



Cardiff  
Metropolitan  
University

Prifysgol  
Metropolitan  
Caerdydd

# Inter-generational transmission in a minority language setting: Stop consonant production by Bangladeshi heritage children and adults

**Authors: Robert Mayr & Aysha Siddika**

**This is a pre-publication version (post peer review).**

**To cite this article:**

**Mayr, R. & Siddika, A. (2016, accepted). Inter-generational transmission in a minority language setting: Stop consonant production by Bangladeshi heritage children and adults. *International Journal of Bilingualism*, DOI:**

**10.1177/1367006916672590**

## **Abstract**

### **Aims and Objectives:**

The purpose of this study was to gain a better understanding of speech development across successive generations of heritage language users, examining how cross-linguistic, developmental and sociocultural factors affect stop consonant production.

### **Design:**

To this end, we recorded Sylheti and English stop productions of two sets of Bangladeshi heritage families: (1) first-generation adult migrants from Bangladesh and their (second-generation) UK-born children, and (2) second-generation UK-born adult heritage language users and their (third-generation) UK-born children.

### **Data and analysis:**

The data were analysed auditorily, using whole-word transcription, and acoustically, examining voice onset time. Comparisons were then made in both languages across the four groups of participants, and cross-linguistically.

### **Findings:**

The results revealed non-native productions of English stops by the first-generation migrants but largely target-like patterns by the remaining sets of participants. The Sylheti stops exhibited incremental changes across successive generations of speakers, with the third-generation children's productions showing the greatest influence from English.

### **Originality:**

This is one of few studies to examine both the host and heritage language in an ethnic minority setting, and the first to demonstrate substantial differences in heritage language accent between age-matched second- and third-generation children. The study shows that current theories of bilingual speech learning do not go far enough in explaining how speech develops in heritage language settings.

### **Implications:**

These findings have important implications for the maintenance, transmission and long-term survival of heritage languages, and show that investigations need to go beyond second-generation speakers, in particular in communities that do not see a steady influx of new migrants.

## **1. Introduction**

A growing body of research has shown that individuals raised in an ethnic minority setting develop different pronunciation patterns from the generation of their foreign-born parents (e.g., Evans, Mistry, Moreiras, 2007; McCarthy, Evans & Mahon, 2011, 2013; Nagy, 2015; Nagy & Kochetov, 2013; Sharma & Sankaran, 2011). Accordingly, second-generation heritage speakers commonly exhibit non-native features in the heritage language, and therefore tend to be perceived as foreign-accented in it (Lein, Kupisch, van de Weijer, 2015; Kupisch, Barton, Hailer, Klaschik, Stangen, Lein, van de Weijer, 2014; Nagy, 2015; Nagy & Kochetov, 2013). At the same time, they usually behave much like their monolingual peers in the host language (Evans, Mistry & Moreiras, 2007; Khattab, 2002; McCarthy et al., 2011; McCarthy, Mahon, Rosen & Evans, 2014), although this is not always the case (Stangen, Kupisch, Ergün & Zielke, 2015), and heritage-language markers, such as retroflex realisations of English /t/ (Alam & Stuart-Smith, 2011; Heselwood & McChrystal, 1999, 2000; Kirkham, 2011; Sharma & Sankaran, 2011), may be retained to fulfil socio-indexical functions. In contrast, first-generation migrants, in particular those who arrived in the host country as adults, tend to have a distinct foreign accent in their L2 (Abrahamsson & Hyltenstam, 2009; Flege, Munro & McKay, 2005; Flege, Schirru & McKay, 2003; McCarthy et al., 2011; Piske, MacKay & Flege, 2001), whilst retaining a relatively authentic accent in their native language (Khattab, 2002; McCarthy et al., 2013).

Little is known, however, about the speech development of subsequent generations of heritage language users (but see Hrycyna, Lapinskaya, Kochetov & Nagy, 2011 and Nagy & Kochetov, 2013). Do second- and third-generation children in language minority settings differ from each other in their pronunciation of the heritage language and the host language? If so, how do the differences manifest, and can they be explained on the basis of their parents' production patterns? The present study aims to address these questions by

investigating the stop consonant productions of two sets of Bangladeshi heritage families: (1) first-generation female migrants from the Sylhet area of Bangladesh and their UK-born children, and (2) second-generation UK-born female Sylheti heritage language users and their (third-generation) children. In so doing, it aims to disentangle the effects of cross-linguistic, developmental and sociocultural factors.

### ***1.1. Background***

There is a general consensus that the earlier one starts to learn a second language, the less foreign-accented it will be (Au, Knightly, Jun & Oh, 2002; Flege et al., 1995, 2003; Flege, Yeni-Komshian, & Liu, 1999; Yeni-Komshian, Flege & Liu, 2000). Accordingly, individuals acquiring an L2 in adolescence or adulthood virtually always end up with some degree of foreign accent (see Hansen Edwards & Zampini (2008) for an overview), while this is much less likely in those with early exposure to the language (MacLeod, Stoel-Gammon & Wassink, 2009; Kupisch et al., 2014). A number of explanations have been given for these findings. Some have argued for a maturationally-defined critical period (Lenneberg, 1967; Long, 1990; Scovel, 1988, 2000). However, contrary to these claims, there is evidence that native-like proficiency, while rare, is not impossible for late L2 learners (Bongaerts, Mennen & Slik, 2000; Birdsong, 2007; Mennen, 2004). Moreover, the correlation between age of onset of learning and degree of foreign accent is linear without any marked discontinuities (Flege, 1995; Flege et al., 1995). This has led many to abandon maturation-based accounts, and instead to explain age effects on the basis of extra-linguistic factors, such as L1 and L2 usage patterns (Flege, Frieda & Nozawa, 1997; Flege, MacKay & Piske, 2002; Piske et al., 2001; Yeni-Komshian et al., 2000).

Generally, heritage language users are at an advantage over L2 learners in terms of the accuracy of their pronunciation patterns (Au et al., 2002; Chang, Yao, Haynes & Rhodes,

2011; Kupisch et al., 2014; Oh, Jun, Knightly & Au, 2003). Chang et al. (2011), for example, showed that Mandarin heritage speakers in the United States consistently produced greater contrastivity in cross-linguistically similar back vowels, stops and fricatives than native American L2 learners of Mandarin. Similarly, Kupisch et al. (2014) demonstrated that heritage language speakers were perceived to be significantly less foreign-accented in their minority language than L2 learners, although their accent in the majority language was more native-like than in the minority language.

These patterns have been explained on the basis of differences in linguistic experience. While the L1 sound system of late L2 learners is fully in place when L2 learning starts, heritage language children usually have experience with the minority language from birth, or shortly thereafter, and the majority language by the time they start compulsory education. They are often initially dominant in the minority language, in particular if the language is also widely used in the community. However, with the onset of mainstream education in the majority language, there is typically a shift in dominance, with the use of the minority language frequently becoming more restricted (de Houwer, 2009). McCarthy (2015) and McCarthy et al. (2014), for example, showed that Bangladeshi heritage children's perception and production of English /p b k g/ was heavily influenced by Sylheti during their first year in an English-speaking nursery, but was much more like that of their monolingual peers a year later.

Nevertheless, early language exposure does not guarantee native-like accents in the heritage language (cf. Kupisch et al., 2014; McCarthy et al., 2013; Oh et al., 2003). For example, Oh et al. (2003) showed that childhood speakers of Korean who had stopped using the heritage language upon school entry were foreign-accented in it. Similarly, Kupisch et al. (2014) reported that heritage language users in Germany, France and Italy with exposure to both languages from birth were rated as foreign-accented in the minority language. These

patterns are also reflected in studies examining speech production. McCarthy et al. (2013), for example, showed that second-generation London Bengalis produced non-native VOT patterns in their minority language, and Nagy & Kochetov (2013) revealed incremental changes in the heritage language VOT patterns of successive generations of Russian and Ukrainian speakers in Toronto in the direction of English (see also Hrycyna et al., 2011 and Nagy, 2015). Interestingly, Italian heritage speakers in the study did not show this pattern. The authors speculate that these differences may be a result of the greater community support offered to ethnic Italians than Russians and Ukrainians in Toronto, including dedicated language classes.

In contrast to the minority language, heritage language users are usually native-like in the host language. For example, Kupisch et al. (2014) found no difference in an accentedness rating between their monolingual speakers and the heritage language speakers in the host language. Likewise, the Gujarati heritage speakers in Evans et al., (2007) produced their English vowels much like their monolingual English peers, and the second-generation Bengali heritage speakers in McCarthy et al. (2013) did not differ from monolingual controls in their production of English vowels and VOT. Nevertheless, the host language is not always immune to non-native patterns. For example, in Darcy & Krüger's (2012) study, 10-year-old Turkish heritage children living in Germany whose first exposure to German was between 2 and 4 years of age were less accurate in the perception of some German vowel contrasts than monolingual German-speaking children. Similarly, Stangen et al. (2015) found highly variable patterns in their study on global foreign accent in Turkish-German heritage language users from Germany: the majority were perceived to have a foreign accent in either the host language or the heritage language (n=13), while some were foreign-accented in neither language (n=3), and others in both (n=5).

Where non-native forms occur in heritage language speakers, they may be a result of inadvertent cross-linguistic interactions. According to the *Speech Learning Model* (SLM; Flege, 1995; Flege et al., 2003), this happens when cross-linguistically similar L1 and L2 sounds are perceptually equated with each other, a phenomenon termed *equivalence classification*. One of the best-known examples of this phenomenon is the difficulty that Japanese learners face with the perception and production of English /l/ and /r/, which they tend to assimilate to their single Japanese category /r/ (e.g., Aoyama, Flege, Guion, Akahane-Yamada & Yamada, 2004). Alternatively, where bilinguals are able to perceive a difference between L1 and L2 categories, they may strive to increase cross-linguistic distinctiveness. For example, the early Italian-English bilinguals in Flege et al.'s (2003) study produced English /eɪ/ with exaggerated vowel-inherent spectral change to keep it maximally distinct from monophthongal Italian /e/. Both mechanisms may lead to patterns that differ from those produced by monolingual speakers. According to the SLM, the likelihood that cross-linguistically similar sounds are distinguished is greater in early than late bilinguals since the L1 sound system is less established in younger learners, and hence more amenable to reorganisation. This may explain why heritage language speakers tend to outperform L2 learners. Similar explanations are offered by other theories of L2 speech learning. The *Perceptual Assimilation Model* (PAM: Best, 1995; PAM-L2: Best & Tyler, 2007), for instance, predicts difficulties in L2 perception on the basis of the assimilability of non-native contrasts to native categories.

In addition to cross-linguistic interactions, non-native forms in heritage language settings may arise from socio-cultural factors and form part of contact varieties (Alam & Stuart-Smith, 2007; Heselwood & McChrystal, 1999, 2000; Hirson & Sohail, 2007; Kirkham, 2011; Kirkham & Wormald, 2015; Lambert, Alam & Stuart-Smith, 2007; Sharma & Sankaran, 2011). Kirkham (2011), for example, argued that British Asians from Sheffield

used retroflex realisations of English /t/ to signal their Asian identity, rather than inadvertently as a result of cross-linguistic interactions, since the use of these forms could not be predicted on the basis of their language use patterns, with even monolingual English speakers from the community using them. Sharma & Sankaran (2011), in turn, examined the acquisition of a native feature, /t/ glottaling, and a non-native feature, /t/ retroflexion, in British Asians from London. They found that younger second-generation speakers used /t/ retroflexion in English to signal their Asian identity, while older second-generation speakers followed first-generation speakers' non-native use of /t/ retroflexion, but unlike them, used /t/ glottaling natively. The authors argue that these patterns are consistent with a socially oriented model that allows for incremental changes to take place, rather than a cognitively oriented one which claims that non-native forms are either innately blocked by an accent filter (Chambers, 2002), or reused by native speakers to mark their identity.

### ***1.2. The present study***

This study investigated stop consonant production in Sylheti-English bilingual children and adults from Bangladeshi heritage backgrounds in Cardiff, South Wales, and as such is the first to examine the speech of ethnic minorities in Wales. Compared with the London Bengali communities in Tower Hamlets, where 30% of the population are of Bangladeshi origin, and Camden, where they constitute the largest minority ethnic group (cf. McCarthy et al., 2013, 2014 for details), the Bengali community in Cardiff is relatively small. In the 2011 Census, some 0.3% of the population of Wales (or 10,687 individuals) considered themselves British Bangladeshis, with 5207 individuals indicating Sylheti as their main home language (Office for National Statistics, 2011). About half of these live in Cardiff, in particular in the areas of Riverside and Grangetown. These communities have a close-knit social structure, including

shops, restaurants and community centres, but, unlike those in Tower Hamlets and Camden, do not witness a steady influx of new arrivals from Bangladesh.

Of the approximately 500,000 British Bengalis, some 95% originate from the rural area of Sylhet in north-eastern Bangladesh (Chalmers & Miah, 1996), where Sylheti is spoken. Sylheti is typologically related to Standard Bengali (SB), but the two languages are not mutually intelligible (Chalmers & Miah, 1996; Rasinger, 2007). While native speakers of Sylheti, including first-generation migrants, are largely competent in SB, the language of education, this is not the case for most UK-born heritage speakers (Chalmers & Miah, 1996).

On the whole, Sylheti has a less complex phonological system than SB, with fewer consonant and vowel categories (Chalmers & Miah, 1996; Gope & Mahanta, 2015; Khan, 2010). Hence, while SB contains sixteen stop categories that systematically contrast in voicing and breathiness<sup>1</sup>, Sylheti only contains nine (Gope & Mahanta, 2015; McCarthy et al., 2013, 2014). These include the voiced breathy stops /b<sup>h</sup>/ and /g<sup>h</sup>/, the voiced non-breathy stops /b̥ d̥ d̥ g/ and the voiceless stops /t̪ t̪ k/.

A small number of acoustic studies have been carried out on Sylheti stops. Gope and Mahanta (2015) examined voiced stop productions with and without underlying breathiness by adult native speakers of Sylheti from India. They found no differences in VOT as a function of breathiness, with all categories realised with a voicing lead. McCarthy et al. (2013) revealed similar patterns for first-generation Bangladeshis who arrived in the UK in their late teens or in adulthood, and for their native Sylheti control speakers; in contrast, early arrivals (i.e. arrival in the UK < 16;0) and second-generation heritage speakers produced Sylheti voiced stops with significantly longer VOT values and less prevoicing, and hence more English-like. Voiceless stops, in turn, were produced within the short-lag range by all speaker groups (native control mean: /t̪/: 26 ms, /t̪/: 14 ms, /k/: 16 ms). This contrasts with

other varieties, such as Dhaka Bengali, where /t<sup>h</sup>/ and /t̪<sup>h</sup>/ are realised with long-lag VOT values, with mean values between 50 and 100 ms (Maxwell, Baker, Bundgaard-Nielson & Fletcher, 2015).

No previous work on children's acquisition of Sylheti stops is available. Studies from other languages suggest that monolingual children acquire the stop voicing contrast earlier in languages that distinguish short-lag and long-lag VOT categories, like English, than in languages with a distinction between lead voice and short-lag VOTs, like Sylheti (Allen, 1985; Bortolini, Zmarich, Fior & Bonifacio, 1995; Gandour, Petty, Dardarananda, Dechongkit & Munkgoen, 1986; Khattab, 2000; Macken & Barton, 1979, 1980). Indeed, the acquisition of lead voicing seems to be a particularly protracted process, perhaps due to its aerodynamic challenges (Cho & Ladefoged, 1999; Van Alphen & Smits, 2004), with children as old as 7;0 struggling to use it consistently (Khattab, 2000). Studies on bilingual and multilingual children, in turn, have shown cross-linguistic interactions (Deuchar & Clark, 1996; Fabiano-Smith & Bunta, 2012; Heselwood & McChrystal, 2000; Kehoe, Lléo & Rakow, 2004; Khattab, 2000; Mayr & Montanari, 2015; Simon, 2010). For example, the Dutch-English bilingual child studied by Simon (2010) realised Dutch /p/ and /t/ with long-lag VOT values, instead of target short-lag ones, after extensive exposure to English. Similarly, Heselwood & McChrystal's (2000) study revealed greater use of prevoicing in English voiced stops produced by Punjabi-English bilingual children than by their monolingual English peers. The only study to examine stop consonants in Bangladeshi heritage children from a Sylheti-speaking community is McCarthy et al. (2014; see also McCarthy, 2015). This study revealed changes during the first year of school in the children's production and perception of English /p b k g/ in the direction of their monolingual peers' patterns. No data on the children's Sylheti stops were collected, however.

The purpose of the present study was to extend existing work on the speech of Bangladeshi heritage speakers in the UK by investigating the production of stop consonants in Sylheti and English by second- and third-generation children and their mothers.

## **2. Method**

### **2.1. Participants**

Two sets of Bangladeshi heritage families participated in the study. All demographic data were established using a parental questionnaire. The first family consisted of first-generation female immigrants from the Sylhet area of Bangladesh (GEN 1 MUMS: N=5) and their children (GEN 2 CHILDREN: N=10). The GEN 1 MUMS had a mean age of 35 years (SD: 7.65) at the time of data collection and had been residing in the UK for an average of 12.2 years (SD: 7.29), having arrived in the UK in their late teens or in early adulthood. Sylheti was their predominant home and community language, and they reported mainly consuming Sylheti-language media (cf. Table 1 for details). Their children, in contrast, ranged in age between 3;7 and 5;0 (mean age: 4;5) and were all born in the UK. Their use of Sylheti and English in the home was approximately equal, while they only spoke Sylheti in the community 23% of the time, and never consumed Sylheti-language media. These families live in the Riverside and Grangetown areas of Cardiff, which have a large concentration of Bangladeshi heritage speakers, and spend an average of one week per year in Bangladesh to visit relatives.

The second set of Bangladeshi heritage families encompassed second-generation female heritage speakers of Sylheti (GEN 2 MUMS: N=6) and their children (GEN 3 CHILDREN: N=10). The GEN 2 MUMS had all been born in the UK to first-generation migrants from Bangladesh. They had a mean age of 31 years (SD: 3.08) at the time of data collection and did not differ significantly in age from the GEN 1 MUMS ( $t(5.078) = 1.189$ ,

p=.287). They reported using Sylheti in the home 23% of the time and in the community 13% of the time. Their children (GEN 3 CHILDREN) were also all born in the UK, and were the same age as the GEN 2 CHILDREN (Mean: 4;3; range: 3;8-4;11;  $t(18) = .831, p = .417$ ). This group exhibited the lowest Sylheti usage patterns, with only 5% of conversations in the home and 9% in the community carried out in the minority language. In contrast to the first-generation mothers and their children, these families do not live in communities with a large concentration of Bangladeshi heritage speakers, and virtually never travel to Bangladesh to visit relatives. None of the participants reported any speech, language and communication difficulties. All had normal hearing.

	Age (years; months)	Age of arrival (years)	Length of residence (years)	*Mean % Sylheti at home	*Mean % Sylheti in community	*Mean % Sylheti media
GEN 1 MUMS	35 (27-44)	23 (16-40)	12 (4-24)	93	60	92
GEN 2 CHILDREN	4;5 (3;7-5;0)	Born in UK	-	50	23	0
GEN 2 MUMS	31 (27-36)	Born in UK	-	24	13	0
GEN 3 CHILDREN	4;3 (3;8-4;11)	Born in UK	-	5	9	0

\*Estimated language use patterns; mothers reported for their children.

Table 1: *Participant information.*

## 2.2. Materials

This study aimed to assess all stop consonant categories that occur word-initially in Sylheti and English. Table 2 depicts the materials used in the study. They include monosyllabic and bisyllabic words starting with a singleton bilabial, coronal or velar stop in the onset. Words were chosen with which young children and adults were expected to be familiar, and which

could be elicited via pictorial representation. The English dataset comprised the categories /p b t d k g/, the Sylheti dataset the categories /p p<sup>h</sup> b b<sup>h</sup> t̪ t̪̪ d̪ d̪̪ k k<sup>h</sup> g g<sup>h</sup>/. Note that the latter included three categories that previous research had shown to be realised as fricatives by Sylheti speakers, but that historically constitute stops, i.e. /p/, /p<sup>h</sup>/ and /k<sup>h</sup>/ (Gope & Mahanta, 2015; McCarthy et al., 2013).

Sylheti			English		
TARGET SOUND	TARGET WORD	TRANSCRIPTION WITH GLOSS	TARGET SOUND	TARGET WORD	TRANSCRIPTION
<i>Bilabial</i>					
/p/	ফাটা	/paṭa/ ‘leaf’	/p/	park	/pak/
/p <sup>h</sup> /	পুল	/p <sup>h</sup> ul/ ‘flower’			
/b/	বাগ	/bag/ ‘tiger’	/b/	bees	/biz/
/b <sup>h</sup> /	বাড়ি	/b <sup>h</sup> ari/ ‘heavy’			
<i>Coronal</i>					
/t̪/	তিন	/t̪in/ ‘three’	/t/	teeth	/tiθ/
/t̪̪/	টুট	/t̪̪ut/ ‘lips’			
/d̪/	দাত	/d̪at/ ‘teeth’	/d/	deep	/dip/
/d̪̪/	ঢাথ	/d̪̪ax/ ‘call’			
<i>Velar</i>					
/k/	কুড়া	/kutta/ ‘dog’	/k/	keys	/kiz/
/k <sup>h</sup> /	খাম	/k <sup>h</sup> am/ ‘work’			
/g/	গাল	/gal/ ‘cheek’	/g/	geese	/gis/
/g <sup>h</sup> /	ঘাস	/g <sup>h</sup> af/ ‘grass’			

Table 2: *Experimental materials*

### 2.3. Procedure

Data collection took place in a quiet room in the participants’ homes. Each participant was recorded twice, once in a Sylheti session, and once in an English one, with the two sessions separated by several days. This procedure was adopted to minimise the likelihood of dual

language activation (Grosjean, 2001). Recordings were made using a Zoom H2 Handy Recorder with integrated condenser microphone, which was positioned a few centimetres from the participant's mouth (sampling rate: 44.1 kHz; 16-bit resolution). Each session commenced with a brief conversation in the target language with the experimenter, a UK-born Sylheti-English bilingual. This was followed by a picture-naming task that aimed to elicit three instances of each target word produced at a natural pace in a carrier phrase (English: *X I say*; Sylheti:  $\text{হে} \text{খই}$  /heə xɔɪ/ ('X he says')). This procedure yielded  $3 \times 6 = 18$  tokens of the English stops and  $3 \times 12 = 36$  tokens of the Sylheti stops from each participant, giving a total of 1674 tokens. No formal assessment of the children's lexical knowledge was carried out, but almost all items could be elicited spontaneously. In the few instances where this was not possible, semantic prompts were given, and if these were unsuccessful, the target words were modelled by the experimenter. No attempts were made to elicit stop consonants in isolation.

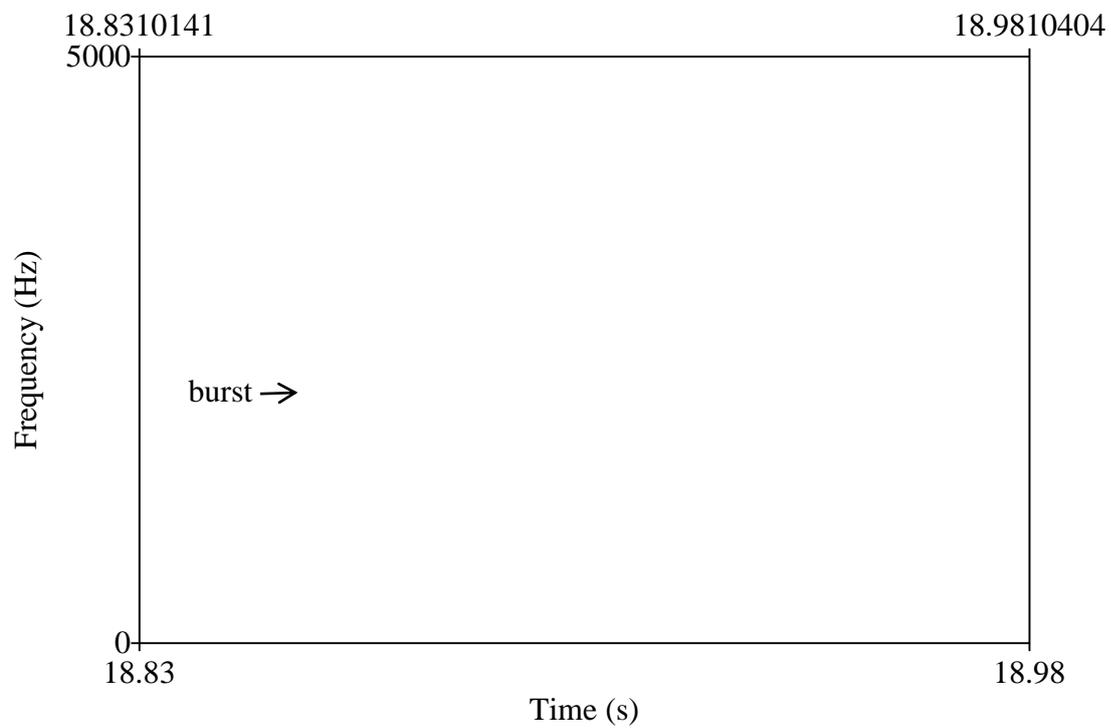
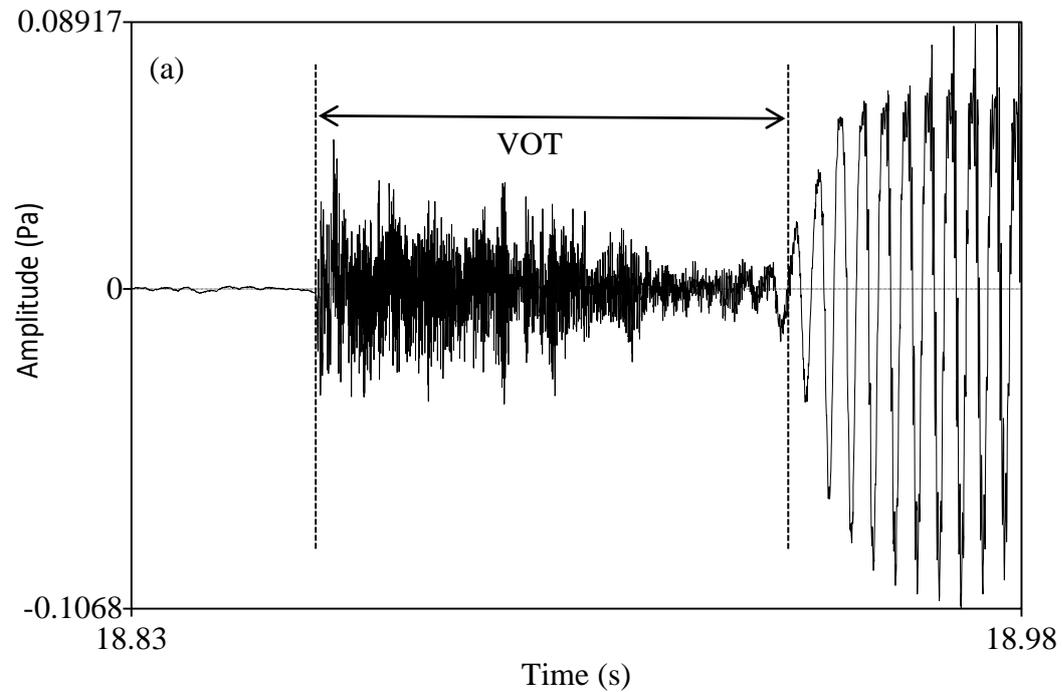
#### **2.4. Analysis**

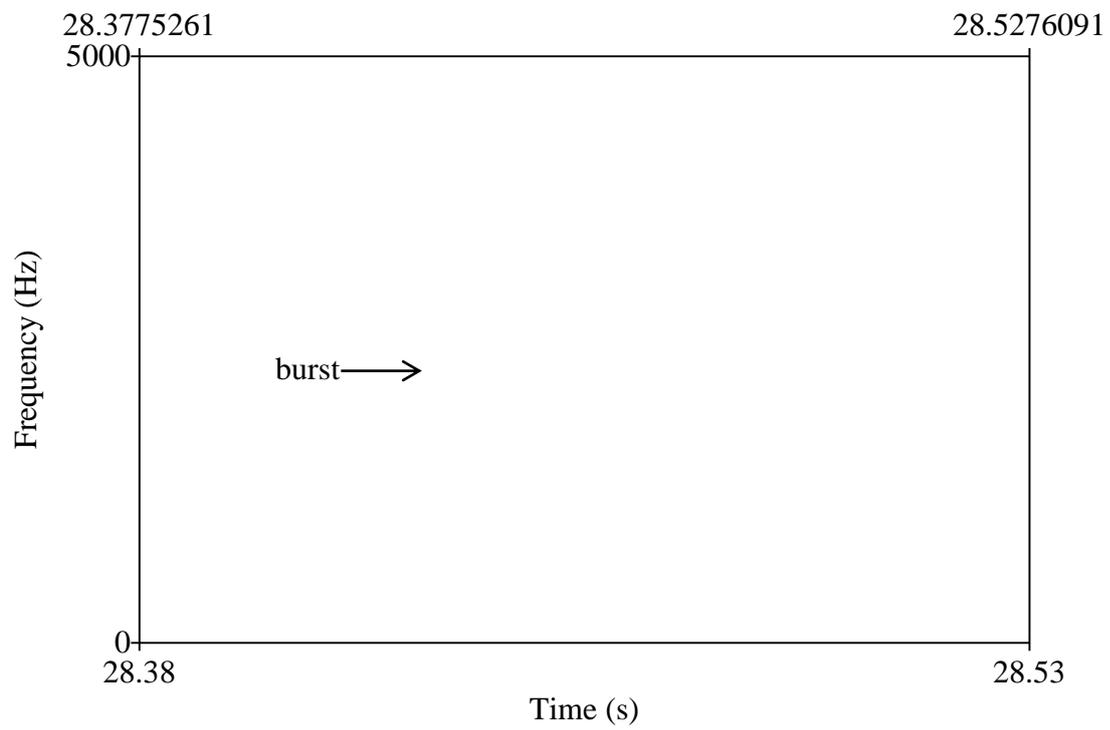
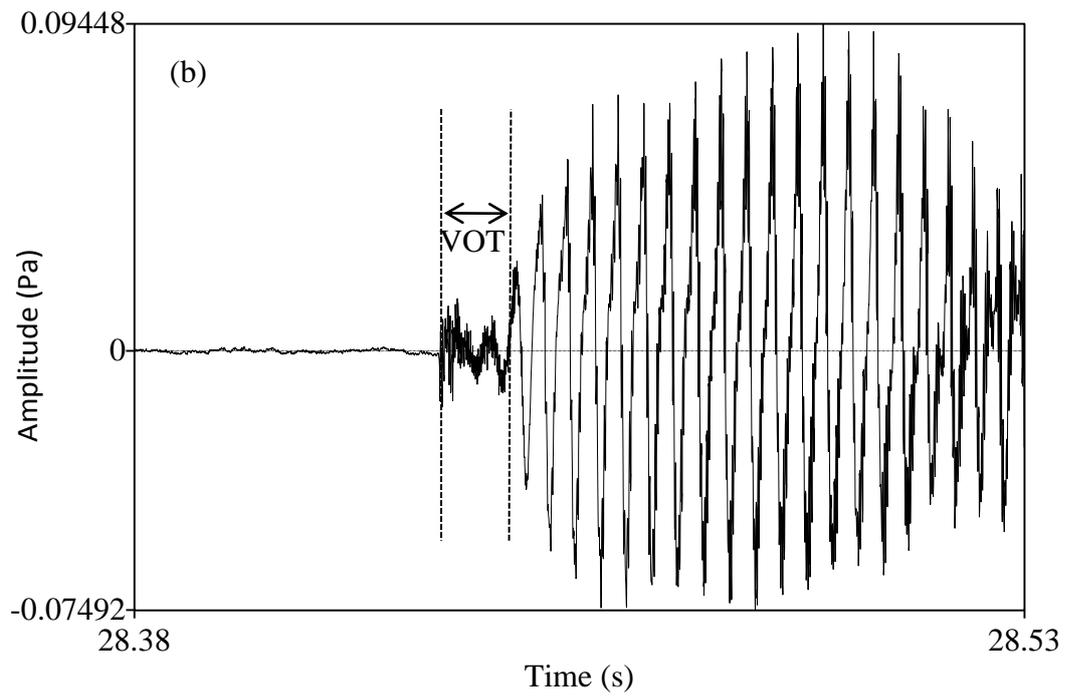
Many studies have examined stop consonants acoustically (e.g., Bortolini et al., 1995; Kirkham, 2011; Mayr & Montanari, 2015; Sundara, 2005). While temporal measures, such as voice onset time (VOT), allow for direct comparisons between child and adult participants, this is not the case for spectral measures that aim to assess differences in place of articulation as they vary with vocal tract size. Moreover, the relation between acoustic properties for place of articulation and breathiness, and their articulatory and perceptual correlates is complex. For example, differences in spectral shape may be due, not to differences in place of articulation, but variations in the degree of damping of the active articulator (Sundara, 2005).

For these reasons, we opted for a two-way approach in the present study. First, the materials were analysed auditorily. This involved all target words being transcribed in broad phonetic transcription by a phonetically-trained Sylheti-English bilingual, using the symbols of the International Phonetic Association (International Phonetic Association, 2005). This analysis focused on establishing the place of articulation of each stop production. This was particularly critical for coronal stops. In addition, it assessed the presence or absence of breathiness in voiced stops. Only tokens that conformed to the reported adult forms in native Sylheti and English stops (Chalmers & Miah, 1996; Docherty, 1992; Gope & Mahanta, 2015; McCarthy et al. 2013) were classified as target-like<sup>2</sup>. As a measure of reliability, the entire dataset was independently reanalysed by a second phonetically-trained researcher with no prior knowledge of Sylheti or related languages. Cohen's  $\kappa$  was run to determine if there was agreement between the two sets of transcriptions. The results revealed substantial agreement ( $\kappa = .682$  (95% CI, .629 to .735),  $p < .0005$ ), based on Landis & Koch's (1977) classification. Any differences in the two sets of transcriptions were resolved by consensus. Uncertainty remained on one token of English /g/ and one token each of Sylheti /b<sup>h</sup>/, /t̪/ and /d/. These tokens were removed from further analysis.

To assess voicing, we analysed the participants' VOT patterns acoustically, using PRAAT software (Boersma & Weenink, 2010). Measurements were taken from the release burst of each token, signalled by a sharp peak in waveform energy, to the onset of voicing of the following vowel, as marked by the zero crossing of the first glottal pulse for modal voicing (cf. Figure 1a, 1b). Tokens that displayed more than one transient were measured from the first visible release burst. If voicing occurred during the closure period, VOT was measured from the point at which vocal fold vibration could be discerned in the waveform, together with aperiodic wide-band energy in the spectrograms, up to the first release burst (cf. Figure 1c). The onset of lead voicing was established visually. Tokens where this could not

be determined clearly were excluded from the VOT analysis, as were tokens without a visible release burst. In total, 28 Sylheti tokens (= 2.51%) and 22 English tokens (= 3.9%) were excluded.





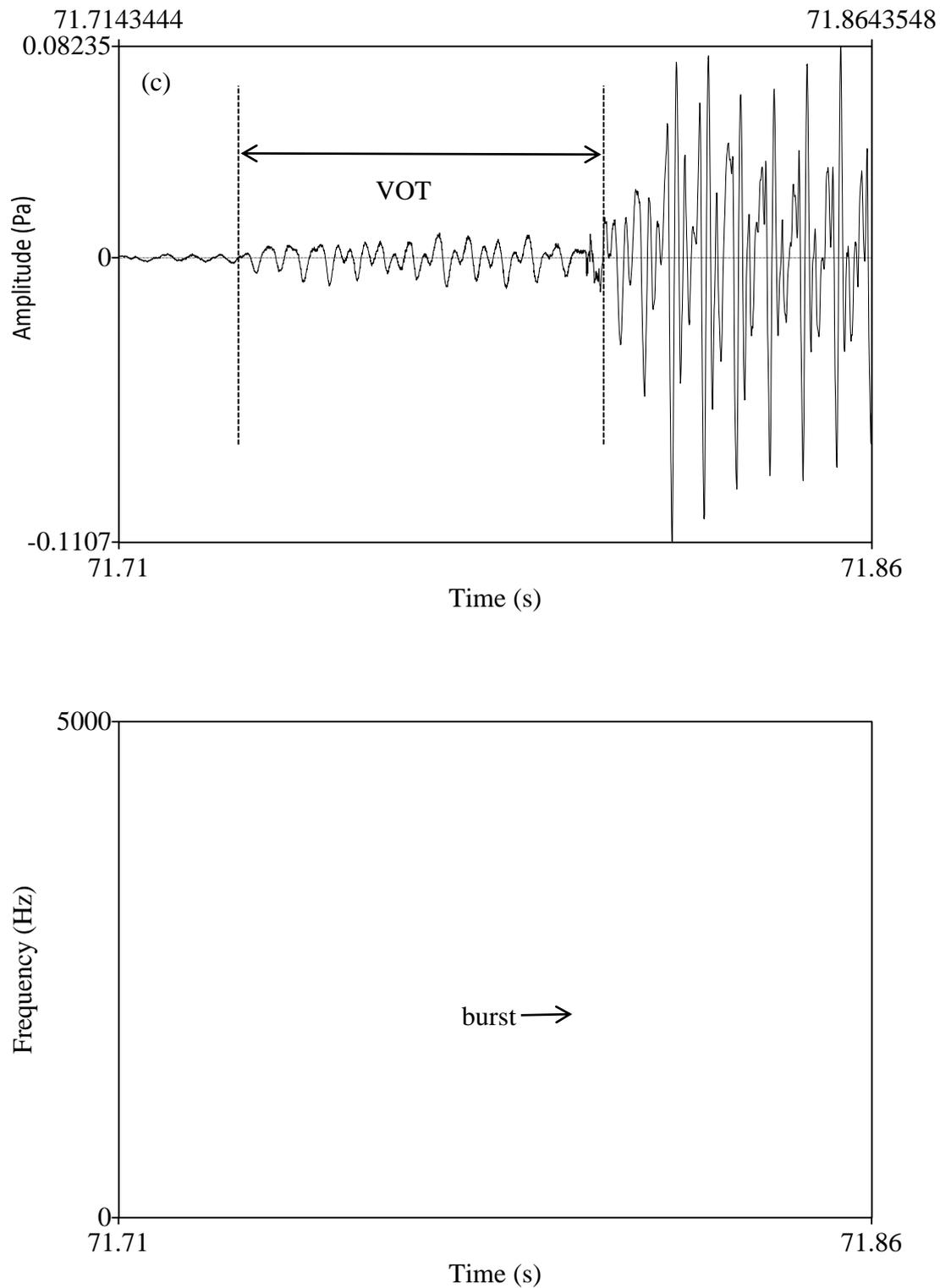


Figure 1: Waveform and spectrogram of stop realised with long-lag VOT (a), short-lag VOT (b) and lead VOT (c); all 150 ms in duration.

### 3. Results

In line with previous studies (Gope & Mahanta, 2015; McCarthy et al., 2013), we found that Sylheti /p/ and /p<sup>h</sup>/ were realised as [f], and Sylheti /k<sup>h</sup>/ as [x] in virtually all instances. As a result, these categories were not analysed further. All other categories in Sylheti and English were realised as stops.

In what follows, the results are organised in three parts according to place of articulation (i.e. bilabial, coronal and velar). Each part is further divided by language, first presenting intra-linguistic comparisons for Sylheti and English stops and then a cross-linguistic comparison. The auditory and acoustic results are integrated within each section.

To determine differences between the groups and stops, linear mixed-effects models were run separately in R (R Core Team, 2016) for the auditory and acoustic data, and for each place of articulation, using all analysed tokens. In each model, *stop* and *group* were entered as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. Note that *stop* and *group* were coded around zero. This made it possible to interpret the fixed factors as main effects. Using the LmerTest function in R (Bates, Maechler, Bolker & Walker, 2015), degrees of freedom were obtained via the Satterthwaite approximation with which p-values could be generated.

#### 3.1. Bilabial stops

##### 3.1.1. Sylheti

Figure 2a depicts the percentage of bilabial Sylheti stops that were produced accurately in terms of place of articulation and breathiness, as assessed in the auditory analysis. The results show that the GEN 1 MUMS managed to produce /b/ and /b<sup>h</sup>/ entirely accurately in terms of these dimensions, while the GEN 3 CHILDREN had the lowest accuracy score overall, with

all non-target-like tokens of /b<sup>h</sup>/ realised as [b], lacking breathiness (cf. Table 3 for a confusion matrix of all Sylheti stops<sup>3</sup>).

To determine whether the between-group differences are significant, a linear mixed-effects model was run with *stop* (2 levels: /b/; /b<sup>h</sup>/) and *group* (4 levels: GEN 1 MUMS; GEN 2 CHILDREN; GEN 2 MUMS; GEN 3 CHILDREN) as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 4, revealed no significant main effect of *stop*, but a significant main effect of *group* and a significant *group*\**stop* interaction. To examine these results further, we compared each of the groups with each other in separate regression models, run separately for /b/ and /b<sup>h</sup>/, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .025, using the Holm-Bonferroni method (Holm, 1979). Only one of the models revealed a significant effect of *group*, with the GEN 2 CHILDREN outperforming the GEN 3 CHILDREN on /b<sup>h</sup>/ (p= .006).

	[b]	[b <sup>h</sup> ]	[t]	[ṭ]	[t̥]	[d]	[ḍ]	[d̥]	[k]	[g]	[g <sup>h</sup> ]	[ʎ]
/b/	85	5				3						
/b <sup>h</sup> /	16	76										
/ṭ/			38	51					3			
/t̥/			25	7	61							
/ḍ/						30	63					
/d̥/						23	15	54				
/k/									93			
/g/						1				90		1
/g <sup>h</sup> /										13	79	1

Table 3: Confusion matrix: Sylheti stops

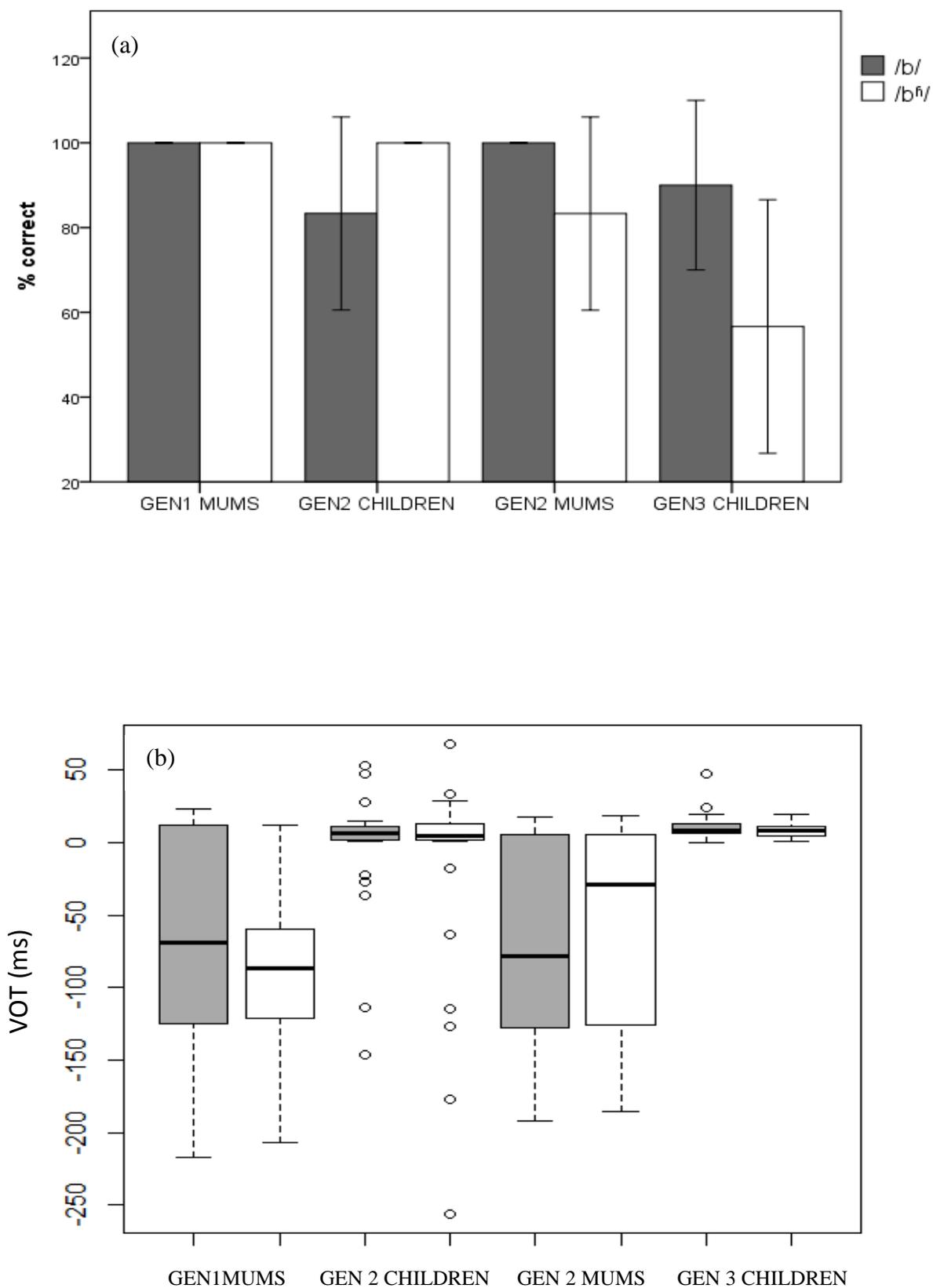


Figure 2: Percent correct productions (a) and VOT distributions (b) for Sylheti /b/ (grey) and /bʰ/ (white).

Figure 2b shows the results of the VOT analysis for Sylheti /b/ and /b<sup>h</sup>/. In line with previous work (Gope & Mahanta, 2015; McCarthy et al., 2013), the adult participants predominantly exhibited prevoicing (GEN 1 MUMS: 68%, i.e. 19/28 tokens; GEN 2 MUMS: 56%, i.e. 19/34 tokens). In contrast, the GEN 2 CHILDREN only realised 21% of their tokens with a voicing lead (i.e. 12/58 tokens), and the GEN 3 CHILDREN prevoiced none, instead realising all tokens with short-lag VOT values.

Model		$\beta$	SE	t	p
<i>Percent correct (Sylheti)</i>	Intercept	0.77155	0.07838	9.844	<.001***
	Group	-0.17062	0.07105	-2.401	0.02251*
	Stop	-0.11655	0.12609	-0.924	0.36244
	Group*Stop	-0.32018	0.11432	-2.801	0.00869**
<i>VOT (Sylheti)</i>	Intercept	-27.220	9.059	-3.005	0.00522**
	Group	19.224	8.226	2.337	0.02601*
	Stop	-5.901	6.750	-0.874	0.38832
	Group*Stop	5.065	6.208	0.816	0.42010
<i>VOT (English)</i>	Intercept	23.378	5.446	4.292	<.001***
	Group	16.426	4.925	3.335	0.002242**
	Stop	-108.839	11.701	9.302	<.001***
	Group*Stop	16.238	10.584	1.534	0.135178
<i>Percent correct (Cross-linguistic)</i>	Intercept	0.88518	0.03912	22.628	<.001***
	Group	-0.08607	0.03548	-2.426	0.0213*
	Stop	-0.02635	0.02989	-0.882	0.3787
	Group*stop	-0.06404	0.02714	-2.359	0.0190*
<i>VOT (Cross-linguistic)</i>	Intercept	-2.4522	6.9640	-0.352	0.72713
	Group	17.6279	6.3133	2.792	0.00888**
	Stop	8.2429	3.0540	2.699	0.00744***
	Group*stop	0.7684	2.7611	0.278	0.78104

Table 4: Results of mixed-effects models for bilabial stops

To determine whether the between-group differences in VOT are significant, a linear mixed-effects model was run with *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 4,

revealed no significant main effect of *stop*, and no significant *group\*stop* interaction. However, it did find a significant main effect of *group*. To examine this effect further, we compared each of the groups with each other in separate regression models, run separately for /b/ and /b<sup>h</sup>/, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .008. The results revealed significantly longer VOT values for the GEN 3 CHILDREN on both Sylheti stops than the GEN 1 MUMS (/b/:  $p = .003$ ; /b<sup>h</sup>/:  $p < .001$ ), and the GEN 2 MUMS (/b/:  $p = .001$ ; /b<sup>h</sup>/:  $p = .002$ ). No differences were observed between the adult participants and the GEN 2 CHILDREN.

### 3.1.2. English

The auditory analysis revealed that English /b/ and /p/ were consistently produced at the correct place of articulation. Moreover, there were no breathy tokens of English /b/ (cf. Table 5 for a confusion matrix of all English stops). However, an analysis of the participants' VOT patterns showed differences in voicing across the groups (cf. Figure 3). Accordingly, the GEN 1 MUMS prevoiced 73% (11/15 tokens) of their English /b/ productions, while the GEN 2 MUMS only prevoiced 35% (6/17 tokens) and the GEN 2 CHILDREN only 18% (5/28 tokens). As in Sylheti, the GEN 3 CHILDREN produced no prevoiced tokens at all, instead realising English /b/ with short-lag VOT values throughout.

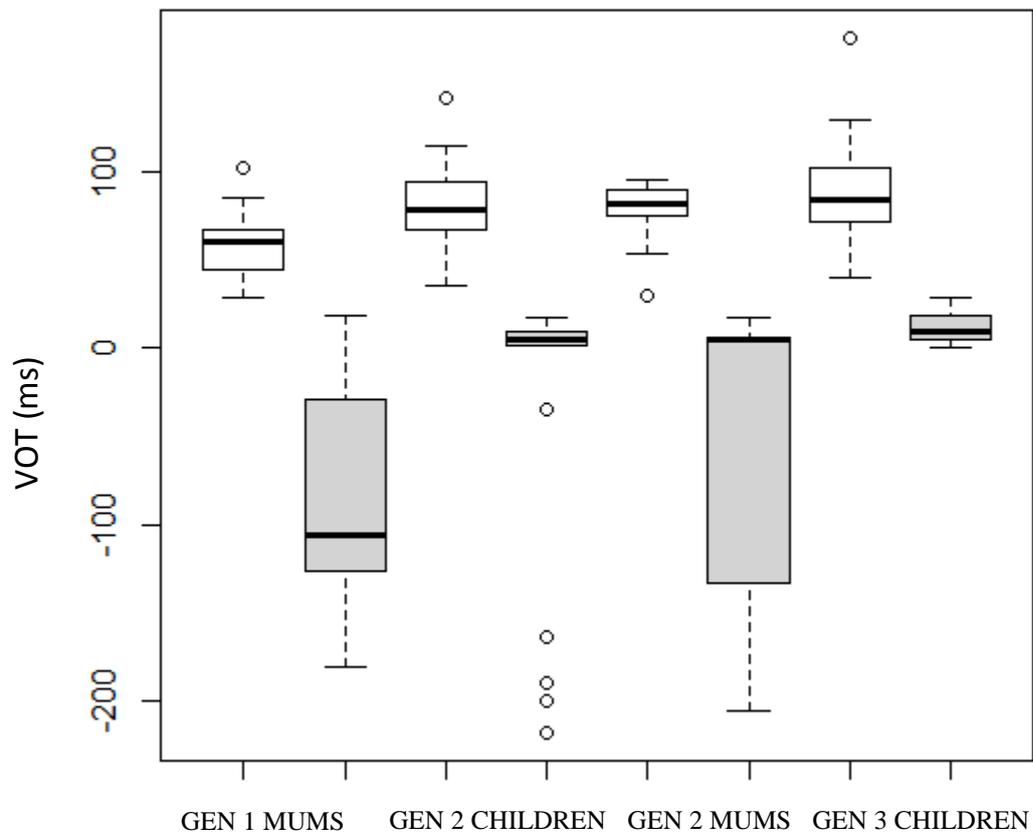


Figure 3: VOT distributions for English /p/ (white) and /b/ (grey).

	[p]	[b]	[t]	[ʈ]	[d]	[ɖ]	[k]	[g]	[j]
/p/	93								
/b/		93							
/t/			81	12					
/d/					81	12			
/k/							93		
/g/								88	4

Table 5: Confusion matrix: English stops

To determine whether the between-group differences are significant, a linear mixed-effects model was run with *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 4, revealed significant main effects of *group* and *stop*, but no significant *group\*stop* interaction. To examine this effect further, we compared each of the groups with each other in separate regression models, run separately for /p/ and /b/, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .01. The results revealed significantly longer VOT values for the GEN 3 CHILDREN on English /b/ than the GEN 1 MUMS ( $p < .001$ ). No other between-group differences were significant.

### 3.1.3. Cross-linguistic comparison

To determine whether the participants produced bilabial stops differently in Sylheti and English, two linear mixed-effects models were run, one on the percent correct scores, and one on VOT. Both models had *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 4, revealed significant main effects of *group* for both models, as well as a significant effect of *stop* for the VOT model, and a significant *group\*stop interaction* for the percent correct model. To examine these effects further, we compared each combination of stops across the two languages for each group in separate regression models with *stop* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .017 for the percent correct scores and .006 for VOT. The results on the percent correct scores revealed significantly greater accuracy on English /b/ than Sylheti /b/ for the GEN 2 CHILDREN ( $p = .007$ ), and significantly greater accuracy on English /b/ than Sylheti /b<sup>h</sup>/ for the GEN 3 CHILDREN ( $p < .001$ ). Moreover, with respect to VOT all four groups exhibited significantly longer VOT values for English

/p/ than Sylheti /b/ and /b<sup>h</sup>/ (all:  $p < .001$ ). There were no significant differences in VOT between Sylheti /b/ and /b<sup>h</sup>/, and English /b/.

### 3.2. Coronal stops

#### 3.2.1. Sylheti

Figure 4a shows the percentage with which the Sylheti coronal stops /t̪ t̪̪ d̪ d̪̪/ were produced at the correct place of articulation. The GEN 1 MUMS were the most accurate while performance by the other groups was variable, resulting in lower accuracy rates, in particular for the GEN 3 CHILDREN. The majority of errors involved realising dental and retroflex stops as alveolars (82%, 116/ 141 tokens). However, confusion between retroflex and dental categories was also common, accounting for 16% of errors overall (22/ 141 tokens). See Table 3 for further details.

To determine whether the between-group differences on Sylheti coronal stops are significant, a linear mixed-effects model was run with *stop* (4 levels: /t̪ t̪̪ d̪ d̪̪/) and *group* (4 levels: GEN 1 MUMS; GEN 2 CHILDREN; GEN 2 MUMS; GEN 3 CHILDREN) as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 6, revealed a significant main effect of *group*, but no significant main effect of *stop* and no significant *group\*stop* interaction. To examine these results further, we compared each of the groups with each other in separate regression models, run separately for each of the stops, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .01. The results revealed significantly greater accuracy on /t̪/ for the GEN 1 MUMS than the GEN 3 CHILDREN ( $p = .005$ ). No other between-group differences reached significance.

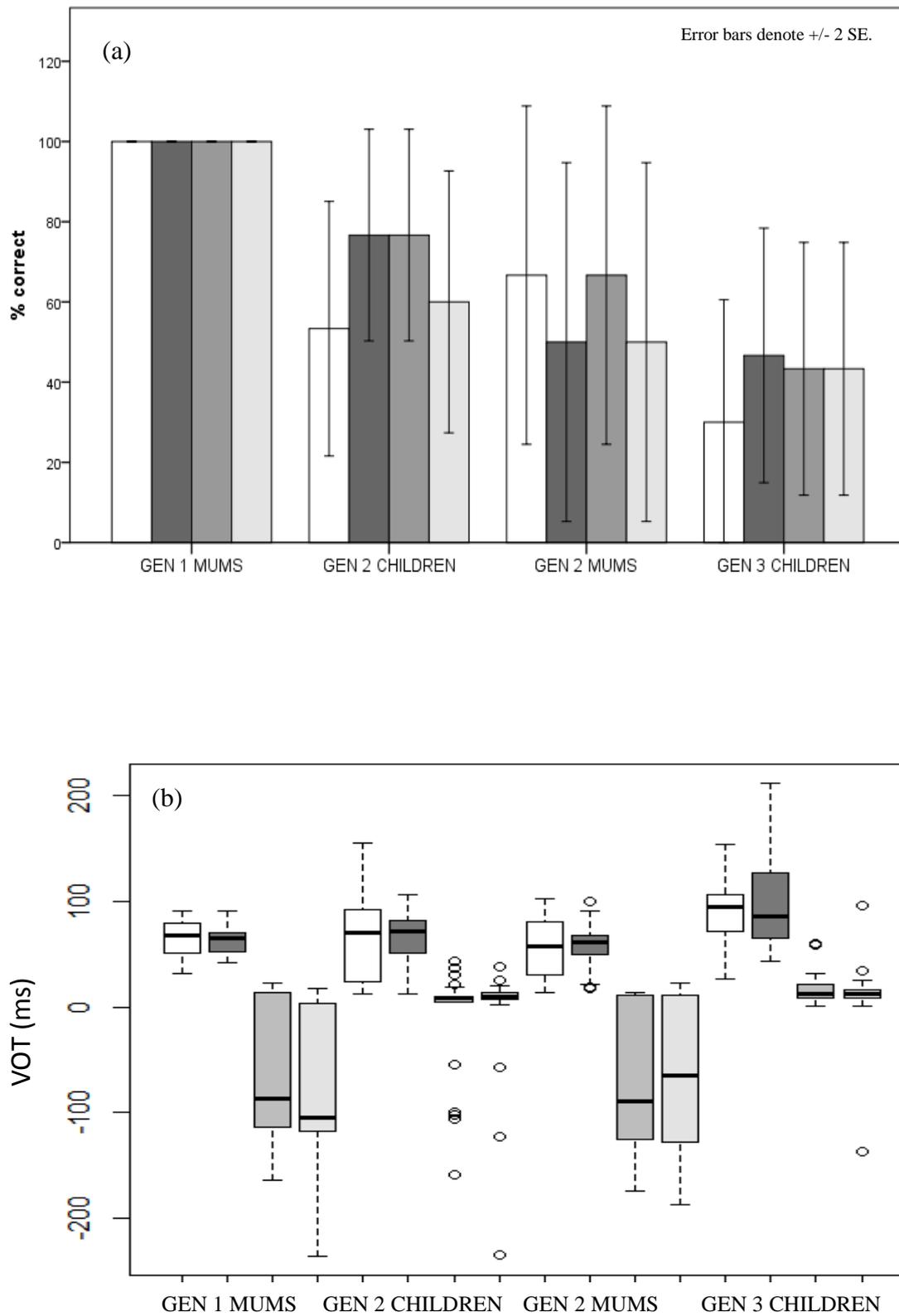


Figure 4: Percent correct productions (a) and VOT distributions (b) of Sylheti /t̪/(white), /t/(dark grey), /ɖ/(grey), /ɖ̪/(light grey).

Model		$\beta$	SE	t	p
<i>Percent correct (Sylheti)</i>	Intercept	0.312865	0.086375	3.622	0.001031 **
	Group	-0.339181	0.078273	-4.333	<.001***
	Stop	0.011637	0.044463	0.262	0.795259
	Group*Stop	0.007135	0.040292	0.177	0.860606
<i>VOT (Sylheti)</i>	Intercept	22.427	5.311	4.223	<.001***
	Group	14.841	4.813	3.084	0.004268 **
	Stop	-39.244	3.689	-10.639	<.001***
	Group*Stop	4.174	3.341	1.249	0.220866
<i>Percent correct (English)</i>	Intercept	0.66537	0.09700	6.859	<.001***
	Group	0.37108	0.08791	4.221	<.001***
	Stop	0.02372	0.02184	1.086	0.279138
	Group*Stop	-0.01222	0.01985	-0.616	0.538874
<i>VOT (English)</i>	Intercept	34.513	5.851	5.899	<.001***
	Group	12.322	5.302	2.324	0.027 *
	Stop	-101.221	10.442	-9.694	<.001***
	Group*Stop	13.907	9.463	1.470	0.152
<i>Percent correct (Cross-linguistic)</i>	Intercept	0.71806	0.03197	22.464	<.001***
	Group	0.03965	0.02897	1.368	0.1810
	Stop	0.03122	0.01303	2.396	0.0228*
	Group*stop	-0.06942	0.01181	-5.877	<.001***
<i>VOT (Cross-linguistic)</i>	Intercept	24.394	5.316	4.589	<.001***
	Group	13.364	4.913	2.720	0.0105 *
	Stop	-26.958	2.585	-10.428	<.001***
	Group*stop	2.801	2.385	1.174	0.2501

Table 6: Results of mixed-effects models for coronal stops

Figure 4b presents the VOT values of the Sylheti coronal stops. Previous data from Sylheti speakers in Asia (Gope & Mahanta, 2015) suggest consistent prevoicing in voiced stops. The two groups of adults in the present study broadly followed this pattern, with the GEN 1 MUMS prevoicing 66% (19/29 tokens) of their voiced coronal stops, and the GEN 2 MUMS 61% (22/36 tokens). In contrast, the GEN 2 CHILDREN only prevoiced 16% (9/55 tokens) of their /d/ and /d̪/ productions, and the GEN 3 CHILDREN fewer than 2% (1/60

tokens). The voiceless coronal stops, in turn, were realised within the long-lag VOT range by all groups.

To determine whether the between-group differences in VOT for Sylheti coronal stops are significant, a linear mixed-effects model was run with *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 6, revealed significant main effects of *group* and *stop*, but no significant *group\*stop* interaction. To examine these results further, we compared each of the groups with each other in separate regression models, run separately for each of the stops, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .004. The results showed that the GEN 3 CHILDREN produced /t̪/, /d̪/, and /d̪/ with significantly longer VOT values than the GEN 1 MUMS (all:  $p < .001$ ), and /d̪/ with significantly longer VOT values than the GEN 2 MUMS ( $p < .001$ ). The GEN 2 CHILDREN produced Sylheti /d̪/ with significantly longer VOT values than the GEN 1 MUMS ( $p = .003$ ).

### 3.2.2. English

Figure 5a shows the percentage of correct productions of English /t/ and /d/. Inspection of the figure shows that the child participants and the GEN 2 MUMS exhibited high degrees of accuracy on these categories. In contrast, the GEN 1 MUMS largely produced them inaccurately. An examination of their error patterns revealed that all non-target like tokens of /t/ were realised as [t̪] and all non-target like tokens of /d/ as [d̪] (cf. Table 4 for details).

These differences were tested in a linear mixed-effects model with *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 6, revealed a significant main effect of *group*, but no

significant main effect of *stop* and no significant *group\*stop* interaction. To examine these results further, we compared each of the groups with each other in separate regression models, run separately for English /t/ and /d/, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .004. The results revealed that the GEN 1 MUMS were significantly less accurate on English /t/ and /d/ than the GEN 2 CHILDREN, the GEN 2 MUMS and the GEN 3 CHILDREN (all:  $p < .001$ ).

Figure 5b depicts the VOT patterns for English /t/ and /d/ across the groups.

Inspection of the figure shows similar patterns for English /t/, with realisations in the long-lag VOT range throughout. In contrast, the VOT patterns for English /d/ show stark differences across the groups: the GEN 1 MUMS mainly realised this category with a voicing lead (67%, i.e. 10/15 tokens); in contrast, the GEN 2 MUMS only exhibited prevoicing in 35% of instances (6/17 tokens) and the GEN 2 CHILDREN in 7% of instances (2/30 tokens), while the GEN 3 CHILDREN did not prevoice any of their English /d/ tokens.

To determine whether these differences are significant, a linear mixed-effects model was run with *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 6, revealed significant main effects of *group* and *stop*, but no significant *group\*stop* interaction. To examine these results further, we compared each of the groups with each other in separate regression models, run separately for each of the stops, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .01. The results showed that the GEN 3 CHILDREN produced English /d/ with significantly longer VOT values than the GEN 1 MUMS ( $p < .001$ ) and the GEN 2 MUMS ( $p = .008$ ). The GEN 2 CHILDREN also produced

/d/ with significantly longer VOT values than the GEN 1 MUMS ( $p=.005$ ). No significant differences were observed for /t/.

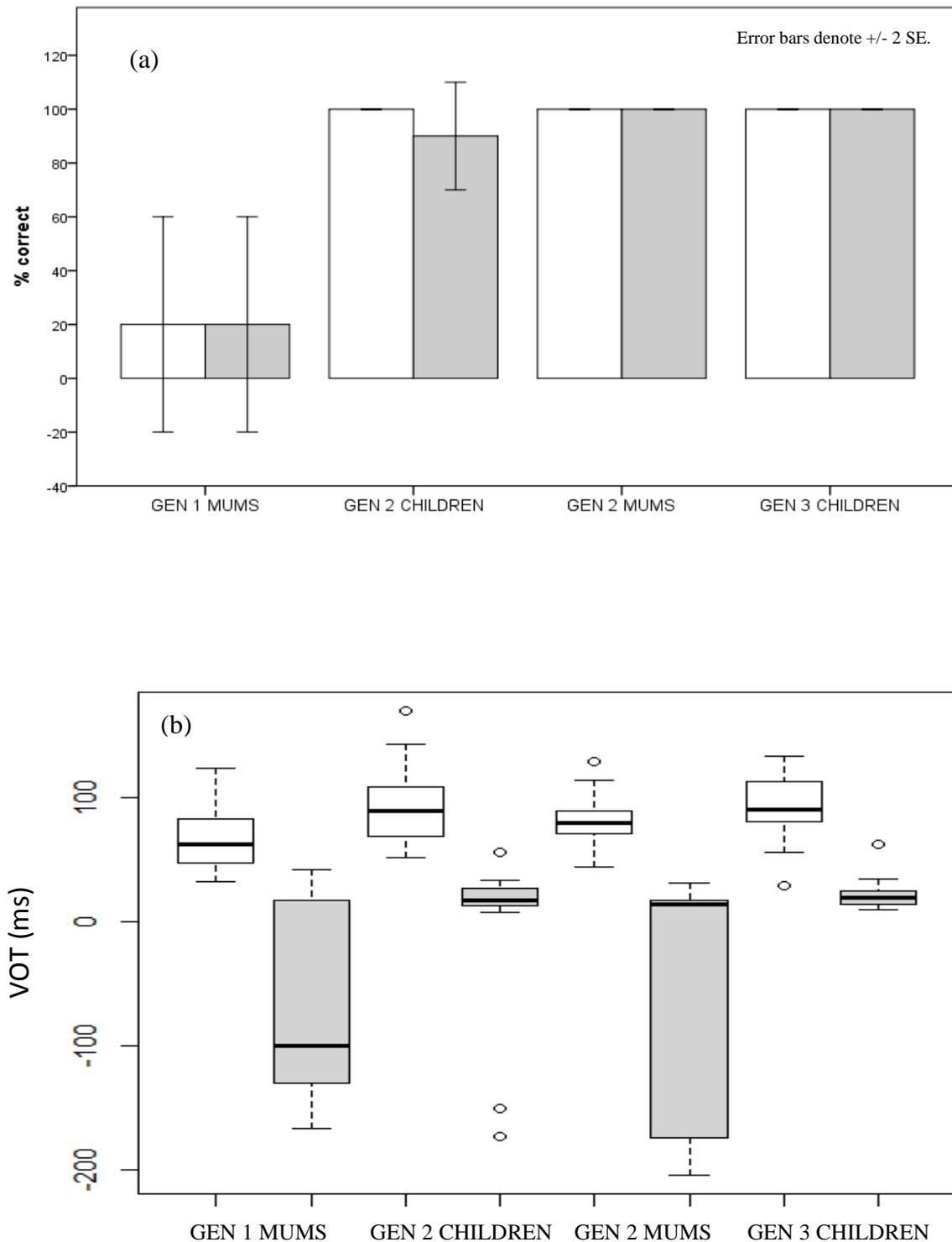


Figure 5: Percent correct productions (a) and VOT distributions (b) for English /t/ (white) and /d/ (grey).

### 3.2.3. Cross-linguistic comparison

To determine whether the participants produced coronal stops differently in Sylheti and English, two linear mixed-effects models were run, one on the percent correct scores, and one on VOT. Both models had *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 6, revealed a significant main effect of *stop* and a significant *group\*stop* interaction for the percent correct model, as well as significant main effects of *group* and *stop* for the VOT model. To examine these effects further, we compared each combination of stops across the two languages in separate regression models, run separately for each group, with *stop* as fixed factor and *speaker* as random factor. The  $\alpha$ -level for the percent correct scores was adjusted to .002, and for VOT to .01. The results showed that the GEN 1 MUMS had significantly higher percent correct scores on Sylheti /t̪/, /t̪̃/, /d̪/ and /d̪̃/ than English /t/ and /d/, while the GEN 2 CHILDREN, the GEN 2 MUMS, the GEN 3 CHILDREN showed the reverse pattern with significantly higher percent scores for English coronal stops than Sylheti ones (all:  $p < .001$ ). There were only two exceptions to this pattern: the GEN 2 CHILDREN did not differ significantly in their accuracy of Sylheti /t̪/ and English /d/ ( $p = .009$ ), and of Sylheti /d̪̃/ and English /d/ ( $p = .009$ ). The results for VOT, in turn, showed that the GEN 2 CHILDREN had significantly longer VOT values on English /t̪/ than Sylheti /t̪/ ( $p = .001$ ) and /t̪̃/ ( $p < .001$ ). Similarly, the GEN 2 MUMS had significantly longer VOT values on English /t̪/ than Sylheti /t̪̃/ and /t̪/ (both:  $p < .001$ ). No other cross-linguistic differences were observed.

## 3.3. Velar stops

### 3.3.1. Sylheti

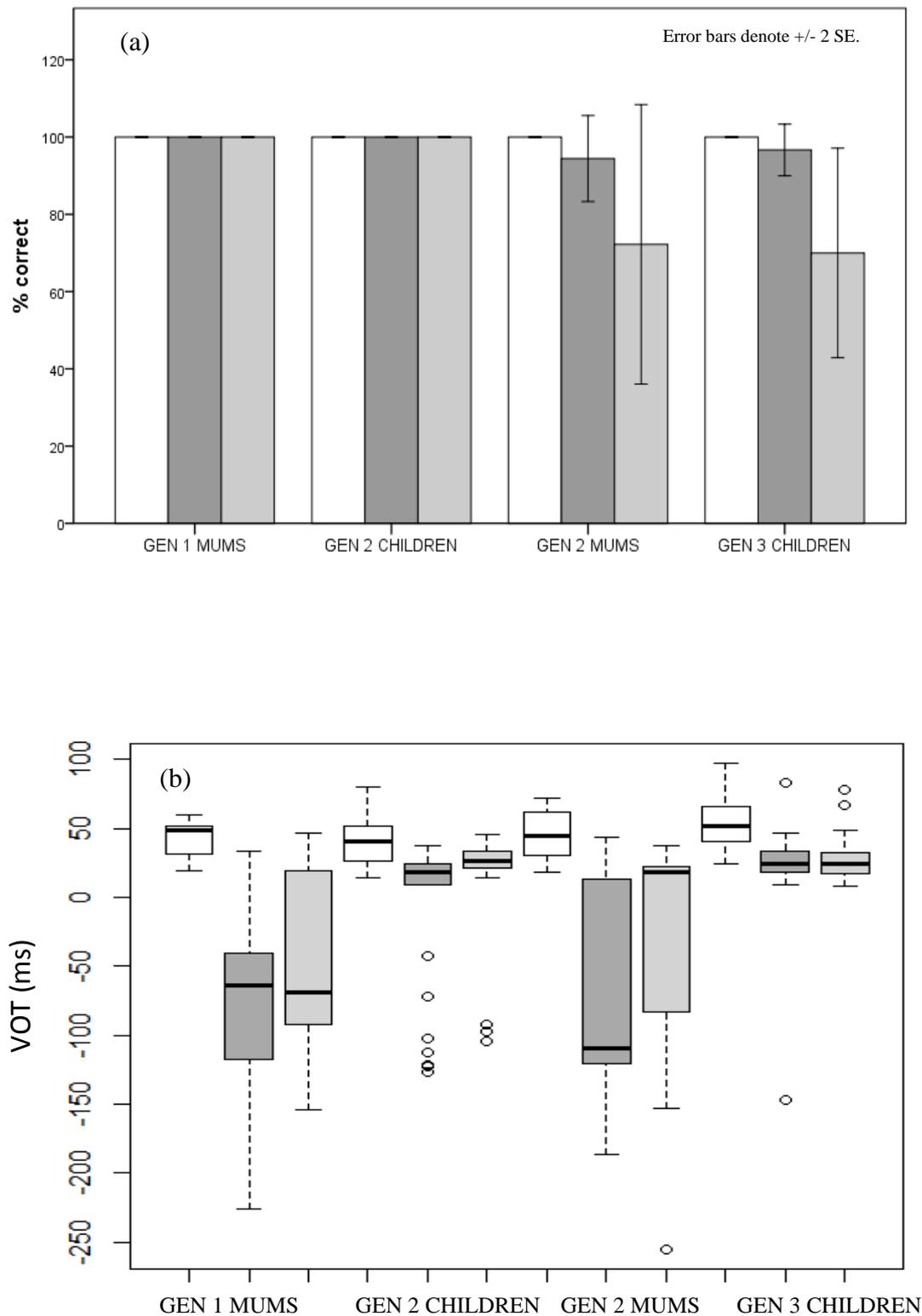


Figure 6: Percent correct productions (a) and VOT distributions (b) of Sylheti /k/ (white), /g/ (dark grey) and /gʱ/ (light grey).

Figure 6a presents the percentage of correct productions of the Sylheti velar stops /k g g<sup>h</sup>/ in terms of place of articulation and breathiness. The figure shows largely accurate performance overall, with slightly lower accuracy scores for the GEN 2 MUMS and the GEN 3 CHILDREN. All erroneous productions of /g<sup>h</sup>/ were realised as [g], lacking breathiness (cf. Table 3 for details). To examine if these differences are significant, a linear mixed-effects model was run with *stop* (3 levels: /k/; /g/; /g<sup>h</sup>/) and *group* (4 levels: GEN 1 MUMS; GEN 2 CHILDREN; GEN 2 MUMS; GEN 3 CHILDREN) as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 7, revealed significant main effects of *group* and *stop* and a significant *group\*stop* interaction. However, further regression models, run separately for each stop with *group* as fixed factor and *speaker* as random factor, and an adjusted  $\alpha$ -level of .025, revealed no significant between-group differences.

<b>Model</b>		<b><math>\beta</math></b>	<b>SE</b>	<b>t</b>	<b>p</b>
<i>Percent correct (Sylheti)</i>	Intercept	0.90156	0.04239	21.267	<.001***
	Group	-0.09162	0.03842	-2.385	0.0233 *
	Stop	-0.12880	0.05573	-2.311	0.0275 *
	Group*Stop	-0.12251	0.05050	-2.426	0.0212 *
<i>VOT (Sylheti)</i>	Intercept	4.907	5.679	0.864	0.39409
	Group	15.756	5.150	3.060	0.00448 **
	Stop	-23.346	4.607	-5.067	<.001***
	Group*Stop	4.685	4.183	1.120	0.26896
<i>Percent correct (English)</i>	Intercept	0.95100	0.03364	28.267	<.001***
	Group	0.03362	0.03049	1.103	0.279
	Stop	-0.09799	0.06729	-1.456	0.155
	Group*Stop	0.06724	0.06098	1.103	0.279
<i>VOT (English)</i>	Intercept	51.233	5.579	9.183	<.001***
	Group	9.793	5.137	1.906	0.0662
	Stop	-101.705	11.858	-8.577	<.001***
	Group*Stop	11.602	10.898	1.065	0.2954

<i>Percent correct</i> ( <i>Cross-linguistic</i> )	Intercept	0.92125	0.03113	29.590	<.001***
	Group	-0.04160	0.02821	-1.475	0.1501
	Stop	-0.12272	0.04902	-2.503	0.0177 *
	Group*stop	-0.08464	0.04442	-1.905	0.0659
<i>VOT</i> ( <i>Cross-linguistic</i> )	Intercept	19.141	5.899	3.245	0.0028 **
	Group	14.139	5.464	2.588	0.0146 *
	Stop	-29.097	3.461	-8.406	<.001***
	Group*stop	4.819	3.207	1.503	0.1437

Table 7: Results of mixed-effects models for velar stops

Figure 6b depicts the VOT values for the Sylheti velar stops. Inspection of the figure shows similar values for /k/ across the groups, but differences in the degree of prevoicing in /g/ and /g<sup>h</sup>/. The GEN 1 MUMS realised 73% of their voiced velar stops with a voicing lead (22/30 tokens) and the GEN 2 MUMS 52% (17/33 tokens). In contrast, the GEN 2 CHILDREN only prevoiced 18% of their voiced velar stops (11/62 tokens), and the GEN 3 CHILDREN fewer than 2% (1/57 tokens), instead realising Sylheti /g/ and /g<sup>h</sup>/ within the short-lag VOT range.

To determine whether the between-group differences in VOT for Sylheti velar stops are significant, a linear mixed-effects model was run with *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 7, revealed significant main effects of *group* and *stop*, but no significant *group\*stop* interaction. To examine these results further, we compared each of the groups with each other in separate regression models, run separately for each of the stops, with *group* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .005. The results showed that the GEN 3 CHILDREN produced Sylheti /g/ and /g<sup>h</sup>/ with significantly longer VOT values than the GEN 1 MUMS (both:  $p < .001$ ), and the GEN 2

MUMS (/g/:  $p < .001$ ; /g<sup>h</sup>/:  $p = .004$ ). The GEN 2 CHILDREN, in turn, produced Sylheti /g<sup>h</sup>/ with significantly longer VOT values than the GEN 1 MUMS ( $p = .002$ ).

### 3.3.2. English

All tokens of English /k/ were produced at the correct place of articulation, and only four tokens of English /g/ were not target-like. Moreover, there were no breathy tokens of English /g/ (cf. Figure 7a and Table 5 for details). However, an analysis of the participants' VOT patterns showed differences in voicing across the groups (cf. Figure 7b). The adult participants exhibited substantial prevoicing of English /g/ (GEN 1 MUMS: 60%, 6/10 tokens; GEN 2 MUMS: 50%, 8/16 tokens), while the child participants produced all their tokens within the short-lag VOT range. English /k/, on the other hand, was consistently produced within the long-lag VOT range by all groups.

To determine whether the VOT differences are significant, a linear mixed-effects model was run with *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 7, revealed a significant main effect of *stop*, but no significant main effect of *group* and no significant *group\*stop* interaction, suggesting that the English velar stops were produced in much the same way by each of the groups.

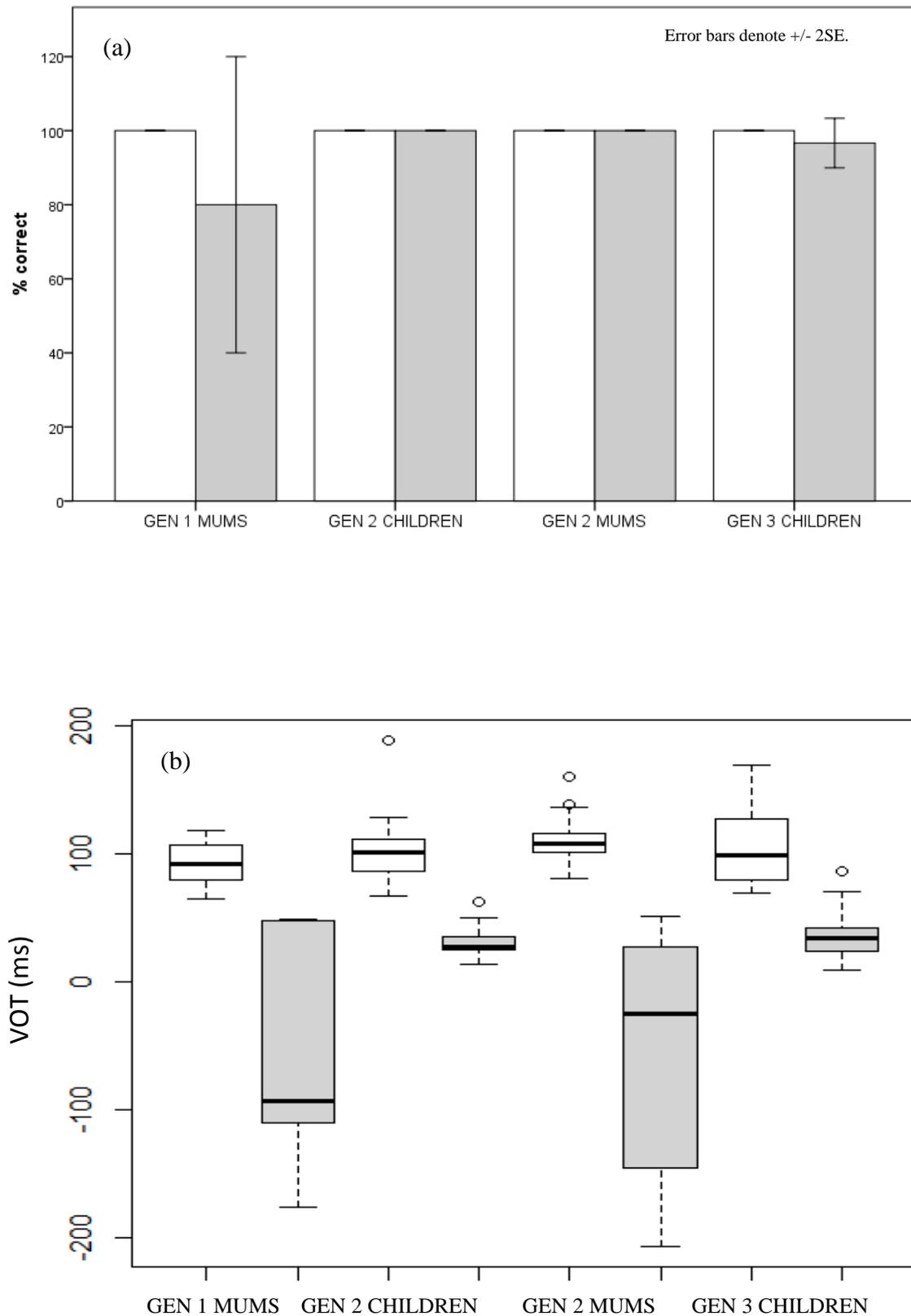


Figure 7: Percent correct productions (a) and VOT distributions (b) of English /k/ (white) and /g/ (grey).

### 3.3.3. Cross-linguistic analysis

To determine whether the participants produced velar stops differently in Sylheti and English, two linear mixed-effects models were run, one on the percent correct scores, and one on VOT. Both models had *stop* and *group* as fixed factors (including interaction) and *speaker* as a random factor with random slopes for *stop*. The results, displayed in Table 7, revealed a significant main effect of *stop* for the percent correct and VOT models, and a significant main effect of *group* for the VOT model. To examine these effects further, we compared each combination of stops across the two languages in separate regression models, run separately for each group, with *stop* as fixed factor and *speaker* as random factor. The  $\alpha$ -level was adjusted to .025 for the percent correct scores and .007 for VOT. The results for the percent correct model revealed significantly greater accuracy on English /g/ than Sylheti /g<sup>h</sup>/ for the GEN 2 MUMS ( $p=.004$ ) and the GEN 3 CHILDREN ( $p<.001$ ). With respect to VOT, all four groups produced English /k/ with significantly longer VOT values than Sylheti /k/ (all:  $p<.001$ ). Moreover, the GEN 2 CHILDREN produced English /g/ with significantly longer VOT values than Sylheti /g/ ( $p<.001$ ).

## 4. Discussion

The purpose of this study was to gain a better understanding of cross-generational transmission in heritage language settings. To this end, we examined the Sylheti and English stop consonant productions of two sets of Bangladeshi heritage families: (1) first-generation migrants from the Sylhet area of Bangladesh who arrived in the UK in adulthood, and their UK-born (second generation) children, and (2) second-generation UK-born adults and their (third-generation) children. The results revealed significant differences in both the host

language and the heritage language across the generations, and between the child and adult participants. In what follows, the adult and child participants' acquisition patterns will be discussed, followed by an examination of socio-cultural factors. Finally, we will consider the implications of our findings for the maintenance and transmission of heritage languages.

#### **4.1. Acquisition patterns: adults**

To begin with, an investigation of the GEN 1 MUMS' L2 English stops revealed a number of non-native patterns. For example, English /d/ was commonly realised as [d̥] and English /t/ as [t̥]. The GEN 1 MUMS also predominantly produced /b d g/ with a voicing lead, rather than with short-lag VOT values. While the use of lead voicing in English is not non-native per se, its occurrence tends to be marginal. In Docherty (1992), for example, it accounted for 7 % of voiced stops. In contrast, the GEN 1 MUMS produced 27/ 40 tokens, i.e. 68% with a voicing lead. These patterns conform to those found in McCarthy et al.'s (2013) study and suggest an influence of the participants' native language – a finding that is expected in L2 learners (e.g., Colantoni, Steele & Escudero, 2015; Flege, 1995; Hansen Edwards & Zampini, 2008; Mayr & Escudero, 2010). Note, however, that the GEN 1 MUMS also showed evidence of successful acquisition. For example, they realised English /p t k/ within the long-lag VOT range, and made a clear cross-linguistic distinction between Sylheti and English /k/ (with average VOT values of 40ms and 91ms, respectively).

The GEN 1 MUMS' Sylheti stop productions, in turn, largely conformed to those of native Sylheti speakers in Asia (Gope & Mahanta, 2015) and recent arrivals in the UK (McCarthy et al., 2013). Accordingly, their /b<sup>h</sup>/ and /g<sup>h</sup>/ were consistently realised as voiced breathy stops, and their coronal stops appropriately as dentals and retroflexes. Moreover, they produced the majority of their voiced stops with a target-like voicing lead. Interestingly,

however, there are some indications that their Sylheti stop productions may not be entirely native-like. For example, unlike McCarthy et al.'s (2013) study, in which first-generation Sylheti speakers and native controls produced Sylheti voiceless stops with short-lag VOT values, the GEN 1 MUMS' productions were much longer. To some extent, methodological differences between the two studies can explain these patterns: in McCarthy et al. (2013) the target words were embedded in the middle of a carrier sentence, while they occurred at the beginning of a carrier sentence in the present study. This may have made it more likely for participants to treat them like items in citation form, resulting in longer VOT values (Auzou et al., 2000). However, their mean values for Sylheti /t̪/ and /t̪/ (both 63ms) are much too long for this to be the only credible explanation. Moreover, despite predominantly prevoiced realisations of Sylheti voiced stops, the GEN 1 MUMS produced a fair amount of tokens within the short-lag VOT range (27/ 87 tokens, i.e. 31%), and hence differently from native control speakers (Gope & Mahanta, 2015; McCarthy et al., 2013). These patterns indicate that not only their L2 stops were non-native, but also some of their L1 categories, suggesting bi-directional interactions. While current theories of bilingual speech learning, such as the SLM (Flege, 1995; Flege et al., 2003), take account of these effects, they cannot explain why L2-to-L1 transfer only affected the first-generation migrants in the present study, but not in McCarthy et al. (2013). A possible explanation might be differences in social structure across the two communities studied: first-generation migrants in Tower Hamlets are regularly exposed to native Sylheti speech by new arrivals, reinforcing homeland norms, while there is virtually no influx of new arrivals from Bangladesh in the Cardiff community. However, social variables of this kind do not currently form part of formal bilingual speech learning models.

The GEN 2 MUMS, in turn, produced all English stops at the correct place of articulation, including the coronals, and were hence more accurate than the GEN 1 MUMS.

Note, however, that they realised 40% (i.e. 20/50 tokens) of their voiced stops with a voicing lead, which suggests a subtle influence of their heritage language. As such, the findings obtained here add to the few existing studies that have shown non-native patterns in the host language of second-generation speakers (Darcy & Krüger, 2012; Stangen et al., 2015). In these studies, non-native speech was associated with continued high use of the heritage language. In contrast, the GEN 2 MUMS had relatively low use of Sylheti in the home and community, and clearly considered themselves dominant in English. This suggests that high L1 use may not be an absolute prerequisite for non-native patterns in the host language, in particular in subtle areas of pronunciation, such as prevoicing, which has limited perceptual salience (Van Alphen & Smits, 2004) and can occur in native English speech, albeit in smaller proportions.

The GEN 2 MUMS' Sylheti stop productions also showed evidence of successful acquisition. For example, they exhibited similar mean prevoicing values for their voiced stops as the GEN 1 MUMS. On the other hand, in contrast to the latter, they commonly realised Sylheti coronals as alveolars, and produced some tokens of /g<sup>h</sup>/ as [g]. Differences of this kind between first-generation migrants and second-generation heritage language users are well attested (e.g., Evans et al., 2007; McCarthy et al., 2011, 2013; Nagy & Kochetov, 2013; Sharma & Sankaran, 2011), and have been explained in a number of ways. Chambers (2002) claimed that second-generation speakers have an innate accent filter that blocks non-native features in the host language. However, this claim is undermined by evidence that non-native features do occur in the host language (Darcy & Krüger, 2012; Stangen et al., 2015), and has been largely discredited by Sharma & Sankaran's (2011) work. A more plausible explanation is a socially oriented approach according to which an individual's speech patterns are the result of "network, demographic and intergroup forces" (Sharma & Sankaran, 2011: 403). While these factors have not been investigated in a detailed ethnographic study here, there are

clear differences in linguistic experience and language use across the generations.

Accordingly, the GEN 1 MUMS spent their formative years in Bangladesh, live in close-knit communities with many other Sylheti speakers, and use Sylheti as the main language in the home on a daily basis. In contrast, the GEN 2 MUMS have either never been to Bangladesh, or only spent short periods of time there to visit family members. They live in areas of Cardiff that are ethnically heterogeneous with few opportunities to use Sylheti, and they predominantly use English in the home.

#### ***4.2. Acquisition patterns: children***

Both sets of children produced the English stops accurately, with /t/ and /d/ consistently realised at the alveolar place of articulation, and voiceless stops with long-lag VOT values (cf. Whiteside, Henry & Dobbin, 2004). Interestingly, the GEN 2 CHILDREN prevoiced some of their English /b/ and /d/ productions (7/58, i.e. 12%), while the GEN 3 CHILDREN produced all their English voiced stops within the short-lag VOT range. Since the extent of prevoicing conforms closely to that reported in previous work on English monolinguals (e.g. Docherty, 1992; Lisker & Abramson, 1964), the GEN 2 CHILDREN's patterns do not suggest cross-linguistic interactions.

Their Sylheti stop productions, in contrast, showed substantial differences. For example, the GEN 2 CHILDREN were entirely accurate in their production of /b<sup>h</sup>/ and /g<sup>h</sup>/, while the GEN 3 CHILDREN commonly produced these categories without breathiness. The GEN 2 CHILDREN also produced more target-like coronal stops than the GEN 3 CHILDREN. Finally, the two sets of children differed in their voicing patterns. Specifically, although both groups predominantly realised Sylheti voiced stops with short-lag VOT values, the GEN 2 CHILDREN exhibited a moderate level of prevoicing (32/176 tokens, i.e. 18%),

while it was virtually absent in the GEN 3 CHILDREN (3/170 tokens, i.e. <1%). How can these patterns be explained?

To begin with, developmental factors could be at work. Indeed, it has been shown that prevoicing is acquired late in monolingual and bilingual development (Bortolini et al., 1995; Fabiano-Smith & Bunta, 2012; Gandour et al., 1986; Kehoe et al., 2004; Khattab, 2000; Macken & Barton, 1979, 1980) since it has limited perceptual salience (Van Alphen & Smits, 2004) and is articulatorily complex (Ohala, 1997). Hence the children's lesser degree of prevoicing compared with that of the adults points to a developmental explanation. However, the prevoicing patterns observed cannot solely be explained in this way. After all, if that was the case, the two sets of children should have exhibited similar patterns, considering they were matched in age. The GEN 3 CHILDREN's virtual absence of prevoicing coupled with a number of other English-like patterns in their Sylheti stops suggest that other factors are at work, as well.

The most likely factor is linguistic experience. Specifically, the GEN 2 CHILDREN are growing up in a home where Sylheti is the dominant language and they live in an area that is densely populated with other Sylheti speakers. In contrast, the GEN 3 CHILDREN mainly hear English in their homes and there is substantially less Sylheti spoken in their immediate environment as they live in an ethnically more heterogeneous area. Finally, the differences between the two sets of children may be related to the input they receive. While input was not assessed directly in this study, based on an analysis of the children's mothers' productions, the GEN 2 CHILDREN may largely hear target-like productions from their mothers, while the input that the GEN 3 CHILDREN receive is likely to include a number of non-native features. The latter may not be significant in contexts where there is sufficient native-like input from other speakers. However, in contexts of reduced input, as in the present case, non-native patterns may be influential for the next generation of speakers. Mayr & Montanari

(2015), for instance, showed that multilingual children who only receive input in one of their languages from a single speaker are highly responsive to their patterns and home in on speaker-specific phonetic information. Hence, since the GEN 3 CHILDREN have restricted exposure to Sylheti, their mothers' non-native productions may be partly responsible for their own non-native realisations in the heritage language.

It is important to note that the differences between the two sets of children cannot be explained on the basis of current models of bilingual speech learning. The PAM-L2 (Best & Tyler, 2007) does not take any social variables into account, and the only one that is formalised in the SLM (Flege, 1995; Flege et al., 2003) is age of learning<sup>4</sup>. However, the GEN 2 CHILDREN and the GEN 3 CHILDREN did not differ on this variable as they had both been exposed to Sylheti from birth.

### ***4.3. Socio-cultural factors***

As reviewed in the introduction, a growing body of research has shown that heritage language features may occur in the host language to fulfil socio-indexical functions (Alam & Stuart-Smith, 2011; Heselwood & McChrystal, 2000; Hirson & Sohail, 2007; Kirkham, 2011; Kirkham & Wormald, 2015; Lambert et al., 2007; Sharma & Sankaran, 2011). For example, Kirkham (2011) showed that British Asians used retroflex realisations of /t/ in English to signal their Asian identity. In the present study, only the GEN 1 MUMS used clearly non-native forms in their English. It is uncertain whether they have only arisen from inadvertent interaction between the L1 and L2 sound systems, in line with previous research on L2 learners (cf. Flege, 1995; Flege et al., 2003), or whether they have also been mediated by social factors.

But what about the other participant groups? How can the absence of heritage language features in their productions be explained? To begin with, they have been shown to

emerge for the first time in adolescence and young adulthood, the most critical developmental periods for identity formation (see Steinberg & Morris, 2001 for an overview). The absence of heritage language features in the speech of the 3-5 year-old children is hence not surprising. It is less obvious, however, why the GEN 2 MUMS showed no evidence of heritage language forms in their English. One possibility is that they do use them, but only in informal contexts. Since the present study only assessed their productions in a formal experimental setting, this possibility cannot be addressed by the data gathered here. Alternatively, the use of these forms may be related to socio-economic status (SES). Hence, anecdotal evidence from members of the community suggests that heritage language forms, in particular retroflex realisations of /t/ and /d/, may be associated with low levels of education and SES. The GEN 2 MUMS, however, were all well-educated with the majority holding university degrees and employed in professional posts. Nevertheless, in the absence of detailed ethnographic data, this explanation remains speculative, and requires systematic investigation in future research.

#### ***4.4. Language maintenance and transmission***

In the present study, only first-generation migrants were identified as clearly non-native in the host language. In contrast, the heritage language showed incremental changes across successive generations: the GEN 1 MUMS' Sylheti stop productions, while not identical, were close to those of Sylheti speakers in Asia, those of the second-generation participants showed an increase in non-native forms, while the GEN 3 CHILDREN's productions were the least target-like. These patterns are in line with those observed in the Russian and Ukrainian heritage speakers described by Nagy and her associates (Hrycyna et al., 2011; Nagy, 2015; Nagy & Kochetov, 2013). As in the present study, they found a cross-generational trend away from homeland norms and towards those of the host language.

These patterns could suggest the emergence of a new contact variety, as has been argued for other British Asian communities (e.g. Heselwood & McChrystal, 1999). However, considering the changes observed across successive generations, Cardiff Sylheti would be a highly unstable variety with unclear norms. Moreover, its long-term survival is uncertain. Accordingly, English is the GEN 3 CHILDREN's predominant home language and there are few opportunities to use Sylheti in the neighbourhood. In addition, unlike the close-knit communities described elsewhere (McCarthy et al., 2011, 2013, 2014), there are virtually no new first-generation migrants joining the community that could help maintain the heritage language, and (re)introduce homeland norms.

On the other hand, the GEN 3 CHILDREN were able to converse in Sylheti and carry out a picture-naming task, suggesting reasonable linguistic abilities in the language overall, although this would need to be confirmed in a systematic study of their lexical and grammatical proficiency. They also showed clear evidence of speech learning despite converging patterns. For example, some of their /b<sup>h</sup>/ and /g<sup>h</sup>/ tokens were produced with target-like breathiness, and some of their coronal stop tokens were produced at the correct place of articulation. It remains to be seen whether these factors are sufficient to ensure the long-term survival of Sylheti in the community.

## **5. Conclusion**

This study is one of the few to investigate the speech patterns of heritage language speakers in both the heritage language and the host language. It showed differences in the production of Sylheti and English stops across generations, and between child and adult participants. As such, it constitutes an extension of previous work on the speech of Bangladeshi heritage children and adults in the UK (McCarthy, 2015; McCarthy et al., 2011, 2013, 2014), and is the first to examine the Sylheti stop productions of UK-born children. It also demonstrates

that non-target-like stop productions are not only manifest in voicing patterns, but also other areas of pronunciation, most notably the place of articulation of coronal stops, and breathiness in voiced categories. In the present study, these areas were assessed auditorily. Future work could complement the findings obtained here with additional acoustic measures, e.g. spectral analyses of stop bursts, and measures of intensity. These will require normalisation procedures to adjust for differences in vocal tract size (see Johnson, 2005 for an overview), and need to include native control groups.

This study is also the first to reveal substantial differences between second- and third-generation children in heritage language settings, with the latter exhibiting an increasing drift towards the patterns of the host language. These findings have important implications for the maintenance, transmission and long-term survival of heritage languages, and show that investigations need to go beyond second-generation speakers, in particular in communities that do not see a steady influx of new migrants. Future work is needed that builds on this research and examines systematically what factors contribute to successful transmission and maintenance of speech patterns in heritage language settings.

Finally, this study has important implications for theory and demonstrates that current models of bilingual speech learning cannot fully account for the speech patterns found in heritage language settings. Future work will need to extend these from their current focus on psycholinguistic processes and incorporate the social variables that mediate them.

## Notes

<sup>1</sup> In some of the literature on Sylheti and similar Indo-Aryan languages, the distinction in voiced stops is referred to as one of aspiration (e.g. Gope & Mahanta, 2015; Khan, 2010). In the present paper, we use the term breathiness, however, to distinguish it from the aspiration found in long-lag voiceless categories.

<sup>2</sup> Note that dental realisations of English /t/ and /d/ were not classified as target-like since they do not occur in Cardiff English (Mees & Collins, 1999).

<sup>3</sup> In the confusion matrices in Tables 3 and 5, the stops in slanted brackets on the left denote the intended categories, while those in square brackets denote target-like and non-target-like realisations.

<sup>4</sup> Note that although Flege and his associates have demonstrated the importance of language use for L2 speech learning (e.g. Flege et al., 1997), this variable has never been formalised in the SLM.

## References

- Abrahamsson, N. & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59, 249-306.
- Alam, F. & Stuart-Smith, J. (2011) Identity and ethnicity in /t/ in Glasgow-Pakistani high-school girls. *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 17-21 August 2011, pp. 216-219.
- Allen, G. (1985). How the young French child avoids the prevoicing problem for word-initial voiced stops. *Journal of Child Language* 12, 37-46.
- Aoyama, K., Flege, J.E., Guion, S., Akahane-Yamada, R. & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /ɾ/ and English /l/ and /r/. *Journal of Phonetics* 233-250.

Mayr & Siddika (2016, accepted)

Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science* 13, 238-243.

Auzou, P., Özsancak, C., Morris, R.J., Jan, M., Eustache, F. & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: a review. *Clinical Linguistics and Phonetics* 14, 131-150.

Bates, D., Maechler, M. , Bolker, B., Walker, S.(2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In Strange, W. (ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: commonalities and complementarities. In Bohn, O.-S. & Munro, M. (eds.), *Language experience in second-language speech learning. In honor of James Emil Flege* (pp. 13–34). Amsterdam/ Philadelphia: John Benjamins.

Birdsong, D. (2007). Native-like pronunciation among late learners of French as a second language. In Bohn, O.-S. & Munro, M. (eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 99-116). Amsterdam/ Philadelphia: John Benjamins.

Boersma, P., & Weenink, D. (2010). *PRAAT: doing phonetics by computer* (version 5.1.31). Institute of Phonetic Sciences, University of Amsterdam, online: <http://www.fon.hum.uva.nl/praat> (last accessed 5 April 2010).

Mayr & Siddika (2016, accepted)

Bongaerts, T., Mennen, S. & Slik, F. van der (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica* 54, 298-308.

Bortolini, U., Zmarich, C., Fior, R., & Bonifacio, S. (1995). Word-initial voicing in the productions of stops in normal and preterm Italian infants. *International Journal of Pediatric Otorhinolaryngology* 31, 191-206.

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

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

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *Journal of the Acoustical Society of America* 129, 3964-3980.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27, 207-229.

Colantoni, L., Steele, J. & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge: Cambridge University Press.

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

de Houwer, A. (2009). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics* 28, 411-424.

Mayr & Siddika (2016, accepted)

Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics* 24, 351-365.

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

Evans, B., Mistry, A. & Moreiras, C. (2007). An acoustic study of first- and second-generation Gujarati immigrants in Wembley: evidence for accent convergence? *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, Germany, 6-10 August 2007, pp. 1741-1744.

Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics* 26, 148-163.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). 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., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics* 23, 567-98.

Flege, J.E., Munro, M. & MacKay, I.R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97, 3125-3134.

Mayr & Siddika (2016, accepted)

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

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

Gandour, J., Petty, S., Dardarananda, R., Dechongkit, S., & Munkgoen, S. (1986). The acquisition of the voicing contrast in Thai: a study of voice onset time in word-initial stop consonants. *Journal of Child Language* 13, 561-572.

Gope, A. & Mahanta, S. (2015). An acoustic analysis of Sylheti phonemes. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*. Glasgow, ISBN 978-0-85261-941-4. Paper number 630, retrieved from <http://www.icphs2015.info/pdfs/Papers/ICPHS0630.pdf>

Grosjean, F. (2001). The bilingual's language modes. In Nicol, J. (ed.), *One mind, two languages: bilingual language processing* (pp. 1-22). Oxford: Blackwell.

Hansen Edwards, J. & Zampini, M.L. (eds.) (2008). *Phonology and second language acquisition*. Amsterdam/ Philadelphia: John Benjamins.

Heselwood, B. & McChrystal, L. (1999). The effect of age-group and place of L1 acquisition on the realisation of Panjabi stop consonants in Bradford: An acoustic sociophonetic study. *Leeds Working Papers in Linguistics and Phonetics*, 7, 49-69.

Heselwood, B. & McChrystal, L. (2000). Gender, accent features and voicing in Punjabi-English bilingual children. *Leeds Working Papers in Linguistics and Phonetics* 8, 45-70.

Mayr & Siddika (2016, accepted)

Hirson, A., Sohail, N. 2007. Variability of rhotics in Punjabi-English bilinguals. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, Germany, 6-10 August 2007, pp. 1501-1054.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.

Hrycyna, M., N. Lapinskaya, A. Kochetov & N. Nagy. (2011). VOT drift in 3 generations of heritage language speakers in Toronto. *Canadian Acoustics* 39, 166-167.

International Phonetic Association (2005). International Phonetic Alphabet, online: <https://www.internationalphoneticassociation.org/> (last accessed 22 December 2015).

Johnson, K. (2005) Speaker normalization in speech perception. In Pisoni, D.B. & Remez, R. (eds.), *The handbook of speech perception* (pp. 363-389). Oxford: Blackwell.

Kehoe, M. M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German–Spanish children. *Bilingualism: Language and Cognition* 7, 71-88.

Khan, S. (2010). Illustrations of the IPA: Bengali (Bangladeshi Standard). *Journal of the International Phonetic Association* 40, 221-225.

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

Khattab, G. (2002). /l/ production in English-Arabic bilingual speakers. *International Journal of Bilingualism* 6, 335-353.

Kirkham, S. (2011). The acoustics of coronal stops in British Asian English. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 17-21 August 2011, pp. 1102-1105.

Mayr & Siddika (2016, accepted)

Kirkham, S. & Wormald, J. (2015) Acoustic and articulatory variation in British Asian English liquids. *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow, ISBN 978-0-85261-941-4. Paper number 640, retrieved from <http://www.icphs2015.info/pdfs/Papers/ICPHS0640.pdf>.

Kupisch, T., Barton, D., Hailer, K., Klaschik, E., Stangen, I., Lein, T. & Weijer, J. van de (2014). Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal* 11, 123-150.

Lambert, K., Alam, F., Stuart-Smith, J. (2007). Investigating British Asian accents: Studies from Glasgow. *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, Germany, 6-10 August 2007, pp.1509-1511.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159-174.

Lein, T., Kupisch, T. & Weijer, J. van de (2015). Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism*, first published on June 17, 2015 as doi:10.1177/1367006915589424.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

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

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12, 251-285.

Mayr & Siddika (2016, accepted)

Macken, M. A., & Barton, D. (1979). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 7, 41-74.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language* 7, 433-458.

MacLeod, A., Stoel-Gammon, C. & Wassink, A.B. (2009). Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers. *Journal of Phonetics* 37, 374-387.

Maxwell, O., Baker, B., Bundgaard-Nielson, R. & Fletcher, J. (2015). A comparison of nonsense and real-word stimuli: Coronal stops in Bengali. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow, ISBN 978-0-85261-941-4. Paper number 900, retrieved from <http://www.icphs2015.info/pdfs/Papers/ICPHS0900.pdf>

Mayr, R. & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition* 13, 279-297.

Mayr, R. & Montanari, S. (2015). Cross-linguistic interaction in trilingual phonological development: the role of the input in the acquisition of the voicing contrast. *Journal of Child Language* 42, 1006-1035.

McCarthy, C. (2015). Speech perception and production in sequential bilingual children: A longitudinal study. *Journal of the Acoustical Society of America* 138, 1894.

- McCarthy, K., Evans, B. & Mahon, M. (2011). Detailing the phonetic environment: A sociophonetic study of the London Bengali community. *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 17-21 August 2011, pp. 1354-1357.
- McCarthy, K., 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.
- McCarthy, K., Mahon, M., Rosen, S. & Evans, B. (2014). Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition. *Child Development* 85, 1965-1980.
- Mees, I. & Collins, B. (1999). Cardiff: A real-time study of glottalization. In Foulkes, P. & Docherty, G. (eds.), *Urban voices: Accent studies in the British Isles* (pp. 185-202). London: Arnold.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32, 543-563.
- Nagy, N. (2015). A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua* 164B, 309-327.
- Nagy, N. & Kochetov, A. (2013). VOT across the generations: A cross-linguistic study of contact-induced change. In Siemund, P., Gogolin, I., Schulz, M. & Davydova, J. (eds.), *Multilingualism and language diversity in urban areas: Acquisition, identities, space, education* (pp. 19-38). Amsterdam/ Philadelphia: John Benjamins.

Office for National Statistics (2011). Key statistics for Wales, retrieved on 22 December 2015 from <http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-unitary-authorities-in-wales/stb-2011-census-key-statistics-for-wales.html>

Oh, J., Jun, S., Knightly, L., & Au, T. (2003). Holding on to childhood language memory. *Cognition* 86, 53–64.

Ohala, J. J. (1997). The relation between phonetics and phonology. In Hardcastle, W.J. & Laver, J. (eds.), *The handbook of phonetic sciences* (pp. 674-694). Oxford: Blackwell.

Piske, T., MacKay, I.R.A. & Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, 191-215.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.

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* (Vol.11). Oxford: Peter Lang.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics* 20, 213-223.

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

Simon, E. (2010). Child L2 development: a longitudinal case study on Voice Onset Times in word-initial stops. *Journal of Child Language* 37, 159-173.

Mayr & Siddika (2016, accepted)

Stangen, I., Kupisch, T., Ergün, A.L. & Zielke, M. (2015). Foreign accent in heritage speakers of Turkish in Germany. In Peukert, H. (ed.), *Transfer effects in multilingual language development* (pp. 87-108). Amsterdam/ Philadelphia: John Benjamins.

Steinberg, L. & Morris, A.S. (2001). Adolescent development. *Annual Review of Psychology* 52, 83-110.

Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *Journal of the Acoustical Society of America* 118, 1026-1037.

Van Alphen, P., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics* 32, 455-491.

Whiteside, S., Henry, L. & Dobbin, K. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America* 116, 1179-1183.

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition* 3, 131-149.