This is an Accepted Manuscript of an article published by Taylor & Francis in Asian Englishes on 21 September 2024, available at: https://doi.org/10.1080/13488678.2024.2391234*. Please see published version for final changes.*

Variation in VOT in English child-directed speech of English-Mandarin and English-Malay early bilinguals in Singapore

Jasper Hong Sim[a]

[a]*National Institute of Education, Nanyang Technological University, Singapore*

jasper.sim@nie.edu.sg

Variation in VOT in English child-directed speech of English-Mandarin and English-Malay early bilinguals in Singapore

Variability has been observed in the phonetic/phonological properties of child-directed speech from late-second language bilingual caregivers, but less is known about input variation among early bilinguals in multi-dialectal/multilingual contexts. This study examined English stop voicing contrasts in the adult-directed and child-directed speech of 30 Singaporean early bilingual mothers who differed in their other language (Mandarin/Malay) and language dominance. Results showed that, while Malay and Chinese mothers used similar voice onset time (VOT) for voicing contrast, they differed in where on the VOT continuum these contrasts were made, regardless of speech style and dominance. Compared to their English-dominant peers, Malay/Mandarin-dominant mothers also produced smaller contrasts overall. These patterns could be attributed to cross-linguistic influence, as well as effects of long-term language contact and the mothers' linguistic experiences. Sub-phonemic variation adds complexity to child bilingual phonological acquisition, emphasising the need to consider input properties in sociolinguistically complex contexts.

Keywords: caregiver input; adult-directed speech; variable input, new Englishes; voice onset time; stop voicing contrast

Introduction

There is general consensus that considerable between-speaker and within-speaker variability exists in the speech of adult bilinguals (e.g., Amengual, 2019; Bosch & Ramon-Casas, 2011). Yet studies on child bilingual production often assume a homogeneous input, and input properties are less often cited as a potential contributor, much less directly explored as a primary variable (Johnson, 2018; Kehoe, 2015). A small but growing body of work has begun to foreground the phonetic and phonological variation in the input that children receive (Fish et al., 2017; Mayr & Montanari, 2015; Sim, 2021) and the effects of such variability on phonological development and outcomes (Mayr & Siddika, 2018; Ramon-Casas et al., 2021; Sim, 2023; Sim & Post, 2021, 2023; Stoehr et al., 2019; for a review, see Sim & Post, 2024). The present study contributes to the underexplored area of phonological acquisition in multilingual and multi-dialectal contexts. It investigates Singaporean bilingual caregivers' production of voice onset time (VOT) of English word-initial stops, which is the time interval between the release of a stop closure and the onset of vocal fold vibration. The phonemic contrast between stops in many language varieties can be described in terms of the presence (long-lag VOT) or absence (short-lag VOT) of aspiration and/or prevoicing (lead VOT; Cho & Ladefoged, 1999; Lisker & Abramson, 1964), and VOTs of bilinguals may differ from their monolingual peers partly due to cross-linguistic influence (CLI; Kehoe, 2015). Past studies that examined caregiver input and child production were primarily interested in differences between monolinguals and bilinguals, and between 'native'/first language (L1) and 'non-native'/second language (L2) speakers. This present study examines variation between bilingual caregivers who acquired their two languages early but who differed in their other language and in their

language dominance, which is arguably more representative of societies characterised by widespread individual bilingualism and societal multilingualism.

Variation in caregiver input

Variability in speech may be observed in caregivers who acquired their L2 later in life and/or those who may not have achieved proficiency, such as those in bi-national families. Phonetic input from these caregivers can be inconsistent, and it may exhibit phonetic characteristics that differ from their monolingual peers due to the assimilation of or interactions between their two phonological systems (Barlow et al., 2013; Flege et al., 2003), which is further modulated by factors such as the amount of continued L1 use (Piske et al., 2001). These inter-speaker differences may manifest in the acoustic-phonetic enhancements caregivers make in their child-directed speech (CDS). In their examination of the VOT of late-L2 Spanish-English caregivers, Fish et al. (2017) found that bilingual caregivers produced English /p t/ with shorter VOT than monolinguals in both adult-directed speech (ADS) and infant-directed speech (IDS). Even with the exaggeration of VOT in IDS, bilingual caregivers' overall VOT for English /p t/ in IDS was shorter than monolinguals' overall VOT for the same plosives in monolingual ADS. Moreover, while monolinguals produced significantly longer VOT for English voiceless /p t/ than voiced /b d/ in IDS, which may help enhance infants' perception of voicing contrast, the effect was not observed in the bilingual caregivers, who increased the VOT of all stops to similar extents. Children are sensitive to sub-phonemic information in the input, and such fine-grained variation can influence language outcomes (Cristià, 2011; McMurray & Aslin, 2005; Sim & Post, 2021). Stoehr et al. (2019), for instance, investigated the effects of non-nativelike and attrited maternal input by examining the VOT production of Dutch-German preschoolers. These children acquired the heritage language, German, predominantly from their mothers who spoke L1 German, and acquired the majority language, Dutch, from their fathers who were L1 speakers of Dutch, and from their mothers, who were L2 speakers. They reported individual variation in the VOT of the child bilinguals, which was associated with individual variation of VOT in their mothers' non-native speech in Dutch and their attrited speech in the heritage language, German.

Variable or differential production is not restricted to late acquirers. In multi-dialectal, multilingual contexts, variation can arise from caregivers speaking two different L1s, and even those who speak the same languages may differ in their language dominance and experiences, thereby exhibiting accent differences (Amengual, 2019; Amengual & Chamorro, 2015; Bosch & Ramon-Casas, 2011; Kirkham & McCarthy, 2021). Moreover, in communities that have undergone long-term language contact or shift, as is the case of speakers of New Englishes or heritage languages, differential features that once emerged from effects of bilingualism in one generation may be transmitted to and retained by subsequent generations of proficient L1 speakers (Kirkham, 2017; Mayr & Siddika, 2018; Sim & Post, 2023; Nguyen, 2020; Li et al., 2023), and variably used as part of their ethnolinguistic repertoire for socially meaningful work (Gnevsheva, 2020; Sharma & Sankaran, 2011; Sim, 2021; Nguyen, 2018). In other words, what may appear as effects of individual bilingualism may in fact be learnt through the input, although these are often difficult to disentangle (Mayr et al., 2017). Sim (2019) examined the Singaporean English (SgE) accents of 10 educated English-Malay early bilinguals and found

that participants who were more Malay-dominant exhibited features that were explicated to be a likely result of Malay influence. Those who were English-dominant, contrarily, had an English accent that lacked Malay-specific markers to the extent that they were identified by naïve listeners to be ethnically Chinese. Sim argued that since these bilinguals acquired both languages early, the maintenance and use of ethnically marked features could be learnt from their Malay-dominant families and social circles, and not solely due to CLI. While sociolinguistic studies focusing on ethnic variation in ADS are not uncommon, much less is known about variability in the CDS of caregivers in these contexts.

Stop voicing contrast

English stop voicing contrast between fortis /p t k/ and lenis /b d g/ is traditionally described in terms of the presence or absence of aspiration and not in terms of phonetic voicing (Lisker & Abramson, 1964). However, in certain non-utterance-initial contexts, such as following sonorants or nasals, /b d g/ can be at least partially voiced (lead VOT; Abramson & Whalen, 2017; Davidson, 2016). The fortis–lenis contrast as described may not apply to the English stops of all bilinguals owing to potential influence from or assimilation with the stop contrast system of their other language (Fish et al., 2017; Kehoe, 2015). In Singaporean English (SgE), for example, impressionistic accounts described /p t k/ to be generally aspirated to a far lesser degree than in varieties such as British English, thereby overlapping with the VOT range for /b d g/ produced with positive VOT (Low & Brown, 2005). Deterding & Poedjosoedarmo (1998), however, noted that voicing contrast in SgE might be achieved by prevoicing /b d g/ stops, especially by Malay speakers, owing to potential influence from the Malay language that has a two-way contrast that involves short-lag /p t k/ and fully voiced /b d g/ stops (Clynes & Deterding, 2011; Deterding & Poedjosoedarmo, 1998, p. 175). Sim (2019) also found that, across speech styles, the VOT of English phonologically voiceless stops produced by English-dominant Malays was significantly much longer than those produced by their Malay-dominant counterparts. The initial stops of Mandarin, contrastingly, are all voiceless and there is a two-way contrast between short-lag /p t k/ and long-lag /p^h t^h k^h/ stop. Mandarin stops therefore have similar phonetic realisations along the positive VOT continuum to English stops, although (Taiwanese) Mandarin aspirated stops have been reported to have longer VOT than the same stops in (British) English (Chao & Chen, 2008).

The present study

This study explores variation in the implementation of stop voicing contrast by examining the VOT of English word-initial stops in the ADS and CDS of Singaporean Chinese and Malay mothers. The study seeks to ascertain whether there are differences in VOT within and between speech styles (i.e., ADS versus CDS) according to ethnicity (or 2L1/early L2) and language dominance. It is anticipated that the phonetic contrasts between voiced and voiceless stops will be greater in CDS than in ADS (e.g., Fish et al., 2017). Based on the assumptions that there may be CLI between the stop contrast systems of the mothers' two languages, it is also predicted that, especially for Malay-dominant mothers, the positive VOT of voiceless stops of Malay caregivers will be shorter than the same stops of their Chinese peers, but their lead VOT

will be longer, and that they will prevoice lenis stops more frequently (e.g. Deterding & Poedjosoedarmo, 1998; Sim, 2019).

Methodology

Participants

The participants were 15 ethnically Chinese (English-Mandarin bilinguals) and 15 Malay (English-Malay bilinguals) Singaporean mothers between the ages of 28 and 38 years ($N = 30$, $M = 33.5$, $SD = 2.97$). All caregivers were simultaneous or early sequential bilinguals who were exposed to both languages by the age of seven (age of acquisition of English: $M = 1.83$, $SD = 2.38$, $Mdn = 0$; Mandarin/Malay: $M = 0.83$, $SD = 1.80$, $Mdn = 0$). They had undergone the same education system that required them to learn English as an L1 and an ethnic mother tongue (EMT; Malay or Mandarin, depending on ethnicity) as a second. The Bilingual Language Profile (BLP; Birdsong et al., 2012), a self-report tool, was used to assess language dominance. Questions in the BLP fall under four sub-components: language history (includes age of acquisition), use, proficiency and attitudes. The BLP generates a composite score, which ranged from –218 (EMT-dominant) to +218 (English dominant). The scores of the Chinese mothers ranged from –58.66 to +150.8 ($M = 62.09$, $SD = 59.17$) and those of the Malays were between –30.78 and +127.8 ($M = 34.43$, $SD = 41.88$). Their preschoolers were on average 48 months old ($SD = 11.9$), of which 15 were girls (eight Malays) and 15 were boys (seven Malays). All participants were recruited through word of mouth and social media as part of a larger project. Ethical approval was obtained from the Ethics Committee for the School of the Humanities and Social Sciences, University of Cambridge (no.19/199).

Materials

The stimuli were English monosyllabic and stress-initial disyllabic words with word-initial stop /p t k b d g/ that preceded a close vowel /i, u/ (Table 1). Stimuli used to elicit ADS were embedded in the carrier phrase 'Please say _ again'. Many of the same target words and others with word-initial stops in the same vocalic contexts were elicited in CDS. This was done through a picture description task that involved a park scene and the reading of a children's storybook, 'Duck and Goose' (Hills, 2006).

Table 1. Experimental stimuli.

Stop	Target word	
	Adult-directed speech	**Child-directed speech**
/p/	Peacock, pea, people, pool	Peacock, peas, people
/t/	T-shirt, tea, tickling, two	T-shirt, tea, tickling
/k/	Key, kicking, kitten, cooed, cook, could,	Key, kicking, kitten
/b/	Bead, bee, bid, boo	Big, book, busy
/d/	'D', do	Did, do
/g/	Geese, good	Geese, good, goose

Recording procedures

The recording took place in a quiet room with minimal reverberation and noise in the respective homes of the participants. Mothers carried out the picture description task and book reading with their child without the presence of any other adult. They were instructed to use only English to interact with their child, to avoid a bilingual mode, and to speak as they would normally with them. CDS was recorded using a NAGRA ARES-MII recorder through an omni-directional lapel microphone that was pinned on the collar of the mother. The elicitation of ADS occurred in the quiet room with only the author. Mothers were asked to read each sentence twice. This was recorded by the author using a Zoom H5 recorder. Both recorders recorded at a sampling rate of 44.1 kHz at 16 bit.

Acoustic analysis

VOT boundaries for all word-initial stops were placed manually based on acoustic cues in the waveforms and spectrograms (Figure 1) on Praat (v. 6.2.17; Boersma & Weenink, 2022). VOT was defined as the time interval between the release burst as signalled by a sharp peak in waveform energy and the onset of voicing, which was taken to be the nearest upward-going zero-crossing of the start of periodicity in the waveform. In cases of multiple stop bursts, measurements were taken from the earliest release burst that was followed by continuous frication to the onset of voicing. If voicing occurred during the closure, as in the case of lead voicing, VOT was measured from the onset of glottal pulses in the waveform or discernible visual cues in the spectrogram that indicate voicing onset up to the first release burst. In non-utterance-initial positions, word-initial stops may be partially voiced (see Davidson, 2016). In such cases, following Abramson & Whalen (2017), stops were considered to have lead VOT equal to the closure duration if the closure was voiced for at least half of its duration; otherwise, they were considered to have positive VOT. 8% of all tokens (n = 226) were excluded from further analysis because VOT could not be reliably measured due to noise and overlapping speech. The final dataset consists of 1222 stops in ADS and 1141 in CDS.

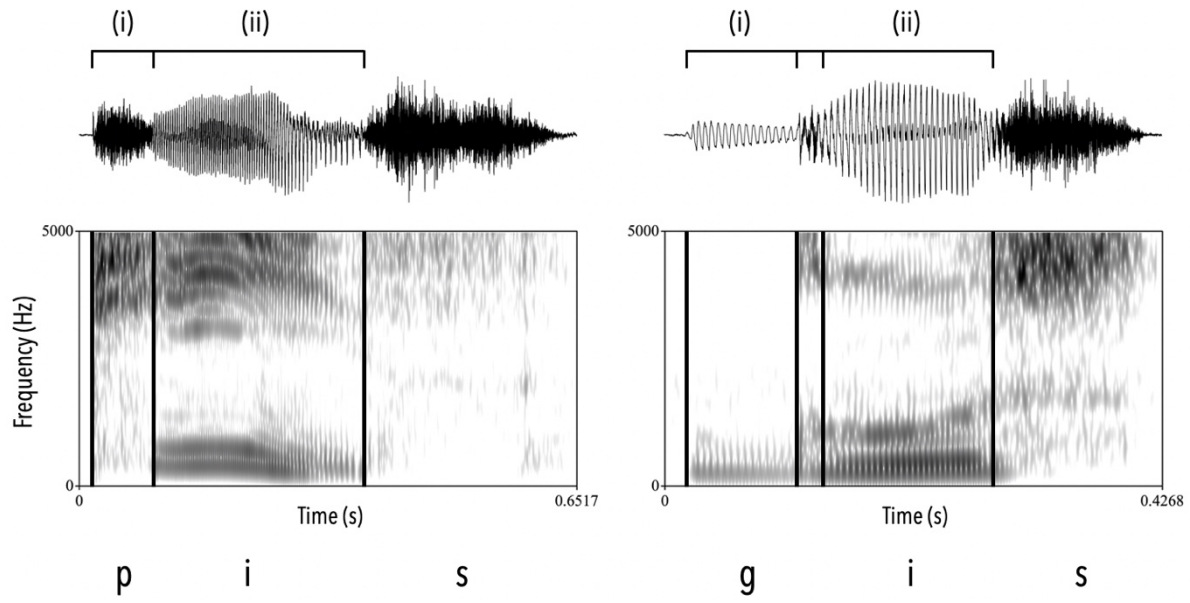

Figure 1. Representative waveforms and spectrograms of *peas* (left, positive VOT) and *geese* (right, lead VOT). (i) VOT (ii) vowel duration.

Various sources of VOT variation were considered. Faster speaking rate is associated with shorter VOT, especially for aspirated voiceless stops and prevoiced stops (Kessinger & Blumstein, 1997; Miller et al., 1986). Following previous studies (e.g., Chodroff & Wilson, 2017; Fish et al., 2017), vowel duration was used in this study as a proxy for speech rate, which was calculated as the latency between voicing onset and offset. Stops with anterior articulations generally have shorter positive VOT and longer negative VOT (Cho & Ladefoged, 1999; Herd, 2020), and so place of articulation was considered as a factor in the statistical models. VOT can also be moderated by prosodic and contextual factors. In their analysis of VOT in American English, Chodroff & Wilson (2017) found that in connected speech, stops in utterance-initial, post-pausal, and pre-pausal positions had longer positive VOT, and stops in utterance-final position had shorter VOT. In the present study, stops were coded for utterance/phrase position, namely initial, medial, or final. The segment preceding the stop was also noted; they were either a nasal, sonorant (vowel and glides), voiced or voiceless fricative, voiced or voiceless stop, or pause (breath or silence of longer than 150 ms). As syllable-final voiced obstruents in SgE tend to be devoiced (Bao, 2003), whether they were marked as 'voiced' was contingent on the presence of a voice bar in the frication noise for fricatives and presence of voicing at the beginning of closure for voiced stops. Finally, syllable number was recorded, as VOT is generally longer in monosyllabic than polysyllabic words (Chodroff & Wilson, 2017).

Statistical analysis

Mixed-effects regression analyses were conducted using R software (v. 4.2.1; R Core Team, 2022) and the 'lme4' package (Bates et al., 2015). In all models, the random effect structure included random intercepts for subject and word and, for variables of interest, by-subject and by-word slopes, as justified by the data. Fixed effects that were included in the models are

specified in the following. All continuous predictors were z-standardised. Categorical predictors were weighted effect coded (te Grotenhuis et al., 2017). To evaluate the contribution of each predictor, and to arrive at a more parsimonious model, pairwise model comparisons between a full model that included all the explanatory variables and a more restricted model that excluded the predictor under consideration were performed using likelihood ratio tests. Interaction terms were further investigated using the 'emmeans' package (Lenth, 2018). Outliers in the raw measurements were detected using the interquartile range method (n = 0). After checking the normality of residuals, influential outliers were removed using Cook's distance, before the final, optimised models were obtained and reported in the following.

Results

An examination of raw VOT values revealed bimodal distributions for both Chinese and Malay mothers (Figure 2). Figure 3 further shows VOT patterns according to ethnicity, style, and place of articulation. Voiceless and voiced stops with positive VOT were analysed separately from voiced stops with negative VOT.

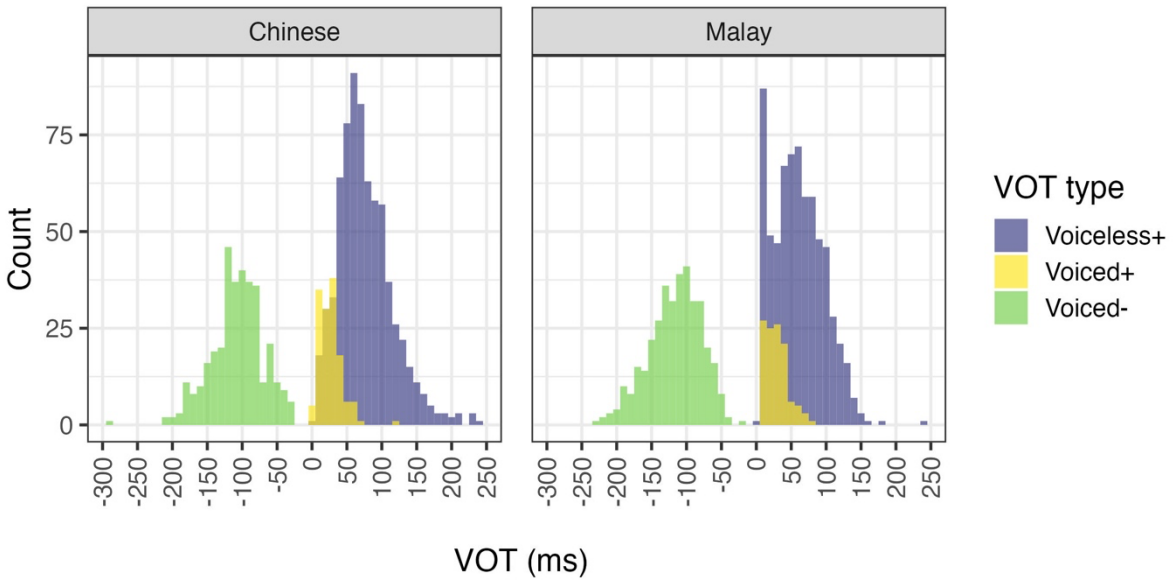

Figure 2. Distributions of raw VOT values.

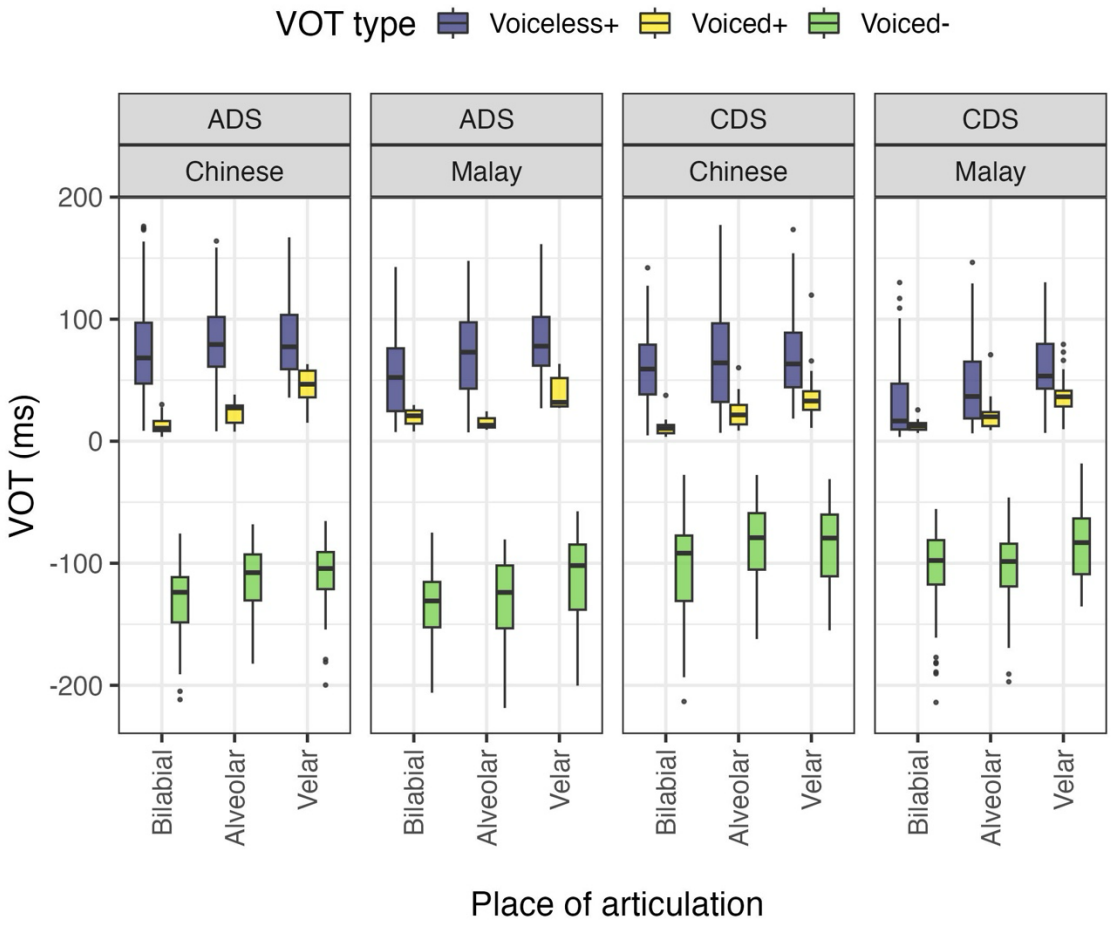

Figure 3. VOT patterns by function of ethnicity, style, VOT type and place of articulation. Outliers beyond +180 ms and -220 ms (n = 18) were excluded from the plots for clarity.

Positive VOT

Means of positive VOT according to ethnicity and style are presented in Table 2. The fixed effects included in the maximal model were style (contrast weights: CDS = -0.96, ADS = 1), ethnicity (Malay = -1.07, Chinese = 1), number of syllables (disyllabic = -1.02, monosyllabic = 1), vowel (back = -3.35, front = 1), place of articulation (velar-alveolar: -0.81, 1; velar-bilabial: -0.74, 1), phonological voicing (voiced = -5.50, voiceless = 1), preceding segment (pause, nasal, sonorant, voiced fricatives, voiced stops, voiceless fricatives, voiceless stops[1]), phrase position (medial-initial: -0.30, 1; medial-final: -0.26, 1), gender of child (female = -1.08; male = 1), vowel duration, BLP, and age of child. All combinations of three-way interaction between voicing with ethnicity, style, and BLP were also added in the full model.

[1] 'Pause' was the omitted level. Coding (level/pause): nasal (1/-0.10), sonorant (1/-3.34), voiced fricatives (1/-0.01), voiced stops (1/-0.03), voiceless fricatives (1/-0.31), voiceless stop (1/-0.16).

In the reduced model[2] (observations = 1520, marginal $R^2 = 0.63$, conditional $R^2 = 0.80$), the main effects of style, ethnicity, BLP, voicing, vowel duration, syllable number, place of articulation, and the three two-way interactions: ethnicity and voicing, style and voicing, and BLP and voicing, were significant predictors. The interaction between ethnicity and voicing (β = 0.03, *SE* = 0.31, *t* = 3.75, *p* < .001) revealed that both Chinese and Malay mothers exhibited phonemic contrast between phonologically voiced and voiceless stops by producing significantly longer VOT for voiceless stops regardless of style and BLP. The contrast, however, was smaller for Malays as their voiceless stops had shorter VOT than those produced by Chinese mothers (*b* = -16.14, *SE* = 3.31, *t* = -4.88, *p* < .001). In the interaction between style and voicing (β = 0.04, *SE* = 0.37, *t* = 3.50, *p* < .001), VOT was longer in ADS than CDS regardless of ethnicity and BLP, but only significantly so for phonologically voiceless stops (*b* = 15.51, *SE* = 2.75, *t* = 5.64, *p* < .001). There were also effects of BLP on voicing contrasts (β = 0.03, *SE* = 0.32, *t* = 2.86, *p* = .004). Spotlight analysis at mean, lower and upper quantiles of BLP and pairwise comparisons (Bonferroni adjusted) revealed that, regardless of style and ethnicity, as dominance in English increases, VOT of voiceless stops only is significantly longer, and therefore also the voiced-voiceless contrast (*b* = 8.45, *SE* = 2.30, *t* = 3.67, *p* = .003). Additionally, compared to the average, VOT of stops in disyllabic words was shorter than monosyllabic ones (β = 0.34, *SE* = 1.97, *t* = 6.01, *p* < .001), and VOT was positively associated with vowel duration (β = 0.21, *SE* = 0.72, *t* = 11.05 *p* < .001). Velar stops had longer VOT than both alveolar stops (*b* = 13.70, *SE* = 4.32, *t* = 3.17, *p* = .01) and bilabial stops (*b* = 19.17, *SE* = 4.12, *t* = 4.65, *p* = .0003), but the difference between alveolar and bilabial stops was not significant.

In sum, the analysis of positive VOT revealed that mothers overall produced longer VOT for phonologically voiceless stops than for voiced stops, thereby showing phonemic contrasts in the positive VOT dimension. However, for voiceless stops only, VOT was overall shorter for Malay mothers (than Chinese), in CDS (than ADS) and for more EMT-dominant mothers (compared to more English-dominant mothers).

Table 2. VOT means of stops produced with positive VOT (*n* = 1652).

Stop	Chinese						Malay					
	ADS			CDS			ADS			CDS		
	n	***M***	***SD***	***n***	***M***	***SD***	***n***	***M***	***SD***	***n***	***M***	***SD***
/p/	105	74.66	36.24	112	61.68	33.02	98	53.28	30.82	112	33.16	35.13
/t/	115	85.51	38.63	110	67.52	42.79	116	71.70	34.44	92	45.62	32.86
/k/	163	88.48	37.82	110	74.32	38.46	161	82.30	28.22	104	62.92	27.21
/b/	17	14.19	8.11	19	11.38	7.57	3	19.49	10.93	13	13.14	5.10
/d/	11	23.88	9.63	40	23.17	11.02	3	15.66	7.88	37	20.87	11.35
/g/	11	45.62	14.68	42	35.64	17.83	7	40.43	15.08	51	37.78	13.44

[2] Model syntax:
lmer(vot~style+ethnicity+blp+voicing+vowel_dur+syllable+poa+voicing*ethnicity+voicing*style+voicing*blp+blp*ethnicity*style+(1+style+voicing|subject)+(1+style|word))

Negative VOT

An analysis was first performed to ascertain whether style, ethnicity, and BLP moderated the likelihood of whether lenis stops were produced with lead voicing. The proportions of prevoiced stops according to style, ethnicity and preceding segment are presented in Figure 4. There was considerable individual variation that could not be reflected in the figure; proportions of voiced stops that were prevoiced ranged from 28% to 90% (*Mdn* = 76.7%) for Chinese mothers, and from 58% to 93% (*Mdn* = 75.3%) for Malay mothers.

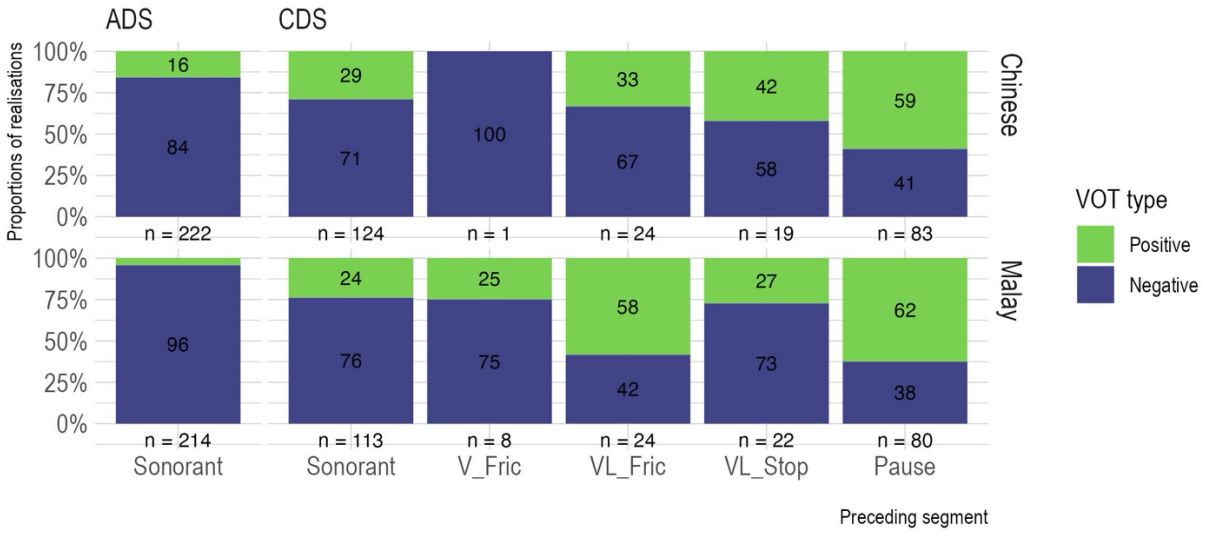

Figure 4. Proportions of voiced stops with positive and negative VOT according to style, ethnicity, preceding segment (V = Voiced, VL = Voiceless, Fric = Fricative).

Generalised linear mixed-effects modelling was run on all phonologically voiced stops (*n* = 934). The binary response variable was VOT type (positive = 0, negative = 1). The fixed effects in the maximal model included style (CDS = -0.88, ADS = 1), ethnicity (Malay = -1.03, Chinese = 1), number of syllables (disyllabic = -39.61, monosyllabic = 1), place of articulation (velar-alveolar = -0.88, 1; velar-bilabial = -1.03, 1), preceding segment (pause, sonorant, voiced fricatives, voiceless fricatives, voiceless stops[3]), vowel duration and BLP. Phrase position was excluded due to high collinearity with preceding segment type.

In the reduced model (observations = 934, marginal R^2 = .33, conditional R^2 = .51), the main effects of style, ethnicity, preceding segment, and the two-way interaction between style and ethnicity were significant predictors. The interaction between style and ethnicity (*OR* = 0.66, 95% CI [0.48, 0.91], *p* = .01) revealed that stops of Chinese mothers were less likely to be voiced compared to average, but only so in ADS (*b* = -1.53, *SE* = 0.72, *z* = -2.13, *p* = .03). Additionally, compared to average, stops following sonorants were more likely to be prevoiced (*OR* = 1.42, 95% CI [1.24, 1.62], *p* < .001), but those following voiceless fricatives were less likely to be voiced (*OR* = 0.42, 95% CI [0.23, 0.79], *p* = .007).

[3] 'Pause' was the omitted level. Coding (level/pause): sonorant (1/-4.13), voiced fricatives (1/-0.06), voiceless fricatives (1/-0.29), voiceless stops (1/-0.25).

VOT length

Means of negative VOT according to ethnicity and style are presented in Table 3. The data of Mi16 (n = 11) were excluded from the regression analysis because she did not complete the book reading activity. The data of C35 (n = 9) were also excluded due to poor recording quality during book reading. The only one stop that preceded a nasal consonant was also removed. The fixed effects included in the maximal model were style (contrast weights: CDS = -1.32, ADS = 1), ethnicity (Malay = -0.95, Chinese = 1), number of syllables (disyllabic = -42.1, monosyllabic = 1), gender of child (female = -0.87, male = 1), vowel (back = -0.95, front = 1), place of articulation (velar-alveolar: -0.91, 1; velar-bilabial: -1.3, 1), preceding segment (pause, sonorant, voiced fricatives, voiceless fricatives, voiceless stops[4]), phrase position (medial-initial: -0.13, 1; medial-final: -0.12, 1), vowel duration, BLP, and age of child. The three-way interaction between ethnicity, style, and BLP was also added to the full model.

In the reduced model[5] (observations = 642, marginal R^2 = .36, conditional R^2 = .74), the main effects of style, ethnicity, vowel duration and preceding segment were significant predictors. Compared to average, negative VOT in CDS was shorter than in ADS (β = 0.24, SE = 2.23, t = -3.80, p < .001). Malay mothers produced longer negative VOT than average, regardless of style or BLP (β = -0.16, SE = 1.80, t = -3.10, p = .002). Additionally, negative VOT was positively associated with vowel duration (β = 0.26, SE = 1.07, t = 8.79, p < .001). Compared to the weighted grand mean, negative VOT of stops following sonorants was longer (β = 0.14, SE = 0.54, t = 8.89, p < .001), but shorter for those after voiced fricatives (β = -1.45, SE = 10.37, t = -4.94, p < .001) and voiceless fricatives (β = -0.65, SE = 4.41, t = -5.17, p < .001).

Table 3. VOT means of stops produced with negative VOT (n = 711).

Stop	Chinese						Malay					
	ADS			CDS			ADS			CDS		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
/b/	99	130.27	28.86	41	107.16	52.34	110	137.02	29.27	42	108.49	41.67
/d/	48	113.87	29.79	47	82.87	33.60	55	131.18	35.55	49	104.39	33.91
/g/	49	109.89	28.21	62	82.94	28.85	51	110.33	32.05	58	84.63	28.54

In sum, the analyses on phonologically voiced stops revealed that stops produced by Malay mothers were more likely to be prevoiced than those by Chinese mothers in ADS. Across styles, Malay mothers' prevoiced stops also had longer negative VOT than those produced by Chinese mothers. Mothers overall produced prevoiced stops with shorter negative VOT in CDS than ADS.

[4] 'Pause' was the omitted level. Coding (level/pause): sonorant (1/-8.84), voiced fricatives (1/-0.11), voiceless fricatives (1/-0.41), voiceless stops (1/-0.42).

[5] Model syntax:
lmer(vot~style+ethnicity+blp+vowel_dur+preceding_segment+(1+style|id)+(1|word)).

Discussion

In communities characterised by widespread multilingualism and those that are experiencing or have undergone long-term language contact and shifts, there may be variation in the phonetic input that children receive because of, *inter alia*, the caregivers speaking a different other language and varying in their language experience. This study explored the variation in the implementation of English stop voicing contrast in the ADS and CDS of 30 Singaporean mothers, 15 of whom were English-Mandarin bilinguals and ethnically Chinese, and the others ethnically Malay and English-Malay bilinguals. These caregivers were early bilinguals who were born and raised in the same broader sociolinguistic context but differed in their other language and language dominance. Key findings revealed that for stops produced with positive VOT, mothers overall produced longer VOT for voiceless /p t k/ stops than /b d ɡ/ stops. However, for voiceless stops only, VOT was shorter for Malay mothers, across mothers in CDS, and also shorter for less English-dominant mothers. The analyses on phonologically voiced stops revealed that voiced stops were more likely to be produced with lead VOT by Malay mothers but only significantly so in ADS. Malay mothers' prevoiced stops also had longer lead VOT than those produced by Chinese mothers. Mothers overall produced longer negative VOT in ADS than in CDS.

ADS versus CDS

Caregiver input facilitates language learning as it conveys language-specific information, and modulations in CDS may involve the enhancement of phonetic contrasts that could aid in category formation (Englund, 2005; Fish et al., 2017; Werker et al., 2007). An exaggeration of English stop voicing contrast in CDS would entail longer positive VOT for voiceless stops and/or longer lead voicing for prevoiced stops. In this study, such phonetic enhancement was not observed; instead, phonetic contrasts were bigger in ADS for mothers overall. This cannot be explained by differences in speech rate alone, since some of its effects were controlled for by considering vowel duration in the analyses, and speech rate in CDS tends to be slower than in ADS, which would result in longer VOTs. One reason could be the age of the preschoolers. Some studies proposed that phonetic enhancements are age-mediated and coincide with the child's linguistic development (e.g., Cristià, 2010). Enhancements in stop voicing contrasts were reported in infant-directed speech (Englund, 2005; Fish et al., 2017), and towards preschoolers who were on average 48 months old; mothers in this study may no longer perceive the need for an exaggeration in this aspect, especially since their child would have begun to produce English stops consistently. The absence of enhancement could also be due to the methodology. Variation in styles was investigated by comparing semi-controlled CDS that involved teaching of target words in a picture description task and in the reading of a storybook, and ADS in the form of careful reading of target words in carrier phrases, in a bid to achieve better control over linguistic factors that modulate VOT. However, mothers in this study could have produced canonical forms in their ADS that do not reflect their natural speech or local dialectal norms, especially since Singaporeans are socialised through formal education to approximate an exonormative standard in their very self-conscious speech. That both styles involve read speech also poses a limitation in this study; it remains unclear whether the input

patterns observed reflect the informal, unselfconscious input that children are more commonly exposed to, which is expected to exhibit a higher degree of variability.

Variation between mothers

The findings of this study support the predictions that the positive VOT of voiceless stops of Malay caregivers would be shorter than their Chinese counterparts, and that Malay mothers would produce more and longer truly voiced stops in English. The data also supported the prediction that EMT-dominant mothers would produce shorter positive VOT. While Malay and Chinese mothers in this study were similar in using VOT for voicing contrast (longer positive VOT for voiceless stops and shorter positive/lead VOT for voiced stops), they were dissimilar in where in the VOT continuum the contrasts were made: regardless of language dominance and style, Malay mothers employed shorter positive VOT but longer lead VOT, whereas Chinese mothers had longer positive VOT but shorter lead VOT. This difference was not unexpected and could partially be explained by cross-linguistic influence. Recall that Malay employs a two-way contrast between truly voiced and unaspirated stops. Mandarin, by contrast, employs a two-way phonetic distinction between unaspirated and aspirated stops in the positive VOT dimension, and therefore has similar phonetic realisations to English lenis/fortis stops produced with positive VOT. Previous work has shown that although early bilinguals show autonomy of two phonological systems, their two languages may interact in terms of transfer or convergence of phonetic categories and details that result in new, 'hybrid' accents (Barlow et al., 2013; Kehoe, 2015; Kirkham & McCarthy, 2021; Sim & Post, 2023). In the same way, the short-lag and truly voiced stops of Malay could have influenced the English stop system of the Malay caregivers such that they were producing shorter aspirated stops and more and longer truly voiced stops in English than their Chinese counterparts.

The linguistic experiences of the mothers in this study could have also contributed to their differential production. Due to the language shift towards English that began in the 1960s, more Singaporeans in later generations are L1 speakers of English and are increasingly proficient, but this is not the case for the parents of the caregivers in this study, who might have acquired English late, or were non-English speaking bilinguals of other heritage languages (Bolton & Ng, 2014). Malay mothers, especially those who were Malay-dominant, might have been raised in homes where Malay was the dominant language, and could have been exposed to Malay-influenced English from significant adults and peers, thereby acquiring these differential features through vertical and horizontal transmission, in addition to effects of CLI (Bosch & Ramon-Casas, 2011; Mayr & Siddika, 2018; Ramon-Casas et al., 2021; Sim, 2019; Sim & Post, 2023; Stoehr et al., 2019). This could also potentially explain Chinese mothers' use of prevoiced stops. It was revealed that there was great interadult variation in the frequency at which phonologically voiced stops were prevoiced by Chinese mothers (28% to 90%), and the high rate of prevoicing is unexpected due to the lack of voiced stops in Mandarin. Instead of CLI, Chinese mothers who consistently produced prevoiced English stops could have learnt this feature through caregivers or peers who spoke an English that was influenced by historical language varieties that have been gradually replaced. These include Chinese languages that have prevoiced stops such as Hokkien and Teochew, which were widely spoken within the

Chinese communities, and from Bazaar Malay, a Malay-lexified pidgin that served as a *lingua franca* before it was replaced by English.

Implications for language acquisition

The primary aim of this study was to characterise variation in the input that bilingual children in sociolinguistically complex societies are exposed to. The findings of this study contribute to the small body of work that foregrounds the phonetic and phonological variation that exists in the input that children receive from late-L2 bilingual caregivers (Fish et al., 2017; Khattab, 2002; Mayr & Siddika, 2018; Stoehr et al., 2019), by providing evidence of variation in the input of caregivers who were early bilinguals but who differed in their other language and in their language dominance, which is commonplace in many communities that are characterised by widespread bi-/multilingualism. Children are sensitive to sub-phonemic information in the input, and fine-grained variation in the input is not ignored in the acquisition process (Cristià, 2011; McMurray & Aslin, 2005; Sim & Post, 2021; Stoehr et al., 2019). In his preliminary study that examined the VOT production of a subset of the children of the caregivers in this study, Sim (2023) found that, despite being equally highly English dominant in terms of amount of language use (input and output), the children of Chinese mothers also produced significantly longer positive VOT for their English fortis stops than those produced by their Malay peers, reflecting adult norms. Indeed, investigations on child production that are situated within similarly complex social-linguistic contexts should not assume that the input children receive is qualitatively homogeneous, and should consider input properties when explaining variable language outcomes.

This study also highlights the complexity of bilingual phonological acquisition in multi-dialectal, multicultural contexts. Not only could children be exposed to phonetic input that differs between caregivers due to their speaking a late L2 or a different early L2/second L1, as well as in their language dominance, but there may also be phonetic overlap in the contrast systems of the child's two languages. In this study, for instance, the children of more Malay-dominant mothers would be exposed to English and Malay stops that have very similar VOT properties, which may delay category formation and stabilisation (Bosch & Ramon-Casas, 2011; Ramon-Casas et al., 2021). In the preliminary study, Sim (2023) also compared the Malay children's Malay voiceless unaspirated /p t k/ stops with their English fortis stops and found that regardless of languge dominance, most did not differentiate these stops in their VOT, which suggests category assimilation (Barlow et al., 2013; Flege et al., 2003). In schools, children may further be exposed to input of significant adults or peers that qualitatively differs from that received at home. Further work can explore whether and how exposure to other accent models may induce change.

Acknowledgements

An earlier version of this article was presented at UK Language Variation and Change 14. The author is grateful to the families who participated in this study. This work was supported by fieldwork grants from the Faculty of Modern and Medieval Languages and Linguistics, and Jesus College, University of Cambridge. The author would like to thank the National Institute

of Education, Nanyang Technological University, Singapore for the fieldwork grants and PhD funding that made the study possible. The author would also like to thank their PhD supervisor, Brechtje Post, for unwavering support.

Disclosure statement

The author reports there are no competing interests to declare.

References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Amengual, M. (2019). Type of early bilingualism and its effect on the acoustic realization of allophonic variants: Early sequential and simultaneous bilinguals. *International Journal of Bilingualism*, *23*(5), 954–970. https://doi.org/10.1177/1367006917741364

Amengual, M., & Chamorro, P. (2015). The Effects of Language Dominance in the Perception and Production of the Galician Mid Vowel Contrasts. *Phonetica*, *72*(4), 207–236. https://doi.org/10.1159/000439406

Bao, Z. (2003). Social stigma and grammatical autonomy in nonnative varieties of English. *Language in Society*, *32*(1), 23–46. https://doi.org/10.1017/S0047404503321025

Barlow, J. A., Branson, P. E., & Nip, I. S. B. (2013). Phonetic equivalence in the acquisition of /l/ by Spanish–English bilingual children. *Bilingualism: Language and Cognition*, *16*(1), 68–85. https://doi.org/10.1017/S1366728912000235

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.3.14) [Computer software]. http://www.praat.org/

Bolton, K., & Ng, B. C. (2014). The dynamics of multilingualism in contemporary Singapore. *World Englishes*, *33*(3), 307–318. https://doi.org/10.1111/weng.12092

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, *39*(4), 514–526. https://doi.org/10.1016/j.wocn.2011.02.001

Chao, K.-Y., & Chen, L. (2008). A Cross-Linguistic Study of Voice Onset Time in Stop Consonant Productions. *International Journal of Computational Linguistics & Chinese Language Processing, Volume 13, Number 2, June 2008*, 215–232. https://aclanthology.org/O08-4005

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30–47. https://doi.org/10.1016/j.wocn.2017.01.001

Clynes, A., & Deterding, D. (2011). Standard Malay (Brunei). *Journal of the International Phonetic Association*, *41*(02), 259–268. https://doi.org/10.1017/S002510031100017X

Cristià, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *The Journal of the Acoustical Society of America*, *128*(1), 424–434. https://doi.org/10.1121/1.3436529

Cristià, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of the Acoustical Society of America*, *129*(5), 3271–3280. https://doi.org/10.1121/1.3562562

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, *54*, 35–50. https://doi.org/10.1016/j.wocn.2015.09.003

Deterding, David., & Poedjosoedarmo, G. R. (1998). *The sounds of English: Phonetics and phonology for English teachers in Southeast Asia*. Prentice Hall; /z-wcorg/.

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Language*, *25*(2), 219–234. https://doi.org/10.1177/0142723705050286

Fish, M. S., García-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2017). Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *Journal of Phonetics*, *63*, 19–34. https://doi.org/10.1016/j.wocn.2017.04.003

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0

Gnevsheva, K. (2020). The role of style in the ethnolect: Style-shifting in the use of ethnolectal features in first- and second-generation speakers. *International Journal of Bilingualism*, *24*(4), 861–880. https://doi.org/10.1177/1367006920902520

Herd, W. (2020). Sociophonetic voice onset time variation in Mississippi English. *The Journal of the Acoustical Society of America*, *147*(1), 596–605. https://doi.org/10.1121/10.0000545

Hills, T. (2006). *Duck & Goose* (1st ed). Schwartz & Wade Books.

Johnson, E. K. (2018). Putting the terms “monolingual” and “bilingual” under the microscope. *Applied Psycholinguistics*, *39*(4), 753–756. https://doi.org/10.1017/S0142716418000206

Kehoe, M. (2015). Cross-linguistic interaction: A retrospective and prospective view. *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015*. International Symposium on Monolingual and Bilingual Speech.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, *25*(2), 143–168. https://doi.org/10.1006/jpho.1996.0039

Khattab, G. (2002). /L/ production in English-Arabic bilingual speakers. *International Journal of Bilingualism*, *6*(3), 335–353. https://doi.org/10.1177/13670069020060030701

Kirkham, S. (2017). Ethnicity and phonetic variation in Sheffield English liquids. *Journal of the International Phonetic Association*, *47*(1), 17–35. https://doi.org/10.1017/S0025100316000268

Kirkham, S., & McCarthy, K. M. (2021). Acquiring allophonic structure and phonetic detail in a bilingual community: The production of laterals by Sylheti-English bilingual children. *International Journal of Bilingualism*, *25*(3), 531–547. https://doi.org/10.1177/1367006920947180

Lenth, R. (2018). *Estimated marginal means, aka least-squares means.* [R package].

Li, K. K., Nguyen, L., Bryant, C., & Yoo, K. (2023). Lexical tonal effects in code-switching: A comparative study of Cantonese, Mandarin, and Vietnamese switching with English. *International Journal of Bilingualism*, 13670069231181508. https://doi.org/10.1177/13670069231181508

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. WORD, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Low, E. L., & Brown, A. (2005). *English in Singapore: An introduction*. McGraw-Hill.

Mayr, R., & Montanari, S. (2015). Cross-linguistic interaction in trilingual phonological development: The role of the input in the acquisition of the acquisition of voicing contrast. *Journal of Child Language*, *42*(5), 1006–1035. https://doi.org/10.1017/S0305000914000592

Mayr, R., Morris, J., Mennen, I., & Williams, D. (2017). Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in Welsh and English. *International Journal of Bilingualism*, *21*(3), 245–267. https://doi.org/10.1177/1367006915614921

Mayr, R., & Siddika, A. (2018). Inter-generational transmission in a minority language setting: Stop consonant production by Bangladeshi heritage children and adults. *International Journal of Bilingualism*, *22*(3), 255–284. https://doi.org/10.1177/1367006916672590

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, *95*(2), B15–B26. https://doi.org/10.1016/j.cognition.2004.07.005

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking Rate and Segments: A Look at the Relation between Speech Production and Speech Perception for the Voicing Contrast. *Phonetica*, *43*(1–3), 106–115. https://doi.org/10.1159/000261764

Nguyen, L. (2018). Borrowing or Code-switching? Traces of community norms in Vietnamese-English speech. *Australian Journal of Linguistics*, *38*(4), 443–466. https://doi.org/10.1080/07268602.2018.1510727

Nguyen, L. (2020). *Cross-generational linguistic variation in the Canberra Vietnamese heritage language community: A corpus-centred investigation* [Apollo - University of Cambridge Repository]. https://doi.org/10.17863/CAM.65721

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*(2), 191–215. https://doi.org/10.1006/jpho.2001.0134

R Core Team. (2022). *R: A language and environment for statistical computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ramon-Casas, M., Cortés, S., Benet, A., Lleó, C., & Bosch, L. (2021). Connecting perception and production in early Catalan–Spanish bilingual children: Language dominance and quality of input effects. *Journal of Child Language*, 1–22. https://doi.org/10.1017/S0305000921000787

Sharma, D., & Sankaran, L. (2011). Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change*, *23*(3), 399–428. https://doi.org/10.1017/S0954394511000159

Sim, J. H. (2019). "But you don't sound Malay!": Language dominance and variation in the accents of English-Malay bilinguals in Singapore. *English World-Wide*, *40*(1), 79–108. https://doi.org/10.1075/eww.00023.sim

Sim, J. H. (2021). Sociophonetic variation in English /l/ in the child-directed speech of English-Malay bilinguals. *Journal of Phonetics*, *88*, 101084. https://doi.org/10.1016/j.wocn.2021.101084

Sim, J. H. (2023). Influence of bilingualism or caregiver input? Variation in VOT in simultaneous bilingual preschoolers in Singapore. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2344–2348). Guarant International.

Sim, J. H., & Post, B. (2021). Variation in quality of maternal input and development of coda stops in English-speaking children in Singapore. *Journal of Child Language*, 1–26. https://doi.org/10.1017/S0305000921000593

Sim, J. H., & Post, B. (2023). Influence of caregiver input and language experience on the production of coda laterals by English–Malay bilingual preschoolers in multi-accent Singapore. *Journal of Child Language*, 1–26. https://doi.org/10.1017/S0305000923000375

Sim, J. H., & Post, B. (2024). Early phonological acquisition in multi-accent contexts. In E. Babatsouli (Ed.), *Multilingual Acquisition and Learning: An ecosystemic view to diversity* (pp. 192–215). John Benjamins Publishing Company.

Stoehr, A., Benders, T., van Hell, J. G., & Fikkert, P. (2019). Bilingual preschoolers' speech is associated with non-native maternal language input. *Language Learning and Development*, *15*(1), 75–100. Scopus. https://doi.org/10.1080/15475441.2018.1533473

te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & Konig, R. (2017). When size matters: Advantages of weighted effect coding in observational studies. *International Journal of Public Health*, *62*(1), 163–167. https://doi.org/10.1007/s00038-016-0901-1

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*(1), 147–162. https://doi.org/10.1016/j.cognition.2006.03.006