F1, $F(1,411) = 6.975$, $p = .009$; F2, $F(1,210) = 4.421$, $p < .001$] and interaction between group and vowel [F1, $F(1,411) = 7.98$, $p = .005$; F2, $F(1,172) = 5.52$, $p = .02$]. All other main effects and interactions were not significant ($p > .05$). Sequential Sidak post hoc analyses revealed that overall, *boot* /uː/ was produced with a higher F1, lower F2 and longer duration than *book*

Table 4.7. Monolingual-English, and Sylheti-English speaking bilingual children's vowel duration values in normalized z-scores at Time 1 (Time 1) and Time 2 (Time 2). A higher z-score reflects a longer vowel duration.

		Monolingual								***Bilingual***							
		/iː/	**/ɪ/**	**/ɑː/**	**/ʌ/**	**/uː/ I**	**/ʊ/ I**	**/uː/ II**	**/ʊ/ II**	**/iː/**	**/ɪ/**	**/ɑː/**	**/ʌ/**	**/uː/ I**	**/ʊ/ I**	**/uː/II**	**/ʊ/ II**
Time 1	*mean*	.42	-.62	.88	-.61	.10	-.59	12	-.45	0.53	-.58	.75	-.66	.10	-.59	.80	-.60
	median	.03	-.70	.85	-.67	.4	-.70	.8	-.63	.32	.37	.81	-.06	.4	.58	-.07	.-65
	SD	1.10	.99	.78	.49	.80	.64	.54	63	1.02	1.0	.90	.75	.80	.65	67	.45
Time 2	*mean*	.61	-.69	1.07	-.74	.27	-.79	.16	-.67	.61	-.79	.75	-.82	.19	-.62	.76	-.55
	median	.55	-.75	1.13	-.76	.21	-.67	.7	-.70	-.67	-.78	.77	-.85	.07	-.64	.72	-.58
	SD	1.04	1.03	.70	.28	.62	.56	.89	.70	.67	.44	.82	.38	.41	.54	.65	.44

/ʊ/. The bilinguals produced *book* with a significantly higher F1 and lower F2 than did the monolinguals ($p < .05$). There was no significant difference in production between Time 1 and Time 2 ($p > .05$).

/uː/-/ʊ/ II (pool – pull)

For /uː/-/ʊ/ I, the GLMM analyses revealed a significant main effect of vowel [F1, $F(1, 365) = 112.88$, $p < .001$; F2, $F(1, 379) = 14.65$, $p < .001$; duration, $F(1, 143) = 109.45$, $p < .001$], time [F1, $F(1, 365) = 9.589$, $p = .002$; F2, $F(1, 368) = 18.93$, $p < .001$; duration, $F(1, 404) = 4.87$, $p = .028$] and a main effect of group for duration, $F(1, 404) = 6.72$, $p = .010$. There was a significant interaction between vowel and group for F1 and F2 [F1, $F(1, 371) = 6.86$, $p = .009$; F2, $F(1,383) = 15.55$, $p < .001$], and a significant interaction between group and time for F2, $F(1, 368) = 8.75$, $p = .003$. All other main effects and interactions were not significant ($p > .05$).

Sequential Sidak post hoc analyses revealed that overall /uː/II was produced with a significantly higher F1 and F2, and a longer duration than /ʊ/II ($p < .05$). The bilinguals produced /ʊ/II with a significantly higher F1 than did the monolinguals ($p < .05$). For F2, the bilinguals production of /uː/II was not significantly different from /ʊ/II ($p > .05$) at Time 1 and Time 2. In contrast, the monolinguals produced *pull* with a lower F2 than they did for *pool* ($p < .05$), and demonstrated an increase in F2 from Time 1 to Time 2 ($p < .05$).

Summary of experiment 2

All children were able to distinguish between the vowel contrasts in terms of formant values and duration. Minimal differences were found between the bilingual and monolingual children. Specifically, the bilingual children produced /ɑː/ with a higher F2 than did the monolingual children. Although this difference was significant, it is important to note that in relation to the rest of the vowel space this difference was relatively small i.e., -0.62 (bilingual) vs. -0.70 (monolingual). The main differences were found in the children's production of the back vowels, and, as expected differences were found in two phonetic environments i.e., /uː/-/ʊ/ I and /uː/-/ʊ/ II. All children displayed GOOSE fronting for /uː/I *boot,* however for /ʊ/I *book*, the monolingual children used a significantly higher F2 than did the bilingual children i.e., FOOT fronting. The bilinguals showed an increase in F2 from Time 1 to Time 2 however the significant difference between the monolingual and bilinguals remained. For /uː/-/ʊ/ II, the bilingual children produced /ʊ/II with a higher F1 than did the monolinguals, yet in relation to the whole vowel space, this was relatively small i.e., -0.77 vs. -0.95 Hz (z-score transformed).

4.4. DISCUSSION

The aim of Chapter 4 was to track the acquisition of English monophthongal vowel contrasts by Sylheti-English sequential bilingual children. To do so, the children's perception and production of English monophthongs was assessed and compared to that of their monolingual English speaking peers at two time points: when they were in nursery (Time 1), and one year later during the first year of primary school (Time 2). For perception, the children's identification of naturally produced English vowel contrasts that were hypothesized to be difficult for Sylheti speaking listeners was assessed. To investigate how robust the children's vowel categories were, perception abilities were assessed in quiet and noise. Vowel production was investigated through an acoustic-phonetic analysis of the children's production of vowels that were elicited during a picture naming game.

At Time 1, the bilingual children displayed overall significantly lower vowel identification scores than did their monolingual English peers in both the quiet and noise condition. As expected, the monolingual English controls consistently identified all vowels, with an average of 92% correct. Closer inspection of the vowel confusions revealed that the bilinguals misidentified the vowel contrasts that do not exist in their first language (Sylheti) e.g., /ʊ/ was identified as /uː/ 35% of the time, whereas, they consistently correctly identified the control vowels in both noise and quiet. A possible explanation for these findings is that the bilingual children are initially sensitive to the vowels of their ambient input. Before entering nursery, all of the bilingual children were predominately exposed to Sylheti from their main caregivers. Furthermore, based on the findings from chapter two, it is likely that any English that the children would have been exposed to is likely to have been Sylheti-accented. Developmentally, vowels have been shown to be the first speech sounds to become language-specific during infancy (see e.g., Kuhl et al. 2008), and recent research has shown that infants are sensitive to fine-grained variation in caregiver's speech (Cristià, 2011). Thus, a possible explanation could be that on entry to nursery the sequential bilinguals had started to develop Sylheti vowel categories, and that in addition, any developing English vowel categories may have been influenced by the potential Sylheti-accented speech input from their main caregivers i.e., collapsed English phonetic categories that reflect Sylheti.

Interestingly, closer inspection of the individual vowel group scores at Time 1 showed that both the monolingual and bilingual children had the highest identification scores for the /iː/-/ɪ/ contrast, followed by /ɑː/-/ʌ/ and /uː/-/ʊ/ respectively. For example, the bilinguals misidentified /uː/ for /ʊ/ 31% of the time, in contrast to /iː/ that was misidentified as /ɪ/ only 10% of the time. Although this pattern was mainly significant in the bilingual children's scores in quiet and in noise, the monolinguals also had lower scores for the /uː/-/ʊ/ contrast, particularly in the noise condition, e.g., /uː/ was identified as /ʊ/ 24% of the time, in contrast to /ɪ/ that was identified as /iː/ only 2% of the time. Although one minimal pair per contrast and the differing phonetic environments limits a comparison between the separate vowel contrasts, such findings may suggest a general developmental pattern. To test this hypothesis, further research is needed either using the same phonetic environment across vowel contrasts or multiple minimal pairs. Yet, as found in the current study, minimal pairs for child experiments are often limited by age of acquisition and imageability of target word.

Alternatively, the pattern in the data may reflect a stimulus effect, specifically for the back vowels /uː/-/ʊ/. Firstly, the /p/V/l/ phonetic environment may have affected the phonetic realization of these vowel. That is, the final /l/ will have caused the back vowels to be retracted, preventing the fronting of the vowels and possibly making less spectral distinction in the contrast. Further, the stimuli measurements in Table 4.1 show that although the contrasts differ for duration, for formant values, the front vowels have a larger distinction between /iː/ and /ɪ/ for F2 in comparison to the other vowel groups. If children are particularly sensitive to the F2 for distinguishing contrasts, then such differences may have affected the children's perceptual accuracy.

For production, the results show fewer differences between the monolinguals and bilinguals. Importantly, the bilingual children distinguished the target vowel contrasts using significantly different formant and duration values. Thus, unlike the findings from the adult speakers in Chapter 2, these children distinguish between the vowel contrasts, albeit different from their monolingual peers. It could therefore be suggested that such observational patterns reflect the variation between speakers, rather than a difference in vowel category.

The only significant differences in production were evident in the back vowels /ʊ/ and /ɑː/, namely, the bilingual children produced /ʊ/ with a significantly lower F2 than did the monolinguals, who displayed fronting of the /ʊ/ vowel, a phenomenon known as FOOT fronting (Hawkins & Midgley, 2005; Torgersen et al., 2006). These findings are in line with the Torgersen et al., (2000) study that showed a similar difference between "Anglo" (i.e., white-British) and 'non-Anglo' (i.e., of ethic origin) from the London Borough of Hackney. Interestingly, both groups displayed GOOSE fronting for the /uː/ vowel. Although /uː/ was not investigated in the Torgersen et al., (2000) study, these findings suggest that the bilingual children may have acquired this phonetic variant since starting nursery. The findings from the second-generation speakers in Chapter 2 suggest that /uː/ fronting is also present in speakers from the London Bengali community, thus it is likely that this variant will eventually be acquired by the children in the current study. Therefore, it could be suggested that the bilingual children in this study are in the process of acquiring this particular variant.

By Time 2, all children displayed an increase in VID scores, suggesting a general developmental trend. The bilingual children showed a significant increase in VID scores for all test vowel contrasts, however, the bilingual children had overall lower VID scores than did the monolinguals. Investigation of the specific vowel groups revealed that this difference was driven by the bilinguals perception of the /ɑː/-/ʌ/, and /uː/-/ʊ/ contrast. Specifically, at Time 2 the bilinguals no longer performed significantly different from the monolinguals for the /iː/-/ɪ/ contrast in quiet and in noise. Although the bilinguals displayed a significant increase in perception scores for all contrasts, a significant overall difference between bilinguals and monolinguals remained in both the noise and quiet condition.

Similar patterns of vowel perception have been found for 11 year old Turkish-German sequential bilingual children, who had started to acquire German around 2;9 years old (Darcy & Krüger, 2012). The study showed that the bilinguals performed significantly worse on a German vowel discrimination task than did their monolingual German speaking peers. Darcy & Krüger partly attribute this finding to the maintenance of the children's L1, not only was Turkish the language spoken at home, these children were enrolled in a dual language Turkish-German school. The authors suggest that the high use of the L1, even for the case of early learners, is likely to influence the L2 (see also, Pallier et al., 1997). Similarly, in the current research, Sylheti was still dominant within the home and the local community for the bilingual children. It could therefore be suggested that these children had a similar pressure to maintain their L1, and in turn this affected the children's perception of L2 vowels, even after one year of English experience at school.

For production, the only significant difference between the monolinguals and bilinguals at Time 2 was for the /ʊ/ vowel. Similar to Time 1, the bilingual children used a significantly lower F2 than did the monolinguals. Interestingly, although the difference was still significant at Time 2, the bilinguals' produced /ʊ/ with a significantly higher F2 at Time 2 than at Time 1, suggesting some change in production i.e., they had started to front this variant.

5. The acquisition of the English voicing contrast

5.1. INTRODUCTION

This chapter reports the second part of the longitudinal study. The aim of the experiments reported in this chapter was to track the acquisition of the English voicing contrast by Sylheti-English sequential bilinguals. As shown in Chapter 2, in Sylheti, although VOT varies with place of articulation, voiced plosives are produced with voicing lead (mean: -75ms), and voiceless plosives with short-lag (0-30ms). In contrast, English voiced plosives fall into the short-lag range and voiceless plosives fall into the long-lag range, with an average of 21ms and 56ms VOT, respectively (Docherty, 1992). What is particularly interesting for the purpose of this study is the overlap between the voiced plosives in English and the voiceless plosives in Sylheti: the English voiced plosives (/b/, /d/, /g/) fall into the same VOT region as Sylheti voiceless plosives (/p/, /t/, /k/). In addition, Chapter 2 demonstrated that first-generation immigrants use Sylheti-like voicing patterns when producing English plosives, and so it is likely that the children in this study will be exposed to foreign-accented variants in English.

This study focused on the acquisition of English bilabial and velar plosives. Two experiments were conducted to track the children's acquisition of the English voicing contrast. Experiment 1 investigated the children's perception of the English voicing contrast was assessed using a real word identification task. Specifically, the children's categorization of a /p/-/b/ and /k/-/g/ voicing continua was assessed. Experiment 2, investigates the children's production of English plosives.

5.2. Experiment 1: perception

5.2.1 MATERIALS AND METHODS

Participants

The same children who participated in the experiments in Chapter 4 also participated in the experiments reported in Chapter 5 (40 Sylheti-English and 15 monolingual English). As in Chapter 4 all children were tested twice: in nursery (mean: 52.7 months) and one year later in the first year of Primary school education, reception class. The experiments reported in this chapter were conducted in a different session and day to the experiments reported in Chapter 4.

Stimuli

The same *pea-bee* and *coat-goat* continua used in Hazan, Messaoud-Galusi, Rosen, Nouwens and Shakespeare, 2009, and Ramus et al., 2003. See Figure 5.1. Stimuli were generated by copy-synthesis using the cascade branch of the Klatt synthesizer (Klatt, 1980). The aim of copy-synthesis is to obtain stimuli that are controlled in order to focus on specific features, but are also natural sounding, as all parameters are based on utterances produced by a single speaker. For each minimal pair, initial values for fundamental and formant frequency, vowel duration and burst characteristics were measured from a natural *bee* /bi/ and *goat* /gəʊt/ tokens recorded by a native female British English speaker. For *pea-bee* the total syllable duration was 390ms. For *bee*, F1 was set at 390 and reached 185Hz at the end of the vowel. The F2 and F3 transition increased from 1400Hz and 2500Hz respectively to 2540Hz and 2970Hz, and reached 2760Hz and 3377Hz at the end of the vowel. F4 was set at 3950Hz. For *goat* the total syllable duration was 459ms. The F1 transition increased from 477Hz to 640Hz, and

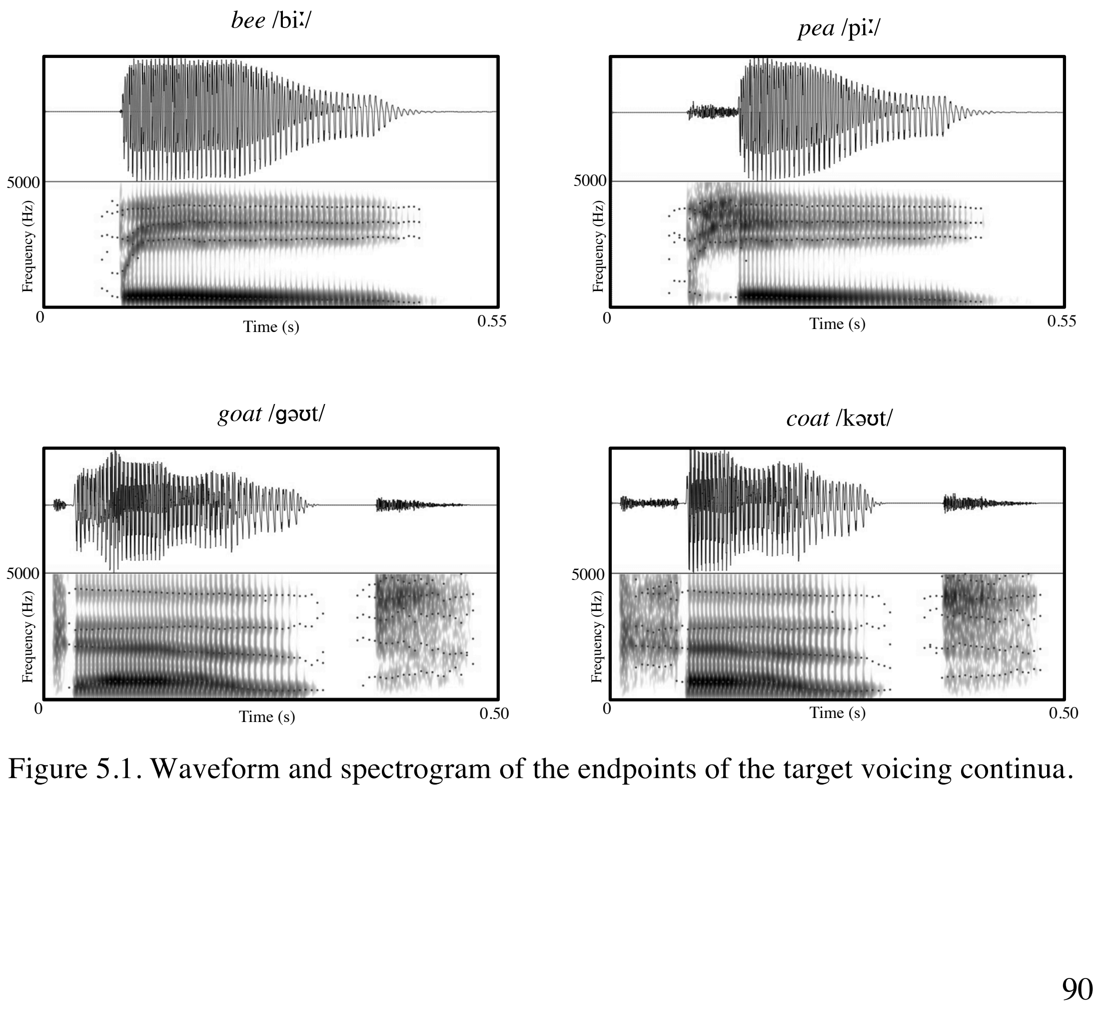


Figure 5.1. Waveform and spectrogram of the endpoints of the target voicing continua.

decreased to 306Hz at the end of vowel, F2, F2 and F4 were set at 2080, 2900 and 4380Hz respectively, and reached 1645, 2800, 4130Hz. To obtain stimuli differing in voice onset time (VOT), the continua were generated by delaying the onset of the voicing whilst simultaneously increasing the duration of aspiration. All continua varied across VOT in 1ms steps. For *pea-bee*, the endpoints ranged from 0ms VOT for /bi/, to 60ms VOT for /pi/. For *coat-goat*, the endpoints ranged from 20ms VOT for /gəʊt/ to 70ms VOT for /kəʊt/. For both continua, the first formant onset frequency covaried with VOT, as it naturally does.

To avoid lexical bias (see e.g., Walley & Flege, 1999; Thompson & Hazan, 2010) all of the target words were checked against age appropriate vocabulary lists i.e., Oxford Communicative Development Inventory Database (Hamilton, Plunkett & Schafer, 2000), MRC Lexical database (Wilson, 1988), Bristol Norms (Stadthagen-Gonzalez & Davis, 2006), as well as, classroom vocabulary list, class teachers, and parents. See Table 5.1 for the monolingual English age of acquisition and written word frequency (Thorndike-Lorge, 1944). In addition, children had to identify the words in a 3-forced choice picture-pointing task, and correctly produce the target words in a picture-naming task. See Appendix 3a for target pictures.

Table 5.1. Age of acquisition (in years) and written word frequency of target words.

Target words	***Age of acquisition (years)***	***Written word frequency***
Bee	3.91	172
Pee	3.90	*Not available*
Goat	3.90	57
Coat	2.97	896

Procedure

The tests were presented to the children in a computer game format on a laptop in a quiet room in the nursery/school. See Figure 5.2 for an illustration of the test procedure. The stimuli were presented at 65 dB over Sennheiser HD 25-1 II headphones. The participants were instructed that *Panda is learning to say new words, and because you already know these words you're the best person to help him.* Participants were instructed to *listen to Panda and point to what he says*. An on-screen reward (bamboo) was given after each trial. Verbal feedback i.e. *Well*

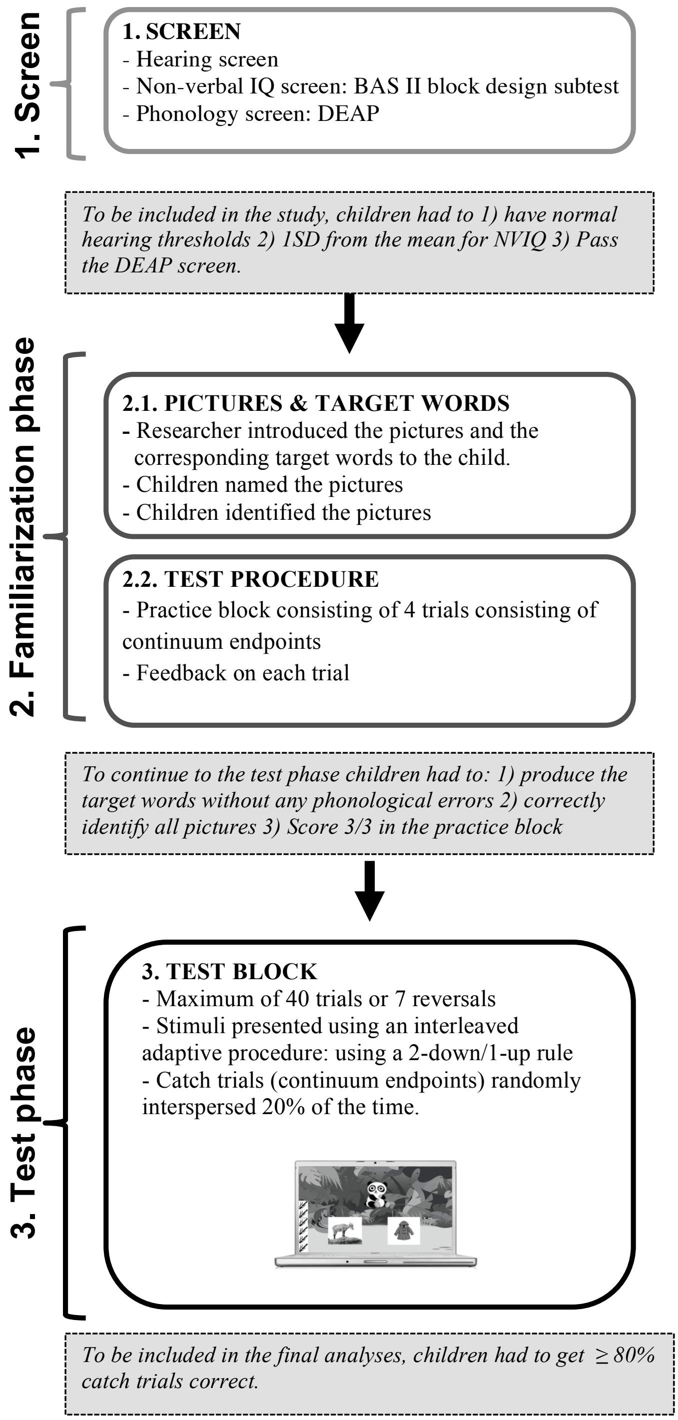


Figure 5.2. Flow diagram illustrating the test procedure for the vowel identification testing. Each phase was conducted in separate 15-minute blocks.

done! and a tick (correct response), or dong sound and a cross (incorrect response), was only given for catch trials (continuum endpoints). The task consisted of a familiarisation block and test trials. A two-alternative forced-choice task was used to assess category identification. The participants identified the stimulus by pointing to a picture of the target word e.g. a pea or a bee. The familiarisation block consisted of 4 trials (2 per endpoint) and the participants received feedback after each trial. To continue to the test block, participants had to score 100% correct (i.e. 4/4) in the familiarization block. Test stimuli were presented using an interleaved adaptive procedure as described in Ramus et al., (2003). The advantages of using an adaptive procedure are: (1) trials are concentrated in the region most crucial for estimating the phoneme boundary and slope of the function, making an efficient use of a small number of presentations, (2) the level of difficulty is consistent across participants as the level of performance (71% *coat* or *goat* responses) is tracked for each listener.

Two independent tracks were used, starting at opposite ends of the continuum. Each track operated under the same rules, and was designed to track 71% of voiced (*bee/goat)* or voiceless (*pea/coat)* responses using a 2-down/1-up rule (Levitt, 1971). The initial step size was 10ms, reducing linearly over the first 3 reversals. To move quickly to the region of interest, the initial track used a 1-down/1-up rule, after the first reversal the 2-down/1-up rule was used. To track attention, continuum endpoints (catch trials) were randomly interspersed 20% of the time. The task ended after 7 reversals or a maximum of 40 trials. See Appendix 3b for an example track of the adaptive procedure.

For each test, the responses to all trials (including catch trials) were aggregated and logistic regression was used to obtain a best-fit sigmoid function, giving estimates of the slope and phoneme boundary. The boundary locates the 50% point on the continuum, that is, the point at which the percept changes from one phonemic category to the other for a particular listener. The slope of the identification function indicates the listener's sensitivity to variations in the particular acoustic feature used in the continuum. A shallower slope indicates a lower degree of consistency in the labeling of the continuum, and in turn less refined phonological categories.

To track attention, the interspersed endpoint trials were analyzed separately. For a session to be included in the final analyses, performance on these interspersed endpoint trials had to be 80% or better (as in Nittrouer, 2005).

5.2.2. RESULTS

To investigate the influence of socioeconomic status (SES) on the target dependent variables (categorization slope and phoneme boundary), a preliminary analyses comparing the means of the different SES groups (indexed by maternal and paternal level of education) was ran. Separate one-way ANOVAs revealed that there was no significant difference in slope ($p > .05$) or phoneme boundary ($p > .05$) between SES groups. Thus, all of the participants were included in the final analyses.

To test for differences in perception, separate Generalized Linear Mixed Model (GLMM) analyses, using the identity link function, were conducted in SPSS. For all analyses, language group and time were treated as fixed factors. To account for individual differences between children, participant was treated as a random factor in all analyses. Significant interactions were explored with sequential Sidak post-hoc analysis. Slope values were log transformed for statistical tests because of their skewed distribution.

/p/-/b/ (pea-bee)

A total of 18 children (3 monolingual, 15 bilingual) did not meet the attention criterion leaving results for 37 children (12 monolingual, 25 bilingual).

Figure 5.3 shows the averaged data in the monolingual and bilingual group. These graphs reveal an increase in the steepness of the slope from Time 1 (nursery) to Time 2 (reception) for all children.

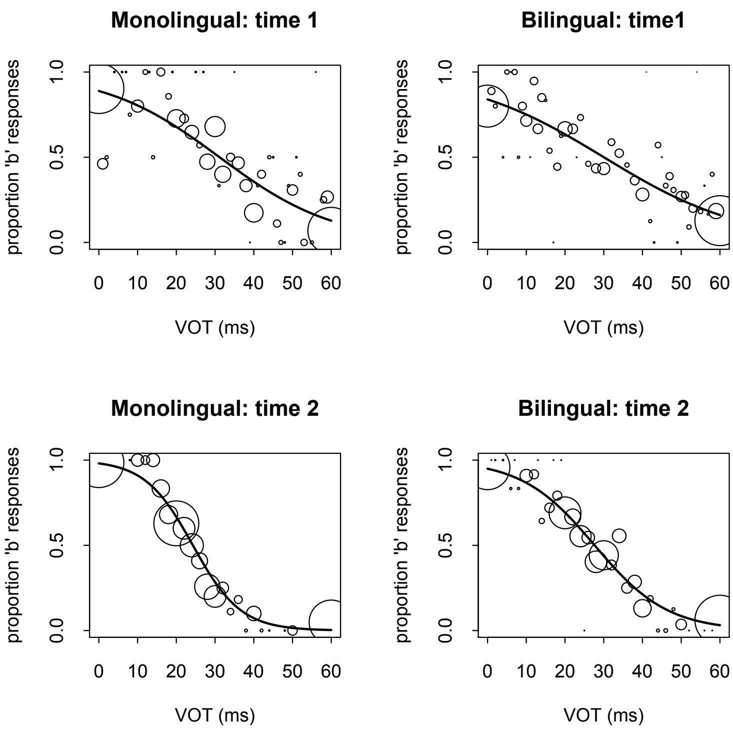


Figure 5.3. Identification functions for the /p/-/b/ contrast aggregated over listeners, for bilingual and monolingual children at Time 1 (nursery) and Time 2 (reception). The size of the circles is proportional to the number of trials at a given point on the continuum i.e., larger circle indicate to more trials.

The slopes at Time 1 reveal a slight difference in the sharpness of categorization slope between bilingual and monolinguals, where the bilinguals have a slightly shallower slope. Individual differences, and the adult target slope value (mean: -0.25) can be observed in Figure 5.4. The range of slope values was larger for both groups at Time 2, however their slope values were closer to that of the adult target. GLMM analyses revealed no significant interaction between group and time, $p > .05$. The main effect of group, $F(1, 37) = 4.02$; $p < .05$, was not significant, confirming that the monolinguals and bilinguals did not differ at Time 1 or 2. The main effect of time, $F(1,37) = 71.28$; $p < .001$, was significant, confirming that the steepness of the slope increased significantly between Time 1 and 2 for both groups.

For phoneme boundary, Figure 5.5 reveals a difference between Time 1 (mean: 31ms) and 2 (mean: 24ms) for the monolingual children, but only slightly for the bilinguals (Time 1 mean: 30ms, Time 2 mean: 27ms). The box plots suggest that overall there was within group variation for both groups, and no difference between the groups at Time 1 and 2. At Time 2, the children's boundary was closer to that of the adult target (mean: 23ms).

GLMM analyses revealed no significant interaction between group and time, $p > .05$. The main effect of group was not significant, $p > .05$, confirming no difference between the two groups. The main effect of time, $F(1,49) = 5.477$; $p < .05$, was significant. Thus, overall there was a change in phoneme boundary between Time 1 and 2. Observation of the box plots (Figure 5.5) suggests that this effect was driven by the monolingual group, who displayed a much larger change in the location of their phoneme boundary from Time 1 to Time 2 than did the bilinguals.

/k/-/g/ (coat-goat)

A total of 13 children (3 monolingual, 10 bilingual) did not meet the attention criterion, giving results for 41 children (12 monolingual, 29 bilingual).

Figure 5.6 shows the averaged data for the monolingual and bilingual groups. These graphs reveal an increase in the steepness of the slope from Time 1 to Time 2 for the bilingual, but not for the monolingual children. Individual differences and the adult target slope value (mean: -0.25) can be observed in Figure 5.4. Both groups

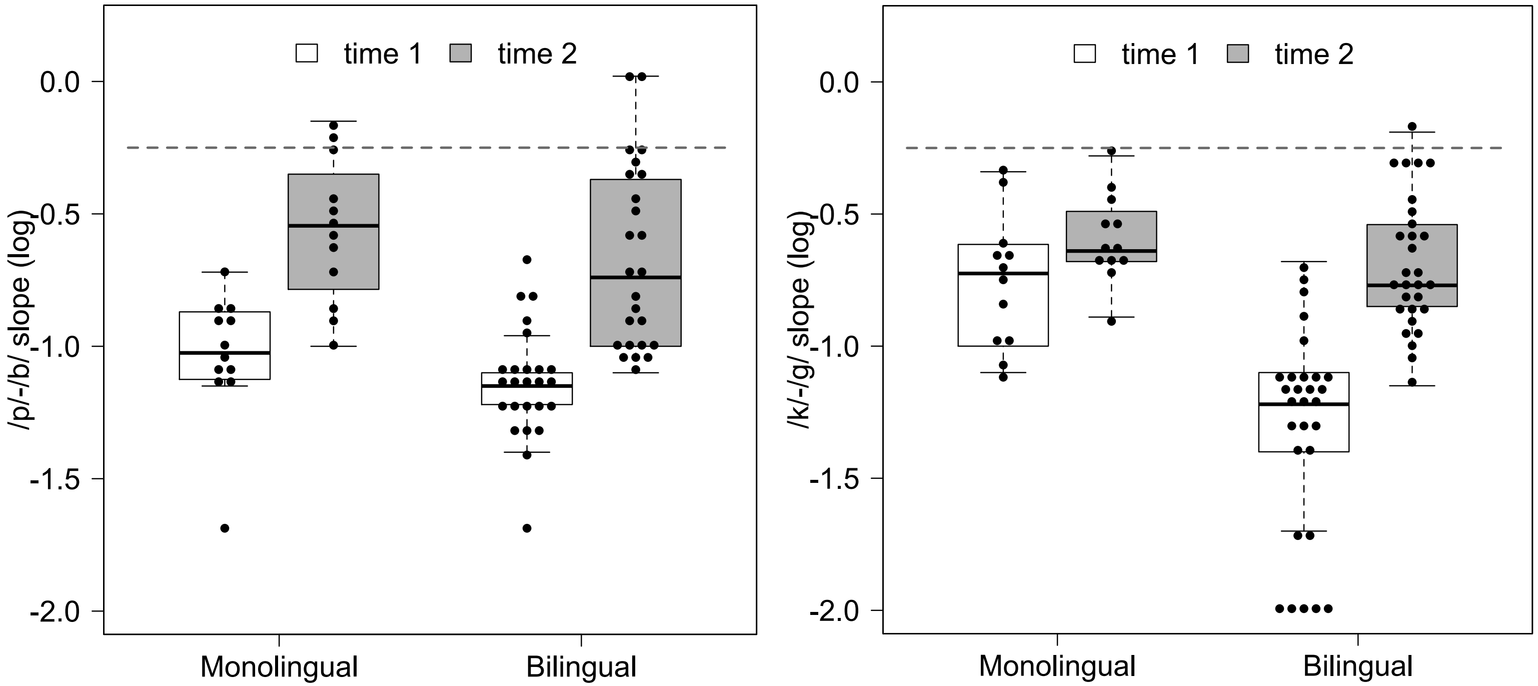


Figure 5.4. Boxplots of slope values for monolingual and bilingual children's /p/-/b/ and /k/-/g/ identification function. The white boxes represent Time 1, and the grey boxes represent Time 2. The black dots represent individual data points. The dashed grey line represents the adult monolingual English mean slope value.

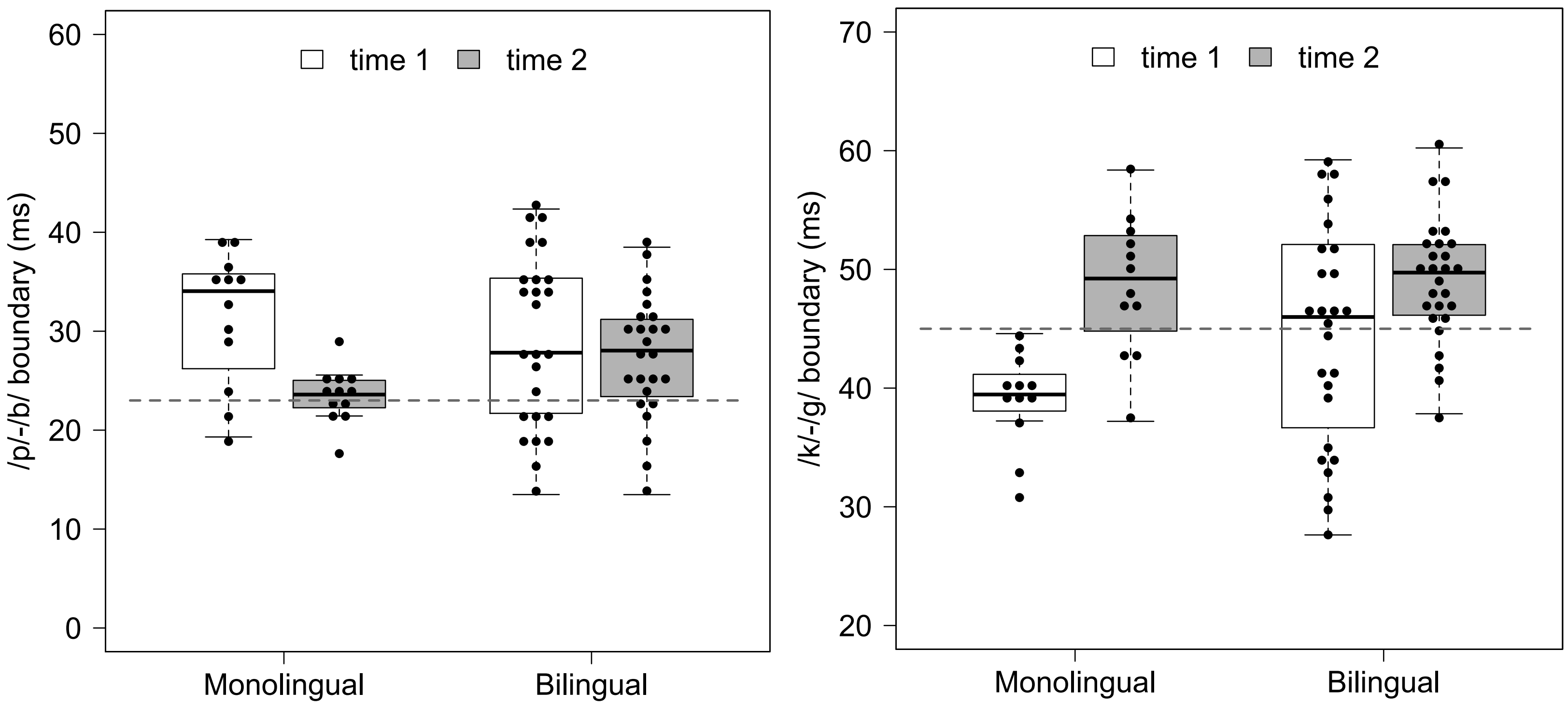


Figure 5.5. Boxplots of monolingual and bilingual children's phoneme boundary (VOT, milliseconds) for /p/-/b/ and /k/-/g/. The white boxes represent Time 1 and the grey boxes represent Time 2. The black dots represent individual data points. The dashed grey line represents the adult monolingual English listeners.

display within-group variation at Time 1 and 2, but at Time 2 all children have values that are closer to that of the adult target.

GLMM analyses revealed that the main effect of group, $F(1,44) = 9.03$; $p < .005$, time, $F(1,41) = 40.22$; $p < .001$, and an interaction between group and time, $F(1,41) = 5.53$; $p < .03$, were all significant. Sequential Sidak post-hoc tests confirmed a significant difference in slope between monolinguals and bilinguals at Time 1 ($p < .05$), but not for Time 2 ($p > .05$). The difference in the steepness of the slope between Time 1 and 2 was significant for the bilingual children ($p < .05$), but not for the monolinguals ($p > .05$). In sum, the bilingual children improved significantly between Time 1 and 2, and in turn did not differ from the monolinguals at Time

Figure 5.5 reveals a difference in phoneme boundary between Time 1 (mean: 43ms) and Time 2 (mean: 49ms) for all children; all children had a longer VOT boundary is evident at Time 2. The box plots suggest that there was no difference between the groups either at Time 1 or Time 2. GLMM analyses revealed no

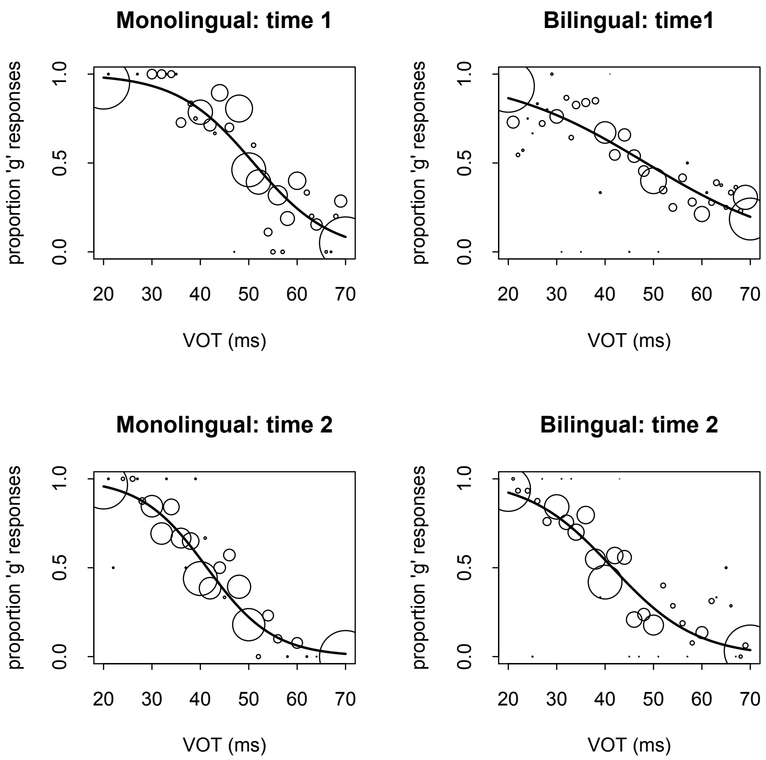


Figure 5.6. Identification functions for the /k/-/g/ contrast aggregated over listeners, for bilingual and monolingual children at Time 1 (nursery) and Time 2 (reception). The size of the circles is proportional to the number of trials at a given point on the continuum i.e., larger circle indicate to more trials.

significant interaction between group and time, $p>0.05$. The main effect of group was not significant, $p>0.05$. The main effect of time, $F(1,39) = 10.612$; $p<0.01$, was significant, confirming the change in phoneme boundary between Time 1 and Time 2.

Summary

Overall, the results indicate that there are changes in perceptual acuity for both bilingual and monolingual children. That is, from nursery testing to reception testing (1 year later) all children display an increase in the steepness of the slope of their identification functions for the bilabial *pea-bee* continua, indicating that they have refined their underlying phonemic categories. For the *coat-goat* continua, all children displayed an increase in the steepness of the slope, but this was only significant for the bilingual group. Further, the bilinguals and monolinguals had similar phonetic boundaries for both continua at Time 1 and Time 2, but all children shifted the location of their phonetic boundary from Time 1 to Time 2, such that informal inspection revealed it was closer to that of the monolingual English-speaking adults at Time 2 than at Time 1.

However, monolinguals and bilinguals differed in the steepness of the slope of the identification function at Time 1, and these effects were different for the bilabial and velar continua. For /k/-/g/, bilinguals had a significantly shallower slope at Time 1 than did monolinguals. By Time 2 however, the bilinguals' slope sharpness had increased such that they no longer differed from their monolingual peers. In contrast, though there were some differences between monolinguals and bilinguals in slope values for the /p/-/b/ continuum at Time 1, these did not reach significance. There were also no differences in slope values between monolinguals and bilinguals at Time 2 for the /p/-/b/ continuum; both groups displayed a significant increase in slope values from Time 1 to Time 2 but this did not differ between groups.

5.3. Experiment 2: production

5.3.1. MATERIALS AND METHODS

Participants

The same children who participated in Experiment 1 also participated in Experiment 2 (40 Sylheti-English and 15 monolingual English). As in Experiment 1 all children were tested twice: in nursery (mean: 52.7 months) and one year later in the first year of Primary school education, reception class. Experiment 2 was conducted on a different day to Experiment 1.

Target sounds

English bilabial and velar plosives: /p/, /b/, /k/, /g/ were elicited in word-initial stressed position. For each consonant two words were elicited (see Table 5.2).
As for the perception tests, all of the target words, they were checked against age appropriate vocabulary lists i.e., Oxford Communicative Development Inventory Database (Hamilton, Plunkett & Schafer, 2000), MRC Lexical database (Wilson, 1988), Bristol Norms (Stadthagen-Gonzalez & Davis, 2006), as well as, classroom vocabulary list, class teachers, and parents. See Table 4.1 for the monolingual English age of acquisition and written word frequency (Thorndike-Lorge, 1944). In addition, prior to the production session, children had to name the words and identify them in a picture-pointing task. See Appendix 3c for target pictures.

Table 5.2. Target words for plosive production, including age of acquisition and written word frequency.

Target	***Target word***	***Age of acquisition (years)***	***Word frequency***
/b/	*bee* /bi/	3.91	172
	bear /beə/	3.92	605
/p/	*pea* /pi/	3.90	*Not available*
	pear /peə/	4.00	62
/g/	*goat* /**g**əʊt/	3.90	57
	good /**g**ʊd/	3.56	5122
/k/	*coat* /**k**əʊt/	2.97	896
	cat /**k**aet/	3.23	306

Procedure

All recordings were made in a quiet room in the nursery/school using a H2 Zoom recorder with a sampling rate of 44.1 kHz, 16-bit resolution. The words were elicited twice in a randomised order using a picture naming technique.

Missing and excluded tokens

A total of 8 tokens: 3 /b/, 2 /p/ and 3 /g/ were missing from the analysis. This was either due to the child incorrectly naming the picture (e.g., using a given name for a bear), or poor recording quality.

Acoustic analysis

A total of 872 tokens were analyzed. The acoustic measurements were made in Praat (Boersma & Weenink, 2012). The VOT measurements were obtained from the waveform and checked against the corresponding spectrogram. The lag VOT (i.e. above 0ms), was measured as the time between the release of the plosive closure and the first periodic part of the waveform from the following vowel (in milliseconds). For prevoiced tokens, the voicing lead was measured using the waveform and the spectrogram. See Figure 5.7, for examples.

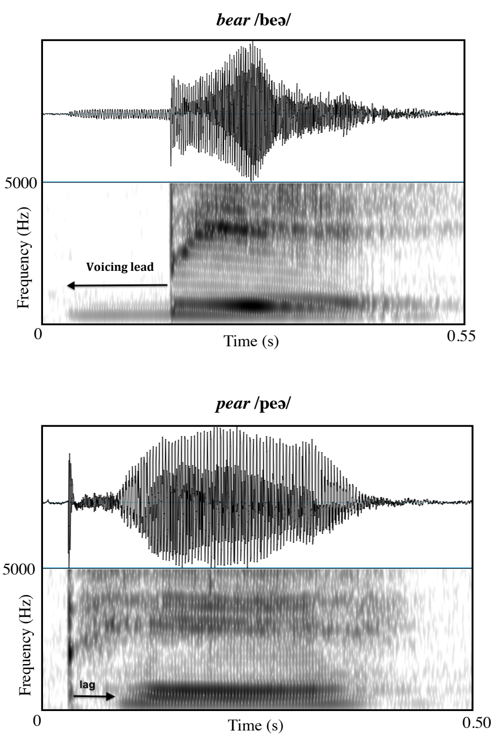


Figure 5.7. Waveform and spectrogram illustrating VOT measurement for a bilingual child's production using voicing lead (*bear*) and lag (*pear*) VOT

5.3.2. RESULTS

As for perception, a preliminary analysis comparing the production of the different SES groups (indexed by maternal and paternal level of education) was ran. Separate one-way ANOVAs revealed that there was no significant difference in VOT values between SES groups ($p > .05$). Thus, all of the participants were included in the final analyses.

To test for differences in production, separate Generalized Linear Mixed Model (GLMM) analyses, using the identity link function, were conducted in SPSS. For all analyses, language group, time and voicing were treated as fixed factors. To account for potential individual differences between children, participant is treated as a random factor in all analyses. Significant interactions were explored with sequential Sidak post-hoc analyses. The production results from each phoneme category (bilabial, velar) are presented separately. For each phoneme, VOT measures for bilingual and monolingual children at Time 1 (nursery) and Time 2 (reception) are tested.

/p/-/b/

Inspection of Figure 5.8 indicates that there are differences between the bilingual and monolingual children at nursery age (Time 1). At Time 1, the bilingual children produced /b/ using short-lag and prevoiced values (mean: -6ms), whereas the monolinguals only used a short-lag (mean: 11ms). For /p/, there was no difference between monolinguals and bilinguals at Time 1 or 2.

GLMM analyses revealed a significant interaction between group and time, $F(1,769) = 10.208$; $p = .001$, and group and stop, $F(1,779) = 4.868$; $p = .028$. The main effect of group, $F(1,51) = 7.784$; $p < .01$, stop, $F(1,779) = 103.453$; $p < .001$, and time, $F(1,769) = 9.148$; $p < .001$, were all significant. Sequential Sidak post-hoc tests revealed that bilinguals used an overall significantly shorter VOT than did the monolinguals at Time 1 ($p < .05$), but not at Time 2 ($p < .05$). Specifically, the bilinguals used a significantly shorter VOT than did the monolinguals for /b/ ($p < .05$). There were no significant differences in the production of /p/, ($p > .05$).

/k/-/g/

Inspection of Figure 5.8 suggests differences between monolingual and bilingual children at Time 1. That is, at Time 1, the bilingual children used a shorter VOT for

/k/ (mean: 59 ms) and /g/ (mean: 0ms), than did the monolinguals (/k/ mean: 79ms, /g/ mean: 17.5ms). At Time 2, both the bilingual and monolingual children used a long-lag when producing /k/ (mean: 72ms) and a short-lag for /g/ (mean: 16ms).

GLMM analyses revealed a significant interaction between group and time, $F(1{,}838) = 8.199$; $p < 0.1$, and main effect of group, $F(1{,}838) = 19.772$; $p < .001$, and stop, $F(1{,}838) = 960{,}659$; $p < .001$. Sequential Sidak post-hoc tests confirmed that the difference in VOT between monolinguals and bilinguals for

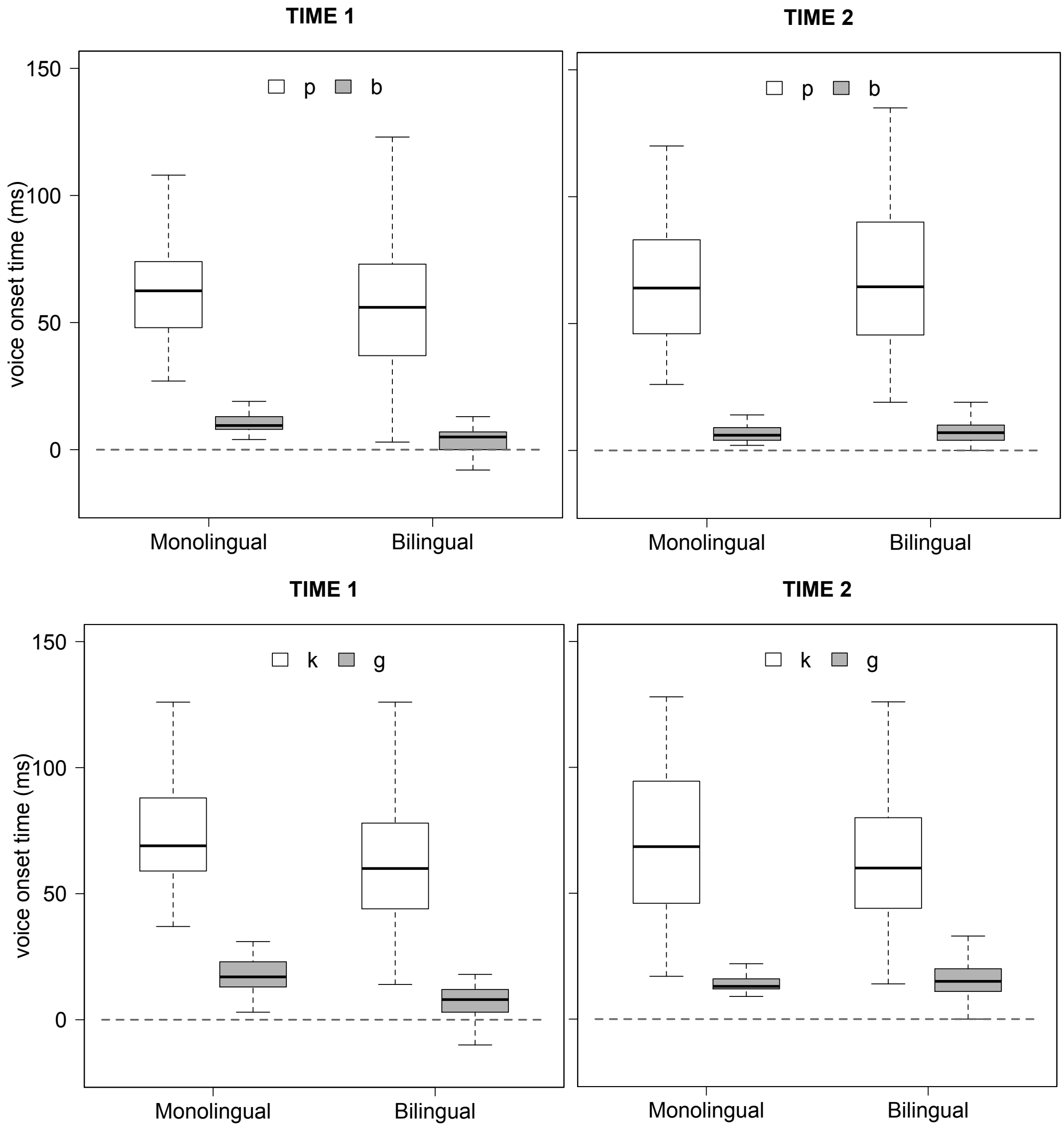


Figure 5.8. Boxplots representing the monolingual and bilingual childrens' production of /k/-/g/ and /p/-/b/ at Time 1 and at Time 2. Voice onset time measures for each consonant are given in milliseconds. The white boxes refer to the voiceless plosive /p/ or /k/ and the grey boxes refer to the voiced plosive /b/ or /g/.The dashed grey line represents 0ms. Any value below this represents voicing lead.

/k/ and /g/ was significant at Time 1 ($p < .05$) but not at Time 2 ($p > .05$). That is, at Time 1, the bilingual children used a significantly shorter VOT for /k/ and /g/ than did the monolinguals ($p < .05$). The bilingual children showed a significant increase in VOT for /k/ and /g/ between Time 1 and 2 ($p < .05$), whereas the monolinguals did not change ($p > .05$).

Summary

The production results reveal significant differences between the monolingual and bilingual VOT measures for bilabial and velar plosives at Time 1, but not at Time 2. For bilabials, this difference was only evident for voiced plosives. The bilingual children used a significantly shorter VOT for /b/ than did the monolinguals. Specifically, the bilinguals used a prevoicing and short-lag, whereas the monolinguals used a short-lag. The children did not differ in VOT values for /p/.

For velar plosives, the bilingual children used a significantly shorter VOT for /k/ and /g/ than did their monolingual peers. The bilinguals produced /g/ using prevoiced and short-lag VOT values, whereas the monolinguals only produced /g/ using a short-lag. For /k/ both groups produced VOT values that fall into the long-lag range, however the bilinguals produced /k/ using significantly shorter VOT values.

At Time 2, no difference was observed between groups. The bilingual children used significantly longer VOT values than they did at Time 1, i.e., /b/ and /g/ were produced using a short-lag, and were thus in line with their monolingual peers.

5.4. DISCUSSION

The aim of this was to track the acquisition of the English voicing contrast by Sylheti-English sequential bilingual children. To do so, the children's perception and production of English monophthongs was assessed and compared to that of their monolingual English speaking peers at two time points: when they were in nursery, and one year later during the first year of primary school.

At Time 1, differences were found between the monolingual English and Sylheti-English bilingual children. For perception, the monolingual children categorized the *coat-goat* voicing contrast more consistently than did the Sylheti-

English bilingual children. In turn, the bilinguals displayed shallower categorization slopes for *coat-goat* than did their monolingual peers. For production, children differed in their production of the voiced plosives /b/ and /g/: Bilingual children produced these with significantly shorter VOT values than did their monolingual peers. Although all children produced voiced plosives using a short-lag VOT, the bilingual children also produced some prevoiced variants. However, there were some similarities between the bilinguals and monolinguals at Time 1. Children did not differ in their categorization of the bilabial voicing contrast *pea-bee*; both monolingual and bilingual children had a similar slope steepness and phoneme boundary. Likewise, there were no differences in the phoneme boundary location for the /k/-/g/ continua. In production, there was no difference in VOT for the voiceless bilabial plosive, /p/; all children produced this with a long VOT.

How might this pattern of results be explained? Although it is possible that attentional factors may have affected children's performance on the tasks, this is an unlikely explanation for this pattern of results. Attention was tracked using interspersed continuum endpoints, and the data that were used in the final analyses were from children that met the attention criteria (more than 80% of the endpoints correct). It is also unlikely to be due to target word effects. All words were checked for familiarity with the children's teachers and caregivers. In addition, the children had to name and identify the target words correctly in order to take part in the study.

A more plausible explanation for the results at Time 1 is that they reflect children's sensitivity to the acoustic properties in their ambient language. Although it is difficult to fully understand the children's underlying perceptual representations without assessing their abilities in both Sylheti and English, the shallower categorization slopes, combined with shorter VOT values in production suggest that the children have less refined categories for English plosives, and may be using their existing (i.e. Sylheti-like) phonemic categories when producing and perceiving English plosives. This interpretation is supported by previous studies investigating monolingual language acquisition. Such studies have shown that early language experience is crucial in the development of speech perception skills, and that this experience shapes the acquisition and structure of underlying phonetic categories (see e.g. Native Language Magnet Theory, Kuhl, 2004). Children are thought to develop a "neural commitment" to their native language (L1) such that when acquiring a second language later in life, their native phonetic categories interfere with their production and perception of non-native target sounds (see e.g., Iverson, et al, 2003).

However, the children in this study did not differ in all aspects of their perception and production of English plosives at Time 1. Bilinguals and monolinguals chose similar phoneme boundaries in perception for both the *pea-bee* and *coat-goat* continua. Sylheti plosives have a shorter VOT than English plosives (McCarthy et al., in press), and so it was expected that at least at Time 1, Sylheti-English children would have had a shorter phoneme boundary for English plosives. Instead, the bilingual children selected boundaries that were similar to those of their monolingual English peers.

Despite the differences in language background, why were there no pervasive differences in VOT boundaries for the monolingual and bilingual children at Time 1? One possibility is that the structure of the synthetic continua affected the results. The stimuli only covered the English VOT range and the children were not presented with voicing lead stimuli. However, this explanation seems unlikely as, had they been using adult-like Sylheti categories, the bilingual children still could have behaved differently from their monolingual English peers, e.g., by placing the boundary at a significantly lower VOT value than the monolingual children, or at the extreme, categorizing all the stimuli as /p/.

Another possibility is that the bilingual children had not yet fully developed the voicing lead in their L1. The development of the voicing lead has been shown to take longer than the short-long lag distinction (Macken & Barton, 1980). Furthermore, stimuli within the voicing lead region of the VOT continuum are less accurately discriminated than stimuli in the short-lag range (Aslin & Pisoni, 1980). Lastly, it is possible that the bilingual children had started to acquire English bilabial plosives prior to testing. The children had been in an English-speaking nursery for 7 months, and it is possible that children were not tested early enough to see the transition from Sylheti- to English-dominant categorization. Furthermore, as there is an increase in slope steepness at Time 2 for all children, it is likely that at Time 1, both monolingual and bilingual children were still establishing and refining their phonemic categories (see e.g., Aslin & Pisoni, 1980).

This leads to the second aim of the study: to investigate how sequential bilingual children's perception and production of plosive consonants changed as a result of an increase in L2 experience. Bilinguals showed a significant increase in the steepness of their categorization slope for /k/-/g/ from Time 1 to Time 2, indicating a more consistent labeling of the continuum, and thus more refined phonemic categories. Indeed, at Time 2, they were no longer significantly different from their monolingual peers. Bilinguals also showed changes in their production of voiced plosives: they used voicing lead at Time 1 but by Time 2 they used a VOT that was the same as that of their monolingual English peers. This indicates that although initially the bilingual children may have been using their L1 (Sylheti) categories when perceiving and producing English plosives, they had retained a level of plasticity that meant that after relatively little experience with their L2 (English) they were able to establish L2 phonemic categories that matched those of native speakers with relative ease.

Additionally, there was evidence for a developmental trend common to both monolingual and bilingual children. For perception, all children displayed a shift in phoneme boundary for both /p/-/b/ and /k/-/g/, resulting in a more adult-like English phoneme boundary at Time 2. These findings are in line with previous studies that have shown that phonemic categories become more established with age (e.g. Hazan & Barrett, 2000; Mayo & Turk, 2004; Nittrouer, 2005). The findings thus support the idea that language specification continues to develop beyond the first year of life, and that this is facilitated by linguistic experience (e.g. Attunement Theory, Aslin & Pisoni, 1980).

However, though all children shifted their phoneme boundary location over time, English monolingual children had a similar categorization slope for /p/-/b/ at Time 1 to the bilingual children. This is surprising, because as for the /k/-/g/ continuum, we would have expected the monolinguals to have had a steeper categorization slope, particularly as the acquisition of bilabials precedes that of velars (see e.g. Fabiano-Smith & Goldstein, 2010). Why did we find this pattern of results for the perception of the /p/-/b/ continuum, but not the /k/-/g/ continuum? One possibility is that with the growing number of similar sounding lexical items children are required to have more refined phonemic categories to enable them to distinguish between the items in their growing lexicon (e.g. Nittrouer, 1996; Jusczyk, 1993). In the data, what we might be seeing is a pattern specific to the children, and which reflects the structure of their lexicon. That is, these children may have more words in their lexicon that contain word-initial /g/ or /k/, than word-initial

/p/ or /b/. In turn, this may have resulted in more refined phonemic categories for velar plosives.

Alternatively, it may be the case that the monolingual children are exposed to an input that contains a high amount of intra- and inter-speaker variability for bilabial plosives. For example, VOT values for voiced and voiceless plosives in adult- and child-directed speech have been shown to overlap considerably (e.g., Malsheen, 1980). Although, L2 phonetic training research suggest that variability in input supports perceptual learning (see e.g., Holt & Lotto, 2006), variability may play a different role for children, who are in the process of acquiring their phonetic system. It is possible that the ability to consistently identify continua that vary in VOT reflects a child's ability to deal with inter- and intra-speaker variability. Thus, the shallower bilabial categorization slope at Time 1 could reflect children's developing ability to cope with variable input in the ambient environment.

A more plausible explanation for the differences in categorization observed between *pea-bee* and *coat-goat* though, is the phonetic environment. If, as suggested by the Developmental Weighting Shift Hypothesis (DWS, Nittrouer, 1996), children are initially more sensitive to formant transitions, the difference in the following vowel between the two contrasts may have affected the acoustic salience of the formant transitions. *Pea* and *bee* contain the high front vowel /i/, which is characterized by a high F2. Consequently, the relatively high frequency of the onset of F2 and F3 for these consonants in the /i/ vowel environment, likely makes the formant transition less acoustically salient. This hypothesis is supported by studies of adult speech perception that have shown bilabial and velar plosives are often misclassified in the context of high front vowels such as /i/ (e.g., Blumstein & Stevens, 1979). The improvement in categorization observed at Time 2 could be due, in part, to the children gaining more linguistic experience, and in turn attending to acoustic properties that do not involve spectral change, such as VOT (see e.g., Nittrouer, 2005).

In sum, the findings reported here suggest overall group differences between the bilingual and monolingual children at Time 1, but by Time 2 the bilingual children have caught up with their monolingual peers. At a broad level, the experiments reported in Chapter 4 and 5 have enabled a detailed analysis of group differences between the bilinguals and monolinguals. As with all child development data, variability was observed in both the monolingual and bilingual data. Yet, group analysis provides little explanation for the individual differences. Specifically, at Time 1, some of the bilingual children had vowel identification scores and phoneme categorization slopes similar to their monolingual peers, where as other bilingual children had much lower scores and shallower slope. Likewise, for, although an overall group difference showed a significant difference between the bilinguals and monolinguals for voiced plosives at Time 1, observation of the individual data shows that not all of the bilingual children used prevoicing. Thus, an alternative analysis could be to run separate analyses on prevoiced vs. short-lag tokens to compare the proportion of token type (i.e., prevoiced or short-lag) to the monolinguals' production. Such analyses may reveal relatively fewer differences between the groups i.e., bilinguals may produce a higher proportion of short-lag voiced plosives, closer to that of their monolingual peers.

Based on the previous research on bilingual and second-language acquisition, such variability in performance could be accounted for by differences in language exposure as well as variation in the speech input that the children received before starting full-time education, i.e., main caregivers, and after entering nursery, i.e., the child's social networks. The chapters to follow aim to explore the influence of the amount of English language exposure (Chapter 6), as well as explore the possible influence of the acoustic-phonetic features of caregiver speech (Chapter 7) on the children's phoneme acquisition outcomes.

6. The influence of language exposure and social networks on speech perception and production

6.1. INTRODUCTION

The aim of this chapter is to investigate the influence of the children's social networks on their English speech perception and production measures. Specifically, the investigation concerned the relationship between the children's language exposure before and after entering a full-time English speaking nursery, and their English speech perception and production measures. Previous research on simultaneous bilinguals has shown that the specific amount of language exposure influences children's speech and language outcomes in each language. In general, results indicate that children display better outcomes for vocabulary, grammar, phonology, and more recently for speech perception in their dominant language (see e.g., De Houwer, 2009; Garcia-Sierra et al., 2011; Hoff 2012; Pearson et al, 1997; Place & Hoff, 2011). As shown in Chapters 4 and 5, the sequential bilingual children in this research, are initially mainly exposed to Sylheti, and are then gradually immersed in English, normally through older siblings and when they enter English-speaking education. Although the children were predominately exposed to Sylheti by their main caregivers before entering education (at least 80% Sylheti), as will be shown in this chapter, the children varied in the amount of exposure they had to Sylheti and to English before starting full-time education

These differing amounts of exposure to each language are greatly determined by the background of the children's social network ties i.e., people with whom they interact regularly. For all of the children in this study however, English became more dominant once they started full-time education (estimated attendance of about 30 hours per week). Children varied in terms of whom they interacted with regularly (e.g., monolingual English speaking neighbours vs. first-generation Sylheti speaking grandparents). This resulted in a variation in the amount of English that the children were exposed to during their regular interactions. Such differences in the exact amount of exposure pose interesting questions regarding the children's acquisition of the English phonemic contrasts. That is, although significant group differences were found between the monolingual English and Sylheti-English sequential bilinguals at Time 1, was this influenced by the individual differences in initial

English language exposure? Further, do the differences in English exposure once the children start full-time education influence their outcomes at Time 1 and 2?

In addition to the amount of exposure to each language, previous research on monolingual and simultaneous bilingual language development has shown that other factors in children's language environment, such as having older siblings, can play a key role in language acquisition (see Hoff, 2006 for review). In an immigrant community setting this is particularly relevant given that the background of the children's interlocutors is likely to influence the type of input the child is exposed to. For example, as shown in Chapter 2, any Late arrival Bangladeshi network ties (e.g., grandparents) are likely to use accented variants when producing English vowels and plosives. In contrast, a child's second-generation older sibling is more likely to speak the host country language and may provide a more native-like English speech model for the child thus being the primary source of English input for the younger children (see e.g., Hoff, 2006; Gregory, 1998).

To investigate the influence of these factors on the sequential bilingual children's speech perception and production outcomes, a social network approach was adopted (see Wei, 1992; Sharma, 2011). Data regarding the children's regular interlocutors, that is, their network ties, was collected during interviews with the children's main caregivers. Specifically, for each network tie the following information was collected: the language/s used, the amount of time spent with the child in hours, the context of the interaction and the speaker's background e.g., age of arrival in the UK. To investigate the influence of these factors on the children's speech perception and production measures reported in Chapter 4 and 5, two analyses were conducted. For language exposure, children were grouped according the amount of English exposure for both before and after entering full-time education. These groups were subsequently compared on their speech perception and production measures. To investigate the relationship between the children's network characteristics (i.e., background of network members) and their speech perception and production outcomes correlation analyses were conducted.

6.2. MATERIALS AND METHOD

6.2.1. Participants

38 caregivers (6 fathers, 32 mothers, aged 26 – 36 years-old) of the children from Chapter 4 and 5 participated in the study. Eighty five percent of the caregivers were born in Sylhet, Bangladesh, of which 34% arrived in the UK before the age of 16 (mean 7.5 years, range 5 – 12 years old). The remaining 66% arrived in the UK after 16-years-old (mean 22 years, range 19-27). One additional caregiver was recorded but not included due to background noise in the recording.

6.2.2. Social Network Measures

Measures were taken from two time points as reported by the caregivers: 1) A retrospective measure for the time before the child started nursery (pre-nursery score), and, 2) a current measure, once the child had started nursery (post nursery score). For each time point, data was obtained using the Social Network section of the language environment interview (see Appendix 2a, part 2). Caregivers were asked to list all of the people with whom their child has regular interactions (i.e., at least every 2 weeks); the amount of time each of the interlocutors spend with the child in hours; the context of each interaction e.g., Quran class; and each interlocutors' language background. The latter included age of arrival, first language, place of birth and ethnicity. Based on this data the following two separate measures were calculated: ***i.*** language exposure index, and, ***ii.*** social network ties:

i. Language exposure index

Language exposure was indexed by the amount of English exposure. Two separate measures were taken: ***pre*** and ***post***:

1) <u>Pre-nursery language exposure measure (***pre***)</u>

The pre-score covers the period before the children started nursery i.e., from 0 months to ~42 months old. The pre measurement refers to the percentage of English exposure from the child's main caregivers[2]. This percentage was based on two factors: 1) proportion of the child's waking hours that the main caregiver spent with the child and, 2) the language/s spoken to the child. For each main caregiver, Sylheti

[2] Here, a main caregiver is defined as spending no less that 70% of the child's waking hours with the child.

only, English only or Both (including sub-options: *equal amounts of both language* i.e., 50%, *more English*, i.e., 25% or *more Sylheti* i.e., 75%). See Table 6.1 for an example of this calculation. The final measures were normally distributed ranging from 0% to 20%, with a group mean of 8% (SD: 6%).

Table 6.1. Example of pre exposure calculation.

Main caregivers	**% time spent with child***	**English used with child**	**Total English exposure**
Mother	50%	25% (i.e., *Both, more Sylheti*)	**12.5%**
Grandmother	50%	0% (i.e., *only Sylheti*)	

* based on the % of the child's waking hours.

2) Post language exposure measure: after starting full-time education (***post***):
The post-measure covers the time after entering full-time education. The total amount of English, Sylheti and Both (Sylheti and English) language exposure for each child was calculated based on the proportion of the child's total weekly waking hours spent with each network member and the language/s they used with the child. The ties were broadly grouped into three contexts: *school ties* e.g., teacher, *home ties*, e.g., sibling, and *other* e.g., neighbour. From this, the percentage of the child's weekly hours of exposure to each language was calculated for each context. The individual context measures were combined to calculate separate overall language exposure percentages for English, Sylheti and Both. An example of this calculation can be found in Table 6.2 below. The 'combined English percentage' was used in all subsequent analyses. The observed percentages were normally distributed ranging from 27% to 77% English exposure, with a group mean of 51% (*SD*: 15%)

ii. *Social network ties*

Each interlocutor in the child's social network (i.e., each tie) was broadly categorized as belonging to one of five groups based on their language background: 1) Late arrivals (Late), i.e., Sylheti-speaking first-generation who arrived in the UK after 16 years old, 2) Early arrivals (Early), i.e., Sylheti-speaking first-generation who arrived before 16 years old, 3) Second-generation, i.e., of Bangladeshi-heritage, born in the UK, 4) UK born monolingual-English speakers, 5) Other e.g., Somali-

Table 6.2. Example of a post exposure calculation.

Post language exposure				
	Context			
Language	*School ties*	*Home ties*	*Other*	***Combined percentage***
English	36	0	5	**41***
Sylheti	0	26	5	31
Both	0	26	2	28
% hours	36	52	12	100

*The final post exposure measure was based on the combined English measure

heritage. For each child the percentage of ties for each category was calculated. For example, 10% of a child's social network tie might be made up of second-generation speakers. In addition, information on the number of older siblings, their average age and the language used with the child was collected. Information on the numbers of extended family living in the home was also collected, however group statistical analysis was not possible as only 6/40 children had extended family living in the home. Table 6.2 summarizes the Social Network data for all of the children.

6.2.3. Child outcome measures

The outcome measures reported in Chapters 4-6 are reported again here. Namely, speech perception measures: total vowel identification score (i.e., score for all vowel contrast groups combined) and voicing categorisation (i.e., VOT categorization slope and phoneme boundary), and acoustic speech production measures.

For plosive production, voice onset time (VOT) was measured. For vowels, Euclidean distances between the Lobanov z-score values were calculated. As the aim of this study was to investigate how the children's vowel productions related to their language exposure the decision was made to calculate the Euclidean distance rather than the English vowel space measures from Chapter 4. The Euclidean distance values were used to measure the separation between the children's vowel productions. That is, the larger the Euclidean distance, the greater the size of the contrasts that were produced (i.e., less accented). The following equation was used to calculate the Euclidean distance:

$$\text{Euclidean Distance} = \sqrt{}\,(F1vi - F1vii)^2 + (F2vi - F2vii)^2$$

Where vi and vii refer to vowel 1 (e.g., ship) and vowel 2 (e.g., sheep) respectively. The mean Euclidean distance values obtained for the four tokens were calculated.

Children were tested twice: after six months in English speaking full-time education and approximately one year later. Details of these measures, testing procedures and scores can be found in Chapters 4 and 5.

6.2.4. Procedure

The child's main caregivers were interviewed in their home (*n* = 35) or a quiet room in their child's school (*n* = 4). All interviews were recorded using a H2 Zoom audio recorder, using a bi-directional microphone, with a sampling rate of 44.1 kHz, 16-bit resolution. The interviews lasted an average of 40 minutes each. All interviews were conducted 3 months after the first testing tranche (Time 1), when the children were in nursery. Thus, the caregivers were required to give retrospective estimates for the questions regarding pre-measures (i.e., *before your child entered nursery*), and current estimates for questions regarding the post-measures.

For the families where the main caregivers had little or no English (*n* = 4) either a Sylheti-speaking teaching assistant from the local community or a UK born English-speaking family member acted as an interpreter for the interview. The interpreters were briefed before the interview. The briefings consisted of explaining the purpose of the interview as well as each section and question the interview questionnaire. The interviewer transcribed the answers directly and subsequently crosschecked all of the data using the recordings.

6.3. RESULTS

The results will be presented in two parts: 1) The relationship between the children's network factors and English speech perception measures and, 2) The relationship between the children's network factors and English speech production. A summary of the children's network measures can be found in Table 6.3.

Two separate analyses were conducted. First, to investigate the relationship between language exposure and the children's outcome measures (at Time 1 and Time 2), the children were divided into high and low English exposure groups based on a median split of the language exposure scores (i.e., pre and post measures separately). The groups were subsequently compared using Generalized Linear

Table 6.3. Summary of the children's language exposure measures, siblings, and social network ties. The shaded cells are measurements before the children entered full-time education. All other measures refer to once the children had started education (i.e., after ~3.5 years old).

	Language exposure (English)				**Siblings**			**Network ties***			
	Pre low	*Pre high*	*Post low*	*Post high*	*No.*	*Age (years)*	*Language*	*BL*	*BE*	*UK B*	*UK*
Mean	.2	.14	.40	.65	1.5	9	.42	.26	.10	.52	.07
Median	0	.15	.40	.66	1	9	*n/a*	.24	.10	.52	.06
SD	.2	.4	.05	.07	1.17	2.9	*n/a*	.09	.08	.12	.05
Range	.0 - .5	.8 - 20	.27 - .50	.52 - .77	0-5	7-11	*n/a*	.13 - .47	.0 - .22	.42 - .76	.0 - .17

* **BL**: Bangladeshi Late-speakers, **BE**: Bangladeshi Early-speakers, **UK B**: UK-born second-generation speakers, **UK**: UK born monolingual English speakers

Mixed Model (GLMM) analyses, with exposure group, vowel/plosive, time, and condition (for vowel ID only i.e., quiet or noise) treated as fixed factors.

To account for individual differences, participant was treated as a random factor in all analysis. Significant effects were explored with post hoc sequential Sidak tests. All tests were corrected for multiple comparisons. Separate analyses were conducted for pre-nursery and after starting nursery English exposure scores, hereafter referred to as ***pre*** and ***post*** English exposure score, respectively.

Second, to investigate the relationship between the children's outcomes and their social network characteristics e.g., older siblings, correlation coefficients were calculated between the children's speech perception and production measures and the social network factors.

6.3.1. Is there a relationship between the children's language exposure and their speech perception measures?

i. Vowel identification (VID)

Figure 6.1 and Table 6.4 show the children's vowel identification scores at Time 1 and Time 2, grouped according English exposure (i.e., high vs. low) for the pre and post measure.

Table 6.4. Bilingual children's total VID scores, in quiet (Q) and noise (N). Children are grouped according to the amount of English exposure (i.e., low or high), for the pre (i.e., before nursery) and post (after entry to nursery) measures.

		Pre low	***Pre high***	*Post low*	*Post high*
Time 1	*Mean Q*	.80	.83	.81	.82
	Median Q	.78	.81	.75	.79
	SD Q	.09	.09	.11	.07
	Mean N	.70	.77	.68	.76
	Median N	.72	.75	.70	.78
	SD N	.10	.80	.11	.07
Time 2	*Mean Q*	.91	.91	.92	.90
	Median Q	.93	.93	.94	.93
	SD Q	.07	.05	.08	.06
	Mean N	.86	.86	.87	.85
	Median N	.85	.87	.89	.85
	SD N	.06	.07	.07	.06

The GLMM analysis revealed a main effect of time and condition for pre and post measures [time, pre: $F(1,85) = 84.65, p < .001$; post: $F(1,82) = 100.44, p < .001$; condition, pre: $F(1,82) = 33.74, p < .001$; post: $F(1,78) = 37.184, p < .001$], confirming that the children displayed an increase in scores from Time 1 to Time 2 and had significantly lower scores in the noise condition. The main effect of group for pre and post measures was not significant ($p > .05$).

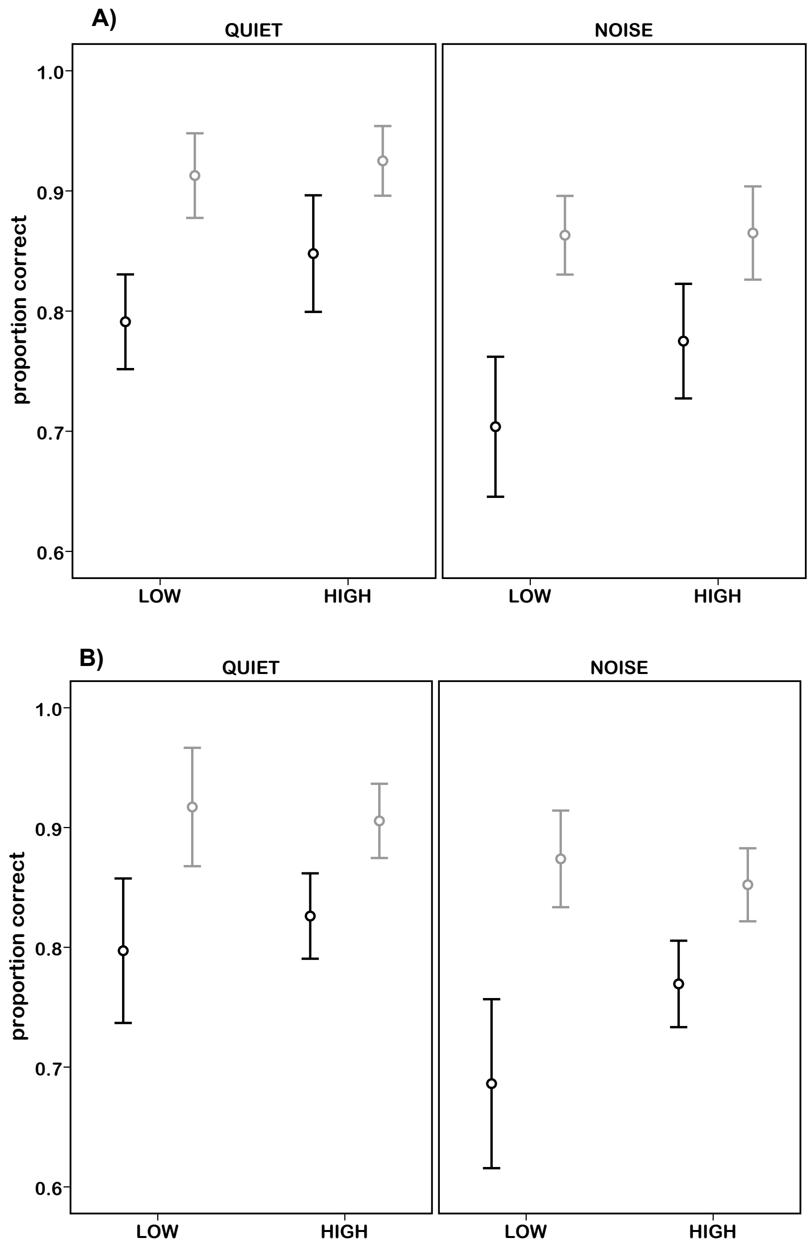


Figure 6.1. 95% confidence intervals for the children's vowel identification scores in quiet and noise condition, at Time 1 (black bars) and Time 2 (grey bars). Children are grouped according to the pre (A) and post (B) language exposure measures: low English exposure and high English exposure.

Overall, the high exposure and low exposure group were not significantly different. There was a significant interaction between group and time for pre and post scores [pre: $F(1,85) = 4.075$, $p = .04$; post: $F(1,82) = 9.43$, $p = .003$]. Sequential Sidak tests revealed that low English exposure group (pre and post) had significantly lower VID scores than did the high exposure group (pre and post) at Time 1 ($p < .05$), but not at Time 2 ($p > .05$). Specifically, at Time 1 the low exposure group had significantly lower VID scores than did the high exposure in the noise condition ($p < .05$), but not in quiet ($p < .05$).

ii. VOT categorization

Figure 6.2 and Table 6.5 show the children's categorization slope values at Time 1 and Time 2, grouped according English exposure (i.e., high vs. low) for the pre and post measure. The GLMM analyses revealed a significant main effect of time for /p/-/b/ [*pre*: $F(1,20) = 38.458$, $p < .001$; *post*: $F(1,30) = 33.450$, $p < .001$] and /k/-/g/ [*pre*: $F(1,46) = 11.295$, $p = .002$; *post*: $F(1,85) = 4.075$, $p < .001$], where all children showed an increase in slope steepness from Time 1 to Time 2. There was a significant main

Table 6.5. Bilingual children's categorization slope (log) for /b/-/p/ and /g/-/k/ at Time 1 and Time 2. Children are grouped to the amount of English exposure (i.e., low or high), for the pre (i.e., before nursery) and post (after entry to nursery) measures.

		Pre low	*Pre high*	*Post low*	*Post high*
Time 1	*Mean* /b/-/p/	-1.2	-1.02	-1.13	-1.07
	Median /b/-/p/	-1.2	-1.09	-1.22	-1.15
	SD /b/-/p/	.16	.16	.19	.15
	Mean /g/-/k/	-1.4	-1.1	-1.34	-1.18
	Median /g/-/k/	-1.2	-1.0	-1.22	-1.15
	SD /g//k/	.40	.30	.44	.27
Time 2	*Mean* /b/-/p/	-.74	-.49	-.70	-.54
	Median /b/-/p/	-.92	-.45	-.91	-.56
	SD /b/-/p/	.35	.30	.42	.24
	Mean /g/-/k/	-1.16	-.82	-1.18	-.84
	Median /g/-/k/	-1.15	-.81	-1.02	-.76
	SD /g/-/k/	.61	.22	.56	.37

effect of group for the pre language exposure measures [/p/-/b/: $F(1,20) = 6.358$, $p < .001$; /k/-/g/: $F(1,46) = 4.505$, $p < .001$], confirming that overall the low exposure group had significantly shallower slopes than did the high exposure group. The main effect of group for the post language exposure measure was not significant for /p/-/b/ ($p > .05$) and /k/-/g/ ($p = .096$). The interaction between group and time was not significant for /p/-/b/ and /k/-/g/ for the pre ($p > .05$) and post ($p > .05$) language exposure measures.

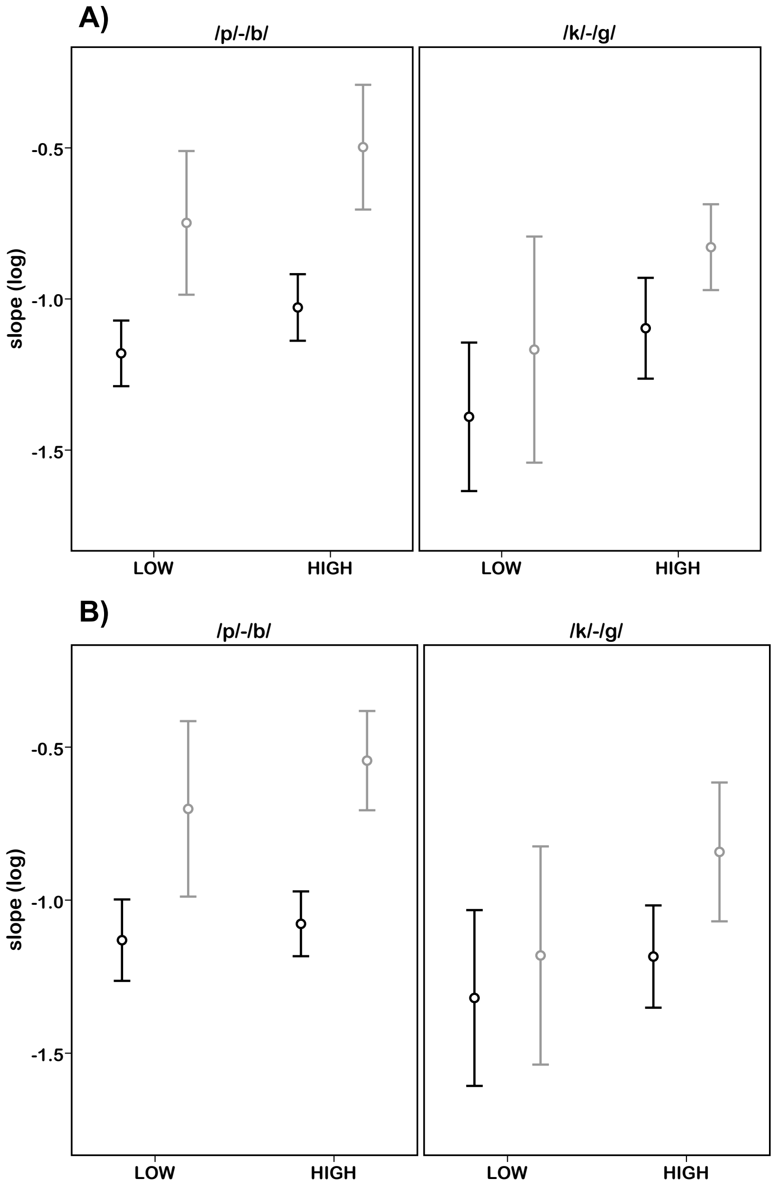


Figure 6.2. 95% Confidence intervals for the children's categorization slope values for the /p/-/b/ and /k/-/g/ voicing contrast. Measures are given in log for Time 1 (black bars) and Time 2 (grey bars). Children are grouped according to the pre (A) and post (B) language exposure measures: low English exposure and high English exposure.

For phoneme boundary, Table 6.6 suggests no difference between the high and low English exposure groups for the pre and post measures for /p/-/b/ and /k/-/g/, at Time 1 or Time 2. The GLMM revealed no significant main effect of group, time, or interaction between group and time for /p/-/b/ ($p > .05$) and /k/-/g/ ($p > .05$) pre and post measures confirming no difference between the high and low English exposure groups, and no difference between Time 1 and 2.

Table 6.6. Mean phoneme boundary values in milliseconds for the children's categorization of /p/-/b/ and /k/-/g/ at Time 1 (Time 1) and Time 2 (Time 2). Children are grouped according to their pre and post (shaded) language exposure measures: low English exposure and high English exposure.

	VOT phoneme boundary (ms)			
	language exposure group			
contrast	*Pre low*	*Pre high*	*Post low*	*Post high*
/p/-/b/ Time 1	29	30	29	30
/p/-/b/ Time 2	29	25	30	24
/k/-/g/ Time 1	47	45	46	47
/k/-/g/ Time 2	48	48	48	49

6.3.2. Is there a relationship between the children's social network factors and their speech perception measures?

The relationship between the children's speech perception scores and their social network factors was investigated using Pearson product-moment correlation analysis for all factors apart from *sibling language* (categorical: English only = 1, Both = 2, Sylheti only = 3; note: no siblings used *Sylheti only*), where Spearman rank order correlation analysis was conducted. Table 6.7 displays the correlation coefficients obtained for the social network factors and the children's speech perception measures at Time 1 and 2.

i. Vowel identification (VID)

The results show that the language exposure from the older siblings is the only factor that is related to children's vowel identification scores. Specifically, at Time 1 the children who had siblings who spoke more English in the home had significantly higher VID scores (quiet: $r = -.349$, $p < .05$; noise: $r = -.628$, $p < .001$), this was

Table 6.7. Correlation coefficients (*r*) relating the children's speech perception measures at Time 1 and 2 and social network factors.

	Sibling measures			***Network ties ❖***			
Speech perception measures	*No. older siblings*	*Age of siblings (mean)*	*Sibling language*	*BL*	*BE*	*UKB*	*UK*
VID-Q Time 1	.165	-.012	**-.349***	-.047	-.085	.182	-.214
VID-N Time 1	-.054	-.018	**-.628****	.176	-.074	.073	-.025
VID-Q Time 2	-.027	-.110	.256	.252	-.119	.211	-.117
VID-N Time 2	.151	.142	.300	.020	-.056	.046	-.093
/p/-/b/ slope Time 1	-.258	-.131	-.167	-.046	-.100	.190	.010
/p/-/b/ slope Time 2	-.430	.256	**-.772****	-.110	-.88	-.416	-.044
/k/-/g/ slope Time 1	-.070	-.117	-.117	-.70	.062	.060	-.148
/k/-/g/ slope Time 2	**.672****	-.73	-.073	-.232	-.173	**.504***	-.281
/p/-/b/ boundary Time 1	-.038	.322	.322	-.027	-.060	.135	.016
/p/-/b/ boundary Time 2	.279	.392	.392	**-.491***	.126	.30	.171
/k/-/g/ boundary Time 1	.011	-.180	-.232	-.049	.002	.000	-.174
/k/-/g/ boundary Time 2	.004	-.058	-.192	.118	-.299	.398	-.328

*Note: shaded cells refer to significant correlation coefficients: *p < .05, ** p < .001*

❖**BL**: Bangladeshi Late-speakers, **BE**: Bangladeshi Early-speakers, **UK B**: UK-born second-generation speakers, **UK**: UK born monolingual English speakers highly significant in the noise condition. For all other social network factors, there was no significant correlation with the children's English vowel identification scores at Time 1 and 2.

ii. VOT categorization

The results show that the sibling measures are significantly correlated with some of the VOT categorization measures. Specifically, the language exposure from the older siblings is significantly related to their /p/-/b/ categorization slope values at Time 2 ($r = -.772$, $p < .001$). That is, the children who had older siblings who only spoke English in the home had steeper categorization slopes at Time 2. For /k/-/g/, the children's categorization slope values were significantly related to the number of older siblings in the home at Time 2 ($r = .672$, $p < .001$). The children who had more older siblings in the home had significantly steeper categorization slopes at Time 2. Likewise, the children's /k/-/g/ slope values at Time 2 were significantly related to the proportion of UK born Bangladeshi-origin ties to their network ($r = .504$, $p < .05$). The children who had more UK born ties had higher categorization slope values.

The only social network factor that was related to the children's phoneme boundaries was the proportion of Late arrival first-generation ties in their social network, for /p/-/b/ ($r = -.491$, $p < .05$). Specifically, the children who had more Late arrival ties in their social network had a significantly lower phoneme boundary, shorter than that of the Standard Southern British English adult target (see Chapter 5 adult control data for detail), and therefore closer to that of Sylheti.

Summary

The results show that the children who were exposed to more English before entering nursery (i.e., pre measure) had significantly higher vowel identification scores and steeper categorisation slopes for /p/-/b/ and /k/-/g/ than did the low English exposure group at Time 1. Further, for vowel identification, the children who had more English exposure once starting school (i.e., post measure) had higher scores at Time 1. For phoneme boundary, no difference was found between the groups. At Time 2, the results showed no difference between the high and low exposure groups for all perception measures.

The correlation analyses between the children's social network factors and speech perception outcomes revealed a significant relationship between the amount of

English exposure from older siblings and the children's vowel identification scores at Time 1, and /p/-/b/ categorization slope at Time 2. That is, children who had English speaking older siblings had higher VID scores and a steeper /p/-/b/ slope. Likewise, the number of older siblings and proportion of UK born Bangladeshi-origin network ties was significantly correlated with the children's /k/-/g/ categorization slope. Children who had more older siblings and UK born Bangladeshi network ties had steeper /k/-/g/ categorization slopes at Time 2. For phoneme boundary, the proportion of Late arrival first-generation ties was significantly correlated with the children's /p/-/b/ boundary at Time 2. That is, children who had more Late arrival ties had significantly shorter phoneme boundary.

6.3.3. Is there a relationship between the children's language exposure and their English speech production measures?

i. Vowel production

As in Chapter 4, the two /uː/-/ʊ/ phonetic environments: /b/VC (i.e., /uː/-/ʊ/ **I**) and the /p/V/l/ environment (i.e., /uː/-/ʊ/ **II**) were analyzed and presented separately (see 4.3.2. for details).

Table 6.8 shows the pre and post exposure group mean Euclidean distance values for the children's production of the /iː/-/ɪ/, /ɑː/-/ʌ/ and /uː/-/ʊ/ contrasts, at Time 1 and 2. The data indicate no difference between the low and high exposure groups for the pre and post language exposure measures. The GLMM analyses revealed no significant main effect of group for all contrasts for the pre language exposure measurements ($p > .05$) and the post language exposure measurements ($p > .05$) confirming that there was no significant difference between the low and high language exposure groups.

There was a significant main effect of time for /iː/-/ɪ/ [pre: $F(1,36) = 32.806$, $p < .001$; post: $F(1,71) = 34.036$, $p > .05$] and /ɑː/-/ʌ/ [pre: $F(1,36) = 10.372$, $p =.003$; post: $F(1,36) = 6.954$, $p = .01$], where the children showed an increase in Euclidean distance values from Time 1 to 2. For /uː/-/ʊ/ I and II, the main effect of time was not significant ($p > .05$). There were no significant interactions between group and time for all contrasts for the pre and post language exposure measurements ($p > .05$)

Table 6.8. Mean Euclidean distance values in Hertz (z-score converted) for the children's production of the target vowel contrasts at Time 1 (Time 1) and Time 2 (Time 2). Children are grouped according to the pre and post language exposure measure: low English exposure and high English exposure.

	Euclidean distance (z-score converted Hz)			
	Language exposure group			
Contrast	Pre low	Pre high	Post low	Post high
/iː/-/ɪ/ Time 1	.76	.82	.77	.80
/iː/-/ɪ/ Time 2	1.02	1.02	.99	1.06
/ɑː/-/ʌ/ Time 1	.77	.77	.76	.78
/ɑː/-/ʌ/ Time 2	.94	.96	.87	1.03
/uː/-/ʊ/ I Time 1	1.37	1.32	1.32	1.37
/uː/-/ʊ/ I Time 2	1.43	1.37	1.40	1.39
/uː/-/ʊ/ II Time 1	.55	.63	.57	.60
/uː/-/ʊ/ II Time 2	.67	.69	.64	.71

ii. Plosive production (VOT)

Figure 6.3 show the VOT values for production of bilabial and velar plosives at Time 1 and 2, grouped according to the pre and post language exposure groups (i.e., high and low English exposure).

GLMM analysis revealed a significant main effect of group for /g/ pre measures [$F(1,66) = 5.941$, $p =.025$] and /b/ post measures [$F(1,66) = 5.941$, $p = 0.017$]. Sequential Sidak tests confirmed that the low exposure groups used a significantly shorter VOT than did the high exposure group ($p < .05$). The main effect of group was not significant for all other analyses ($p > .05$). The main effect of time for /b/, /p/ and /g/ was significant [*pre*: /b/, $F(1,65) = 4.781$, $p < .05$; /p/, $F(1,54) = 4.550$, $p < .05$; /g/, $F(1,44) = 20.495$, $p = .002$; *post*: /b/, $F(1,66) = 5.941$, $p = .05$ /p/, $F(1,63) = 0.284$, $p = .002$; /g/, $F(1,34) = 12.386$, $p = .001$], confirming that the children displayed an increase in VOT from Time 1 to 2. There was no significant main effect of time for /k/ ($p > .05$). The interaction between group and time was not significant for all analyses ($p > .05$).

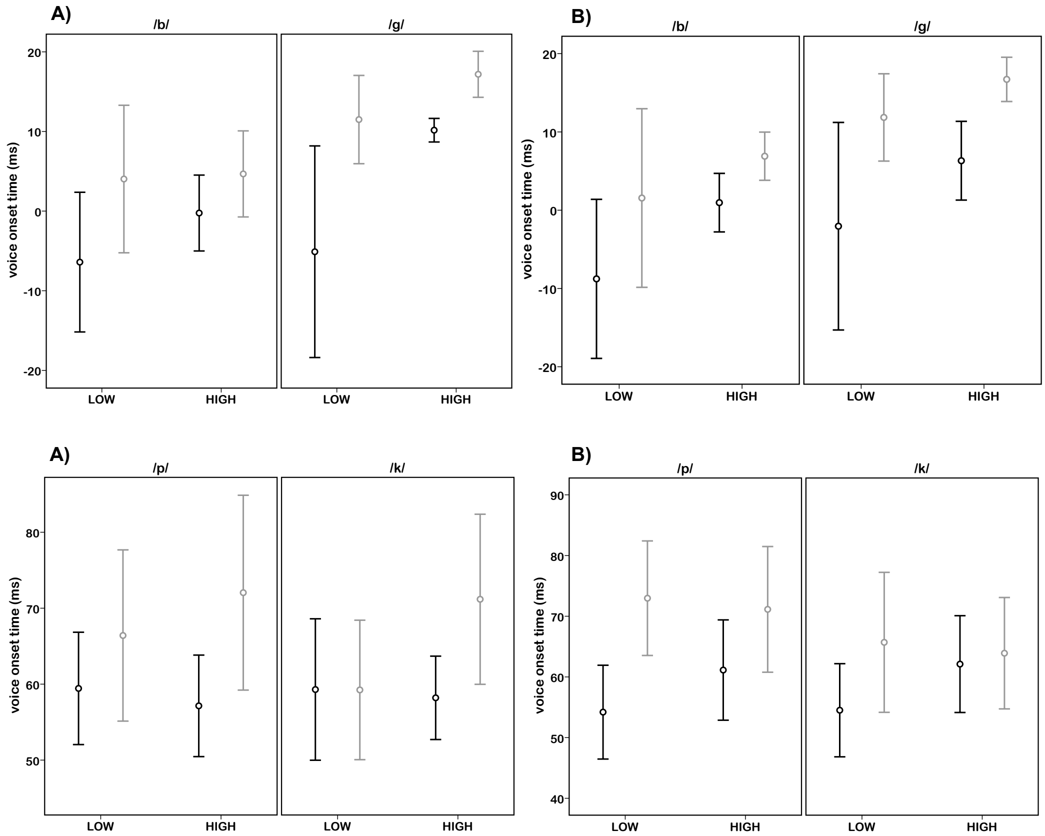


Figure 6.3. 95% confidence intervals for the children's production of English /b/, /g/, /p/ and /k/ at Time 1 (black bars) and Time 2 (grey bars). Voice onset time measurements for each consonant are given in milliseconds (ms). Children are grouped according to the pre (A) and post (B) language exposure measure: low English exposure and high English exposure.

6.3.4. Is there a relationship between the children's social network factors and their speech production measures?

As was the case for perception, the relationship between the children's speech production measures and their social network factors was investigated using Pearson product-moment correlation and Spearman rank order correlation (for language of siblings). Table 6.9 displays the correlation coefficients obtained for the social network factors and the children's speech perception measures at Time 1 and 2.

i. Vowel production

The results show that the proportion of Late arrival Bangladeshi ties in the children's social network was significantly correlated with the children's vowel production measures. Specifically, the children who had fewer Late arrival ties produced the /uː/-/ʊ/ contrast (e.g., *boot, book*) with a larger Euclidean distance than

Table 6.9. Correlation coefficients (*r*) relating the children's speech production measures at Time 1 and 2 and social network factors.

		Sibling measures			***Network ties ❖***			
	Speech production measures	*No. older siblings*	*Age of siblings (mean)*	*Sibling language*	*BL*	*BE*	*UKB*	*UK*
Euclidean distance (Hz)	/iː/-/ɪ/ Time 1	.044	-.031	.064	.075	.030	-.014	-.329
	/iː/-/ɪ/ Time 2	-.211	-.308	-.114	-.015	.039	-.156	-.263
	/ɑː/-/ʌ/ Time 1	-.148	-.006	-.122	.331	-.160	.090	-.207
	/ɑː/-/ʌ/ Time 2	1.211	-.307	-.090	.245	-.05	-.263	.327
	/uː/-/ʊ/ I Time 1	.178	.04	.005	**-.354***	.272	-.153	-.257
	/uː/-/ʊ/ I Time 2	-.210	-.307	-.112	**-.342***	-.015	-.262	.326
	/uː/-/ʊ/ II Time 1	-.079	.04	.045	.299	-.216	-.104	.257
	/uː/-/ʊ/ II Time 2	-.211	-.305	-.55	-.194	-.061	.251	-.127
Voice onset time (ms)	/p/ Time 1	-.54	-.54	-.043	-.275	.150	.207	.004
	/p/ Time 2	-.003	.003	.099	-.009	.019	-.266	.254
	/b/ Time 1	-.173	-.173	-.125	-.275	.150	.245	-.284
	/b/ Time 2	.213	-.213	.482	-.004	.019	.041	-.080
	/k/ Time 1	.079	.79	**-.384***	-.113	.125	.204	-.098
	/k/ Time 2	-.341	-.341	-.080	.119	.128	-.155	.006
	/g/ Time 1	.209	-.209	-.068	.032	-.159	.036	.236
	/g/ Time 2	-.53	-.53	-.265	-.155	-.201	.120	-.159

N

Note: *shaded cells refer to significant correlation coefficients: *p < .05, ** p < .001* ❖**BL**: Bangladeshi Late-speakers, **BE**: Bangladeshi Early-speakers, **UK B**: UK-born second-generation speakers, **UK**: UK born monolingual English speakers did the children who had relatively more Late arrival ties. The correlation was significant at Time 1 ($r = -.354, p < .05$) and Time 2 ($r = .-342, p < .05$). There were no other significant correlations between the social network factors and vowel production.

ii. Plosive production (VOT)

The results show that the sibling language was significantly correlated with the children's /k/ production at Time 1 ($r = -.384, p < .05$). Specifically, the children who had English speaking older siblings used longer VOT values for /k/, i.e., similar to the monolingual English controls in Chapter 3. There were no other significant correlations between the social network factors and the children's plosive production.

Summary

The results show no influence of language exposure on the children's Euclidean distance values for the production of the target vowel contrasts at Time 1 and Time 2. For plosive production, there was a significant correlation between the amount of English exposure before entering full-time education (pre exposure) and their VOT measures for /b/, /p/ and /g/. Children who had the lowest English exposure before entering full-time education, produced the plosives with a significantly shorter VOT, including prevoiced tokens for /b/ and /g/, than did the high exposure children. This difference was no longer significant at Time 2. For the post exposure measure, there was no difference between the high and low group at Time 1 and 2.

The correlation analyses between the children's social network factors and speech production revealed that the only factor related to vowel production was the proportion of Late arrival Bangladeshi-origin ties. Children who had fewer Late arrival ties produced *boot* and *book* with a larger Euclidean distance, demonstrating GOOSE and FOOT fronting. For plosive production, the only factor related to the children's VOT measurements was sibling language for /k/ production at Time 2. The children with English speaking siblings used a longer VOT time, than did the children with siblings who spoke both Sylheti and English to the child.

6.4. DISCUSSION

The aim of the experiments reported in this chapter were to investigate the influence of the children's language exposure and social networks on their English speech perception and production measures. To do so, the children's main caregivers were interviewed regarding the children's social network ties, including information on the language/s used with child, the amount of time spent with the child, the context of the interactions, and the background of the speakers. From this data, the children's language exposure, in hours per week, was calculated for the period before entering full-time education (pre measure) and the once the child started full-time education (post-measure). The children were divided into low and high English exposure groups for both the pre and post English exposure measures. The groups were subsequently compared on their speech perception and production outcomes. In addition, information regarding the children's siblings i.e., number of older siblings in the home, average age, and languages used with the child, and the background of the children's network ties e.g., proportion of second-generation Bangladeshi ties was used in a correlation analyses.

The results showed that the pre language exposure measures were significantly related to the children's perception and production outcomes at Time 1. Specifically, the children who were exposed to less English had significantly lower VID scores and shallower /p/-/b/ and /k/-/g/ VOT categorization slopes at Time 1 than did the high exposure group. Further, for VID, the children in the pre low exposure group performed significantly worse in the noise condition than did the high exposure group. Likewise, for production, the results displayed some differences between the groups at Time 1. The children with a low pre score used a shorter VOT when producing /b/, /p/ and /g/ than did the high pre group. There were no differences between the groups for vowel production. The difference between the pre high and low group was not significant at Time 2.

These findings are in line with previous bilingual research that has shown that the specific amount of exposure to each language can influence the children's speech and language outcomes (Hoff, 2003). The current findings suggest that for speech perception, sequential bilingual children are likely to be sensitive to the specific amount of English exposure; in turn this influences the formation of their phonetic categories in English (see also Garcia-Sierra et al., 2011). The children who had more

English exposure, albeit relatively small, before starting to fully acquire English in nursery may have already started to organise their perceptual space to account for both languages (see e.g., Curtin et al., 2011). This would make it easier to acquire the English phonemic contrasts (Flege, 1995). Whereas the low exposure group, some of whom had no English exposure from their main caregivers before entering school, may have developed more robust Sylheti phonemic categories, in turn making it harder to acquire new English categories as quickly as the high English exposure group (see e.g., Garcia-Sierra et al., 2011). Yet, no differences in vowel production were found between the groups. Both groups produced the vowels using a similar Euclidean difference value. It is important to note however that, as shown in Chapter 4, the bilingual children were not significantly different from their monolingual peers for vowel production when tested at Time 1, after 7 months of English experience at school. It is possible that the children may have differed in terms of other acoustic features (e.g., voice quality) when the children were first tested in nursery (Time 1). Alternatively, the 7 months of English experience in nursery prior to testing may have enabled the bilingual children to acquire the English vowel contrast for production and in turn no differences between the high and low group were observed.

Once the children had entered nursery their language exposure shifted from Sylheti to English dominant. Not only were they exposed to English at nursery and school, the children were also exposed to varying amounts of English from their network ties. Some children had more English speaking ties than others, resulting in differences in English exposure between the children. The second measure, post, accounted for these differences in exposure. However, unlike the pre measure, the only outcome measure that was related to the children's post measure was the VID scores. That is, children who were in the low post exposure group had significantly lower VID scores in the noise condition than did the high exposure group. However, this difference was only present at Time 1 in the noise condition. Thus, it is possible that differences in post English exposure between the children facilitated the acquisition of the English contrasts, resulting in a significant difference between the groups. That is, the children exposed to more English once starting nursery (i.e., high exposure post group) will have been able to establish more robust English vowel categories than the children who were exposed to less English (i.e., low exposure post group).

Another possible explanation draws on the findings from infant research that suggests that vowel categories develop earlier than consonants (see e.g., Kuhl et al.

2008). Specifically, when the children start school their Sylheti vowel categories may be more established than their Sylheti consonant categories. The continuation of high Sylheti exposure (i.e., the low exposure post group) during the initial stages of English acquisition in nursery may have influenced the children's ability to acquire the English contrasts. In turn the children in the low-exposure post group had less robust English vowel categories than did the children in the high-exposure post group, who in turn displayed better English vowel identification. This may also explain the differences in vowel identification in the noise condition at Time 1. If the low-exposure group have less robust categories than the high exposure group they are likely to find vowel identification particularly difficult in hard listening situations such as noise (Mayo, Florentine & Buus, 1997).

In addition to the amount of English exposure, this study explored the children's social networks in terms of the characteristics of their network ties. The correlation analyses showed that some, but not all factors were related to the children's perception and production measures. Sibling factors were significantly correlated with the children's vowel ID scores at Time 1 and plosive categorization at Time 2. The children who had English speaking older siblings had higher vowel ID scores at Time 1 and steeper /p/-/b/ categorization slopes at Time 2. Thus, having English-speaking siblings in the home likely influenced the children's English phoneme perception abilities. Likewise, the number of older siblings and the proportion of UK born siblings was significantly related to the /k/-/g/ slope values at Time 2. The children with more older siblings and UK born ties had steeper slopes at Time 2.

Such findings raise interesting questions regarding the type of input the children are exposed to from their network ties i.e., accented English vs. native-like English speech. As shown in Chapter 2, it is likely that the second-generation older siblings will have provided the children with a more native-like English speech model than a first-generation caregiver. This may explain why the sibling factors are significantly correlated with the child's perception outcomes. Further, the majority of parents in this study reported that although they mainly spoke Sylheti to their child, once the child started to acquire English in nursery, English became dominant amongst siblings. Even in cases where parents reported that they had a "Sylheti-only" rule in the home, the siblings would speak both Sylheti and English to each other. This pattern of language use between siblings has been reported in previous bilingual studies that have highlighted how it is the older siblings who drive the host country language use in the home (see e.g., Hoff, 2006).

Not all social network factors were correlated with the children's perception and production. Average age of siblings was not significantly correlated with any of the children's outcomes. This could be due to the homogeneity in the age of the older siblings: the majority were aged 7-9 years old. For the children's production measures only two social network factors were related to the children's outcomes. First, the children's proportion of Late arrival Bangladeshi ties was significantly correlated with their production of /uː/ and /ʊ/ (*boot* and *book*) at Time 1 and 2. The children who had more Late arrival ties had smaller Euclidean distance values than did the children who had fewer Late arrival ties. The children with fewer Late arrival ties also demonstrated GOOSE and FOOT fronting. Second, sibling language was significantly correlated with the children's /k/ production at Time 1. The children who had English-speaking siblings used longer VOT values for /k/ than did the children who had siblings who spoke both Sylheti and English. As for perception, it is likely that the older siblings provided the child with a native-like English speech model. All other factors were not significantly correlated with the children's outcomes.

It could be the case that these broad measures of network ties are not sensitive enough to illustrate differences between the children. Although it could be hypothesized, based on the findings from Chapter 1, that Late arrival speakers are likely to use more accented- Sylheti variants and use less English than UK born ties, these measures give little information about the exact input or language used by these speakers (see also Fernald, 2006). In contrast, the language exposure measures estimated the amount of time the child is exposed to each language and thus provided a more detailed picture about what language/s the child is being exposed to. The analyses on language exposure revealed some interesting findings. Although all of the children in this study were exposed to no less than 80% Sylheti before entering school, differences in the remaining 20% of English exposure still influenced the children's initial speech perception abilities. These findings are relevant when considering simultaneous bilingual research where bilingualism is often classified as having a proportion of language 1 and a proportion of language 2, but these proportions often vary between studies (e.g., 65% & 35%, Sebastián-Gallés & Bosch, 2003; Bosch, 2009; 80% & 20%, Sundara & Scutellaro; 70% & 30%, Fennell, Byers-Heinlein & Werker, 2007). Thus, any differences in development found between bilingual studies could be attributed to the small differences in the proportion of exposure in each language.

This study demonstrated that the amount of English exposure, particularly before the children entered full-time education, significantly influenced the child's outcomes at Time 1. The next chapter will explore the influence of language exposure further, with the aim to investigate the influence of fine-grained phonetic differences in the main caregivers English speech production.

7. Exploring the relationship between caregiver speech and phonemic acquisition

7.1. INTRODUCTION

Chapter 6 investigated the role of language exposure and social networks on the acquisition of English phonemes. This chapter explores the role of input further by investigating the influence of fine-grained acoustic-phonetic differences in speech input (i.e., caregiver accent) on the children's perception and production of English vowels and plosives. As discussed in Chapter 1, not only are these children likely to be exposed to differing amounts of Sylheti and English depending on their social networks, they are also likely to be exposed to Sylheti-accented English, depending on the language background of the speaker as was set out in Chapter 2. Recent research has shown that infants and toddlers are sensitive to variability in voice and accent (Schmale & Seidl, 2009; Schmale, Cristià & Seidl, 2012), as well as fine-grained phonetic differences in caregivers' speech (Cristià, 2011). This chapter will explore the relationship between the caregivers' production of English and the bilingual children's acquisition of English phonemes.

A detailed acoustic-phonetic analysis of English vowels (formant analysis) and plosives (voice onset time) produced by the bilingual children's main caregivers was conducted. Due to practical constraints it was not possible to record the caregivers' speech in a child-parent dyad, therefore the parents were recorded separately in the child's home. The caregivers were subsequently grouped into "low-accented" and "high-accented" based on their realizations of the English speech sounds. Based on these groupings, their children's English phonemic categorization measures (perception and production) were compared.

7.2. MATERIALS AND METHOD

7.2.1. Participants

38 caregivers (6 fathers, 32 mothers, aged 26 – 36) of the children from Chapters 3 and 4 participated in the study. Eighty five percent of the caregivers were born in Sylhet Bangladesh, of which 34% arrived in the UK before the age of 16 (mean 7.5 years, range 5 – 12 years old). The remaining 66% arrived in the UK after the age of 16

(mean 22 years, range 19-27). One additional caregiver was recorded but not included due to background noise in the recording.

To ensure that the speech reflected the English that the child is likely to be exposed to from his/her main caregiver, in the cases where the main caregiver only spoke Sylheti (*n* = 4), the caregiver who spoke the most English to the child was recorded for this study. These four caregivers spent an average of 60% of the child's waking hours with the child.

7.2.2. Target sounds

The same target words used in the child study (Chapter 4 & 5) were elicited using a picture-naming task (see Appendix 1b, 1c, 2b, 2c for the target pictures). Six monophthongal vowels: /iː/, /ɪ/, /ɑː/, /ʌ/, /uː/, /ʊ/ were elicited in a CVC context. For each target sound two words were elicited, one was identical to the word used in the child vowel identification perception task (see Chapter 4): *sheep* /ʃiːp/*, ship* /ʃɪp/*, cart* /kɑːt/*, cut* /kʌt/*, pool* /puːl/*, pull* /pʊl/. The other was in a /b/VC context: *bead* /biːd/*, big* /bɪg/*, bart* /bɑːt/*, bubble* /bʌbəl/*, boot* /buːt/*, book* /bʊk/, and identical to the words used in the child vowel production study (see Chapter 4).

For plosives, bilabial and velar voiced/voiceless sounds: /p/, /b/, /k/, /g/ were elicited in a word-initial stressed position. Two words were elicited per target sound, one was identical to the minimal pairs used in the child categorization task in (Chapter 5): *bee* /biː/ - *pea* /piː/*, coat* /kəʊt/ - *goat* /gəʊt/. The second word was identical to the words in the child plosive production study (Chapter 5): *pear* /peə/, bear /beə/, cat /kæt/, good /gʊd/.

7.2.3. Procedure

The caregiver recordings were made in a quiet room in the caregivers home (*n* = 33) or in their child's school (*n* = 5) using a H2 Zoom audio recorder with a sampling rate of 44.1 kHz, 16-bit resolution.

All pictures were elicited in the carrier sentence *Say* _______ *again*. As it was not possible to elicit the words in a natural child directed speech environment i.e., a child-caregiver dyad, the caregivers were instructed to name the pictures as if they were naming them to their child. To facilitate this, the researcher demonstrated how she would name a picture to a younger family member. To avoid influencing the caregivers' speech style, the researcher stressed the importance of naming the picture in the way that they would speak their child, and this may be different to the researchers style. If the caregiver named the picture incorrectly e.g., *shoe* for the *boot* picture, the researcher gave semantic cues e.g., *they are bigger than shoes and you may wear them when it's cold*. For vowels, if the semantic cue did not elicit the word, a phonetic cue was given e.g., *it starts with* /ʃ/. Care was taken to not produce the target vowel sound. If the cues did not elicit the target word then the researcher named the picture and asked for a repetition. In such cases, the elicited word was not included in the final analysis. All target words were elicited twice in a randomized order.

7.2.4. Acoustic analysis of the caregiver speech

All of the acoustic analyses were conducted in Praat 5.3.04 (Boersma & Weenink, 2012).

Vowels

A total of 902 tokens were analyzed. The stimuli were located manually, and then F1 and F2 were extracted using hand-corrected LPC analyses. Formant frequencies were measured from the mid-point of the steady-state portion of the vowel. The steady-state portion of the vowel was defined as the part of the vowel that was closest to the midpoint and where the formant frequencies were most stable.

Three steps were taken to quantify the spectral differences between the three vowel contrasts: /iː/-/ɪ/, /ɑː/-/ʌ/, /uː/-/ʊ/. First, so that the data could be compared, Lobanov's z-score transformation (Lobanov, 1971) was used to normalize the F1 and F2 values.

As the aim of this study was to investigate the caregivers level of Sylheti accentedness in their production of English vowel contrasts, the decision was made to calculate the Euclidean distance between the Lobanov z-score values for each vowel contrast e.g., *ship – sheep*. Thus, instead of plotting the caregivers' English vowel space, the Euclidean distance values were used to measure the separation between the caregivers' vowel productions. That is, the larger the Euclidean distance, the greater the size of the contrasts that were produced (i.e., less accented). The following equation was used to calculate the Euclidean distance:

$$\text{Euclidean Distance} = \sqrt{}\,(F1vi - F1vii)^2 + (F2vi - F2vii)^2$$

Where vi and vii refer to vowel 1 (e.g., ship) and vowel 2 (e.g., sheep) respectively. The mean Euclidean distance values obtained for the four tokens were calculated.

Preliminary analysis of the vowel formants showed a significant difference between the two phonetic environments: /p/V/l/ vs. /b/VC for the back vowels: /uː/ and /ʊ/. As noted in previous studies, this is likely due to the phenomenon known as GOOSE and FOOT fronting (see e.g., Hawkins & Midgley, 2005; Wells, 1982), causing these vowels to move forward in the vowel space, i.e., be produced with a higher F2. In the /p/V/l/ context GOOSE and FOOT fronting is prevented due to the velarization caused by the final /l/. In turn, for some of the caregivers, the /uː/ and /ʊ/ in the /b/VC had a higher F2 than in the /p/V/l/ context. A one-way ANOVA confirmed that the F2 values for /uː/ and /ʊ/ were significantly higher in the /b/VC context than in the /p/V/l/ [$F(1,73) = 11.88$, $p = .001$]. Subsequently, the two phonetic environments were analyzed separately and will be hereafter referred to as **/uː/-/ʊ/ I** for the /b/VC context and **/uː/-/ʊ/ II** for the /p/V/l/ context.

Plosives

A total of 602 tokens were analyzed. The voice onset time (VOT) measurements were obtained from the waveform and checked against the corresponding spectrogram. The lag VOT (i.e., above 0 ms) was measured as the time between the release of the plosive closure and to the zero crossing of the first periodic part of the waveform following the vowel (in milliseconds). For prevoiced tokens the voicing lead was measured using the waveform and spectrogram.

7.2.5. Child outcome measures

The child measures reported in Chapter 4 (vowels) and Chapter 5 (voicing) were used in the analyses to follow. A detailed description of the child measures can be found in Chapter 4 (vowels) and Chapter 5 (VOT). All children were tested twice: Time 1, in nursery (after an average of 7 months full-time English speaking education, mean age: 52.7 months) and Time 2, 1 year later (reception class).

For vowel perception, the children completed a three-alternative forced-choice computerized task in two conditions: quiet and 2dB SNR speech-shaped noise, separately. The children were required to identify real word minimal pairs that contained the target vowel contrast e.g., ***sheep***, *shape,* ***ship***, as spoken by a Standard Southern British English (SSBE) female speaker. For vowel production, an acoustic analysis (F1, F2, and duration) of the children's production of the target vowel contrasts was conducted. For the current analyses, the Euclidean distances for the children's normalized F1 and F2 were used (calculated in Chapter 6). For VOT categorization, the children completed a 2 alternative forced-choice adaptive task to assess category identification along a *pea-bee* and *coat-goat* VOT continuum. For each continuum, the children's categorization slope and phoneme boundary was measured. For plosive production, an acoustic analysis of the children's VOT for English word initial /p/, /b/, /k/, /g/ was conducted. The children produced the same words as used in the current experiment.

7.3. RESULTS

The results will be presented in three parts: **1)** a description of the caregivers' speech, **2)** the relationship between caregivers' speech and the bilingual children's vowel categorisation, **3)** the relationship between the caregivers' speech and the bilingual children's plosive categorisation.

7.3.1. Caregivers' speech

For all contrasts, variability in the caregivers' production was found. For vowels, some caregivers had relatively smaller Euclidean Distance values than others. Likewise, for VOT, similar variability was observed between caregivers i.e., some caregivers used a shorter VOT values for the English plosives than others (see Table 7.1 & 7.2 shaded cells for data). Based on a median split of the data, caregivers were classified into "high-accented" if the Lobanov Euclidean distance (ED) scores or voice onset time (VOT) values were below the median, and "low-accented" if the ED or VOT values were at or above the median (see Tables 1 & 2 for the total group values and subsequent high-accented and low accented grouping).

Vowels

Table 7.1 shows the mean Lobanov Euclidean distance scores for each vowel contrast obtained for the caregiver groups: high-accented and low-accented. Overall, the low-accented group produced the larger distance scores than did the high-accented group, in particular for the /uː/-/ʊ/ I contrast. The lowest distance scores were found for the /uː/-/ʊ/ II contrast by the high-accented group.

Table 7.1. Caregiver Euclidean distance scores (z-score transformed Hz) for each vowel contrast, for the high-accented and low-accented group. Shaded grey rows include all data.

Contrast/ group	Mean ED	Range	SD
/iː/-/ɪ/	.46	.08 -1.10	.27
High	.25	.08 - .40	.21
Low	.67	.45 – 1.10	.09
/ɑː/-/ʌ/	.55	.18 – 1.09	.22
High	.38	.18 - .51	.10
Low	.74	.53 – 1.09	.17
/uː/-/ʊ/ I	.71	.12 – 1.93	.52
High	.27	.12 - .60	.15
Low	1.2	.66 – 1.93	.36
/uː/-/ʊ/ II	.38	.05 – 1.07	.24
High	.20	.05 - .33	.09
Low	.57	.35 – 1.07	.21

Table 7.2. Caregiver voice onset time (VOT) in milliseconds for bilabial and velar plosives, for the high-accented and low-accented caregiver group. Shaded grey rows include all data.

Target sound/ group	Mean VOT	Range	SD
/p/	34	4.5 – 68.3	17.2
High	20	38.3 – 68.3	10.3
Low	48	4.5 – 36.8	9.0
/b/	-27	9.5 – -120	36.4
High	-53	-120 – -2	32.6
Low	1	-15 – 9.5	6.3
/k/	39	65.7 – 5.5	14
High	29	5.5 – 39.35	41.5
Low	50	40 – 65.7	27
/g/	-.9	-58.5 – 19	25.5
High	-18	-57.5 – 9.5	68.5
Low	16	10 - 19	24.7

Voice onset time (VOT)

Table 7.2 shows the mean VOT values obtained for the caregivers' productions of /p/, /b/, /k/ and /g/. For the voiced consonants /b/ and /g/ prevoicing is evident. As expected, the longest VOT values are present for /k/. In general the high-accented group used short VOT values i.e., < 30ms, including prevoiced tokens for the English plosives, in contrast the low-accented group only produced lag tokens i.e., >1 ms.

Summary of the caregivers' speech

The acoustic-phonetic analysis of the caregivers' speech revealed variability between the speakers English production. Some caregivers produced English vowel contrasts with relatively small ED values and the plosives with prevoiced and short-lag VOT, whereas other caregivers produced the vowels with larger vowel ED values and plosive with short-long VOT. Based on a median split of the data, the caregivers were subsequently divided into two significantly different accent-groups: high-accented and low-accented. Similar to that of the findings in Chapter 2, the high-accented caregiver group produced English phonetic categories that reflect their first-language (L1), Sylheti. That is, the target vowel contrasts were produced with small Euclidean distances and the plosives were produced with a short or voicing lead. In contrast, the low-accented caregiver group used fewer Sylheti-like features in their speech, using larger Euclidean differences for the vowel contrasts, and longer VOT values than did the high-accented group.

7.3.2. Is there a relationship between caregiver accent and the bilingual children's English vowel categorisation?

A summary of the children's vowel identification (Vowel ID) scores and production Euclidean distance values can be found in Table 7.3 and Table 7.4, respectively. To investigate if there was any relationship between the caregiver speech and children's vowel perception and production, separate analyses were conducted for each vowel contrast, e.g., /i/-/ɪ/.

To investigate the influence of caregiver speech on the children's vowel perception and production, separate Generalized Linear Mixed Model (GLMM) analyses were conducted for each vowel contrast, with time (i.e., Time 1, Time 2), parent group (i.e., high-accented and low-accented) and condition (i.e., quiet, noise, for perception only) treated as fixed factors. To account for individual differences, participant was treated as a random factor. Significant effects were explored using sequential Sidak post-hoc analyses. All tests were corrected for multiple comparisons.

Vowel perception

Table 7.3 shows the bilingual children's Vowel ID scores for each contrast at Time 1 and Time 2. For /iː/-/ɪ/, the GLMM analysis revealed a main effect of group, $F(1,68)$= 4.38, $p < .05$, time $F(1,80)$= 28.82, $p < .001$, and condition $F(1,66) = 5.583$, $p < .05$. As shown in Figure 7.1, overall the children in the high-accented caregiver group had significantly lower Vowel ID scores than did the low-accented group, both groups displayed an increase in scores from Time 1 to Time 2. Sequential Sidak tests revealed a significant difference between the high-accented and low-accented group in the noise condition at Time 1 and 2 ($p < .05$). There was no significant interaction between time and group, $F(1,80) = 0.91$, $p > .05$.

For /ɑː/-/ʌ/, the GLMM analysis revealed a main effect of group $F(1,108)$ = 4.09, $p = .046$, time, $F(1,108) = 23.49$, $p < .001$, and condition, $F(1,108)$ =8.18, p = .005. Figure 1 shows that overall the high-accented group had significantly lower Vowel ID scores than did the low-accented group ($p < .05$). Sequential Sidak tests revealed that the high-accented group had significantly lower scores than did the low group in the noise condition at Time 1 ($p < .05$), but not at Time 2. There was no significant difference between the groups in the quiet condition at Time 1 and 2 ($p >$ .05). The interaction between group and time, $F(1,85) = 1.198$, $p > .05$, was not significant.

For /uː/-/ʊ/, the GLMM analyses revealed a significant main effect of group, $F(1,112) = 5.33$, p = .02, time, $F(1,112) = 12.68$, $p = .001$. Figure 7.1 shows that overall the high-accented group had significantly lower Vowel ID scores than did the high-accented group. Sequential Sidak tests revealed that high-accented group had significantly lower Vowel ID scores in the noise condition at Time 1 ($p < .05$). All children displayed and increase in Vowel ID scores from Time 1 to Time 2. The interaction between group and time, $F(1,112) = 2.07$, $p > .05$, was not significant.

Table 7.3. Children's vowel ID proportion correct scores for the vowel contrasts in quiet (Q) and noise (N), for the high-accented (shaded cells) and low-accented caregiver groups at Time 1 and Time 2.

		Child vowel ID scores					
		High-accented caregiver group			Low-accented caregiver group		
	Contrast	*Mean*	*Range*	*SD*	*Mean*	*Range*	*SD*
Time 1	/iː/-/ɪ/ Q	.88	.67 – 1	.12	.88	.67 – 1	.15
	/iː/-/ɪ/ N	.84	.50 – 1	.16	.84	.67 – 1	.11
	/ɑː/-/ʌ/ Q	.76	.50 – .83	.14	.78	.50 – .83	.18
	/ɑː/-/ʌ/ N	.63	.50 – .83	.11	.77	.50 – .83	.10
	/uː/-/ʊ/ Q	.64	.50 – .83	.15	.68	.50 – .83	.12
	/uː/-/ʊ/ N	.53	.50 – .83	.15	.68	.50 – 67	.07
Time 2	/iː/-/ɪ/ Q	1	1	0	1	1	0
	/iː/-/ɪ/ N	.94	.67 – 1	.11	.99	.83 – 1	.04
	/ɑː/-/ʌ/ Q	.89	.67 – 1	.15	.91	.67 – 1	.10
	/ɑː/-/ʌ/ N	.80	.67 – 1	.08	.83	.67 – 1	.14
	/uː/-/ʊ/ Q	.81	.50 = 1	.14	.90	.67 – 1	.20
	/uː/-/ʊ/ N	.67	.50 – 1	.19	.72	.50 – 1	.19

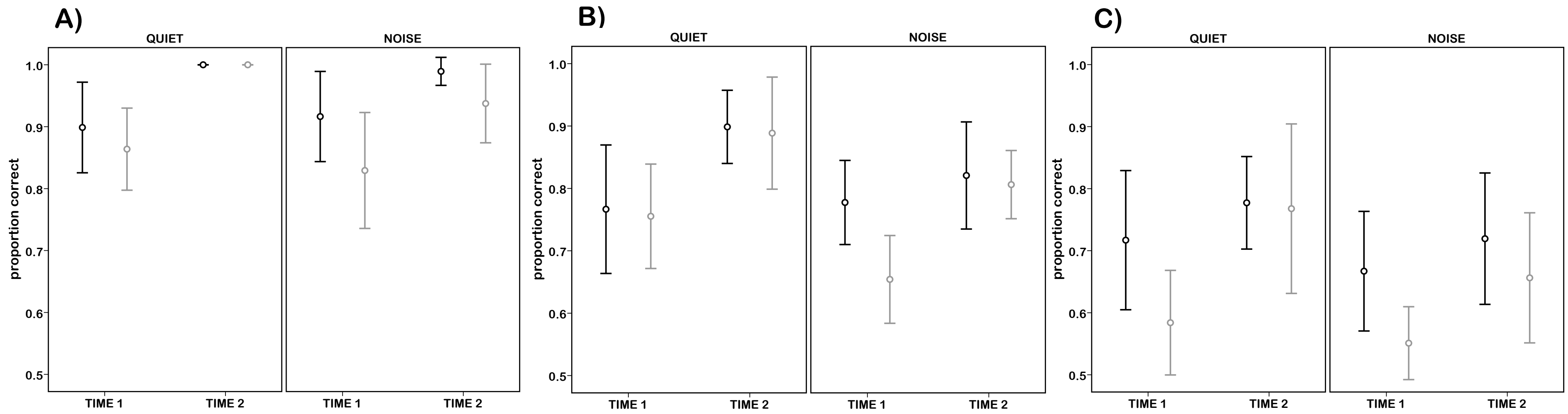


Figure 7.1. Children's vowel identification scores **A)** /iː/-/ɪ/, **B)** /ɑː/-/ʌ/ and **C)** /uː/-/ʊ/ English vowel contrasts at Time 1 (nursery) and Time 2 (school). Children are grouped according to their caregiver group: high-accented (grey bars) or low-accented (black bars). Error bars indicate 95% confidence intervals.

Vowel production

Table 7.4 shows the children's Euclidean distance values for each vowel contrast: /iː/-/ɪ/, /ɑː/-/ʌ/, /uː/-/ʊ/. The child data was grouped according to high-accented and low-accented caregiver speech. At Time 1 and 2 few differences can be observed between the high-accented and low-accented group. The largest contrast difference can be observed at Time 2 for the /uː/-/ʊ/ I contrast (mean ED; high: 1.28, low: 1.53). The GLMM analyses revealed a significant main effect of time for /iː/-/ɪ/, $F(1,67) = 37.19$, $p < .0001$, where all children had larger Euclidean distance values for /iː/-/ɪ/ at Time2. All other main effects and interactions were not significant ($p > .05$).

Summary of vowel results

The analysis revealed a relationship between the caregivers' speech and the children's vowel perception for all vowel contrasts. Specifically, the data showed a significant overall difference between the children of high-accented caregivers and low-accented group caregivers for all vowel contrast. The children of high-accented caregivers i.e., smaller Euclidean distance values for their production of the target English vowel contrast, had lower VID scores than the children of low-accented caregivers. This was particularly evident at Time 1 (nursery) and in the noise condition. An overall significant difference between the children of high-accented and low-accented group was found in the noise condition, but not in the quiet condition. This difference was significant at Time 1 and 2 for /iː/-/ɪ/, but only Time 1 for /ɑː/-/ʌ/ and /uː/-/ʊ/. For production, but no difference between the children of high-accented and low-accented caregiver groups was found.

7.3.3. Is there a relationship between caregiver accent and the bilingual children's English plosive categorization?

A summary of the children's perception (categorization slope & phoneme boundary) and production (VOT) measures can be found in Table 7.5 and 7.6, respectively. The data displayed some within-group variability observed in the children's plosive categorization. For perception, some of the bilingual children had similar slope and boundary values to that of their monolingual peers, whilst others had lower values (see Figures 5.4 & 5.5, Chapter 5). For production, within-group variability was observed, where some children produced pre-voiced tokens and others using lag voicing.

Table 7.4. Children's Lobanov (z-score) normalized Euclidean distance values for the production of English vowel contrasts, at Time 1 and Time 2. Children are grouped according to high-accented (shaded cells) and low-accented caregiver speech.

		High-accented			**Low-accented**		
contrast	**Time**	*Mean*	*Range*	*SD*	*Mean*	*Range*	*SD*
/iː/-/ɪ/	*1*	.74	.53 – 1.05	.18	.80	.52 – 1.07	.14
	2	1.02	.67 – 1.20	.14	1.03	.90 – 1.51	.23
/ɑː/-/ʌ/	*1*	.75	.56 – 1.09	.16	.82	.51 – 1.29	.24
	2	.94	.62 – 1.25	.31	.96	.67 – 1.47	.25
/uː/-/ʊ/ I	*1*	1.32	.31 – 2.08	.46	1.39	1.06 – 2.08	.39
	2	1.28	.45 – 2.16	.43	1.53	1.01 – 2.42	.52
/uː/-/ʊ/ II	*1*	.63	.26 – .93	.20	.58	.38 – 1.04	.22
	2	.70	.33 – 1.31	.31	.67	.56 – 1.13	.29

Perception (VOT categorization)

To investigate the influence of the caregiver's VOT on the children's perception (categorization slope and phoneme boundary) at Time 1 and 2, the caregivers' VOT was averaged over the voiced and voiceless stops to give a single "combined" VOT measure per caregiver, per place of articulation (bilabial and velar). An overall lower VOT

indicates high-accented production, whilst an overall higher VOT indicates low-accented production. For example, a caregiver who produced /g/ with a mean VOT of -58ms and /k/ with a mean VOT of 45ms, would have a single VOT measure of -6.5. In contrast, a caregiver who produced /g/ with a mean VOT of 18ms and /k/ with a mean VOT of 67ms, would have a single VOT measure of 43ms. The final caregiver VOT measures can be observed in Table 7.5.

Table 7.5. Caregiver combined VOT measures in milliseconds for the caregivers' bilabial and velar plosive production. The shaded grey cells include all data.

Group	*Bilabial mean*	*Bilabial median*	*Velar mean*	*Velar median*
All	1.8	8.6	18.2	25
High	-17.7	-13	7.3	8.1
Low	19.7	21.3	30	29.5

Table 7.6 shows the children's plosive categorization measures, grouped according to caregiver group. For /p/-/b/ slope, the GLMM analysis revealed a main effect time, $F(1,37) = 30.77$, $p < .001$. As shown in Figure 7.2, all children demonstrated an increase in slope steepness from Time 1 to Time 2. All other main effects and interactions were not significant ($p > .05$). For /p/-/b/ phoneme boundary, there were no significant main effects or interactions ($p > .05$). There was no significant difference between the children in the high-accented and low-accented caregiver group for phoneme boundary.

For /k/-/g/, the GLMM analysis revealed a main effect of group, $F(1,44) = 6.42$, $p = .01$. Figure 7.2 shows that the children in the low-accented caregiver group had significantly higher slope values than did the high-accented group. Sequential Sidak tests revealed a significant difference between the children in high-accented and low- accented caregiver group at Time 2 (p = .02), but not at Time 1 (p > .05). The main effect of time, $F(1,24)$= 11,60, $p < .001$, was significant. All children showed an increase in slope values from Time 1 to Time 2. For phoneme boundary, there was no significant main effects or interactions (p > .05). The children in the high-accented and low-accented caregiver groups did not have a significantly different /k/-/g/ phoneme boundary.

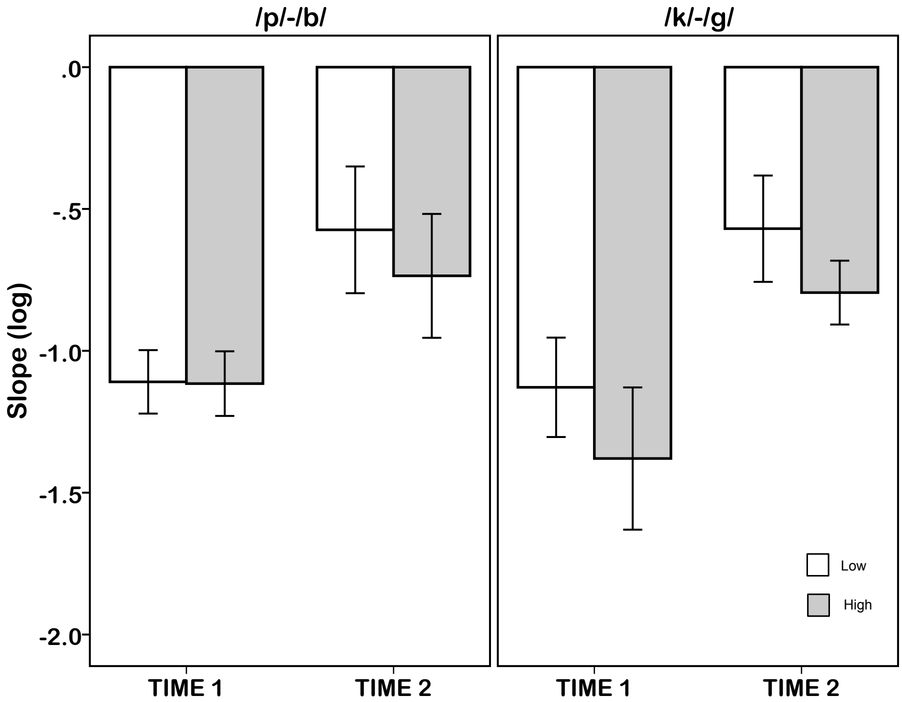


Figure 7.2. Children's categorization slope values, (log) for the *pea-bee* and *coat-goat* VOT continua. A higher log value indicates a steeper categorization slope (i.e., more consistent labeling of the continuum) Children are grouped according to their caregiver group: high (grey) and low (white). The error bars indicate 95% confidence intervals.

Table 7.6. Children's categorization slope and boundary measures for /p/-/b/ and /k/-/g/ at Time 1 and Time 2, for the high-accented (shaded grey) and low-accented group.

	Measures		High-accented			Low-accented		
			Mean	*Range*	*SD*	*Mean*	*Range*	*SD*
Slope (log)	*Time 1*	/p/-/b/	-1.10	-1.40 – -.80	.18	-1.10	-1.27 – -.88	.17
		/k/-/g/	-1.36	-2.0 – -.80	.39	-1.12	-1.70 – -.68	.28
	Time 2	/p/-/b/	-0.80	-1.22 – -.30	.34	-0.57	-1.03 – .01	.35
		/k/-/g/	-1.10	-2.0 – -.29	.58	-.84	-1.40 – -.43	.28
Boundary	*Time 1*	/p/-/b/	34	13 – 59	13	31	16 – 47	10
		/k/-/g/	44	27 – 61	11	43	49 – 37	14
	Time 2	/p/-/b/	35	50 – 19	12	25	30 - 16	6
		/k/-/g/	48	44 – 60	4	46	27 – 63	14

ii. Plosive production

Table 7.7 shows the VOT values for the high and low group. For /p/-/b/, the GLMM analyses revealed a significant main effect of time for /p/, $F(1,35) = 7.29$, $p < .02$, and for /b/, $F(1,36) = 6.23$, $p = .02$, where all children used longer VOT values at

Table 7.7. Children's VOT values in milliseconds for bilabial and velar plosives, for the high (shaded) and low group, at Time 1 and Time 2.

		High-accented			**Low-accented**		
		Mean	*Range*	*SD*	*Mean*	*Range*	*SD*
Time 1	/p/	53	35 – 78	12	59.6	37 – 79	19
	/b/	-8.5	-80 – 7	23	-2.9	-46 – 8	14
	/k/	58	28 – 83	14	59.4	34 – 90	17
	/g/	-3.0	-90 – 14	28	7.6	-2.5 – 14	5
Time 2	/p/	61.7	30 – 111	24	75.5	42 – 110	21
	/b/	3.4	-58 – 10.5	16	5.8	-42 – 13.5	14
	/k/	64	36 – 101	21	64	39 – 125	21
	/g/	11.6	-26 – 24	12	17.3	9 – 29	6

Time 2 than they did at Time 1. There was a main effect of group for /p/, $F(1,35) = 4.94$, $p < .05$, but not for /b/, $F(1,36) = .090$, $p > 0.05$. As shown in Figure 7.3 children in the high-accented caregiver group used significantly shorter VOT values for /p/ than did the low-accented caregiver group. The interaction between group and time was not significant for /b/, $F(1,36) = 0.17$, $p > .05$, or /p/, $F(1,35) = 0.63$, $p > .05$.

For /k/-/g/, the GLMM analyses revealed a significant main effect of group for /g/, $F(1,36) = 4.786$, $p < .05$, but not for /k/, $F(1,65) = 0.03$, $p > .05$. As shown in Figure 7.3, the children in the high-accented caregiver group used a significantly shorter VOT for /g/ than did the children in the low-accented caregiver group. There was a significant main effect of time for /g/, $F(1,36) = 13.34$, $p = .001$, but not for /k/ ($p > .05$). All children showed an increase in VOT values from Time 1 to Time 2 for /g/, but not for /k/. There were no significant interactions ($p > .05$)

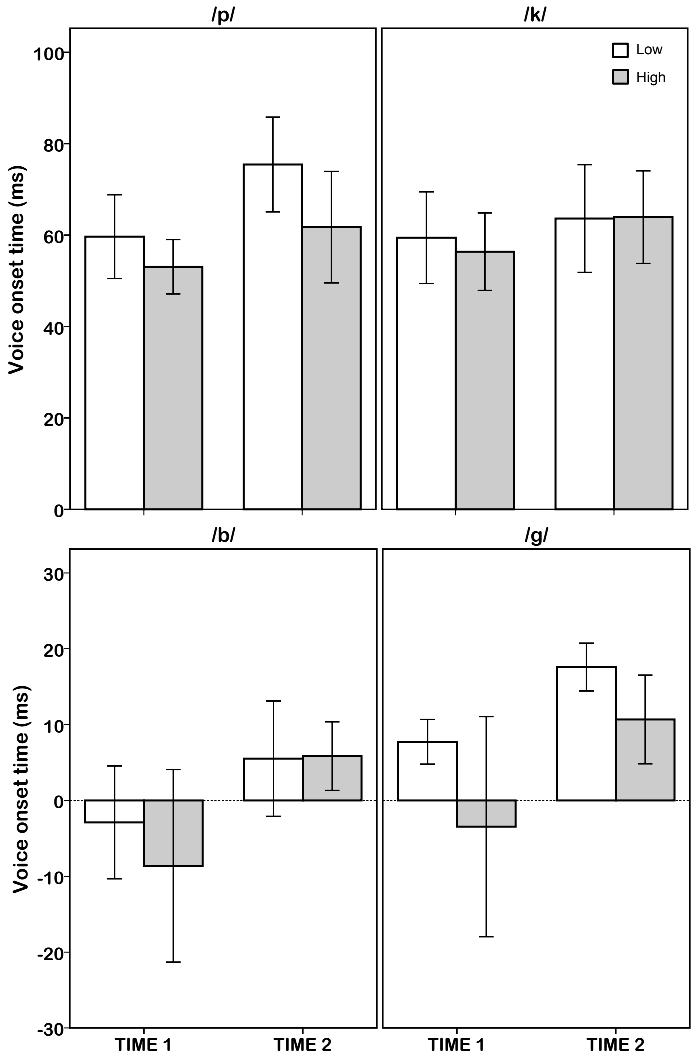


Figure 7.3. Children's voice onset time values in milliseconds for the production of English bilabial and velar plosives. Children are grouped according to caregiver group: high-accented (grey bars) and low-accented (white bars). The error bars indicate 95% confidence intervals.

Summary of VOT results

For perception, the results revealed no group differences for phoneme boundary. The children in the high-accented and low-accented caregiver groups had similar phoneme boundaries for the /p/-/b/ and /k/-/g/ contrast at Time 1 and Time 2. For categorization slope, differences were observed at Time 2, but not at Time 1. Specifically, for /k/-/g/, a significant difference between the children in the high-accented and low-accented caregiver groups was found at Time 1, but not at Time 2. At Time 1, the high-accented caregiver group had significantly shallower categorization slopes than did the low-accented caregiver group. There were no significant differences between the groups for /p/-/b/.

For production, differences were found between the children in the high-accented and low-accented caregiver group at Time 1 and 2 for /p/ and /g/, but not for /k/ and /g/. Specifically, the children in the high-accented caregiver group used significantly shorter VOT than did the low group for /p/ and /g/, including some prevoiced tokens for /g/.

7.4. DISCUSSION

The aim of this study was to investigate the influence of fine-grained acoustic differences in caregiver speech on children's phoneme categorization. To do so, an acoustic analysis of the bilingual children's caregiver production of English monophthong vowel contrasts and plosives was conducted. The caregivers who were recorded were one of the main caregivers who spoke some English to the child before s/he entered full-time education. Children were subsequently grouped according to their caregiver's speech: high- and low accented speech, and further analyses comparing the children's phoneme categorization measure (from Chapter 4 & 5) was conducted.

The acoustic analysis of the caregiver's speech showed variability in the caregivers' production of English vowels and plosives. The caregivers were subsequently divided into two caregiver accent groups: high-accented and low-accented. The high-accented caregivers had smaller Euclidean distance values for the English vowel contrasts, and shorter VOT values for the plosives than did the low-accented caregiver group. Specifically, the high-accented caregivers tended to collapse the vowel contrast, reflecting their Sylheti phonetic categories e.g., /iː/-/ɪ/ produced with a small Euclidean distance, similar to the single Sylheti vowel /i/. Similarly, for plosives, the high-accented caregivers used significantly shorter VOT values, including prevoiced tokens, than did the high group. Similar to the findings in Chapter 2 (adult production study), the high-accented caregivers had arrived in the UK at a later age (i.e., after 16 years old) than did the caregivers' in the low-accented group.

Based on previous research that has shown that infants and children are sensitive to the acoustic-phonetic accent differences in their input (e.g., Cristià, 2011), it was hypothesised that the children of the high-accented caregivers may have developed vowel categories that reflect their accented input thus showing difficulties with the English vowel contrasts. That is, if the small amount of English that these children were exposed to before entering English-speaking education was heavily accented, it is possible that their English categories will also display these Sylheti-accented features, making the English vowel contrasts difficult for the children.

For vowel identification, the children in the high-accented caregiver group had significantly lower identification scores than did the children in the low-accented caregiver group. For /ɑː/-/ʌ/ and /uː/-/ʊ/, this difference was significant at Time 1, particularly in the noise condition, but not at Time 2. For /iː/-/ɪ/, there was a significant difference between the groups at Time 1 and 2. For vowel production, no difference between the groups was observed. The children in the high-accented and low-accented caregiver groups both produced the vowel contrast with similar Euclidean values. For plosives, a significant difference in VOT categorization was observed at Time 2, but not at Time 1. Specifically, at Time 2 the children in the low-accented caregiver group had significantly steeper /k/-/g/ categorization slopes (i.e., more consistent labeling of the continuum) than did the children in the high-accented caregiver group. For /p/-/b/ no significant differences were observed between the groups. For production, significant differences between the groups were found for /p/ and /g/, but not for /k/and /b/. Specifically, the children in the high-accented caregiver group produced /p/ and /g/ with a significantly shorter VOT than did the children in the low-accented caregiver group.

As shown in previous studies, it could be suggested that this pattern of results reflects the children's sensitivity to the fine-grained acoustic-properties in their main caregivers' English production. For vowels, this sensitivity resulted in the children of high-accented caregivers having more difficulties with the English vowel contrasts at Time 1 when they first started to acquire English. In contrast, the children with low-accented caregivers had already started to establish English vowel categories from their caregivers, and in turn had more accurate English vowel identification than the children of high-accented caregivers. By Time 2, all of the children had had a year of English experience in full-time education. This experience would have enabled all of the children to begin to acquire the English vowel contrasts, and this would account for why no difference was observed between the children in the high-accented and low-accented caregiver groups at Time 2. Further, as shown in Chapter 4, no difference was observed between bilinguals and monolinguals at Time 2, suggesting that the children were similar to monolinguals in terms of English vowel development.

For plosive categorization, a difference between the high-accented and low-accented group was observed at Time 2, but not at Time 1. At Time 1 all of the children had shallower categorization slopes than they did at Time 2. The findings from Chapter 6 showed that the amount of English language exposure influenced the children's categorization at Time 1 and 2. In contrast, the results reported here suggest that caregiver accent did not influence the child's categorization at Time 1, but did at Time 2. A possible explanation for this could be that unlike vowels, which are thought to develop earlier than consonants (Kuhl et al., 2008), the children's English and possibly Sylheti plosive development may have been less advanced than their vowel development at Time 1. Thus, despite the differences in their caregiver's speech, at Time 1 all children have difficulties due to their stage in development. By Time 2, all children have more established plosive categories. The difference between the children in the high-accented and low-accented caregiver group may have arisen because the children of high-accented caregivers could have established less robust plosive categories from their accented input. Likewise, the low-accented group English plosive categories may have been reinforced by the more native-like English input.

For plosive production, fewer differences between the low and high group were found. As discussed in Chapter 4 and 5 (the main child study) it may have been the case that 7 months in school may have been enough time for all of the bilingual children to develop native-like English vowel categories (see also Tsukada et al., 2005; Darcy & Krüger, 2012).

Although it was not explored in the current study, the caregiver accent may it may indeed be confounded by many other factors such as English use and Sylheti production. That is, the high-accented caregivers in general may have used less English than did the low-accented group, resulting in more accented-English. This hypothesis is supported by the findings in Chapter 2 that showed that the Late arrival first-generation speakers not only used less English than did the Early arrival first-generation speakers, but also used more Sylheti-accent variants in their English vowel and plosive production. Likewise, the Early arrival first-generation speakers used more English that did the Late arrivals. Thus, the pattern of results could be in part explained by the fact that the children in the high group may have been exposed to less English before entering education than did the low group. Likewise, it is likely that any English that the high group was exposed to from their main caregivers is more accented than the low group. Although further statistical analysis is need, informal observation of the relationship between caregiver accent and English language exposure (see Appendix 4a & 4b) suggests a relationship between caregiver accent and language exposure. Specifically, the scatter plots suggest a positive relationship between caregiver accent and English exposure. Caregivers who used a shorter voice onset time in their stop productions (i.e., more Sylheti-accented), tended to use less English with their child. Likewise, for vowels, caregivers with smaller Euclidean distance values for their English vowel contrasts (i.e., more Sylheti-accented) tended to use less English with their child. Further statistical analyses is needed to understand which of these factors are more strongly related to the children's outcome measures (e.g., Multiple Regression analysis)

In sum, the results reported in this chapter have provided some insight into the possible influence of fine-grained differences in caregiver speech on bilingual children's phoneme acquisition. As shown in Chapter 6 it is likely that the bilingual children in this research will have also received some English input from their siblings, and future investigations could possibly involve analysing the speech of older siblings. Likewise, due to time constraints the caregiver speech data in this experiment was collected in an interview with the researcher. It could be suggested that data collected in this manner, instead of in a caregiver-child dyad (Cristià, 2011; Foulkes & Docherty, 2006), may not fully reflect the speech that the children were exposed to before entering full-time education (see e.g., Foulkes & Docherty, 2006). A future research project could therefore involve crosschecking the data reported in the current experiment with speech data collected in a caregiver-child interaction, to check for any discrepancies in production. However, the current analysis of the caregiver's speech suggests in addition to the influence language exposure has on phoneme acquisition (as shown in Chapter 6), the children in this research may also be sensitive to the fine-grained differences in speech input from their main caregivers. These results suggest that these differences are likely to influence sequential bilingual children's acquisition of the L2, however further research is needed in order to fully understand the weight of this factor in light of other possibly influential factors such as language exposure and features in the children's Sylheti input.

8. Discussion and conclusions

This thesis examined the phonemic development of sequential bilingual children from the London-Bengali community. The overall aim of the thesis was to investigate the effect of language environment and age of acquisition of English phonemes. To date, there is a large body of research on the phonemic development of monolingual infants and children, and a growing body of research on bilingual development, yet this is mainly limited to simultaneous bilinguals i.e., those exposed to both languages from birth (e.g., Bosch & Sebastián-Gallés, 2003a, 2003b; Sundara et al., 2008). Further, second-language research has mainly focused on adults, where the outcomes of L2 speakers with differing ages of acquisition are compared using a cross-sectional design (e.g., Baker & Trofimovich, 2005; Tsukada et al., 2005). However, little is known about the developmental trajectory of children who acquire a second language later in childhood, a pattern of L2 acquisition often found in immigrant communities.

The objective of this thesis was to provide some insight into the developmental patterns of such children. To do so, a series of experiments that aimed to track the acquisition of English phonemes by Sylheti-speaking UK born Bangladeshi children were conducted. Firstly, to better understand the L1 and L2 speech input that the London-Bengali children were likely to be exposed to, an acoustic-phonetic analysis of Sylheti (L1) and English (L2) phonemes produced by adult London-Bengali speakers was conducted. Based on these findings, the acquisition of English monophthongal vowel and plosive contrasts by sequential bilingual children from the London-Bengali community was examined. Children were tested at two time points, after approximately 7 months of English speaking full-time education, and one year later, during the first year of Primary school education. Finally, to investigate the influence of language exposure and caregiver speech on the bilingual children's outcome measures; detailed caregiver interviews and speech production recordings were conducted. The sections to follow will discuss the main findings from the experiments reported in this thesis, in the light of current monolingual, bilingual and L2 speech development research, as well as the current theoretical frameworks.

8.1. Age effects

The study reported in Chapter 2 investigated the production of Sylheti and English monophthongal vowels and plosives by adult speakers from the London Bengali community. All speakers had learned Sylheti as their L1, and had acquired English

(L2) at differing ages i.e., during childhood or as an adult. Not only did this study provide the first acoustic-phonetic description of Sylheti, the findings also showed clear age related differences in the speakers' production of the L1 and L2. Speakers who had acquired the L2 during adulthood demonstrated an influence of their L1 phonetic categories on their production of the L2 vowels and plosives. For example, the Late arrival speakers collapsed the English vowel contrast /iː/-/ɪ/ in terms of spectral information (i.e., F1 and F2) in their production, thus reflecting their single Sylheti category /i/. In contrast, the Early arrival speakers displayed production patterns similar to that of the monolingual English controls i.e., they produced the English vowel contrasts with distinct acoustic properties.

These findings not only reflect the patterns of L2 production that have been observed in previous L2 research (e.g., Flege et al., 2003; Guion, 2003), they also support the current L2 speech perception models (i.e., PAM, Best, 1995, Best & Tyler, 2007; SLM, Flege, 1995; Perceptual Interference account, Iverson et al., 2003). In line with Best's PAM (Best, 1994, 1995; Best & McRoberts, 2003), the L2 Sylheti speakers assimilated the English /iː/-/ɪ/ contrast to the single Sylheti /i/ vowel category (see also, Iverson et al., 2003). Further, as suggested by the SLM (Flege, 1995), the existence of these L1-L2 interactions was directly related to the age of L2 acquisition. The Early arrival speakers and Second-Gen speakers started to acquire the L2 early in development, when their L2 categories were still developing, making it easier to acquire new L2 sounds, and in turn reorganize their perceptual space. In contrast, the Late arrival speakers will have approached the task of acquiring English with their established L1 categories, making it harder to acquire the English contrasts.

This study also provided an interesting insight into the second language learners' production of the L1, which has often been ignored in previous research (but see Baker & Trofimovich, 2005). The findings demonstrated differences between speaker groups L1 production, in particular for plosives. The Late arrival speakers used Sylheti (L1) VOT values that reflected that of the Sylheti monolingual controls. In contrast, the Second-Gen speakers used a longer VOT i.e., closer to their English production. The Early speakers used in-between voicing patterns i.e., a longer VOT than did the Late speakers, but shorter than the Second-Gen speakers. Investigating both languages provides a greater understanding of how these speakers organize their phonetic categories. Interestingly, the cross-language comparison between the speakers' production of Sylheti and English revealed that although the Late speakers did not use native-like English voicing patterns, they still produced distinct contrasts from some plosive categories i.e., alveolar/post-alveolar. Although all speakers distinguished the voiceless stops, the Early arrival speakers were shown to make the largest acoustic distinction between their Sylheti and English production. One possible explanation for the Early speakers pattern of production is the monolingual Sylheti experience in Bangladesh during childhood, combined with the early acquisition of English, plus the continued used of both languages. In contrast, the Late speakers used predominately Sylheti and the Second-Gen speakers mainly used English.

In sum, the findings from the study reported in Chapter 2 showed that within the London- Bengali community, and possibly in similar immigrant communities, a large variability in speech production patterns of both the L1 and L2 is likely to be observed. Further, this likely depends on the speaker's age of acquisition and use of their L1 and L2.

8.2. Developmental patterns: sequential bilinguals

Chapters 4 and 5 investigated the acquisition of English vowels and plosives by Sylheti-English sequential bilinguals from the London-Bengali community. To track their development the children were tested and compared to their monolingual peers at two time points: during their first year of full-time English education (nursery, mean 54.7 months), and around 12 months later when the children were in the first year of Primary school.

From a general development perspective, the findings are in line with previous monolingual research that have shown that perception and production abilities continue to develop beyond the first year of life (Hazan & Barrett, 2000; Mayo & Turk, 2005; Sundara et al., 2006). Similar to the pattern reported in Hazan & Barrett's study on phonemic acquisition (2000), all of the children in this research displayed an increase in their English voicing categorization slopes as well as vowel identification. Although a comparison with adult listeners was a not a direct aim of the research reported in this thesis, monolingual English adult control data was collected in Chapter 5 (acquisition of voicing). Observation of the data suggests that by Time 2 (i.e., mean age 64.7 months), the children had categorization slopes closer to that of the adults than they did at 54.7 months of age. These findings support previous research and developmental theories that suggest that perceptual refinement continues beyond the first year of life (i.e., Nittrouer & Miller, 1997; Sundara et al., 2006). For production, smaller changes were observed between Time 1 and Time 2 for the monolingual children, mainly in the production of the back vowels (i.e., FOOT fronting).

The main aim of the longitudinal study was to investigate the acquisition of English contrasts by the sequential bilingual Sylheti-English children. In general, the findings at Time 1 suggested an influence of the bilingual children's L1 on their perception and production of English phonemes. Specifically, at Time 1, the bilingual children had significantly lower vowel identification scores and shallower voicing categorization slopes than did their monolingual English-speaking peers. For vowel perception, the results showed that the bilingual children consistently confused test English vowel contrasts: /i/-/ɪ/, /ʌ/-/ɑː/ and /uː/-/ʊ/, but not the control pairs e.g., /ei/-/i/. In contrast, the monolingual English-speaking children consistently identified all of the English contrasts. For voicing categorization, the bilinguals were less consistent at labeling the English *pea-bee* and *goat-coat* continua than the monolinguals. In turn they displayed significantly shallower categorization slopes than did the monolinguals, suggesting that they had less refined phonemic categories for the target English voicing contrast.

For production, fewer differences were found between the bilinguals and monolinguals. The children displayed distinct vowel categories for the contrasts, albeit slightly different from the monolinguals (i.e., the bilinguals did not display FOOT fronting). For plosives, significant differences between the bilinguals and monolinguals voicing patterns were found. That is, for voiced plosives /b/ and /g/, the bilingual children used a significantly shorter VOT, including prevoiced tokens, than did the monolinguals, i.e., Sylheti-like voicing patterns. As for the vowels, the children still used VOT to make the voicing distinction e.g., /p/-/b/, yet they displayed differing patterns from their monolingual peers.

As hypothesized in Chapter 1, a possible explanation for these findings is that, like the monolingual English children, the sequential bilingual children, having been exposed to no more than 20% English, will have initially started to develop phonemic categories that reflect Sylheti (see e.g., NLM-e, Kuhl & Iverson, 1995; Kuhl et al. 2008). At the beginning of the study the children were an average of 54.7 months old, and therefore at a more advance stage of language development than many of the infants in previous monolingual and bilingual phonetic development studies (e.g., Sebastián-Gallés & Bosch, 2009; Garcia-Sierra et al., 2011; Sundara & Scutellaro, 2011, but see, Sundara et al., 2006 for a study on older children).

In addition to having more developed L1 phonetic categories than infants, it is likely that the children in this research would have had a developing Sylheti vocabulary, and in turn have started to develop more robust Sylheti phonemic categories (see e.g., PRIMIR, Curtin et al., 2011; Werker & Curtin, 2005). It could be suggested that the bilingual children's developing L1 phonology interfered with their perception and production of the L2. This resulted in the children having difficulties in identifying L2 English contrasts e.g., /i/-/ɪ/ that do not have an exact phonetic counterpart in their L1 (see e.g., SLM, 1995). These children assimilated their existing and more robust Sylheti categories e.g., /i/ to the English /iː/-/ɪ/ contrast (see e.g., PAM, Best 1995). These patterns of perception and production are similar to what might be expected for an adult L2 learner (e.g., Flege et al., 2003).

It is important to note however that the monolingual and bilingual children in this research did not differ on all aspects of phoneme development. Specifically, few differences between the bilinguals and monolinguals were observed for vowel production at Time 1 and Time 2. These findings are in line with previous studies that have investigated L2 vowel perception in children (Baker et al., 2008; Tsukada et al., 2005). These studies showed no difference in L2 vowel production of child learners and their monolinguals peers, but they had significantly lower vowel perception scores than did their monolingual peers, suggesting that L2 vowel categories are acquired in production before perception. Further, the vowel perception data reported in Chapter 5 shows that even at Time 2 some differences between the bilinguals and monolinguals remained. The bilingual children had significantly lower vowel identification scores than did the monolinguals for some vowel contrasts.

One possibility is that this pattern could be related to the children's development of their L1 when they started to acquire English. Based on previous monolingual research that has shown that vowel categories develop earlier than consonants (e.g., Kuhl, 2007), as well as studies that suggest that voicing lead (as in Sylheti) may take longer to develop than the short-long lag VOT (Kewley-Port & Preston, 1974; Zlatin & Koenigsknecht, 1976), it could be suggested that the children's Sylheti vowel categories were more stable and established than their Sylheti voicing categories at Time 1. This would make it harder for the children to acquire the English vowel contrasts.

After an average of 12 months of full-time English experience in school the bilingual children demonstrated a significant increase in perceptual acuity for the

English vowel contrast and voicing categorisation. Some changes were also observed in the children's production of voiced plosives. As a result, the children's vowel identification and voicing categorization slopes increased significantly, so much so that they were either significantly closer to their monolingual peers, or they were no longer significantly different. Likewise for production, the children displayed a significant increase in VOT in their production of voiced plosives, and were no longer significantly different from their monolingual peers. For vowel production, although the bilinguals displayed a significant increase in the F2 dimension for the /ʊ/ (i.e., FOOT fronting), a difference between the monolinguals and bilinguals remained for the back vowel. The monolinguals still had a higher F2 than the bilinguals. However, these differences may reflect a variation that is related to Multicultural London English (Cheshire et al., 2011).

A significant finding in this research is that, although all of the children displayed an increase in perceptual acuity from Time 1 to Time 2, the bilingual children displayed a greater increase in English perception abilities than did their monolingual peers. That is, with as little as 12 months of English experience in school, these children displayed a sharp increase in perception, so much so that they either no longer differed from their monolingual peers in some aspects of speech production and perception, or had significantly reduced the gap between them. From an L2 perspective, these findings support Flege's Speech Learning Model (Flege, 1995, Flege et al., 2003) that posits that the ability to acquire L2 sounds remains intact throughout life, and age affects arise because L1 categories become more established with age (see also, Iverson et al., 2003). Unlike adult L2 learners, the children in this study will have still been developing their L1 Sylheti categories (see e.g., Hazan & Barrett, 2000). In turn, their L1 categories are likely to have been less established and robust than that of adult learners, making it easier to acquire new L2 categories.

The sharp increase in perception and changes in production could be attributed to an overall development in English. For example, some theories of speech development (e.g., PRIMIR, Curtin et al., 2011; Werker & Curtin, 2005), have suggested that child's growing lexicon plays a role in establishing phonemic categories. Specifically, the PRIMIR model of speech development (Curtin et al., 2011; Werker & Curtin, 2005) holds that infant and toddlers phoneme categories are stored in a specific 'Phoneme space'. As a child increases their word-meaning associations, phonemes begin to emerge in the Phoneme space. As vocabulary increases the children's phoneme categories become more refined. The latest version of the PRIMIR model (Curtin et al., 2011) suggests that differences between monolinguals may arise because bilingual children's vocabulary size in each language is likely to be relatively smaller than that of their respective monolingual peers. Although vocabulary was not directly tested in this thesis, it is likely that the bilingual children's development in English vocabulary from Time 1 to Time 2 would have had facilitated the perceptual refinement of English phonemes. As the children gain more experience with English they will have started to expand their English vocabulary, which in turn places greater demands on having distinct phonemic categories.

Although not a direct aim of this thesis, the findings from the voicing acquisition experiments suggest some link between speech perception and production. The link between perception and production is complex and not fully understood, and even more so for children. Moreover, understanding any link between perception and production is difficult, particularly as the data sets are confounded by the different techniques and task demands used to investigate production and perception (see e.g., Werker & Curtin, 2005). However, the data from the current study, from the same children over a period of one year, suggests that there is at least some relationship between perception and production. Like studies of second language acquisition in adults, the results reported in this thesis suggest that perception accuracy is correlated with production accuracy (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997): As the bilingual children's representations became more refined (i.e., steeper categorization slope), they produced more native English-like plosives.

8.3. The role of input and sociocultural factors

In addition to overall group differences between the monolingual English children and the Sylhet-English speaking sequential bilinguals, the experiments conducted in Chapter 6 and 7 provided some insight in to the influence of language exposure, social networks (Chapter 6) and caregiver speech (Chapter 7). The results reported in these experiments are in line with previous bilingual research that has investigated the influence of language exposure on bilingual children's language outcomes (De Houwer, 2009; Hoff, 2006, Hoff et al., 2012; Place & Hoff, 2011). Specifically, the analyses in Chapter 6 showed that children who were exposed to more English before entering nursery had speech perception and production outcomes closer to that of their monolingual peers. Interestingly, all children had had a maximum of 20% English exposure from their main caregivers before entering nursery. Thus, the differences that were found were due to the variations in input within this 20% of the English exposure. Such findings are particularly interesting the light of recent bilingual research that suggests that infants are particularly sensitive to small differences in the ambient input during the first year of life (Garcia-Sierra et al., 2011). The findings from Chapter 6 suggest that even small differences amounts of language exposure (i.e., differences within 20% English) can facilitate the acquisition on L2 phonemes.

These findings also raise questions regarding how bilinguals have been categorized in previous research. To date, simultaneous bilinguals are generally categorized in terms of the specific proportions of each language e.g., 40% of one language and 70% of the other (e.g., Bosch & Sebastián-Gallés, 2003). Although it is well accepted that true "equal" simultaneous bilinguals (i.e., 50/50) do not exist (see e.g., Genesee, 2006 for discussion), the variations in language proportions within and between studies may potentially explain some of the differences found between studies. It could be suggested that variability in input to simultaneous bilingual infants e.g., 70% language one vs. 30% language one, may indeed result in significantly differing patterns of development in the two languages being learned. Although more recent bilingual studies on speech perception development have attempted to account for the exact amount of input (e.g. Garcia-Sierra, 2011), the findings in this research have shown small differences in language exposure may be related to the children's outcomes, and should therefore be accounted for in future bilingual research and developmental theories. Further, with the advances in new technologies e.g., the LENA system, that are able to directly records and analyze the child's language input, such detailed analysis of language exposure is possible.

The final experiment reported in Chapter 7 aimed to explore the role of input in greater detail. In this experiment the relationship between the acoustic-phonetic features of the caregivers' English (L2) speech production the children's acquisition of English phonemes was investigated. Similar to the findings from Chapter 2 (adult production study), the analysis of the caregivers' speech revealed variability in their production. Some caregivers produced more Sylheti-accented variants whereas others had more native English-like speech production patterns. An analysis comparing the children of high-accented vs. low-accented caregivers' revealed significant differences between children. In general, children of high-accented caregivers had lower English vowel identification scores and shallower voicing categorization slopes. Fewer differences were found between the children for production.

These findings are in line recent monolingual infant research that showed that infants are sensitive to fine-grained differences in caregiver speech (e.g., Cristià, 2011). The analyses reported in this thesis suggest that similar perceptual sensitivity might be present in bilingual infants. In a bilingual environment this may mean that children who receive less accented input from main caregivers are able to develop more accurate L2 categories than those of high L2-accented caregivers. Although further research is needed to tease apart the possible effect of accented input vs. L2 exposure, this experiment provides a stepping-stone for future, more controlled, research studies. For example, an experiment that controls for L1 and L2 exposure, with main caregivers with varying degrees of accent.

8.4. Implications for theories of speech development

Overall, the findings from this thesis contribute to current theories of speech development in three main ways: 1) The results have implications beyond the first year of life, 2) The results suggest a link between speech perception and production during development, 3) The results extend current theories of bilingual development (i.e., NLM-e, Kuhl et al., 2008; PRIMIR, Werker & Curtin, 2005) to patterns of development in sequential bilinguals. Further, they provide a baseline of the influencing factors in bilingual development i.e., social networks, language exposure, and caregiver accent

To date, the major theories of speech development primarily focus on development within the first year of life (i.e., NLM-e, Kuhl et al., 2008; PRIMIR, Werker & Curtin, 2005). The findings from this research shed light on the stages of development beyond the first year of life. In line with the few studies on childhood speech development (e.g., Hazan & Barratt, 2000; Nittrouer, 2005), the current findings from the monolingual and bilingual group suggest that this developmental process continues into childhood, where children continue to refine their phonemic categories for both perception and production. The NLM-e (Kuhl, 2008) claims that in the final phase of development in infancy the infants' phoneme representations are relatively stable, yet they still allow for new phonemic category learning. The findings from the child study (Chapter 4 & 5) extend this stage of development beyond infancy. Although infant research has shown that native language phoneme categories are relatively stable by the end of the first year of life, the process of refinement of these phoneme categories is likely ongoing into childhood. Furthermore, although further research is needed, the findings from this thesis suggest that the pattern of phoneme refinement is evident in both perception and production. With further research, developmental theories could be extended to include the link between perception and production.

Of the current theories of speech development, only PRIMIR-*in focus* (Curtin et al, 2011) considers bilingual development, namely simultaneous bilingualism. To date, no theory has accounted for sequential bilingual development. The PRIMIR-*in focus* model claims that bilingual infants posses the same representation spaces, dynamic filters, and learning mechanisms as monolinguals. However bilingual development utilizes a Comparison-Contrast mechanism. This mechanism combined with the multidimensional space described in PRIMIR (Werker & Curtin, 2005) allows the bilingual infant to organise the sounds of his/her languages during the first year of life. However, the case of sequential bilinguals is different. As shown in this thesis, the pattern of language exposure means that the sequential bilinguals are not exposed to their second language until later in childhood. The findings from this research have shown that children are able to successfully acquire the sounds of a second language after infancy. It could therefore been suggested that the Comparison-Contrast mechanism is present for all children and remains intact to enable phoneme acquisition later in childhood. As with monolinguals, sequential bilinguals develop a Phoneme space that is in line with their L1, as exposure to the L2 increase the Comparison-Contrast mechanism is activated to organise the phonemes of the two languages. To fully understand how sequential bilinguals organise their phoneme categories, further research is needed on both the L1 and L2 of sequential bilinguals. Further, it may be the case that as L2 dominance increases the children's Phoneme space is not equally weighted.

In addition, the findings from this thesis suggest that other factors such as the children's social networks, language exposure and parent accent play a role in phoneme acquisition. Although PRIMIR-*in focus* holds that language dominance likely plays a role in the development of bilingual infants Word-Form space, the exact role of language dominance and other input factors (e.g., caregiver accent) is not well defined. In the case of sequential bilinguals, findings from this thesis suggest that bilingual's phoneme development is in part influence by the differences in the language environment. For example, the findings in Chapter 2 demonstrate that children who grow up in such communities are likely to be exposed to accented variants depending on the background of their main interlocutors. The findings from Chapter 7 suggested that sequential bilinguals phoneme categories are likely to be influenced by the variability in their input. Therefore, the PRIMIR-*in focus* model could be extended to incorporate the influence of such factors of phoneme development. Specifically, to include a language exposure mechanism that accounts for the amount and quality (i.e., level of accent) of L1 and L2 input that bilinguals are exposed to. In turn, this would influence the organization of the children phonemic categories in their Phoneme space.

8.5. Limitations and next steps

This research has provided some insight to the developmental trajectories of sequential bilingual children's acquisition of English phonemes. Further, the additional analyses of the children's language exposure patterns have provided possible explanations for the individual differences in development. Yet, a limitation of this research is the lack of investigation of the bilingual children's L1. It could be suggested that tracking one language creates the risk of misinterpreting the child's developmental patterns i.e., they may organise their perceptual space in such a way that accounts for both languages. Although investigating both of the children's language was beyond the scope of the current thesis, future research should attempt to investigate the development of both the L1 and L2. Further, a growing body of research in the area of monolingual phoneme development has suggested that phonemic perception and production is related to later language learning (e.g., Werker & Yeung, 2005). For example, to successfully acquire both languages, bilingual children must also learn how the speech sounds combine in both languages i.e., the phonotactic regularities. Such skills have been linked to later language development for monolingual children (Havy, Bertoncini, Nazzi, 2011). Follow up studies will aim to investigate the bilingual children's sensitivities to L1 and L2 sound patterns, and how they map this skill onto later language development.

8.6. Conclusions

The research reported in this thesis has provided the first detailed analysis of sequential bilingual children's phoneme acquisition. Not only do the findings have implications for children who grow up in the London-Bengali community, they also provide a baseline for the possible developmental patterns for children who grow up in similar immigrant communities. From a theoretical viewpoint, the findings from this thesis contribute to current models of speech perception development. From a research perspective, the findings reported in this thesis provide a starting point for further research on the speech and language development of Sylheti-English sequential bilinguals. In addition, this thesis has provided the first known phonetic description of Sylheti.

From a clinical and educational perspective, understanding the language development of children who grow up in immigrant communities is particularly pertinent given that recent reports have suggested that children from an immigrant background have poorer language outcomes than their monolingual peers (PISA, OECD, 2011). Thus, not only do these findings provide a baseline for typically developing sequential bilingual development, the tests and interview techniques provide a starting point for the development of clinical tools that can used by speech therapists who work in multicultural cities such as London.

References

Adams, C. (1987). *Across Seven Seas and Thirteen Rivers*. London, Tower Hamlets Art Project (THAP) Books.

Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099.

Agar, M. (1996). *The Professional Stranger: An Informal Introduction to Ethnography*. London, Elsevier Academic Press Inc (London) Ltd

Alam, F., Habib, S. M. M., & Khan, M. (2007). *Acoustic analysis of Bangla consonants*. BRAC University CRBLP Technical report.

Allen, G. and Hawkins, S. (1980) Phonological rhythm: definition and development. In G. Yeni-Komshian, J. Kavanagh and C. Ferguson (Eds), *Child Phonology*, Volume 1 (New York: Academic Press), pp.227–256.

Anderson, J. L. Morgan, J. L. & White, K. S. (2003) A statistical basis for sound discrimination. *Language and Speech*, 46, 155-182

Anwar, M. (1979). *The Myth of Return*. London, Heinemann.

Agar, M. (1996). *The Professional Stranger: An Informal Introduction to Ethnography*. London, Elsevier Academic Press Inc (London) Ltd

Asby, P (1995) *Speech Sounds*. Language workbooks: Routledge

Aslin, R. & Pisoni, D. (1980). Some developmental processes in speech perception. In G. Yeni-Komshian, G., J. Kavanagh, J. & Ferguson, C. (eds.). *Child Phonology, vol. 2: Perception*. New York: Academic Press

Bialystok E. 2009. Effects of bilingualism on cognitive and linguistic performance across the lifespan. In I. Gogolin I. & U. Neumann (eds) *Streitfall Zweisprachigkeit – The bilingualism controvery* (pp. 53-68). Wiesbaden, Germany: VS Verlag fur Sozialwissenschaften

Baker, W. & Trofimovich, P. (2005). Interaction of native- and second- language vowel system(s) in early and late bilinguals. *Language and Speech*, 48 (1), 1-27

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child--Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity. *Language and Speech*, *51*(4), 317–342.

Bankson, N. and Bernthal, J., (1998), Analysis and interpretation of assessment data. In J. Bernthal and N. Bankson (eds), *Articulation and Phonological Disorders* (pp.270–298. Boston), MA: Butterworth-Heinemann

Bermant, C. (1975). *Point of Arrival*. London, Eyre Methuen.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*. (pp. 289-304). Dordrecht, The Netherlands: Kluwer Academic.

Best, C. T. (1994) The emergence of native-language specific constraints in perception of non-native speech: A perceptual assimilation model. In J. Goodman, & H. C. Nusbaum (eds). *The development of speech perception: The transition from speech to spoken words*. Cambridge. MIT Press

Best, C. T., (1995) A direct realist view of cross-language speech perception. In W. Strange (Ed) *Speech Perception and Linguistic Experience: Issues in Cross-language research*, pp. 171-206. Baltimore, MD: York Press,

Best, C. T, & McRoberts, G. (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., Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775–794.

Best, C. T., McRoberts, G., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two non-native consonant contrasts. *Infant Behaviour and Development*, 1, 351-353.

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

Best, C.T., & Tyler, M. D. (2007) Nonnative and second-language speech perception: commonalities and complementarties. In M. J. Munro & O. –S. Bohn (eds) *Second language speech learning: The role of language experience in the speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science, 20*, 539-542.

Boersma, P., Escudero, P. & Hayes, R.A. (2003) Learning abstract phonological form auditory phonetic categories: An integrated model of acquisition of language specific sound categories. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 249-278. Adelaide, Australia

Boersma, P. & Weenink, D. (2012). Praat: Doing phonetics by computer. Retreived from http://www.praat.org

Bohn, O. -S., & Flege, J. E, (1990) Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11, 303-328

Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants : is the distributional account all there is to it ? *Developmental Science*, 6, 874–887.

Bosch, L., & Sebastian-Galles, N. (2001) Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy,* 2, 29-49

Bosch, L., & Sebastian-Galles, N. (2003a). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, *46*(2-3), 217–243.

Bosch, L. & Sebastián-Gallés, N. (2003b). Language experience and the perception fo a voicing contrast in fricatives: Infant and adult data. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1987-1990). Barcelona.

Bradley, R. H. & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371-399

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B, and Tohkura, Y. (1999) Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61 (5) 977-985

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6,* 201-251.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, *7*(03), 207.

Burns, T. C., Yoshida, K.A., Hill, K., Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics,* 28, 445-474

Byrne, D., David, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R. et al. (1994) An international comparison of long term international speech spectra. *Journal of the Acoustical Society of America*, 96 (4), 2108-2120

Camden Key Facts (2012). see: http://tinyurl.com/d6jplw6 Figures retrieved: 4 April 2012

Chalmers, R. & Miah, N. (1996). *Learning Sylheti*. London: The Centre for Bangladeshi Studies

Cheshire, J., Fox, S., Kerswill, P. and Torgersen, E. (2008) Ethnicity, friendship network and social practices as the motor of dialect change: linguistic innovation in London. *Sociolinguistica* 22: 1-23

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics*, *15*(2), 151–196.

Conboy, B.T.; & Mills, D.L. (2006). Two languages, one developing brain: Event-related potentials to words in bilingual toddlers. *Developmental Science*, 9, (1), F1-F12

Conboy, B. & Kuhl, P. K. (2007). Listening to language in culturally specific ways: Learning through social experience. In S. Braten (ed.), *Foundations of Preverbal Intersubjectivity: Being Moved by Action-Perception, Music and Speech*. Cambridge University Press: Cambridge, England

Cristià, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of the Acoustical Society of America*, *129*(5), 3271–80.

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*(4), 492–504.

Dalston, R. M. (1975) Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *The Journal of Acoustical Society of America*, 57, 462 – 469

Darcy, I., & Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics*, *40*(4), 568–581.

Dasgupta, P. (2003) Bangla. In G. Cardona & D. Jain (eds.) *The Indo-Aryan languages* (pp. 551-390). London: Routledge

Davis, K. (1994). Stop voicing in Hindi. *Journal of Phonetics*, 22, 177-193

Davis, K. (1995) Phonetic and phonological contrasts in the acquisition of voicing: voice onset time production in the acquisition of Hindi and English. Journal of *Child Language*, 22, 351-365

Davis, B. L., & MacNeilage, P. F. (1995). The Articulatory Basis of Babbling. *Journal of Speech and Hearing Research,* 38, 1199-1211.

de Boysson-Bardies, B. (1993) Ontogeny of language-specific syllabic productions. In B. de Boysson-Bardies, S. Schonen, P. Jusczyk; P. MacNeilage, & J. Morton. *Developmental Neurocognition: speech and face processing in the first year of life*. pp. 353-363. Dordrecht, The Netherlands: Kluwer Academic de Boysson-Bardies, B., Hallé, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16, 1–17

de Boysson-Bardies, B., Vihman, M. M., Roug-Hellichius, L., Durand, C., Landberg, I., & Arao, F. (1992). Evidence of infant selection from target language: A crosslinguistic phonetic study. In C. A. Ferguson & L. Menn & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications*. Monkton, MD: York Press.

De Houwer, A., (2007) Parental language input patterns and children's bilingual use. *Applied psycholinguistics* 28(3): 411-424

De Houwer, A. (2009). Bilingual first language acquisition. Bristol: Multilingual Matters.

De Houwer, A. (2011) Language input environments and language development in bilingual acquisition. *Applied Linguistics Review*, 2: 221-240.

De Houwer, A., M. H. Bornstein & S. De Coster (2006). Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism* 10(3), 331-348.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin: Foris

Dodd, B., (1995). *Differential Diagnosis and Treatment of Children with Speech Disorder*. London: Whurr.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British-English speaking children. *Clinical linguistics and phonetics*, 17 (8), 617-643.

Eade, J. (ed.) (1997). *Living the Global: Globalization as local process*. London, Routledge.

Eade, J., Vamplew, T, Peach, C. (1996) The Bangladeshis: the encapsulated community. In C. Peach (ed) *Ethnicity in the 1991 Census*. London. HMSO

Eguchi, S., & Hirsh, I. J. (1969) Development of speech sounds in children. *Acta Oto-Laryngologica, Sumplementum*, 275, 1-51

Eilers, R. E., Wilson, W. R. & Moore, J. M. (1979) Speech discrimination in the language innocent and the language-wise: a study in the perception of voice onset time. *Journal of Child Language*, 6, 1-18

Eimas, P. D., (1975) Speech perception in early infancy, In L.B. Cohen & P. Salapetex (eds). *Infant perception: From sensation to cognition*. New York: Academic Press

Eimas, P.D., Siqueland, E. R., Jusczyk, P., and Vigorito., J. (1971) Speech perception in Infants. *Science*, 171, 303-306

Elliot, C. D. (1996) *British Ability Scales: 2nd Edition*. GL Assessment, UK

Escudero, P., & Boersma, P. (2004). L2 speech perception research and phonological theory. L2. *Learning*, 551–585.

Fabiano-Smith, L., & Goldstein, B. A. (2010) Early-, middle-, and late-developing sounds in monolingual and bilingual children: An exploratory investigation. *American Journal of Speech-Language Pathology*, 19, 66-77

Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2013). Using speech sounds to guide word learning: the case of bilingual infants. *Child Development*, *78*(5), 1510–25.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories : User's guide and technical manual*. San Diego: Singular Publishing Group, Inc.

Fernald, A. (2006). When infants hear two languages : interpreting research on early speech perception by bilinguals. In P. McCardle & E. Hoff (eds), *Childhood bilingualism: Research on infancy through school-age* (pp. 19-29). Clevedon, UK: Multilingual Matters.

Flege, J. E. (1991) The interlingual identification of Spanish and English vowels: orthographic evidence. *The Quarterly Journal of Experimental Psychology*. 43A, 3, 701-731

Flege, J. E. (1995). Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language research*. (pp. 229-273) Timonium, MD: York Press.

Flege, J. E., Frieda, E. M. & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25, 169-186

Flege, J. E., & Lui, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527-552.

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

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973–87.

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

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

Flege, J. E., Yeni-Komshian, G. & Liu, S (1999). Age constraints on second language learning. *Journal of Memory and Language*, 41, 78-104.

Forman, C. (1989). *Spitalfields: A Battle for Land*. London, Hilary Shipman.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*, 409–438.

Fox, S. (2007). *The demise of the Cockneys? Language change in London's "traditional" East End*. (Unpublished doctoral dissertation). University of Essex, UK.

Fox, S. (2010). Ethnicity , religion, and practices: Adoclescents in the East End of London. In C. Llamas & Watt, D (eds) *Languages and Identities* (pp. 144-156). Edinburgh: Edinburgh University Press

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., et al. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*(4), 546–557.

Genesee, F. (2006) Bilingual first language acquisition in perspective. In P. McCardle & E. Hoff (eds), *Childhood bilingualism: Research on infancy through school-age* (pp. 45-67). Clevedon, UK: Multilingual Matters.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual review of psychology*, *61*, 191–218.

Gibson J.J. & Gibson, E.J. (1955) Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62, 35-41

GLA (2011) *Population projections 2011: Camden*. Greater London Authority, UK

Goldstein, L.G., & Fowler, C.A. (2003). Articulatory Phonology: A phonology for public language use. In A. S. Meye & N.O. Schiller (eds) *Phonetics and phonology in language comprehension: Differences and similarities* (pp. 159-207). Berlin: Mouton de Gruter

Gregory, E (1998) Siblings as mediators of literacy in linguistic minority communities. *Language and Education* 12 (33) p54

Grierson, G. A. (1903). *The Linguistic Survey of India (vol. 5): Indo-Aryan family Eastern group part 1: Specimens of the Bengali and Assamese languages*. Calcutta: Office of the Superintendent of Government Printing, India.

Guion, S. G. (2003). The vowel systems of Quicha-Spanish Bilinguals. *Phonetica*, 60, 8-128

Guion, S. G., Flege, J. E., & Loftin, J. D. (2000). The effect of L1 use on pronunciation in Quichua–Spanish bilinguals. *Journal of Phonetics*, *28*(1), 27–42.

Hai, H. M. (1960) *A phonetic and phonological study of nasals and nasalzation in Bengali*. Dacca. Dacca University Press.

Hamilton, A., Plunkett, K., & Schafer, G., (2000). Infant vocabulary development assessed with a British Communicative Development Inventory: Lower Scores in the UK than the USA. *Journal of Child Language*, 27, 689-705

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*, *125*(1), 469–79.

Havy, M. Bertoncini, J. Nazzi, T. (2011) Word learning and phonetic processing in preschool-age children. *Journal of Experimental Child Psychology*, 108, 25-43

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

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6 - 12. *Journal of Phonetics*, 28 (4), 377–396.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: is there any evidence for a true deficit? *Journal of speech, language, and hearing research : JSLHR*, *52*(6), 1510–29.

Hazan, V., Senema, A., Faulkner, A. (2004) Effect of intensive audiovisual perceptual training on the perception and production of the /l/-/r/ contrast for Japanese learners of English. *Proceedings of the International Conference for Spoken Language Processing*, Jeju, October 2004. Vol.2. 1145 -1148

Heselwood, B. & McChrystal, L. (1999). The effect of age-group and place of L1 acquisition on the realisation of Punjabi stop consonants in Bradford: an acoustic sociophonetic study. *Leeds working papers in linguistics*, 49-69.

Heselwood, B. & McChrystal, L. (2000). Gender, accent features and voicing in Panjabi-English bilingual children. *Leeds working papers in linguistics*. *8*, 45-70.

Hillenbrand, J. M., Clarke, M. J. & Neary, T. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109 (2), 748-763.

Hirson, A. & Sohail, N. (2007). Variability of rhotics in Punjabi-English bilinguals. In Jürgen Trouvain and William Barry (eds.) *Proceedings of the 16th International Congress of Phonetics Sciences* (ICPhS) Dudweiler ,Germany. Pirrot GmbH. 1501-1504

Hoff, E., 2003. The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5), pp.1368-78.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55-88.

Hoff, Erika, Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27.

Hoff-Ginsberg, E. (2008). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, *19*(04), 603.

Holm, A., Dodd, B. & Hua, Z. (2002) *Diagnotic Evaluation of Ariculation and Phonology (DEAP).* Pearson.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, 119, 3059–3071

Howell, P., Rosen, S., Lang, H., & Sackin, S. (1992). The role of F1 transitions in the perception of voicing in initial plosives. In, *Speech, Hearing and Language: Work in Progress*. Vol. 6. University College London.

Husain, J. (1991). The Bengali speech community. In S. Alladina & V. Edwards (eds) *Multilingualism in the British Isles,* p 75-87. Longman

Iverson, P., Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America* 122(5), 2842-2854

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

Jackson-Maldonado, D., Thal, D. J., Fenson, L., Marchman, V., Newton, T. & Conboy, B. (2003). *El inventario del desarrollo de habilidades comunicativas: User's guide and technical manual*. Baltimore: Paul H. Brookes.

Johnson, C. & Wilson, I. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, 6, 271-289

Jusczyk, P. W. (1993). From general to language specific capabilities: the WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3-28

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology*, 23, 648–654.

Kehoe, M. M., Lleó, C. & Rankow, M. (2004). Voice onset time in bilingual German-

Kent, R. D., & Bauer, H. R. (1985). Vocalizations of one-year olds. *Journal of Child Language,* 12, 491-526.

Kern, S. & Davis, B. L. (2009). Emergent complexity in early vocal acquisition: Cross-linguistic comparisons of canonical babbling. In Chitoran, I. Coupé, C., Marsico, E. & Pellegrino, F. *Approaches to Phonological Complexity*, Phonology and Phonetics Series, Berlin: Mouton de Gruyter. 353-376

Kern, S., Davis, B. L. & Zink. I. (2009) From babbling to first words in four languages: common trends across languages and individual differences. In F. d'Errico, Hombert, J. M. (eds). *Becoming Elloquent: Advances in the emergence of language, human cognition and modern cultures*. John Benjamins

Kewley-Port, D., & Preston, M.S. (1974). Early apical stop production: a voice onset time analysis. *Journal of Phonetics*, 2, 195-210

Kershen, A. J. (2000). Mother tongue, assimilation and the Bangladeshi community in Tower Hamlets. Paper presented at *Bangladeshis in Britain: changes and choices, configurations and perspectives,* London, Guildhall University

Khan, S. U. D. (2010). Bengali (Bangladeshi Standard). *Journal of the International Phonetic Association*, 40(02), 221-225.

Kuhl, P. K. (1993), Innate predisposition and the effects of experience in speech perception: the native language magnet theory. In B. deBoysson-Bardies, S de Schonen, P. Juszczyk, P. McNeilage & J. Morton (eds) *Developmental Neurocognition: speech and face processing in the first year of life*, 258-274. Dordrecht, Netherlands: Kluwer Academic Publishers

Khattab, G. (2000). VOT production in English and Arabic bilingual and monolingual children. *Leeds Working Papers in Linguistics,* 8, 95-122

Klatt, D. H. 1980. Software for a cascade/parallel formant synthesizer. *Journal of Acoustical Society of America*, 67, 737-793.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews. Neuroscience*, *5*(11), 831–43.

Kuhl, P. K. (2007). Is speech learning " gated " by the social brain ? *Developmental Science*, *1*, 110–120.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V, Ryskina, V. L., Stolyarova, E. I., et al. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science*, *277* (5326), 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 of London. Series B, Biological sciences*, *363*(1493), 979–1000.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development : Implications for the " Critical Period ". *Language Learning*, 1, 237–264.

Kuhl, P.K. & Miller, J. D. (1975). Speech Perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72

Kuhl, P. K. & Padden, D. M. (1983) Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, 73, 1003-1010.

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*(2), F13–F21.

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

Kuhl, P. K., Williams, K.A., Lacerda, F., Stevens, K. N. & Lindbolm, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell

Lambert, K. Alam, F. & Stuart-Smith, J. (2007). Investigating British Asian accents: studies from Glasgow. In Jürgen Trouvain and William Barry (eds.) *Proceedings of the 16th International Congress of Phonetics Sciences* (ICPhS)

Lee, S., Davis, B. L., & Macneilage, P. F. (2008). Segmental properties of input to infants: a study of Korean. *Journal of child language*, *35*(3), 591–617.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech : Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, *105*, 1455–1468.

Lieberman, P. (1980). On the development of vowel production in young children. In G. H. Yeni-Komshian & J. F. Kavanagh & C. A. Ferguson (Eds.), *Child phonology 1: Production*. New York, NY: Academic Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967) Perception if the speech code. *Psychological Review*, 74, 431 – 461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Levitt, H. (1971). Transformed up-down methods in psycho- acoustics. *Journal of the Acoustical Society of America*, 49, 467–477

Lisker, L. & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422

Lisker, L. & Abramson, A.S. (1971) Distinctive features and laryngeal control. *Language*, 47, 767-785

Li Wei., Milroy, L. & Pong, S. C. (1992). A two-step sociolinguistic analysis of code-switching and language choice. *International Journal of Applied Linguistics 2*: 63-86

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 49, 606-608

Macken, M. A, & Barton, D. (1980). The acquisition of the voicing contrast in English: study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*(1), 41–74.

MacNeilage, P. F., & Davis, B. L. (1993). Motor explanations of babbling and early speech patterns. In B. de Boysson-Bardies & S. de Schonen & P. Jusczyk & P. F. MacNeilage & J. Morton (Eds.), *Changes in Speech and Face Processing in Infancy: A Glimpse at Developmental Mechanisms of Cognition*. Dordrecht: Kluwer.

MacNeilage, P. F., & Davis, B. L. (2000). On the Origin of Internal Structure of Word Forms. *Science,* 288, 527-531.

MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The Motor Core of Speech: A Comparison of Serial Organization Patterns in Infants and Languages. *Child Development,* (1), 153-163.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, 110, 516-528

Maddieson, I (1984) *Patterns of Sounds*. Cambridge University Press.

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. Yeni-Komshian, G., J. Kavanagh, J. & Ferguson, C. (eds.) (1980). *Child Phonology, vol. 2: Perception*. New York: Academic Press

Masica, C. (1991). *The Indo-Aryan languages*. Cambridge University Press.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, 11 (1), 122–34.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82* (3), B101–11.

Mayo, C., Florentine, M. & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language & Hearing Research*, 40 (3), 686-93

Mayo, C., & Scobbie, J. M. (2003). The Influence of Phonemic Acoustic Cue Weighting Strategies in Children’s Speech Perception. *Hearing Research*, 46, 1184–1196.

Mayo, C., & Turk, A. (2004). Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *Journal of the Acoustical Society of America*, *115*(6), 3184.

McCarthy, K. M. (2009). Accounting for input: the development of a home language environment questionnaire (unpublished Masters thesis, University College London)

McCarthy, K.M., Mahon, M., & Evans, B. (in press) Acquiring a second language in an immigrant community: the production and Sylheti & English stops and vowels by London-Bengali speakers. *Journal of Phonetics*

McCarthy, K.M., Mahon, M., Rosen, S., & Evans, B. (under revision) The development of phonemic categorization by sequential bilingual children: a longitudinal study of VOT acquisition.

McCune, L. & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language and Hearing Research* **44,** 670-684.

Mikuteit, S. & Reetz, H. (2007). Caught in the ACT: the timing of aspiration and voicing in East Bengali, *Language and Speech*, 50, 249-279

Moreiras, C. (2006) *An acoustic study of vowel change in female adult speakers of RP*. Unpublished BSc Thesis, University College London

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

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–20.

Nathan, L., Wells, B., & Donlan, C. (2013). Children's comprehension of unfamiliar regional accents : a preliminary investigation. *Journal of Child Language*, 2, 343-365

Nathan, L., & Wells, B. (2001). Can children with speech difficulties process an unfamiliar accent? *Applied Psycholinguistics*, *22*(3), 343–361.

Nearey, T. M. (1978) *Phonetic Feature Systems for Vowels.* Indiana University Linguistic Club. Indiana.

Nearey, T. M. & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association*, 24 (1), 1-19

Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, *118*(4), 2570.

Nittrouer, S. (1993) The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research*, 959-972.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: evidence from spectral moments. *Journal of the Acoustical Society of America*, *97*(1), 520–30.

Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech and Hearing Research*. 39(2) 278-297

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, *112*(2), 711.

Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *Journal of the Acoustical Society of America*, *118*(2), 1072.

Nittrouer, S, & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, *101*(4), 2253–66.

OFQUAL (2004) National Qualifications Framework: level descriptors. *Office for Qualifications and Examinations Regulation*, U.K

Ohde, R. N. (2012) Formant onsets and formant transitions as developmental cues to vowel perception. The Journal of the Acoustical Society of America 130 (3), 1628-1642

Ohde, R N, & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. *The Journal of the Acoustical Society of America*, *102*(6), 3711–22.

Ohde, R N, & Haley, K. L, & McMahon, C. W. (1996) A developmental study of vowel perception from brief synthetic consonant-vowel syllables. *Journal of the Acoustical Society of America.* 100 (6), 3813-3824

Ohde, Ralph N, & German, S. R. (2011). Formant onsets and formant transitions as developmental cues to vowel perception. *Journal of the Acoustical Society of America*, *130*(3), 1628–42.

ONS. (2009). Office of National Statistics: population and migration. Population estimates by ethnic group: http://www.statistics .gov.uk/statbase/

Palethorpe, S., Wales, R., Clark, J. E., & Senserrick, T. (1996). Vowel classification in children. *Journal of the Acoustical Society of America*, *100*(6), 3843–51.

Pallier, C., Dehaene, S., Poline, J-B, LeBihan, Argenti, A-M, Dupoux, E, & Mehler, J. (2003) Brain imagining of second language plasticity in adopted adults: can a second language replace the first? *Cerebral Cortex*, 13 (2), 155-161

Para, M., Hoff, E, & Core, C. (2011) Relations amoung language exposure, phonological memory, and language development in Spanish-English bilingually developing two-year-olds. *Journal of experimental Psychology*, 108, 113-125

Pearson, B. Z., Fernandez, S. C., Lewedeg, V. & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics* 18, 41–58.

PISA (2011) *Pisa in-focus 11: how are schools adapting to the increasing number of immigrant students*. Organisation for Economic Cooperation and Development.

Place, S. & Hoff, E. (2011). Properties of dual language exposure that influence two-year- olds' bilingual proficiency. *Child Development*. 82, 1834-1849

Polka, L. (1992) Characterizing the influence of native experience on adult speech perception. *Perception & Psychophysics,* 52, 37-52

Polka, L, & Bohn, O-S. (1996) A cross language comparison of vowel perception in English-learning and German-learning infants. *Journal of Acoustical Society of America*. 100, 577-592

Polka, L, & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of experimental psychology. Human perception and performance*, *20*(2), 421–35.

Polka, Linda, Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, *109*(5), 2190.

Poole, I., (1934) Genetic development of articulation of consonants sounds in speech. *Elementary English Review*, 11, 159–161.

Prather, E., Hedrick, D., and Kern, C. (1975) Articulation development in children aged 2 to 4 years. *Journal of Speech and Hearing Disorders*, 40, 179–191

Ramon-Casas, M; Swingley, D; Debastian-Galles, N; Bosch, L. (2009). vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, *59*(1), 96–121.

Ramus, F., Rosen, S., Dakin, S., Day, B.L., Castellote, J.M., White, S. & Frith, U. (2003) Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, 126, 841-865.

Rasinger, S. M. (2007). Bengali-English in East London: a study in urban multilingualism. In Davis, G & Bernhardt, K. A (Eds) *Contemporary studies in Descriptive Linguistics: volume 11*. Peter Lang: Oxford

Ray, P. S., Hai, M. A., & Ray, L. (1966). *Bengali language handbook*. Washington, DC: Centre for Applied Linguistics

Rivera-Gaxiola, M., Silva-Pereyra, J. & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, (8), 162-172

Roug, L., Landburg, I., & Lundburg, L. (1989). phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life. *Journal of Child Language*, 17, 19-40.

Samant, S. (2010). Arab Americans and sound change in southeastern Michigan. *English Today*, 103, 27-33

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental psychology*, *46*(1), 106–19.

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

Schmale, R., & Seidl, A. (2009). Accomodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*, 12, 583-601

Sharma, D. (2011). Style repertoire and social change in British Asian English. *Journal of Sociolinguistics*, 15 (4), 464-492

Sharma, D. & Sankaran, L. (2011). Cognitive and social factors in dialect shift: Gradual change in London-Asian speech. *Language Variation and Change*, 23 (3), 399-428

Singer, J. D. & Willet, J. B. (2003) *Applied longitudinal data analysis*. New York. Oxford University Press

Smit, A., Hand, L., Freilinger, J., Bernthal, J. and Bird, A., (1990), The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55, 779–798.

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

Stoel-Gammon, C. & Dunn, C. (1985). Normal and Disordered Phonology in Children. Austin, Texas

Studdert-Kennedy, M., & Goldstein, L., (2003) Launching language: The gestural origin of discrete infinity. In M. H. Christiansen & S. Kirby (eds). *Language Evolution: The states of the art*. Oxford University Press

Streeter, L.A. (1976) Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* 259, 39-41

Sundara, M. Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism*, *9*(01), 97.

Sundara, M. Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-th/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–88.

Sundara, M. Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, 108, 232-242

Sundara, M. & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505–513.

Sussman, J. E., and Carney, A. E. (1989). “Effects of transition length on the perception of stop consonants by children and adults,” J. Speech Hear. Res. 32, 151–160.

Templin, M., (1957), Certain language skills in children: their development and interrelation-ships. Institute of Child Welfare Monographs, Vol. 26 (Minneapolis, MN: University of Minnesota Press.

Thorndike, E.L. and Lorge, I. (1944). The Teacher's Word Book of 30,000 Words. New York: Teachers College, Columbia University

Torgersen, E., Kerswill, P. and Fox, S. (2006). Ethnicity as a source of changes in the London vowel system. In F. Hinskens (ed.) *Language Variation - European Perspectives*. (pp. 249-263) Amsterdam: Benjamins

Thompson, M. & Hazan, V. (2010). The impact of visual cues and lexical knowledge on the perception of a non-native consonant contrast for Colombian adults. *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds* (p. 493-498).

Tower Hamlets (2012). *Spring School Census 2012*. London Borough of Tower Hamlets

Trofimovich, P., Baker, W., Flege, J. E., & Mack, M. (2003). Second-Language Sound Learning in Children and Adults : Learning Sounds, Words, or Both? In B. Beachley, A. Brown, & F. Conlin (eds.), *Proceedings of the 27th Boston University Conference on Language Development* (p. 775–786) Somerville. MA

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, 75, 1067-1084.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child development*, *75*(4), 1067–84.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, *120*(4), 2285.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263–290.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K.,Kphno, M., Menyuk., and Best, C. (1994) Discrimination of English /r-l/ and /w-y/ by Japaese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. In *Proceesings of the International Conference on Spoken Language Processing, Acoustical Society of Japan,* (4), 1695 -1698

Tunga, S. S. (1995). *Bengali and other related dialects of South Assam*. New Delhi: Mittal Publications

Tyler, M. D., Best, C. T., Goldstein, L. M., Antoniou, M., & Krebs-Lazendic, L. (2008). Six-and Twelve-Month-Olds' Discrimination of Native versus Non-Native Between-and Within-Organ Fricative Place Contrasts. *Ninth Annual Conference of the International Speech Communication Association*.

Vihman, M., (1996) Phonological Development (Oxford: Blackwell).

Vihman, M. M. & Croft, W. (2007). Phonological development: Toward a 'radical templatic phonology. *Linguistics*, 45, 683–725.

Vihman, M. M., Ferguson, C. A., & Elbert, M. F. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics,* 7, 3-40.

Walley, A. & Flege, J. (1999). Effects of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, 27, 307- 332.

Wellman, B., Case, I., Mengert, I. and Bradbury, D., (1931), Speech sounds of young children. University of Iowa study, *Child Welfare*, 5(2).

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

Werker, J. F. & Byers-Heinlein, K. (2008). Bilingualism in infancy : First steps in perception and comprehension. *Trends in Cognitive Sciences* 12, 144–51.

Werker, J. F., Byers-Heinlein, K., Fennell, C. T. (2009). Bilingual beginnings to learning words Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society*, 364. 3649–3663.

Werker, J. F, & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language learning and development*, 1 (2), 197-234

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R.C. (1981) Developmental aspects of cross-language speech perception. *Child Development*, 52 (1), 349-355

Werker, J.F., & Tees, R.C. (1983). Developmental change across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, 37, 278-286

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

Werker, J. F., & Yueng, H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, 9 (11), 519-527.

Werker, J. F., Yeung, H. H., & Yoshida, K. a. (2012). How Do Infants Become Experts at Native-Speech Perception? *Current Directions in Psychological Science*, *21*(4), 221–226.

Wilson, M.D. (1988) The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20(1), 6-11.

Zaltin, M., & Koenigsknecht, R. (1976). Development of the voicing contrast: a comparison of voice onset time in perception and production. *Journal of Speech and Hearing Research*, 19, 93-111

APPENDIX 1a: Adult language background questionnaire

Male/Female: DOB: Place of birth:

Age and date of arrival in UK (if outside the UK): Location (e.g. London):

First language (specify dialect):
Can you write/read in your in this language?

Do you speak any other languages?

Have you attended classes for learning English? If yes, where? How long? To what level?

Current occupation: Language/s used at work:

LANGUAGE USE	English	Sylheti	Both (i.e. Sylheti or English)
Husband/Wife			
Parents			
Children at home			
Children outside the home			
Grandchildren			
Your grandparents or their generation			
Aunts/uncles in UK			
Brothers/Sisters			
Cousins in UK			
Cousins/Family in Bangladesh			
Close friends			
Neighbours			
Colleagues / college friends			
Local shop keeper			
Talking to your doctor			
Children's teachers/your teachers			
Family events			
Writing emails			
Writing letters			
Thinking to yourself			
Counting			

APPENDIX 1b: plosive target sounds and words

Sylheti target sound	Target word (NB: Standard Bengali script)	
/p/	পাতা	‘pata’ (leaf)
/pʰ/	ফুল	‘phul’ (flower)
/b/	বাড়ি	‘bari’ (home)
/bʱ/	ভারী	‘bhari’ (heavy)
/t/	টিয়া	‘tiya’ (bird)
/tʰ/	ঠোঁট	‘thoot’ (lips)
/d/	ডিম	‘dim’ (egg)
/dʱ/	ধোয়া	‘dhowa’ (wash)
/k/	কুকুর	‘kukur’ (dog)
/kʰ/	খাম	‘kham’ (envelope)
/g/	গাল	‘gal’ (cheek)
/gʱ/	ঘাস	‘ghas’ (grass)

English target sound	Target word
/p/	pizza
/b/	book
/t/	teeth
/d/	dog
/k/	cat
/g/	goat

APPENDIX 1c: vowel target sounds and words

Sylheti target sound	Target word (NB: Standard Bengali script)	
/i/	ঈট	‘eet’ (brick)
/ɛ/	দেথ	‘dheka’ (to look)
/e/	দেশ	‘desh’ (country)
/a/	পাত্অ	‘pata’ (leaf)
/ɔ/	দশ	‘dosh’ (ten)
/o/	ছোট	‘choto’ (small)
/u/	ভূত	‘bhut’ (ghost)
English target sound	Target word	
/i/	bead	
/ɪ/	kid	
/ɛ/	bed	
/æ/	cat	
/ʌ/	cup	
/ɛ/	bird	
/ɑ/	card	
/ɒ/	cot	
/ɔ/	court	
/u/	boot	
/ʊ/	book	

DEPARTMENT OF DEVELOPMENTAL SCIENCE

APPENDIX 2a: Home Language Environment Questionnaire 2011-12

***Please note: This questionnaire is interviewer-administered*.**

Participant Identifier No: ______________________________ Interpreter present: ___________________________

Part 1: Background Information

Family's Post code:

Caregiver 1: DOB: Place of birth: Arrival in the UK:
Caregiver 2: DOB: Place of birth: Arrival in the UK:
Caregiver 3: DOB: Place of birth: Arrival in the UK:

Child details

Age:_________________________ DOB: ______________________________ Male/Female

Place of birth:____

If place of birth is outside of the U.K, please state age of arrival: _______________________________________

Has your child ever lived outside of the U.K? If so, where? For how long? _______________________________

Age start nursery_________ months Full time/Part time? If part time when did s/he start full-time: __________

Part 2. Language exposure

What language/s has your child heard since birth? ____

Child's age at first exposure to English: _______ months

Child's age at first exposure to Sylheti: _______ months

2a) PRE EXPOSURE (before starting nursery)

Child's waking hours: Bed time: Wake time

Main caregivers:

Caregiver 1: _________________ Hours spent with child: __________ Where? ______ Language use:

Caregiver 2: _________________ Hours spent with child: __________ Where? ______ Language use

Caregiver 3: _________________ Hours spent with child: __________ Where? ______ Language use:

2b) POST EXPOSURE (after starting nursery)

Child's waking hours: Bed time: Wake time:

Hours spent at: Nursery __________________ Reception class ___________________________________(state days per week____)

Hours spent per day at home: week day _______________ weekend _________________________

Hours spent per day with a child minder (e.g. neighbour after school):

____(state with whom and language/s spoken)

Hours spent with friends/youth group/club outside of school/home ______________________________ (please detail)

Does your child spend time in Bangladesh? If yes, how long ______________________ (in days per year)?

Does child talk on phone to relatives from Bangladesh? If yes, how often? What language?

Do family members visit your home? i.e. cousins from Bangladesh If so for how long? Recently? Language/s used

Social Networks (Based on Li Wei et al 1992; Li Wei & Moyer, 2008, Sharma, 2011,) (COMPLETE ON SEPARATE LARGE SHEET)

Please list all the people your child has regular interactions with:

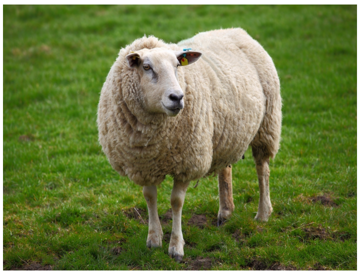
Name	M/F	Relation to child, Bengali?	Place of birth (or arrival in UK, if relevant)	Language/s spoken to child	Hrs spent per wk? (in person? phone?	Diversity index: In what context do they interact? E.g. home, school, playgroup	Diversity index: link people in the child's network: do you share friends/family (list names?

Part 3. Caregiver Education & Occupation:

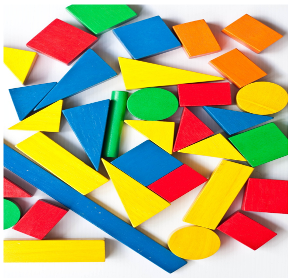
	Mother	Father	Siblings	Other relatives
Completed primary				
High/secondary school				
GCSE/ O-level				
College course/diploma				
A-levels				
University degree				
Post-graduate				
Country and language of education				
Current occupation				

Appendix 2b.

Vowel identification pictures


sheep


ship


shape


cart


cat


cut

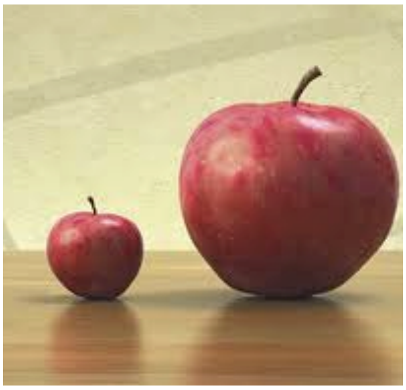

peel


pool


pull

Appendix 2c. Vowel production pictures (note. 2b pictures were also elicited in the vowel production task)

bead



big

bubble

Bart

book

boot

Appendix 2d. Bilingual and monolingual children's vowel formant data.

		Monolingual								Bilingual							
		/iː/	**/ɪ/**	**/ɑː/**	**/ʌ/**	**/uː/ I**	**/ʊ/ I**	**/uː/ II**	**/ʊ/ II**	**/iː/**	**/ɪ/**	**/ɑː/**	**/ʌ/**	**/uː/ I**	**/ʊ/ I**	**/uː/II**	**/ʊ/ II**
Time 1	*Mean F1*	-0.82	-0.27	0.95	1.45	-0.77	-0.50	-1.05	-0.50	-0.86	-0.39	1.08	1.35	-0.83	-0.25	-0.73	-0.52
	Median F1	-0.76	-0.26	0.98	1.37	-0.70	-0.45	-0.88	-0.47	-0.87	-0.45	1.11	1.36	-0.87	-0.25	-0.77	-0.50
	SD F1	0.37	0.31	0.58	0.52	0.30	0.43	0.63	0.41	0.37	0.31	0.58	0.46	0.30	0.35	0.34	0.35
	Mean F2	1.37	0.92	-0.85	-0.50	0.33	-0.18	-1.16	-1.05	1.53	0.94	-0.64	-0.49	0.10	-0.69	-1.16	-1.10
	Median F2	1.44	0.94	-0.84	-0.48	0.39	-0.17	-1.18	-1.20	1.55	0.92	-0.63	-0.47	0.10	-0.85	-1.18	-1.10
	SD F2	0.24	0.25	0.24	0.29	0.45	0.32	0.63	0.50	0.34	0.25	0.20	0.23	0.42	0.49	0.63	0.24
Time 2	*Mean F1*	-0.97	-0.17	1.0	1.31	-0.88	-0.33	-0.86	-0.30	-0.94	-0.24	0.98	1.29	-0.83	-0.18	-0.78	-0.36
	Median F1	-0.99	-0.18	1.02	1.39	-0.88	-0.33	-0.84	-0.33	-0.97	-0.24	1.07	1.27	-0.87	-0.22	-0.80	-0.32
	SD F1	0.29	0.42	0.54	0.56	0.29	0.40	0.30	-0.48	0.30	0.28	0.70	.71	0.41	0.31	0.31	0.36
	Mean F2	1.49	0.73	-0.87	-0.59	0.26	-0.17	-1.09	-0.50	1.49	0.89	-0.71	-0.61	0.34	-0.49	-1.04	-0.92
	Median F2	1.54	0.85	-0.80	-0.57	0.39	-0.16	-1.1	-0.64	1.52	0.91	-0.70	-0.61	0.38	-0.64	-1.13	-0.94
	SD F2	0.43	0.52	0.25	0.19	0.64	0.26	0.40	-0.65	0.30	0.31	0.21	0.21	0.58	0.44	0.38	0.24

Appendix 3a. VOT categorization pictures

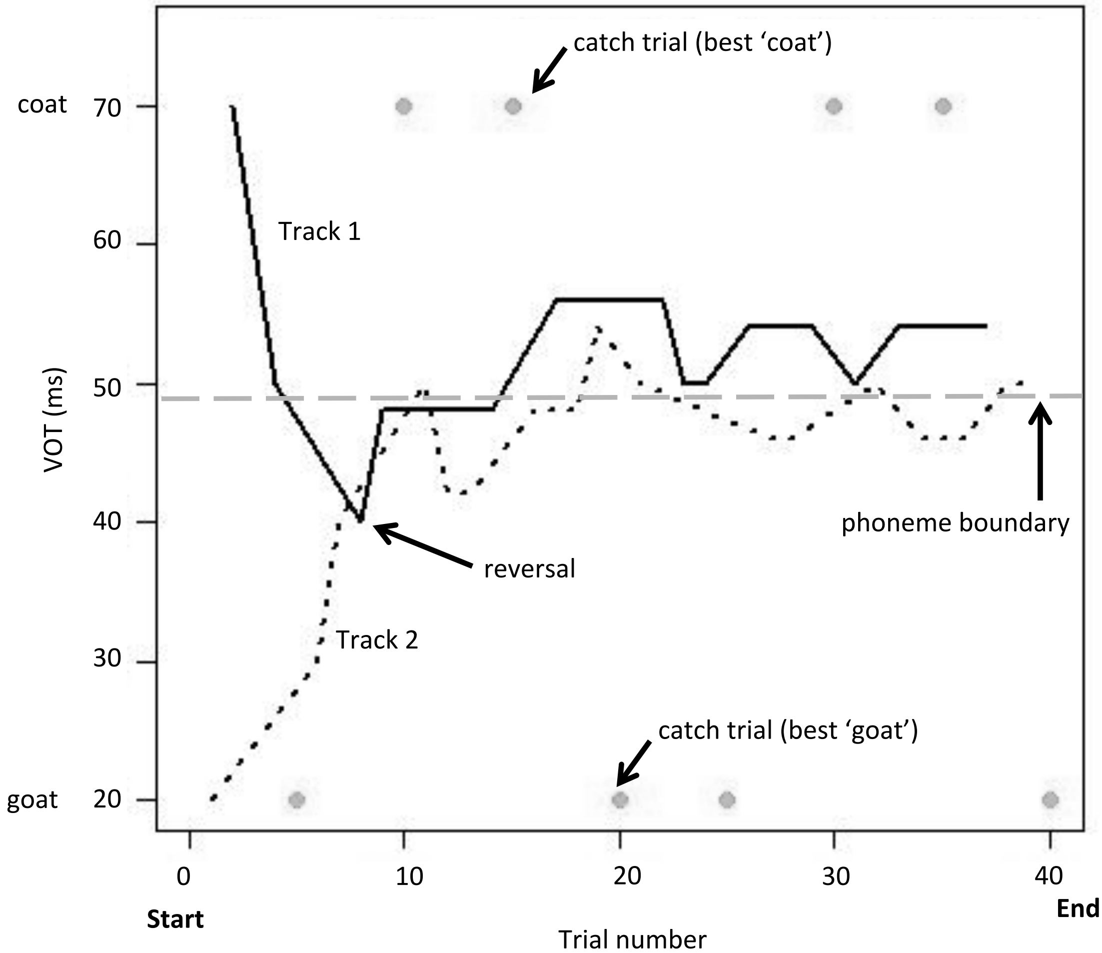
bee

pea

goat

coat

Appendix 3b. Example of adaptive track procedure for categorisation task



A graphical illustration of the operation of the adaptive 1-up/2- down Levitt procedure. Two interleaved adaptive tracks are run independently, with one tracking 71% 'coat' responses (Track 1 -solid line) and one tracking 71% 'goat'' or equivalently, 29% 'coat' responses (Track 2 - dotted line). Continuum end points are also presented every 5 trials, good exemplars of each category, in order to better maintain the stability of the listener's category boundary.

Appendix 3c. Plosive production pictures (note. 3a words were also elicited in the production task)

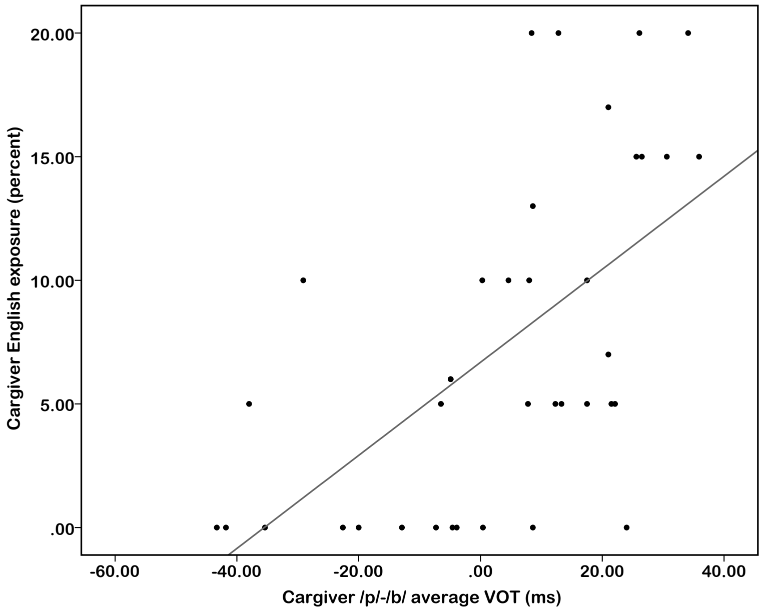
bear

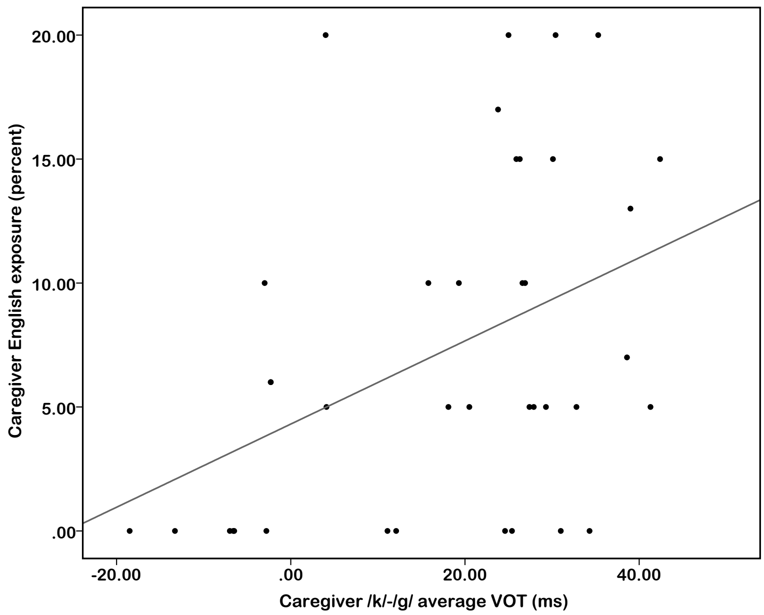
pear

good

cat

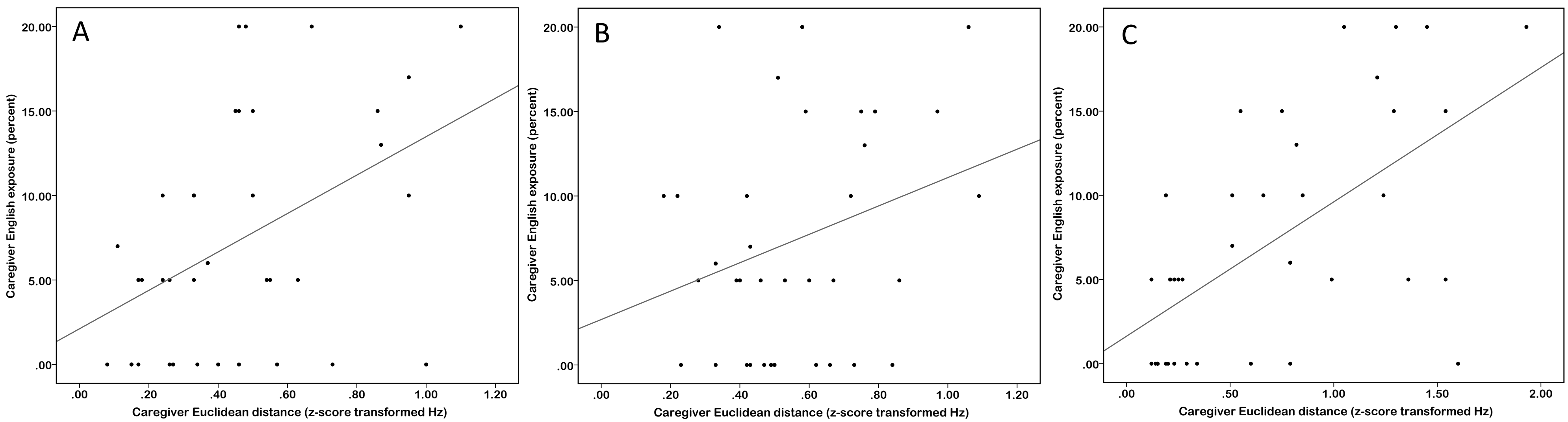
Appendix 4a. Caregiver English stop production vs. English language exposure scatter plots.





Scatter plots illustrating the relationship between caregiver average voice onset time (VOT) for bilabial /p/-/b/, and velar /k/-/g/ stops and caregiver English exposure (percent).

Appendix 4b. Caregiver English vowel production vs. English exposure scatter plots



Scatter plots illustrating the relationship between caregiver English vowel Euclidean distance values (z-score) for front /i/-/ɪ/ (A), central /ɑ/-/ʌ/ (B), and back /u/-/ʊ/ (C) vowel contrasts, and caregiver English exposure (percent).