CHILD DEVELOPMENT

Child Development, September/October 2014, Volume 85, Number 5, Pages 1965–1980

Speech Perception and Production by Sequential Bilingual Children: A Longitudinal Study of Voice Onset Time Acquisition

Kathleen M. McCarthy, Merle Mahon, Stuart Rosen, and Bronwen G. Evans
University College London

The majority of bilingual speech research has focused on simultaneous bilinguals. Yet, in immigrant communities, children are often initially exposed to their family language (L1), before becoming gradually immersed in the host country's language (L2). This is typically referred to as sequential bilingualism. Using a longitudinal design, this study explored the perception and production of the English voicing contrast in 55 children (40 Sylheti-English sequential bilinguals and 15 English monolinguals). Children were tested twice: when they were in nursery (52-month-olds) and 1 year later. Sequential bilinguals' perception and production of English plosives were initially driven by their experience with their L1, but after starting school, changed to match that of their monolingual peers.

To date the majority of speech development research has focused on monolingual infants. However, with the increase in mobility and, in turn, immigration, growing up bilingual is becoming the norm (National Association for Language Development in the Curriculum, 2013; Pupil Level Annual School Census, 2013). Multicultural cities such as London, where there are dense pockets of immigrant communities, are prime examples of this. Children who grow up in such communities are often initially exposed primarily to the family language, and it is not until they enter nursery at around 3 years of age that they gradually become immersed in the host country's language (Darcy & Krüger, 2012; Xu Rattansone & Demuth, 2013). These children are often referred to as sequential bilinguals. Yet research on bilingual speech development in such communities is scarce, being limited to simultaneous bilinguals who are exposed to both languages from birth (e.g., Sundara & Scuttellaro, 2011).

The current study focuses on the acquisition of the English voicing contrast in Sylheti-English sequential bilinguals and their monolingual English peers. Investigating these developmental trajectories sheds light on our understanding of how young children growing up in an immigrant community acquire a second language.

Acquiring the Speech Sounds of Two Languages: Simultaneous and Sequential Bilinguals

It is well established that monolingual infants become attuned to the speech sounds of their ambient language within the 1st year of life (Kuhl, 2004; Werker & Tees, 1984) and that these phonemic categories continue to be refined up until adolescence (Hazan & Barrett, 2000). Research on the phonemic development of simultaneous bilingual infants has shown that they are also able to discriminate the sounds of both of their languages by 12 months old. There appear to be two main accounts of development. On the one hand, some studies have shown that simultaneous bilinguals are able to discriminate the sounds of both of the languages being learned throughout the 1st year of life (e.g., Sundara, Polka, & Molna, 2008). However, others have shown that simultaneous bilingual infants initially differ from their monolingual peers (e.g., Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009). For example, using an ERP paradigm, Garcia-Sierra et al. (2011) showed that 6- to 9-month-old Spanish-English bilinguals showed no neural discrimination of either Spanish or English /t/-/d/ contrasts, but that by 10–12 months they were able to discriminate the contrasts in both languages.

We would like to thank Mike Coleman who designed the testing software used in this study and Steve Nevard for technical support. Many thanks are also due to our interpreter Khadija Kubra. Finally, we would like to thank all of the nurseries, schools, and families who took part in this study.

Correspondence concerning this article should be addressed to Kathleen M. McCarthy, Department of Speech, Hearing & Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London, United Kingdom, WC1N 1PF. Electronic mail may be sent to kathleen.mccarthy@ucl.ac.uk.

DOI: 10.1111/cdev.12275

For many bilingual children who grow up in dense immigrant communities in large multicultural cities such as London, the language environment is typically different from that of simultaneous bilinguals. In such communities the immigrant group is the local majority (see, e.g., Rasinger, 2007), and it is likely that the family language will be dominant within the community. Depending on the family structure, children who grow up in these communities will often acquire the family language as their first language (L1). In some cases, these children may only begin to fully acquire the host language (L2) at a later stage in development, normally when they are in full-time education. Indeed, although it is likely that they will have had some exposure to the L2 (e.g., through the media, older siblings), previous work (McCarthy, 2009) has shown that the majority of these children's interactions are in their family language. Thus, it is likely that a sequential bilingual child will come to the task of acquiring the host country's language with an existing phonology that reflects their L1.

Accounting for the effect of the differing amounts and types of input in such communities is challenging. Current studies of language acquisition in sequential bilinguals differ in terms of how much L2 exposure the children have had, as well as the age at which the children started to acquire the L2 (e.g., Darcy & Krüger, 2012; Tsukada et al., 2005). In general, such studies suggest that there is some L1–L2 influence, at least initially. 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 2.0–4.9 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. However, a recent study of 9-year-old Turkish-German bilingual children, who started to acquire German from 2 to 4 years of age, showed that they had difficulties with certain German vowel contrasts (Darcy & Krüger, 2012). The authors suggest that the children's continued high use of their L1 influenced their ability to perceive and produce the German vowels.

Studying acquisition in sequential bilinguals is further complicated by the fact that these children come to the task of acquiring the host country's language with an existing, but still developing, phonology that reflects their L1. Of the few studies on phoneme categorization beyond the 1st year of life, research on monolingual children has shown that they continue to develop and refine their phonemic categories up until early adolescence (Hazan & Barrett, 2000; Nittrouer, 2005). Similarly, for production, much variability has been found in children's early speech, though this seems to reduce with age (Lee, Potamianos, & Narayanan, 1999; Vihman, 1996). One could imagine that sequential bilinguals acquire their L2 sounds to a native-like level, as when they start learning their L2, their L1 categories are less well established than those of adult learners. This is further supported by L2 research that has shown that early learners show native-like perception and production for L2 contrasts (Baker & Trofimovich, 2005; Flege, Munro, & MacKay, 1995; McCarthy, Evans, & Mahon, 2013; Tsukada et al., 2005).

Theories of second language learning have proposed that the differences between early and late learners are due to differences in neural plasticity, such that L1 phonemic categories become more robust with age, making it harder to acquire new L2 categories in adulthood (Flege, 1995; Iverson et al., 2003). Flege's speech learning model (Flege, 1995) suggests that L1 and L2 phonemes exist in a common phonological space. When the L2 phonemes are similar to those of the L1, learners are likely to use their more established L1 categories. Therefore, L2 sounds are thought to be easier to acquire if they are dissimilar to sounds in the L1. Children are more likely to acquire L2 sounds because they have less well-established L1 categories, and so are better able to reorganize their phonological space. Likewise, for perception, Best's perceptual assimilation model (PAM; Best & McRoberts, 2003; Best & Tyler, 2007) suggests that L2 listeners perceptually assimilate non-native phonemes to native phonemes that they judge to be most similar. PAM suggests that the patterns of assimilation are driven by the relation between the L1 and L2 phonemes. For example, German listeners are relatively good at discriminating the English /i/-/ɪ/ contrast (Bohn & Flege, 1990), whereas Spanish listeners find this contrast difficult (Flege, Bohn, & Jang, 1997). These differences are likely to be due to listeners' L1 vowel inventory; German contains a contrast similar to /i/-/ɪ/ but Spanish does not.

In addition to being exposed to two languages, sequential bilinguals are likely to be exposed to native (e.g., teachers) and non-native (e.g., grandmother) varieties (see e.g., Fernald, 2006; McCarthy et al., 2013). Research on the influence of accented input has only recently started to develop over the

last decade, and has primarily focused on the impact of accent and dialect variation on infant speech perception (e.g., Cristià, 2011; Schmale & Seidl, 2009) with some work on young children (e.g., Nathan & Wells, 2001). To date, this research suggests that infants and children are sensitive to accented variants. For example, Cristià (2011) showed that acoustic features of caregivers' /s/ production were significantly related to English-speaking 12- to 14-month-old infants' ability to discriminate the /s/-/ʃ/ contrast. That is, infants of caregivers who produced an acoustically more extreme /s/, making the /s/-/ʃ/ contrast more distinct, were significantly better at discriminating the /s/-/ʃ/ contrast.

However, studies have shown that infants understand highly variable speech (e.g., Kuhl et al., 2008) as well as foreign-accented input from around 14 months (e.g., Schmale & Seidl, 2009). Such findings raise interesting questions regarding speech perception in sequential bilinguals. These children are likely to be exposed to accented L2 speech input, especially if their main caregivers are late arrival L2 learners (see McCarthy et al., 2013). Possible issues arise when this accented speech affects the production of a phonetic contrast in the L2. For instance, the English /i/-/ɪ/ contrast is often neutralized in the speech of L2 learners, such that *sheep* and *ship* are produced using the same vowel, /i/. In such cases, it is possible that the infant, if only exposed to the L2 from his or her main caregiver, may not initially acquire the L2 contrast but may instead acquire the accented variant.

The Development of the Voicing Contrast

The focus of the current research is on the acquisition of the voicing contrast in English, for example, /b/-/p/. In their pioneering cross-linguistic study of word–initial plosives, Lisker and Abramson (1964) defined voice onset time (VOT) as "the time interval between the burst that marks the release of the stop closure and the onset of quasiperiodicity 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, for example, ***p**ea-**b**ee*. Voicing patterns can be broadly divided into three categories: (a) *voicing lead*, values below 0 ms; (b) *short-lag*, ranging from 0 to 30 ms; and (c) *long-lag*, values over 30 ms (Lisker & Abramson, 1971). 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); plosives with a longer F1 transition and/or lower F1 onset frequency are classified as voiced, and vice versa for voiceless plosives.

Previous research has demonstrated that, at least for monolingual learners, different types of plosives have different patterns of acquisition (e.g., Hazan & Barrett, 2000; Macken & Barton, 1979; Zlatin & Koenigsknecht, 1976). Macken and Barton (1979) outlined three stages of production development for English: (a) children initially produce all plosives in the short-lag region; (b) a distinction between voiced and voiceless plosives is made, though it may not be perceived by adult listeners (i.e., a covert contrast); and (c) 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 voicing lead (e.g., French and Spanish) suggest that voicing lead develops slightly later, perhaps not until 5 years old or later (Macken & Barton, 1979; Zlatin & Koenigsknecht, 1976). It has been suggested that the articulatory demands of voicing lead production and that the coordination of laryngeal control with supralaryngeal articulatory gestures, makes it difficult for children to master (Kewley-Port & Preston, 1974).

For perception, there also appears to be a developmental trend. Previous work using synthetic speech continua suggests that phonemic categorization improves with age (e.g., Hazan & Barrett, 2000; Mayo & Turk, 2004), such that older children are more accurate than younger children, but not as accurate as adults at identifying different speech stimuli. It has been hypothesized that this is because children pay attention to different acoustic cues and weight these cues differently from adult listeners (Nittrouer, 2005; Ohde & German, 2011), a phenomenon often referred to as "perceptual cue weighting" (Nittrouer, 1996; Nittrouer & Miller, 1997). Nittrouer's developmental weighting shift hypothesis (DWS; Nittrouer, 1996) suggests that children are initially more sensitive to global measures such as formant transitions, as they are considered 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 adults (e.g., Howell, Rosen, Lang, & Sackin, 1992; Ohde & Haley, 1997). The DWS (Nittrouer, 1996; Nittrouer & Miller, 1997) proposes that the reason for this developmental pattern is that the aspects of the speech signal that children pay attention to change as they gain experience with their native language.

Such developmental patterns, combined with possible interference effects between the family L1 and host L2 languages, raise interesting questions regarding the phonemic development of sequential bilinguals: Is there evidence for the influence of existing L1 categories when the children begin to acquire their L2? If so, with experience, are children able to perceive and produce, in a native-like way, phonemic contrasts that do not exist in their L1?

The Current Study

Using a longitudinal design, Sylheti-English sequential bilingual children's perception and production of English bilabial and velar plosives was investigated. Children were tested in nursery (52 months old) and 1 year later, in school. The children were all born in the United Kingdom and raised within the London–Bengali community, namely, in the London boroughs of Camden and Tower Hamlets. Continuous migration since 1950s has resulted in a complex multilingual community, made up of first-, second-, and subsequent-generation speakers. This generational mix is further enriched by new first-generation arrivals from Bangladesh. In Tower Hamlets, 64% of primary school aged children (under 10 years of age) are of Bangladeshi origin (Spring School Census; Tower Hamlets, 2012). In Camden, Bangladeshi-origin children are the largest minority group, accounting for 19% of primary school aged children (Camden Council, 2012).

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. While Standard Bengali has a four-way voicing contrast (labial, dental, alveolar, velar), for example, velar: /k/, /kʰ/, /g/, /gʱ/ (Khan, 2010), recent work (McCarthy et al., 2013) has shown that Sylheti has a two-way voicing contrast, and only retains the aspirated/unaspirated series for voiced velar plosives: /g/, /gʱ/. For all other places of articulation, there is a voiced–voiceless pair, for example, /p/, /b/. Although VOT varies with place of articulation, voiced plosives are produced with voicing lead (M = −75 ms) and voiceless plosives with short lag (0–30 ms). In contrast, English voiced plosives are short lag and voiceless plosives are long lag, with an average of 21 and 56 ms 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 the Sylheti voiceless plosives (/p/, /t/, /k/). In addition, previous research has demonstrated that first-generation immigrants use Sylheti-like voicing patterns when producing English plosives (McCarthy et al., 2013), and so it is likely that the children in this study will be exposed to foreign-accented variants in English.

The aims of the current study are (a) to investigate potential language interference effects in children who acquire an L2 at an early stage in childhood and (b) to investigate the developmental trajectory of English bilabial and velar plosive acquisition in Sylheti-English sequential bilinguals. As discussed previously, it is likely that these Sylheti-speaking children will come to the task of learning English plosives with preexisting voicing categories that reflect their L1 Sylheti input. One could imagine that their acquisition of the English voicing contrast will thus be affected by the differences between their L1 and L2, such that initially they will have a different underlying phonemic category structure from their monolingual peers. In production this might mean that children initially use their Sylheti categories when producing English plosives, that is, use voicing lead for voiced and short lag for voiceless plosives, and in perception that they may perform less consistently in identification tasks, and have different phoneme boundaries.

Experiment 1: Perception

Method

Participants

Fifty-five children successfully completed the study: 40 Sylheti-English sequential bilinguals (M_{age} at start = 52.7 months, range = 46–57 months; 25 female, 15 male) and 15 monolingual English controls (M_{age} at start = 54.2 months, range = 47–57 months; 7 female, 8 male). All children were tested twice. At Time 1, the sequential bilingual children had an average of 7 months English language experience in nursery. During this time they also had some language support from Sylheti-speaking teaching assistants. Time 2 was 11–12 months later, in reception class (i.e., 1st year in primary school in the United Kingdom). To have a target adult comparison, an additional 6 monolingual English-speaking adults (2 male, 4 female; M_{age} = 28 years old, range = 24–38 years old) were tested.

To be included in the study, all children had to have (a) normal hearing thresholds, (b) no documented history of chronic middle-ear infections, and (c) no documented history of speech and

language difficulties. All participants had to pass three screening tests (a) a hearing screen of the audiometric frequencies 0.5, 1.0, 2.0, and 4.0 kHz presented at 25 dB HL; (b) a nonverbal IQ screen—within 1 *SD* of the mean for the block design subtest of the British Ability Scales, 2nd edition (Elliot, 1996); and (c) 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 (subsection of the Diagnostic Evaluation of Articulation and Phonology; Dodd, Hua, Crosbie, Holm, & Ozanne, 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 nonverbal IQ screen (n = 3; 2 bilingual, 1 monolingual), failing the phonology screen (n = 2; 1 bilingual, 1 monolingual), failing to complete the test block due to 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 a sequential bilingual, a child needed to have a maximum of 20% exposure to English from the main caregivers from birth to entering preschool education. Language exposure was measured using an adapted version of a language exposure questionnaire developed by McCarthy (2009). Based on previous bilingual questionnaires (Bosch & Sebastián-Gallés, 2001), our questionnaire required the main caregiver to provide an estimation (in hours, per day) of the exposure to English and Sylheti from the child's main caregivers (e.g., parents, grandparents). All children, aside from two, were children of first-generation Bangladeshi-origin parents; that is, both parents were born in Sylhet, Bangladesh, and arrived in the United Kingdom as either children or adults. The remaining two children had one first-generation parent and one second-generation parent, that is, a parent who was born in the United Kingdom. All Sylheti-English bilinguals resided in either Tower Hamlets (n = 21) or Camden (n = 19). To ensure that the monolingual English children had minimal contact with Sylheti, they were recruited from schools in Camden and in the neighboring borough of Hackney. In these schools less than 20% children were of Bangladeshi origin. All caregiver interviews were conducted in the children's home or school.

Socioeconomic status. Information concerning demographic variables such as parent's level of education (in the United Kingdom or Bangladesh), place where parents learned English and Sylheti, and level of English (if attended classes in the United Kingdom), was collected during the interviews. Where possible we matched the children for socioeconomic status (SES), based on maternal and paternal level of education (see, e.g., Bradley & Corwyn, 2002). Categorization of the level of education was based on the National Qualifications Framework (England, Wales, and Northern Ireland; Office of Qualifications and Examinations Regulation, 2004) or the Bangladesh equivalent (based on UK NARIC, http://ecctis.co.uk/naric/), 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), postgraduate university degree (master's or doctorate).

In all, 50 of the 55 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 GCSE level, and 35% up until A-level/college course. Nineteen percent had completed a bachelor's degree and 7% a master's degree.

Stimuli

The continua were the same as those used in previous studies: *pea-bee* (Hazan, Messaoud-Galusi, Rosen, Nouwens, & Shakespeare 2009) and *coat-goat* (Ramus et al., 2003). Stimuli were generated by copy synthesis using the cascade branch of the Klatt (1980) synthesizer. 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 natural tokens recorded by a native female British English speaker. For *pea-bee*, the total syllable duration was 390 ms. For *bee*, F1 began at 390 and reached 185 Hz at the end of the syllable. The F2 and F3 transitions increased from 1400 and 2500 Hz, respectively, to 2540 and 2970 Hz. F2 and F3 then increased to reach 2760 and 3377 Hz at the end of the syllable. F4 was set at 3950 Hz. For *goat*, the total syllable duration was 459 ms. The F1 transition increased from 477 to 640 Hz, and F1 then decreased from 640 to 306 Hz by the end of the syllable. F2, F3, and F4 were set at 2080, 2900, and 4380 Hz,

respectively, and reached 1645, 2800, 4130 Hz at the end of the syllable. To obtain stimuli differing in VOT, the onset of voicing was delayed while simultaneously increasing the duration of aspiration. All continua varied across VOT in 1-ms steps, ranging from 0 ms for /bi/ to 60 ms for /p^hi/, and 20 ms for /gəʊt/ to 70 ms for /k^həʊt/. For the velar continuum, which used a vowel with a relatively high F1, the F1 onset frequency covaried with VOT, increasing with increasing VOT, as it naturally does. For the bilabial continuum, the low F1 of the /i/ vowel precludes any significant transition, so F1 onset frequency varies little, again as would be the case naturally. (See online Appendix S1 for the waveforms and spectrograms of the continua endpoints.)

To avoid lexical bias (see, e.g., Thompson & Hazan, 2010), all the target words were checked with parents, age-appropriate vocabulary lists (e.g., Oxford Communicative Development Inventory; Hamilton, Plunkett, & Schafer, 2000), and classroom vocabulary lists. As an additional check, children had to correctly name the pictures in a picture-naming task and correctly identify the words in a three alternative forced-choice picture-pointing task.

Procedure

The auditory labeling tasks were presented to the children in a computer game format on a laptop in a quiet room in the nursery/school. The stimuli were presented at 65 dB SPL over Sennheiser HD 25-1 II headphones (Wedemark, Hanover, Germany). A two-alternative forced-choice task was used. The instructions to each child were as follows: "Panda is learning to say new words, and because you already know these words, you're the best person to help him. Listen to Panda and point to what he says." The children identified the stimulus by pointing to an on-screen picture of the target word (e.g., a coat or a goat). An on-screen reward (a picture of bamboo) was given after each trial. Corrective feedback (i.e., "Well done!" and a tick for a correct response, or a "dong" and a cross for an incorrect response) was only given for catch trials (continuum endpoints).

The task began with a familiarization block consisting of four trials (two per endpoint) with participants receiving feedback after each trial. To continue to the test block, participants had to score 100% correct.

Test stimuli were presented in a single test using an adaptive procedure (as employed by Hazan et al., 2009; Ramus et al., 2003). Two independent, randomly interleaved adaptive tracks were used, starting at the two endpoints of the continuum (e.g., one at a clear /p/ and another at a clear /b/). A modified Levitt (1971) procedure was then used to estimate the points on the continuum where the stimuli were labeled as one word of the pair (e.g., *goat*) 29% and 71% of the time. When the participant labeled two stimuli in a row as the category from which that track started, the next trial would move further along the continuum, closer toward the phoneme boundary. When a participant identified a sound as coming from the other category, the next trial from that track would be more likely to be identified as an instance of the other category by moving back toward the endpoint of the continuum. The initial step size was 10 ms, reducing linearly to 4 ms over the first three changes in direction of the track (known as reversals). To track attention, and maintain stable phoneme boundaries, continuum endpoints (catch trials) were randomly interspersed 20% of the time. The task ended after seven reversals or a maximum of 40 trials. (See online Appendix S2 for a diagram illustrating the test procedure.)

This adaptive procedure allowed us to efficiently track the children's categorization of the voicing continua, sampling a large range of stimulus values but without testing every step on the continuum. Furthermore, the linear reduction in step size allows for a fine-grained determination of the listener's phoneme categorization. Thus, the advantages of using an adaptive procedure are (a) 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, and (b) the level of difficulty is consistent across participants as the same level of performance (71% *coat* or *goat* responses) is tracked for each listener.

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 categorization 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 on the continuum. A shallower slope indicates a lower degree of consistency in the labeling of the continuum, and in turn less refined phonological categories. The catch trials were analyzed separately. For a session to be included in the final analyses, performance on the catch trials had to be 80% or better (as in Nittrouer, 2005).

Results

For /p/-/b/, 18 children (3 monolingual, 15 bilingual) did not meet the inclusion criterion, leaving results for 37 children (12 monolingual, 25 bilingual). For /k/-/g/, 14 children (3 monolingual, 11 bilingual) did not meet the inclusion criterion, leaving results for 41 children (12 monolingual, 29 bilingual).

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

To test for differences in perception, separate linear mixed model (LMM) analyses were performed for categorization slope and phoneme boundary, using the generalized LMM module in SPSS, and specifying the identity link function. Language group, time, and contrast (bilabial vs. velar) 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.

Categorization Slope

Figure 1 shows the aggregated data for the monolingual and bilingual children, for /p/-/b/ and /k/-/g/. Individual differences and the adult target slope value ($M = -0.25$) can be observed in Figure 2. Both groups display within-group variation at Times 1 and 2. At Time 2, all children have values that are closer to that of the adult target.

Linear mixed model analyses revealed a significant interaction between group and time, $F(1, 150) = 4.37$, $p = .038$, and a main effect of group, $F(1, 150) = 14.71$, $p < .001$, and time, $F(1, 150) = 85.21$, $p < .001$. The main effect of contrast was not significant ($p > .05$). Overall, the bilinguals had significantly shallower slopes than did the monolingual children; however, all children showed a significant increase in slope steepness between Time 1 and Time 2. Sequential Sidak post hoc tests confirmed a significant difference between the bilingual and monolingual children at Time 1 ($p < .05$) but not at Time 2 ($p > .05$). Specifically, at Time 1, the monolinguals had steeper slopes than did the bilinguals for /k/-/g/ ($p < .05$) but not for /p/-/b/ ($p > .05$). By Time 2, there was no significant

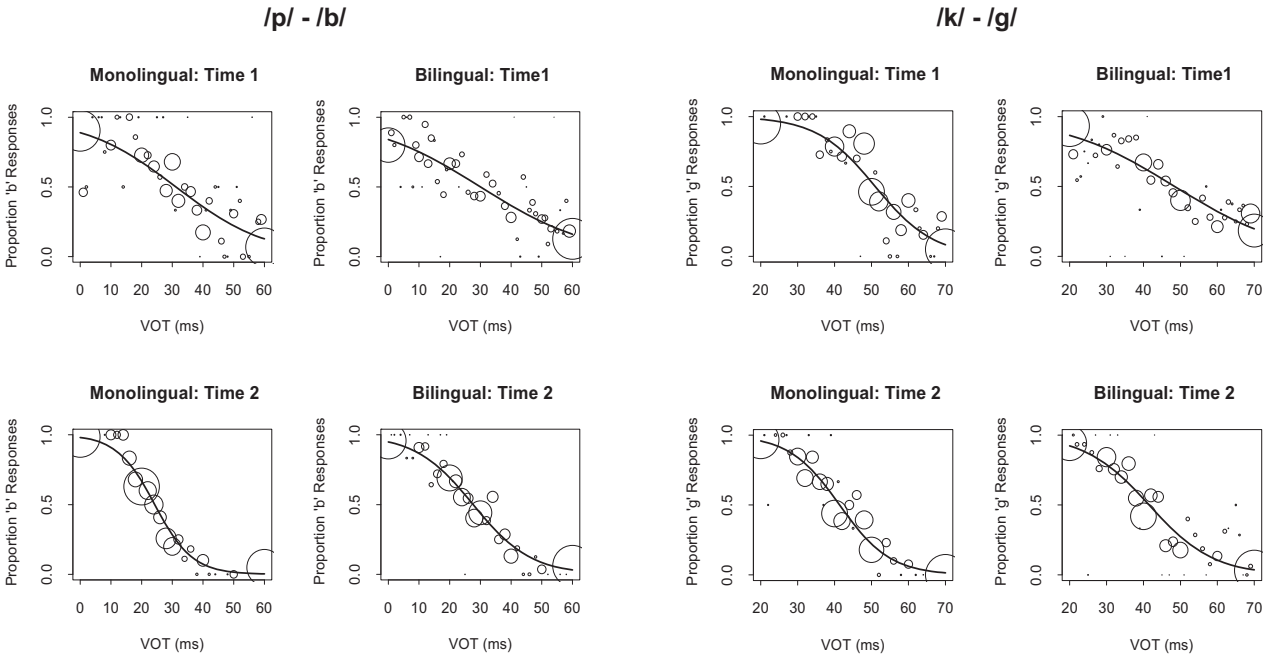

Figure 1. Identification functions for the /p/-/b/ and /k/-/g/ contrasts aggregated over listeners, for bilingual and monolingual children at Time 1 (nursery) and Time 2 (school). The size of the circles is proportional to the number of trials at a given point on the continuum; that is, larger circles indicate more trials.

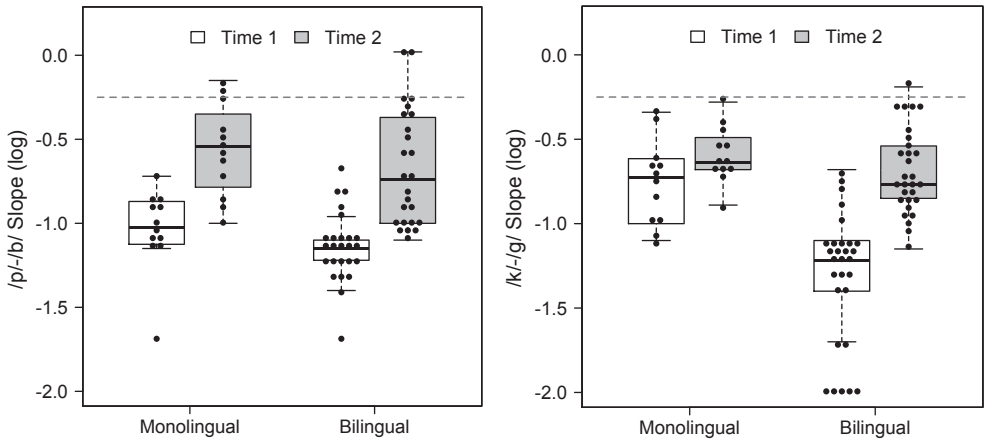

Figure 2. Box plots of slope values for monolingual and bilingual children's /p/-/b/ and /k/-/g/ identification function. A higher log value indicates a steeper slope. The white boxes represent Time 1 and the gray boxes represent Time 2. The black dots represent individual data points. The dashed gray line represents the adult monolingual English mean slope value. Slopes were transformed to a logarithmic scale because the original values were skewed. The range of values shown can be appreciated better by noting first that logistic functions, as seen in Figure 1, have their maximal slope at the phoneme boundary. The adult slope value (at −0.25) corresponds to a slope of about 14 percentage points per ms of voice onset time. Other useful values corresponding to log slope values of −0.5, −1.0, and −1.25 give slopes of 7.9, 2.5, and 1.4 percentage points per ms, respectively.

difference between the monolingual and bilingual children for both contrasts ($p > .05$).

Phoneme Boundary

Figure 3 shows the children's phoneme boundary for /p/-/b/ (Time 1 mean, monolingual = 31 ms, bilingual = 30 ms; Time 2 mean, monolingual = 24 ms, bilingual = 27 ms) and /k/-/g/ (Time 1 mean, monolingual = 40 ms, bilingual = 43 ms; Time 2 mean, monolingual = 49 ms, bilingual = 48 ms). LMM analyses revealed a significant three-way interaction between group, contrast and time, $F(3, 113) = 6.06$, $p = .001$, and a main effect of contrast, $F(1, 112) = 168.10$, $p < .001$. All other main effects were not significant ($p > .05$). Sequential Sidak post hoc tests revealed that /p/-/b/ had a significantly shorter phoneme boundary than /k/-/g/ ($p < .05$), and overall there was no significant difference between the monolinguals and bilinguals ($p > .05$). The monolinguals, however, displayed a significant shift in phoneme boundary from Time 1 to Time 2 for /p/-/b/ and /k/-/g/ such that they were closer to that of the adult target ($p < .05$), whereas the bilinguals showed no significant difference between Time 1 and Time 2 ($p > .05$).

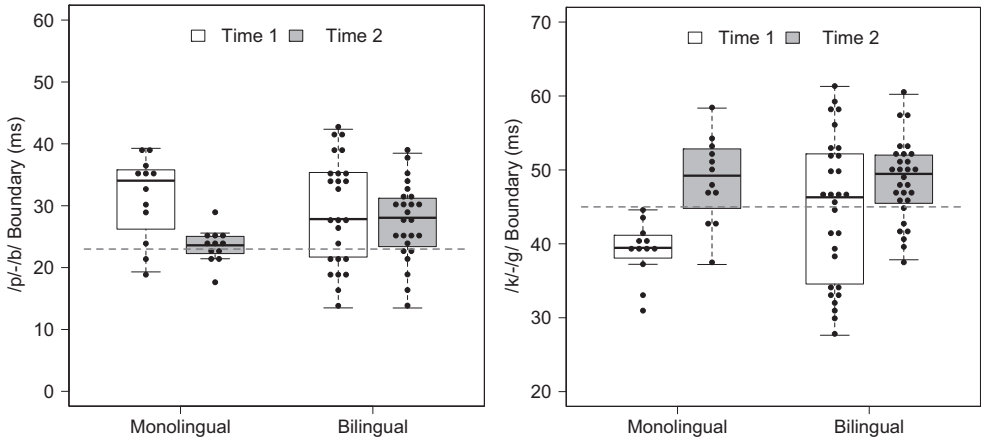

Figure 3. Box plots of monolingual and bilingual children's phoneme boundary (voice onset time, milliseconds) for /p/-/b/ and /k/-/g/. The white boxes represent Time 1 and the gray boxes represent Time 2. The black dots represent individual data points. The dashed gray line represents the adult monolingual English mean boundary.

Summary

Overall, the results showed a significant increase in perceptual acuity from Time 1 to 2 for the monolinguals and bilinguals, such that all children had categorization slopes that were closer to that of the adult target for /p/-/b/ and /k/-/g/ at Time 2. At Time 1, the bilinguals had significantly shallower categorization slopes than did the monolinguals for /k/-/g/, but not for /p/-/b/. For phoneme boundary, there was no significant difference between the monolinguals and bilinguals at Time 1. By Time 2, there was no significant difference between the monolinguals and bilinguals for /p/-/b/ and /k/-/g/ categorization slope. Interestingly, although there was no significant difference in phoneme boundary between the groups at Time 2, the monolinguals displayed a significant shift in phoneme boundary, whereas the bilinguals did not.

Experiment 2: Production

Method

Participants

The same child participants from Experiment 1 took part in Experiment 2. As in Experiment 1, all children were tested twice: after an average of 7 months English language experience in nursery and then 11–12 months later in reception class. The production data were collected on a different day from the perception data.

Target Sounds

English bilabial and velar plosives—/p/ (pea: /**p**^{h}i:/, pear: /**p**^{h}eə/), /b/ (bee: /**b**i:/, bear: /**b**eə/), /k/ (coat: /**k**^{h}əʊt/, cat: /**k**^{h}æt/), /g/ (goat: /gəʊt/, good: /gʊd/)—were elicited in the word-initial stressed position. Two words per consonant were elicited.

As in Experiment 1, all the target words were checked with parents, age-appropriate vocabulary lists (e.g., Oxford Communicative Development Inventory; Hamilton et al., 2000), and classroom vocabulary lists. In addition, the children had to correctly name the pictures and correctly identify them in a picture-pointing three-forced-choice task.

Procedure

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

Missing and excluded tokens. A total of eight tokens—three /b/, two /p/, and three /g/—were missing from the analysis. This was either due to the child incorrectly naming the picture (e.g., using a given name for *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 voicing was measured as the time between the release of the plosive closure and the onset of voicing, defined as the zero crossing of the first glottal pulse. For voicing lead tokens, voicing was measured from the beginning of the prevoicing to the plosive closure.

Results

As for perception, we ran preliminary analyses comparing the production of the different SES groups (indexed by maternal and paternal level of education). 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. As in Experiment 1, to test for differences in production, an LMM analysis was conducted where language group, time, contrast (bilabial vs. velar) and voicing (voiced vs. voiceless) were treated as fixed factors. To account for individual differences between children, participant was treated as a random factor. Significant interactions were explored using sequential Sidak post hoc analysis.

Figure 4 shows the children's VOT values for /p/, /b/, /k/, and /g/. At Time 1, the bilingual children produced /b/ and /g/ using short-lag and voicing lead values (mean /b/ = −6 ms, mean /g/ = 0 ms), whereas the monolinguals used a short-lag (mean /b/ = 11 ms; mean /g/ = 17.5 ms). For voiceless plosives, both the bilinguals and monolinguals used a long lag (mean /p/, monolinguals = 54 ms; bilinguals = 49 ms; mean /k/, monolinguals = 79 ms, bilinguals = 59 ms). By Time 2, both the groups used a long lag when producing /p/ (M = 68 ms) and /k/ (M = 72 ms) and a short lag for /b/ (M = 5 ms) and /g/ (M = 16 ms).

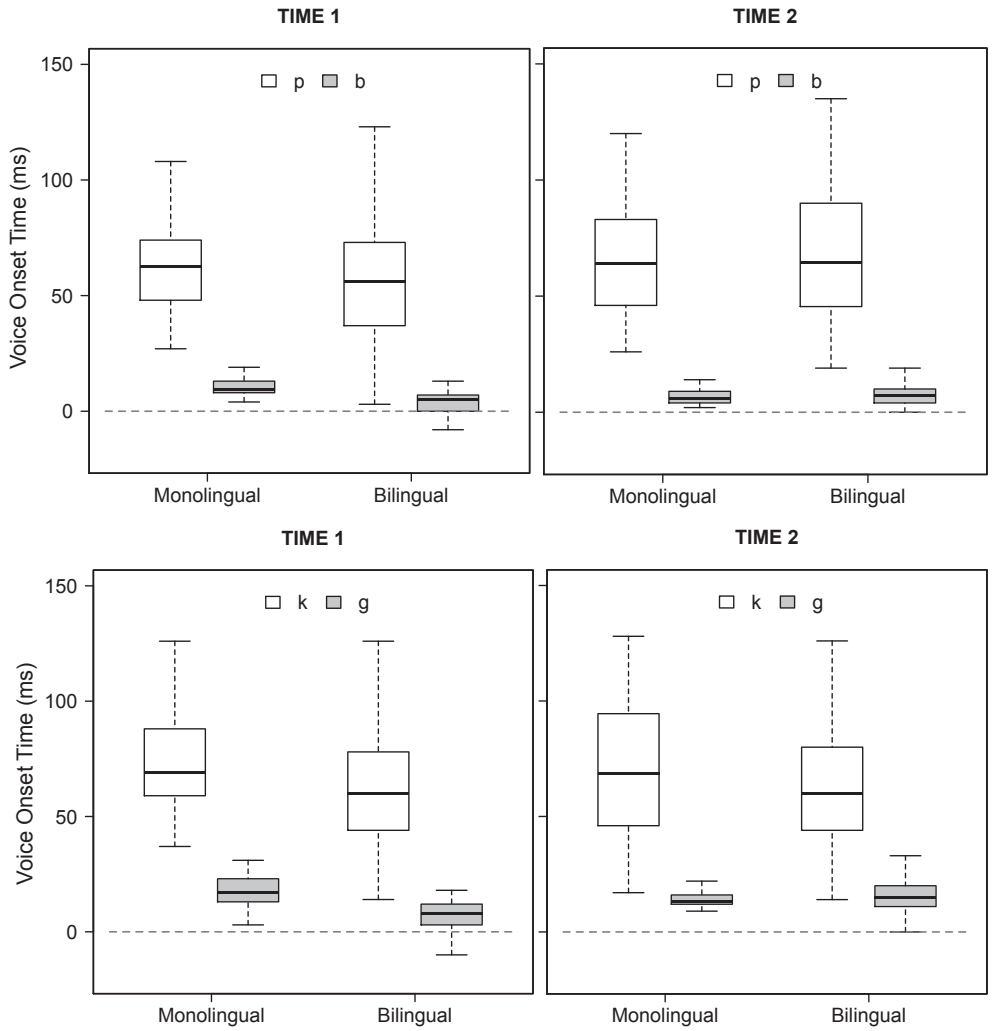

Figure 4. Box plots representing the monolingual and bilingual children's production of /k/-/g/ and /p/-/b/ at Time 1 and at Time 2. Voice onset time measures for each consonant are given in milliseconds. The white boxes refer to the voiceless plosive /p/ or /k/ and the gray boxes refer to the voiced plosive /b/ or /g/. The dashed gray line represents 0 ms. Any value below this represents voicing lead.

Linear mixed model analyses revealed a significant interaction between group and time, $F(1, 672) = 18.068$, $p < .001$, and a four-way interaction between group, contrast, voicing, and time, $F(1, 672) = 3.893$, $p < .001$. There was also a significant main effect of group, $F(1, 677) = 30.877$, $p < .01$; voicing, $F(1, 677) = 473.251$, $p < .001$; contrast, $F(1, 677) = 26.693$, $p < .001$; and time, $F(1, 677) = 9.883$, $p = .002$. Sequential Sidak post hoc tests revealed that overall the bilinguals used a significantly shorter VOT than did the monolinguals at Time 1 ($p < .05$), but not at Time 2 ($p > .05$). Specifically, at Time 1, the bilinguals used a significantly shorter VOT than did the monolinguals for the voiced plosives /b/ and /g/ ($p < .05$), but not for the voiceless plosives /p/ and /k/ ($p > .05$). By Time 2, there was no significant difference between the groups for voiced and voiceless plosives.

Summary

The production results revealed significant differences in voicing between the monolingual and bilingual children, namely, for the voiced plosives. Specifically, at Time 1, the bilingual children used significantly shorter VOT values for voiced plosives /b/ and /g/ that were closer to those of their L1, Sylheti. By Time 2, there was no significant difference between the groups. For voiceless plosives, there was no significant difference between monolinguals and bilinguals, such that both groups used a long-lag VOT at Time 1 and Time 2.

General Discussion

The first aim of this study was to investigate potential language interference effects of the children's L1 (Sylheti) on their categorization of L2 (English) contrasts. To investigate this, we compared the performance of sequential bilingual Sylheti-English-speaking children in perception and production tasks with that of their monolingual English-speaking peers. All children were tested at two time points: during their 1st year of English-speaking nursery (Time 1) and 1 year later (Time 2).

At Time 1, the bilingual children had had an average of 7 months English language experience in nursery, with Sylheti-speaking teaching assistant support. At this time point we found differences between the monolingual English and Sylheti-English bilingual children. For perception, the bilingual children categorized the *coat-goat* voicing contrast less consistently than did the monolingual children, displaying shallower categorization slopes for *coat-goat* than did their monolingual peers. In 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 voicing lead 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 similar slopes and phoneme boundaries. Likewise, there were no differences in the phoneme boundary for the /k/-/g/ continuum. In production, there was no difference in VOT for the voiceless bilabial plosive, /p/; all children produced this with a long VOT.

How might we explain this pattern of results? Although it is possible that attentional factors may have affected children's performance on our tasks, this is an unlikely explanation. 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 (> 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 parents. In addition, the children had to name and identify the target words correctly in order to take part in the study. Our strict inclusion criteria led to a higher proportion of bilingual data sets being excluded from the final analyses in the perception experiment, but this discrepancy is not surprising given that the bilingual children had had less English experience than the monolinguals. Furthermore, research with simultaneous bilinguals has shown that although a typical bilingual's total vocabulary size (both languages) equates to that of a monolingual's single vocabulary, each is smaller than that of their monolingual counterpart (e.g., Hoff et al., 2012). It is thus not surprising that more bilinguals failed our inclusion criteria, given that some may have had less stable English lexical representations at Time 1. This leads us to be confident that the significant differences between the bilinguals and monolinguals found in our final data sets are not due to lexical effects.

A more plausible explanation for our results at Time 1 is that they reflect children's sensitivity to the acoustic properties in their ambient language. The bilingual children would have been exposed predominantly to Sylheti as well as Sylheti-accented variants of English until they entered nursery, 7 months earlier. Therefore, it is likely that these children would have been developing perception and production skills that reflected the phonetics and phonology of their Sylheti input. 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 (Sylheti-like) phonemic categories when producing and perceiving English plosives.

This interpretation is supported by monolingual language acquisition research. 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 L1 phonetic categories interfere with their production and perception of non-native target sounds (see, e.g., Iverson et al., 2003). Likewise, the differences seen here between monolinguals and bilinguals could be explained in terms of language interference. These children may have had difficulties with the English voicing contrast because they were assimilating the L2 English categories into their pre-established L1 categories and were using these categories to perceive and produce English

plosives (see, e.g., Best & Tyler, 2007; Flege, 1995). That is, at Time 1, our children may not yet have developed separate L1 and L2 categories (Johnson & Wilson, 2002).

Furthermore, it is possible that the bilinguals' phonemic categories were also influenced by their caregivers' foreign-accented English input. Although the children in the current study were predominately exposed to Sylheti before starting nursery, previous research on adult speakers from the London–Bengali community shows that first-generation speakers use Sylheti-like VOT values for English plosives (McCarthy et al., 2013). It is likely that our children would also have been exposed to such foreign-accented input and this would have led the bilinguals to develop Sylheti-like English voicing categories. This explanation is supported by research showing that infants are sensitive to subphonemic acoustic–phonetic differences in caregiver speech (e.g., Cristiá, 2011).

However, our children 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., 2013), and so we had expected that at least at Time 1, Sylheti-English children would have had a shorter phoneme boundary for English plosives. Instead, our bilingual children selected boundaries that were similar to those of their monolingual English peers.

Despite the differences in language background, we did not find pervasive differences in VOT boundaries for our monolingual and bilingual children at Time 1. One possible explanation for this is that the structure of the synthetic continua affected the results. Our stimuli only covered the English VOT range and we did not present children with voicing lead stimuli. However, this explanation seems unlikely as, had they been using adult-like Sylheti categories, we would still have expected bilingual children to have behaved differently from their monolingual English peers, for example, 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 our bilingual children had not yet fully developed voicing lead in their L1 when tested at Time 1. The development of voicing lead has been shown to take longer than the short- and long-lag distinction (Macken & Barton, 1979). 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). Finally, it is possible that our 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 we did not catch these children early enough to see the transition from Sylheti- to English-dominant categorization. Furthermore, as we see 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., attunement theory; Aslin & Pisoni, 1980).

This leads us to the second aim of the study: to investigate how sequential bilingual children's VOT perception and production changed as a result of an increase in L2 experience. By Time 2, after an additional year of English experience in school with monolingual English-speaking teachers, the bilinguals showed a significant increase in the steepness of their categorization slope for /k/-/g/ and /p/-/b/, indicating a more consistent labeling of the continuum, and thus more refined phonemic categories. For production, they used a longer VOT for voiced plosives. Indeed, at Time 2, they were no longer significantly different from their monolingual peers. Interestingly, although all children's productions fell into the correct phoneme boundary region for bilabial and velar categorization at Time 1 and Time 2, the results showed a significant shift in phoneme boundary from Time 1 to Time 2 for the monolinguals, but not for the bilinguals. This finding suggests that bilingual phonemic category organization may be different from that of monolinguals, likely as a result of the fact that they have to account for the phonemic categories in both of their languages (Curtin, Byers-Heinlein, & Werker, 2011; Werker, Byers-Heinlein, & Fennell, 2011). Furthermore, the large variability in boundary values observed in the bilingual group may be related to variability in their speech input (see, e.g., McCarthy et al., 2013). That is, the children with shorter phoneme boundaries may have been exposed to more Sylheti or Sylheti-accented input, and vice versa for the children with a more native-like English phoneme boundary (Cristià, 2011).

However, although the findings at Time 1 indicate that initially our bilingual children may have been using their L1 (Sylheti) categories when perceiving and producing English plosives, the results at Time 2 demonstrate that they were able to establish L2 phonemic categories that matched those of native speakers. Yet, to achieve this native-like perception and production the children in our study required more than just the small amount of

ambient English exposure at home and within the community. It seems that an accumulation of English experience, in full-time education with no Sylheti support, between Time 1 and Time 2 was required in order for the bilingual children to acquire the L2 phonemic categories. In addition, there was evidence for a developmental trend common to both monolingual and bilingual children; all children displayed an increase in the steepness of their categorization slopes from Time 1 to Time 2. These findings are consistent with previous studies that have shown that phonemic categories become refined with age (e.g., Hazan & Barrett, 2000; Mayo & Turk, 2004; Nittrouer, 2005). Our findings thus support the idea that language specification continues to develop beyond the 1st year of life and that this is facilitated by linguistic experience (e.g., attunement theory; Aslin & Pisoni, 1980).

Interestingly, monolingual and bilingual children had similar categorization slopes for /p/-/b/ at Time 1. This is surprising, because as for the /k/-/g/ continuum, we would have expected our monolinguals to have had a steeper categorization slopes at Time 1, particularly as the acquisition of bilabials is thought to precede 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; Werker & Curtin, 2005). In our data, what we might be seeing is a pattern specific to our children, and which reflects the structure of their lexicon. That is, the monolingual children may have more words in their lexicon that contain /g/ or /k/ than /p/ or /b/, resulting in more refined phonemic categories for velar plosives. Inspection of tokens produced by 3- to 5-year-old children from the British English CHILDES corpora (MacWhinney, 2000) supports this hypothesis (see online Appendix S3 for token count). Overall, the children produced more tokens containing velar than bilabial plosives, suggesting that the velar plosive category is more advanced for monolingual British English children within our target age group.

Another explanation for the differences in categorization observed between *pea-bee* and *coat-goat*, though, is the phonetic environment. If, as suggested by the 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 acoustically less 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 thus 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).

Finally, while there are differences in performance between monolingual and bilingual children, our data suggest that the development of perception and production of English plosives in sequential bilinguals follows a similar trajectory to that of their monolingual peers. That is, as perceptual categorization becomes more refined, so does the children's realization of English voicing patterns in their 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 1 year, suggest that there is at least some relation between perception and production. Like studies of second language acquisition in adults, our results suggest that perception accuracy is correlated with production accuracy (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997): As our bilingual children's representations became more refined (i.e., steeper categorization slope), they produced more native English-like plosives.

In sum, this study revealed that (a) language experience facilitates the development of phonemic categorization, for both monolingual and sequential bilingual children, and (b) children are initially sensitive to the ambient speech input, but with language experience, are able to acquire new phonemic categories. Further cross-linguistic research on different phonemic contrasts is needed to establish whether such findings are universal across phonemic contrasts, and are common to sequential bilinguals acquiring other languages. In addition, to fully understand how sequential bilinguals organize their phonemic categories, it will be necessary to

investigate the development of production and perception in both the L1 and L2.

References

Aslin, R., & Pisoni, D. (1980). Some developmental processes in speech perception. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Vol. 2. Perception* (pp. 67–96). New York, NY: Academic Press.

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, *48*, 1–27. doi:10.1177/00238309050480010101

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*, 183–216. doi:10/1016/j.bbi.2008.05.010

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

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, *66*, 1001–1017. doi:10.1121/1.383319

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer*. Retrieved 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. doi:10.1017/S0142716400008912

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29–49. doi:10.1207/S15327078IN0201/_3

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*, 217–243. doi:10.1177/00238309030460020801

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399. doi:10.1146/annurev.psych.53.100901.135233

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310. doi:10.3758/2FBF03206911

Camden Council. (2012). *Camden children and young people's profile* (Autumn report).

Cristià, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *Journal of the Acoustical Society of America*, *129*, 3271–3280. doi:10.1121/1.3562562

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*, 492–504. doi:10.1016/j.wocn.2010.12.002

Darcy, I., & Krüger, F. (2012). Vowel production and perception in Turkish children acquiring L2 German. *Journal of Phonetics*, *40*, 568–581. doi:10.1016/j.wocn.2012.05.001

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A., (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. Hove, UK: Psychological Corporation.

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

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

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. doi:10.1044/1058-0360(2009/08-0036)

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. (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., Bohn, O.-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470. doi:10.1006/jpho.1997.0052

Flege, J., Munro, M., & MacKay, I. (1995). The effect of age of second language learning on the production of English consonants. *Speech Communication*, *16*, 1–26. doi:10.1016/0167-6393(94)00044-B

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, R. C., Conoboy, T. B., Romo, H., Klarman, L., . . . Kuhl, K. P. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*, 546–557. doi:10.1016/j.wocn.2011.07.002

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. doi:10.1017/s0305000900004414

Hazan, V., & Barrett, R. (2000). The development of phonemic categorization in children aged 6-12 years. *Journal of Phonetics*, *28*, 377–396. doi:10.1006/jpho.2000.0121

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*, *52*, 1510–1529. doi:10.1044/1092-4388(2009/08-0220)

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*, 1–27. doi:10.1017/S0305000910000759

Howell, P., Rosen, S., Lang, H., & Sackin, S. (1992). The role of F1 transitions in the perception of voicing in initial plosives. *Speech, Hearing, and Language: Work in Progress*, *6*, 118–126.

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. doi:10.1016/S0010-0277(02)00198-1

Johnson, C., & Wilson, I. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, *6*, 271–289. doi:10.1177/13670069020060030401

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

Khan, S. D. (2010). Bengali (Bangladeshi Standard). *Illustrations of the IPA*, *40*, 221–225.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of Acoustical Society of America*, *67*, 737–793. doi:10.1121/1.383940

Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843. doi:10.1038/nrn1533

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*, 979–1000. doi:10.1098/rstb.2007.2154

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.

Levitt, H. (1971). Transformed up-down methods in psycho-acoustics. *Journal of the Acoustical Society of America*, *49*, 467–477. doi:10.1121/1.1912375

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.

Macken, M., & Barton, D. (1979). The acquisition of voicing contrast in English: A study of voice onset time in word initial stop consonants. *Journal of Child Language*, *7*, 433–458. doi:10.1017/S030500090007029

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

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*, *155*, 3184–3194. doi:10.1121/1.1738838

McCarthy, K. M. (2009). *Accounting for input: The development of a home language environment questionnaire* (Unpublished master's thesis). University College London.

McCarthy, K. M., Evans, B., & Mahon, M. (2013). Acquiring a second language in an immigrant community: The production of Sylheti and English stops and vowels by London-Bengali speakers. *Journal of Phonetics*, *41*, 344–358. doi:10.1016/jwocn.2013.03.006

National Association for Language Development in the Curriculum. (2013). *EAL pupils: 1997–2013*. Retrieved from http://www.naldic.org.uk/research-and-information/eal-statistics/eal-pupils

Nathan, L., & Wells, B. (2001). Can children with speech difficulties process an unfamiliar accent? *Applied Psycholinguistics*, *22*, 343–361. doi:10.1017/S0142716401003046

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*, 278–297.

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*, 1072–1088. doi:10.1121/1.1940508

Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, *101*, 2253–2266.

Office of Qualifications and Examinations Regulation. (2004). *National qualifications framework: Level descriptors*. Retrieved from http://ofqual.gov.uk/qualifications-and-assessments/qualification-frameworks/

Ohde, R. N., & German, S. R. (2011). Formant onsets and formant transitions as developmental cues to vowel perception. *Journal of the Acoustical Society of America*, *130*, 1628–1642. doi:10.1121/1.3596461

Ohde, R. N., & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4- year-old children. *Journal of the Acoustical Society of America*, *102*, 3711–3722. doi:10.1121/1.420135

Pupil Level Annual School Census. (2013). *Pupil level annual school census 2013*. Retrieved from http://wales.gov.uk/topics/educationandskills/schoolshome/schooldata/ims/datacollections/pupillevelannualschoolcensus/?lang=en

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. doi:10.1093/brain/awg076

Rasinger, S. M. (2007). Bengali-English in east London: A study in urban multilingualism. In G. Davis & K. A. Bernhardt (Eds.), *Contemporary studies in descriptive linguistics 11*. Oxford, UK: Peter Lang.

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, *12*, 583–601. doi:10.1111/j.1467-7687.2009.00809.x

Sebastián-Gallés, N., & 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*, *12*, 874–887. doi:10.1111/j.1467-7687.2009.00829.x

Stevens, K. N., & Klatt, D. H. (1974). The role of formant transitions in the voiced-voiceless distinction of stops. *Journal of the Acoustical Society of America*, *55*, 653–659. doi:10.1121/1.1914578

Sundara, M., Polka, L., & Molna, L. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*, 232–242. doi:10.1016/j.cognition.2007.12.013

Sundara, M., & Scuttellaro, A. (2011). Rhythmic difference between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*, 505–513. doi:10.1016/jwocn.2010.08.006

Thompson, M., & Hazan, V. (2010). *The impact of visual cues and lexical knowledge on the perception of a non-native contrast for Columbian adults*. Presented at the International Symposium on the Acquisition of Second Language Speech, New Sounds 2010, Poznan, Poland.

Tower Hamlets. (2012). *Spring school census 2012*. London, UK: London Borough of Tower Hamlets.

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*, 263–290. doi:10.1016/j.wocn.2004.10.002

Vihman, M. (1996). *Phonological development*. Oxford, UK: Blackwell.

Werker, J. F., Byers-Heinlein, K., & Fennell, C. T. (2011). Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society*, *364*, 3649–3663. doi:10.1098/rstb.2009.0105

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework for infant speech processing. *Language Learning and Development*, *1*, 197–234. doi:10.1207/s15473341lld0102

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63. doi:10.1016/S0163-6383(84)80022-3

Xu Rattansone, N., & Demuth, K. (2013). The acquisition of coda consonants by Mandarin early child L2 learners of English. *Bilingualism: Language & Cognition*, *17*, 646–659. doi:10.1017/S1366728913000618

Zlatin, 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. doi:10.1044/jshr.1901.93

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Stimuli: Voicing Continua

Appendix S2. Adaptive 1 Up/2 Down Levitt Procedure: Tracking Example

Appendix S3. Token Count Summary of British English CHILDES Corpora