

Acoustic Analysis of Voicing Contrast in Malay Word-Initial Plosives Produced by Mandarin-Speaking Children

Badruzaman Abdul Hamid, Chong Jin Yee and Hasherah Mohd Ibrahim*

Centre for Rehabilitation and Special Needs Studies, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, 50300 Kuala Lumpur, Malaysia

ABSTRACT

Voice Onset Time (VOT) is an acoustic measure widely used to measure the characteristics of plosives across various languages. This study investigates the acquisition of voicing contrast in Malay word-initial plosives produced by Mandarin-speaking children. Fifteen subjects between the ages of 3;00 to 7 and 11 were recruited. The children completed a picture-naming task, and the speech samples were audio-recorded and analyzed using PRAAT. Descriptive VOT values (ms) across ages were presented and compared to previous normative findings. Results revealed that children aged 3;00 produced only a short lag, while children aged 4;00 to 6;00 produced short lag-long lag contrast, which is not similar to the expected VOT pattern of Malay plosives. Additionally, with increasing age, children acquired adult-like VOT productions. The VOT distributions follow the developmental path proposed for languages with two-way contrasts. This finding is the first VOT data on Malay word-initial plosives by Mandarin-speaking children. Maturity of motor development and coordination, as well as cross-linguistic exposure, influenced productions. Thus, further studies examining cross-linguistic influence, particularly controlling for the duration of exposure to L2 in multilingual language acquisition with a larger sample size, are warranted to validate the findings from this study.

Keywords: Bilingual, Malay, Mandarin, plosives, voicing contrast, voice onset time

ARTICLE INFO

Article history:

Received: 20 August 2021

Accepted: 22 July 2022

Published: 10 November 2022

DOI: <https://doi.org/10.47836/pjssh.30.4.11>

E-mail addresses:

badrulhamid@ukm.edu.my (Badruzaman Abdul Hamid)

eunice_chong@outlook.com (Chong Jin Yee)

hasherah@ukm.edu.my (Hasherah Mohd Ibrahim)

* Corresponding author

INTRODUCTION

Malaysia is formed of three main ethnicities (Malay, Chinese, and Indian) and is a multilingual country. Malay is the national language, while Mandarin, Tamil, and other dialects are used in daily communication. Therefore, it is common for non-native speakers to use languages other than their

native language to communicate. For example, Malaysian Chinese speakers may use Mandarin-English-Malay language plus Chinese dialect/s in their daily communication (Lim & Lim, 2018). However, previous studies have shown that their mother tongues influence non-native speakers who use Malay. There have been key research on voicing contrasts in plosives by investigating voice onset time (VOT) in native speakers of the Malay language. Abdul Hamid (2016) studied the acquisition of VOT in Malay children aged 2 to 6, while Shahidi (2010) studied the acquisition of VOT by bilingual Malay English speakers. However, studies examining VOT in multilingual speakers in Malaysia are limited. Due to the scarcity of studies on Mandarin-speaking children and the Malay language, this study aimed to provide information on the acquisition of Malay word-initial plosives by Mandarin-speaking children from ages 3;00 to 7 and 11. Since children with hearing impairment, childhood apraxia of speech, and phonological disorder produce voicing errors in their speech (Park & Byeon, 2015; Umat et al., 2015), it is beneficial to investigate the developmental trend in acquiring voicing contrast by typically-developing children. Additionally, knowing the age of acquisition for voicing contrast could also benefit the teaching and learning of the pronunciation of Malay in schools (Shahidi et al., 2012). The present study investigates the VOT pattern of Malay word-initial plosives produced by Mandarin-speaking children. Additionally,

we hypothesized that as age increases, the VOT values approach adult-like (Kewley-Port & Preston, 1974; Macken & Barton, 1980; Shahidi et al., 2012; Yang, 2018; Yu et al., 2015). We also hypothesized that the voicing contrast between pre-voiced and short lag VOT in Malay word-initial plosives is acquired after age 3;00 (Abdul Hamid, 2016; Khattab, 2002).

LITERATURE REVIEW

Malay and Mandarin Plosives

Malay is a major language of the Austronesian language family and Western Malayo-Polynesian subgrouping. There are 27 consonants in Malay, including six VOICELESS plosives /p, t, k/, and VOICED plosives /b, d, g/. (Abdul Hamid, 2016; Shahidi, 2010). Malay plosives are found in initial, medial, and final word positions. According to Shahidi (2010), Malay VOICELESS plosives are always unaspirated, with no audible breath heard when the plosive is released. However, aspirated plosives are unique in Malay dialects from the Malaysian Peninsula's northern states. In contrast, Mandarin is one of the Sino-Tibetan languages. It is a tonal language with four contrastive tones. In comparison to Malay, there are six plosives in Mandarin: VOICELESS unaspirated /p, t, k/ and VOICELESS aspirated /p^h, t^h, k^h/ (Duanmu, 2007; Shimizu, 2011). Aspiration is the difference between the plosives, not voicing (Lin, 2007; Shimizu, 2011). Unlike Malay, Mandarin plosives only occur at the word-initial position (Ogasawara, 2011).

Voice Onset Time (VOT)

VOT is the duration between the articulator release of a stop and the voicing onset (Lisker & Abramson, 1964). VOT measures the timing of voicing in milliseconds (ms). According to Lisker and Abramson, VOT is useful in separating phonemes into VOICED and VOICELESS categories, regardless of the language studied. Lisker and Abramson's study (1964) categorized plosives into three ranges based on VOT values. A negative VOT value is obtained when the onset of vocal fold vibration or voicing has started before the release of a stop. It is also known as 'voice lead' or 'pre-voiced.' A zero VOT value is obtained when vocal fold vibration or voicing onset coincides with the plosive release. Zero VOT typically occurs in VOICELESS unaspirated plosives. If the voicing has started after the burst release, this results in 'voice lag,' a positive VOT value. All VOICELESS aspirated plosives have positive VOT. The amount of lag separates VOICELESS unaspirated ('short lag' 0– +30msec) from VOICELESS aspirated ('long lag' +30 – +100ms) plosives (Lisker & Abramson, 1964; Shimizu, 1989). Lisker and Abramson (1964) categorized 11 languages into three main categories, depending on the number of plosive categories in the language: (1) two-category languages included American-English, Cantonese, and Tamil; (2) three-category languages that included Korean and Thai; (3) four-category languages that included Hindi and Marathi.

Voicing Contrast of Plosives

The voicing contrast of plosives represents complex acoustic and articulatory coordination (Melo et al., 2015). Voicing contrast in plosives could be studied based on acoustic cues such as voice bar, voicing during the closure, closure duration, vowel duration, and VOT. However, VOT is a reliable acoustic cue for studying voicing contrast in word-initial plosives (Shahidi, 2010). Acquisition of voicing contrast in children has been extensively studied in various languages, such as English (Kewley-Port & Preston, 1974; Macken & Barton, 1980), Hindi (Davis, 1995; Shimizu, 1989), Thai (Shimizu, 1989) and Spanish (Macken & Barton, 1980). Monolingual children typically acquire voicing contrast between short and long lag VOT by age 2;00–2;06 (Kehoe et al., 2004; Kewley-Port & Preston, 1974; Macken & Barton, 1980). However, the contrast between pre-voiced and short lag VOT (e.g., in Spanish, French and Italian) is acquired only after age 3;00. It is due to the difficulty in producing lead voicing, especially for children in mastering its production (Macken & Barton, 1980).

According to Shahidi et al. (2012), the realization pattern for the word-initial voicing contrast in Malay plosives is short-leg for VOICELESS plosives and pre-voiced for VOICED plosives. Shahidi et al. (2016) investigated voicing contrast between VOICED and VOICELESS plosive in the initial position produced by Malay children aged 3;00–to 6 and 11. Findings revealed that children aged 3;00 produced

voicing lag for all plosives, suggesting they have still not acquired voicing contrast in Malay plosives. Children aged 4;00 and above produced pre-voiced for VOICED plosives. The finding showed they have acquired voicing contrast between VOICED and VOICELESS in Malay word-initial plosives. As age increased, the VOT values produced approached adult-like VOT. Abdul Hamid (2016) revealed that children aged 2;00 to 3;00 did not produce voicing lead for VOICED plosives (Table 1). The finding was in line with the findings from Shahidi et al. (2016). Except for velar plosive /g/, children aged 4;00 and above produced pre-voiced for VOICED plosives.

Based on the three-category model proposed by Lisker and Abramson (1964),

Mandarin unaspirated plosives /p, t, k/ occupied the short-leg regions, while Mandarin aspirated plosives /p^h, t^h, k^h/ occupied the long-lag regions. A study by Yang (2018) on the development of stop consonants in 3 to 6-year-old Mandarin-speaking children reported that Mandarin-speaking children at 3 years old had well-developed short-lag VOTs. However, Mandarin-speaking children aged 6 years had not fully established an adult-like long-lag VOT model. It indicates that long-lag VOTs in Mandarin-speaking children in this age range continue to develop. Table 2 shows the VOT Means of Mandarin plosives or Taiwanese Mandarin (TM) reported by past studies. Data from Li (2013) were reported separately for females and males.

Table 1
Mean VOT values for Malay word-initial plosives for children aged 2;00–6; 11

Mean VOT (ms)	2;00	3;00	4;00	5;00	6;00
/b/	11.0	10.7	-50.1	-59.4	-61.8
/d/	13.3	13.3	-38.0	-49.9	-53.6
/g/	18.1	20.3	9.8	12.0	12.0
/p/	8.3	7.9	6.7	8.5	7.9
/t/	14.2	13.2	14.3	13.1	14.1
/k/	23.6	24.7	30.4	29.6	34.0

Table 2
Mean VOT values (ms) for Mandarin word-initial plosives from different studies

	Liao (2005)	Chen et al. (2007) (TM)*	Chao & Chen (2008) (TM)*	Li (2013) (TM)*
/p/	17.9	13.9	14.0	
/t/	18.6	15.3	16.0	12.5 (F), 17.5 (M)
/k/	28.0	27.4	27.0	22.5 (F), 29.5 (M)
/p ^h /	75.4	77.8	82.0	
/t ^h /	71.4	75.5	81.0	93.3 (F), 77.9 (M)
/k ^h /	98.8	85.7	92.0	90.8 (F), 78.8 (M)

Note. *TM - Taiwanese Mandarin

VOT Development in Second Language (L2)

The Second Language Acquisition (SLA) literature provides strong evidence of the transfer of L1 VOT values in the acquisition of L2 VOT, especially at the lower levels of L2 proficiency (e.g., Flege, 1987; Flege & Hillenbrand, 1984). More advanced learners were reported to produce native-like VOT of the language and differentiate L1 and L2 concerning VOT (Flege, 1991). Shimizu (2011) reported that L1 stop categories were used in learning L2 stop categories due to the proximity of the VOT value. Several models, such as the speech learning model (SLM) and perceptual assimilation model (PAM), mentioned that the phonetic distance of plosives in L1 with L2 is the major cue in learning new L2 speech sounds. In order to produce L2 voicing categories, bilinguals try to find L1 sounds whose VOT values are close to their L2 counterparts. Studies showed that bilingual children produced different VOT than monolinguals in both languages (Khattab, 2002). Late L2 learners produced English /t/ with “compromise” VOT values that were intermediate VOT observed in Spanish and English monolinguals (Flege, 1991). Moreover, Balukas and Koops (2015) showed that VOT values for English produced by Spanish-English bilinguals were lower than the VOT production by native English speakers. It could be due to different VOT patterns in Spanish (pre-voiced vs. short lag) and English (short lag vs. long lag).

METHODOLOGY

Participant Characteristics

A total of 15 children residing in the Klang Valley area were recruited. Subjects were divided into 5 age groups: 3 years (hereafter denoted as 3;00) to 3 years and 11 months (hereafter denoted as 3;11), 4;00–4;11, 5;00–5;11, 6;00–6;11 and 7;00–7;11. The children were randomly selected based on the following criteria: (1) native speakers of Mandarin (2) children who are simultaneous/sequential bilinguals, able to understand and use Malay (3) no reported speech and language delay or other disabilities (4) obtained parents’ consent to participate in the study. Those who did not fulfill the inclusion criteria were excluded from the study. Participants were recruited by word of mouth and through kindergarten and tuition centers.

Material and Procedure

A digital audio recorder (Sony IC recorder ICD-SX850) was placed 10 cm from the participants during testing. A picture-naming task was used to elicit words with targeted word-initial plosives /p, b, t, d, k, g/ in Malay. The stimuli used were similar to Abdul Hamid (2016; Table 3). Cues were given if the child had difficulties naming the pictures. The cues were given in the following order: (1) Semantic cue; (2) False choice; (3) Delayed imitation. Direct imitation from the tester was not considered the target response.

Table 3
List of stimulus words

/p/	/t/	/k/
Pisang	Tandas	Kaki
Pokok	Tangan	Kakak
Pensel	Tiga	Kasut
/b/	/d/	/g/
Bola	Datuk	Gajah
Baju	Duduk	Gigi
Buku	Dua	Gula

Recordings and VOT Measurements

The recording took place in a quiet room in the child’s location. Speech samples of the children (3 subjects x 5 age groups x 6 consonants x 3 words x 3 repetitions = 810 tokens) were audio-recorded using a Sony IC recorder ICD-SX850 and transferred into a computer. The PRAAT software version 5.3.56 was used to analyze the samples collected. Speech samples were displayed as spectrograms in the software. VOT, the period between the release of the consonant stop and the onset of voicing, was measured. Burst release indicated stop release, and the first periodic waveform indicated the onset of voicing (Lisker & Abramson, 1964).

RESULTS

Reliability

The internal consistency was measured using Cronbach’s alpha (α) coefficient. Results revealed good internal consistency ($\alpha = .887$). Inter-rater reliability was obtained via the analysis of VOTs on 13% (n=108 tokens) of the audio samples independently by a clinical linguist trained in using the PRAAT software. The intra-class correlation

coefficient (ICC) of .992 indicated high inter-rater reliability.

VOT Values of Malay Word-Initial Plosives Produced by Mandarin-Speaking Children

Table 4 summarizes the VOT values of Malay word-initial plosives produced by Mandarin-speaking children compared to previous local findings. For VOICED plosives, children within the age group of 3 and 6 obtained mean VOT values that fell within the short lag region (0–30ms). Children aged 4;00 obtained mean VOT values of M=23 (SD=6.506) and M=28 (SD=7.506) for phoneme /b/ and /d/, respectively, with both mean VOT values falling within the short lag VOT range. However, the 4 years old children produced long lag (M=50, SD= 23.000) for phoneme /g/. Children aged 5;00 produced long lag for phonemes /b/ and /g/ and short lag for phoneme /d/. Results revealed that children aged 7;00 obtained mean VOT values for phonemes /b/ and /d/ that fell within the pre-voiced region. However, they produced short lag for phoneme /g/ (M= 24, SD=37.041). For VOICELESS plosives, children aged 3;00 obtained mean VOT values that fell within the short lag range. Children aged 4;00, 5;00, and 6;00 years old produced long lag VOT for both phonemes /p/ and /k/. Children aged 7;00 produced short lag VOT range for phoneme /p/ (M=24, SD=2.517) and long lag for phoneme /k/ (M=32, SD= 2.646). Children from all age groups produced short lag VOT for phoneme /t/. Overall, the VOT values

approached adult-like VOT as age increased for VOICED and VOICELESS plosives.

The further the back the place of articulation, the longer the duration of voicing lag, regardless of short or long lag (Figure 1). On the other hand, the further back the place of articulation, the shorter

the pre-voiced duration, which was noted in children aged 7;00.

As the place of articulation moved further back, the mean VOT values of VOICELESS plosives produced by children aged 3;00 and Malay monolingual adults increased (Figure 2). However, the mean

Table 4
VOT for VOICED and VOICELESS plosives (mean VOT (M) in milliseconds (ms) and standard deviation (SD) by age

Phoneme		3; 00	4; 00	5; 00	6; 00	7; 00	Mandarin- Malay bilingual adults (Abdul Hamid et al., 2020)	Malay monolingual adults (Abdul Hamid, 2016)	
VOICED	/b/	M	14	23	32	2	-44	-38	-88
		SD	(1.000)	(6.506)	(0.577)	(5.508)	(5.508)	(55.48)	(3.000)
	/d/	M	25	28	22	21	-6	-12	-76
		SD	(1.528)	(7.506)	(<0.000)	(4.000)	(25.502)	(61.23)	(3.055)
	/g/	M	28	50	46	30	24	23	-54
		SD	(1.000)	(23.000)	(7.506)	(9.000)	(37.041)	(40.68)	(2.887)
VOICELESS	/p/	M	19	54	54	34	24	30	12
		SD	(0.577)	(29.000)	(6.506)	(8.000)	(2.517)	(13.28)	(3.055)
	/t/	M	24	29	24	16	20	21	15
		SD	(1.528)	(1.000)	(5.508)	(1.528)	(1.732)	(5.08)	(2.646)
	/k/	M	30	57	81	65	32	31	32
		SD	(2.000)	(5.508)	(9.504)	(14.503)	(2.646)	(9.60)	(6.557)

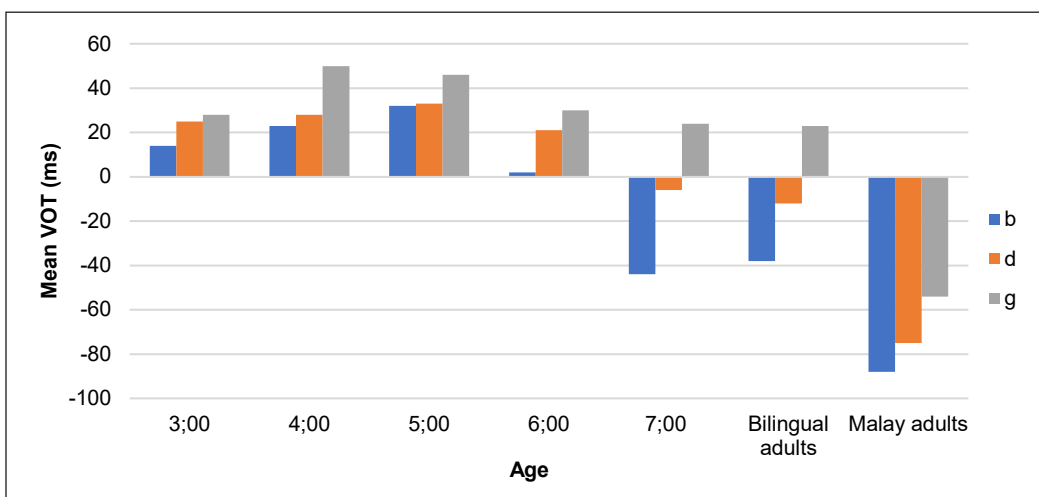


Figure 1. Mean VOT values for VOICED plosives across age group

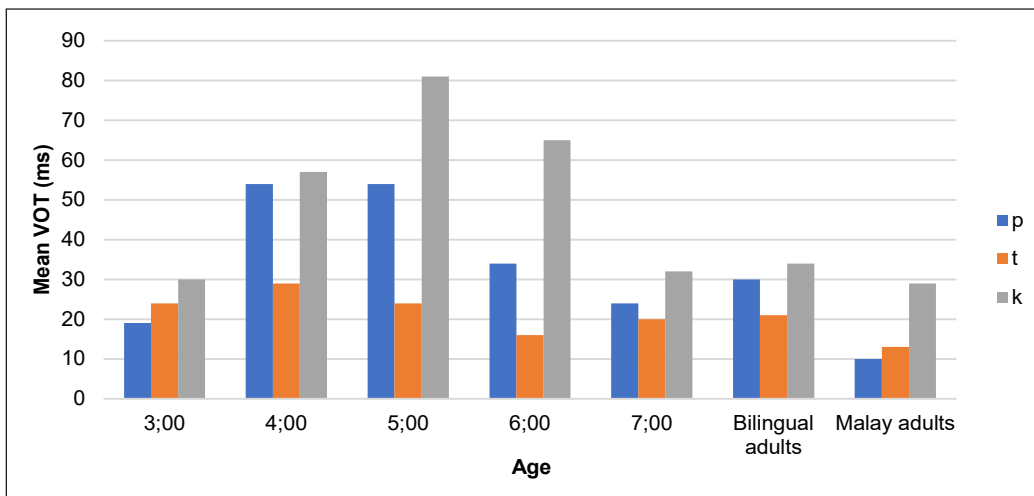


Figure 2. Mean VOT values for VOICELESS plosives across age group

VOT values of children aged 4;00 to 7;00 and bilingual adults did not follow the same pattern. The alveolar plosive /t/ had shorter VOT value compared to bilabial plosive /p/ and velar plosive/k/.

Age of Acquisition of Voicing Contrast

All children aged 3;00, 4;00, and 5;00 did not produce a pre-voiced VOT pattern for VOICED plosives /b, d, g/. In the age group of the 6-year-old, 2 children (66.67%) did not produce the pre-voiced phoneme /b/, and all the children (100%) did not produce pre-voiced phonemes /d/ and /g/. All children aged 7; 00 (n=3) produced pre-voiced for plosives /b/; however, not all produced pre-voiced for /d/ and /g/. It was noted that voicing contrast in plosives was not fully established by 7 years old, especially for VOICELESS plosives /d/ and /g/.

The mean VOT value for phoneme /p/ was similar to phoneme /b/ at age 3 (Figure 3). It could be seen that the mean VOT values for both phonemes/p/ and

/b/ increased from age 3;00 to 5;00, and both phonemes demonstrated a decrease in the mean VOT values at age 6;00. The negative mean VOT value indicated that children aged 7;00 produced pre-voiced for phoneme /b/. Children aged 7;00 obtained mean VOT values similar to bilingual adults but did not achieve the VOT range produced by Malay adults.

Children aged 3;00 to 6;00 obtained positive and similar mean VOT values for both phonemes /t/ and /d/ (Figure 4). However, for the phoneme /d/, children aged 7;00 obtained a negative VOT value, indicating that they produced a pre-voiced VOT pattern for this phoneme.

Unlike native Malay speakers, children aged 3;00 to 7;00 and bilingual adults did not produce negative mean VOT values for the phoneme /g/. Instead, they produced voicing lag (positive VOT) for both phonemes /k/ and /g/ (Figure 5). In addition, children did not produce pre-voiced for velar plosive /g/.

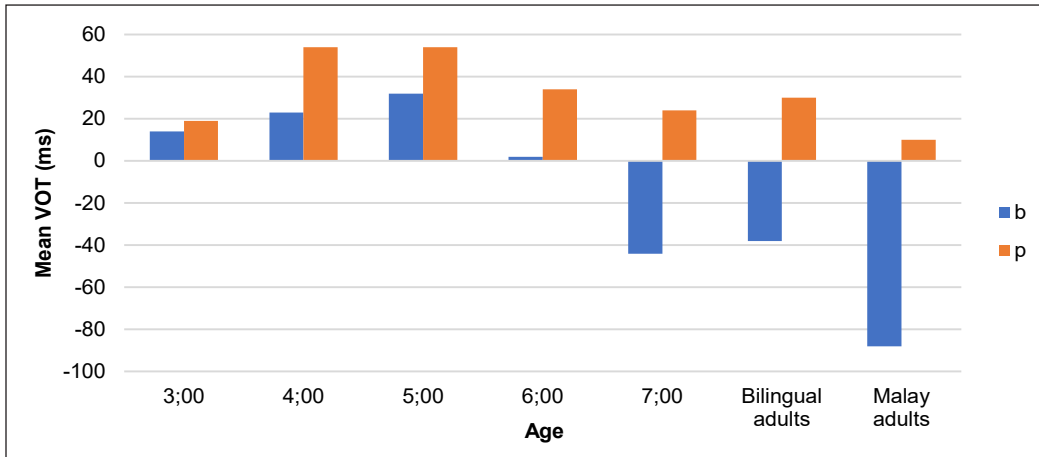


Figure 3. Mean VOT values for bilabial plosives /p/ and /b/ across age groups

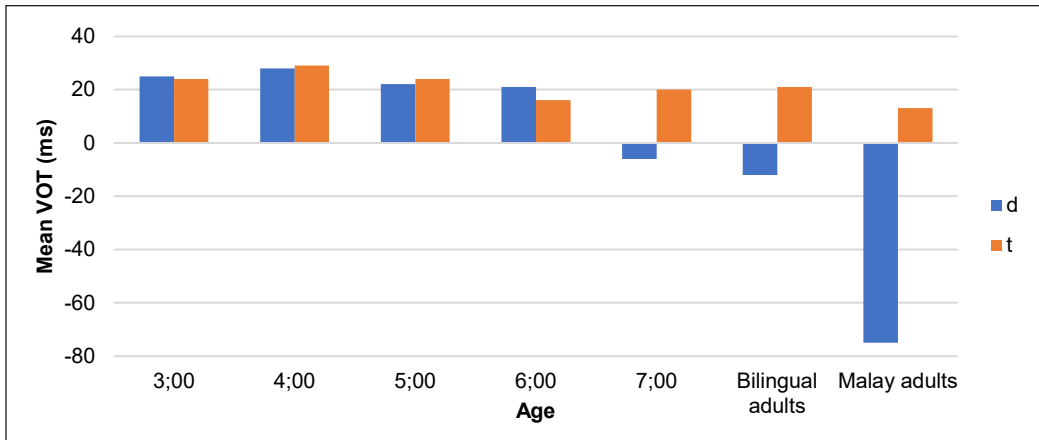


Figure 4. Mean VOT values for alveolar plosives /t/ and /d/ across age groups

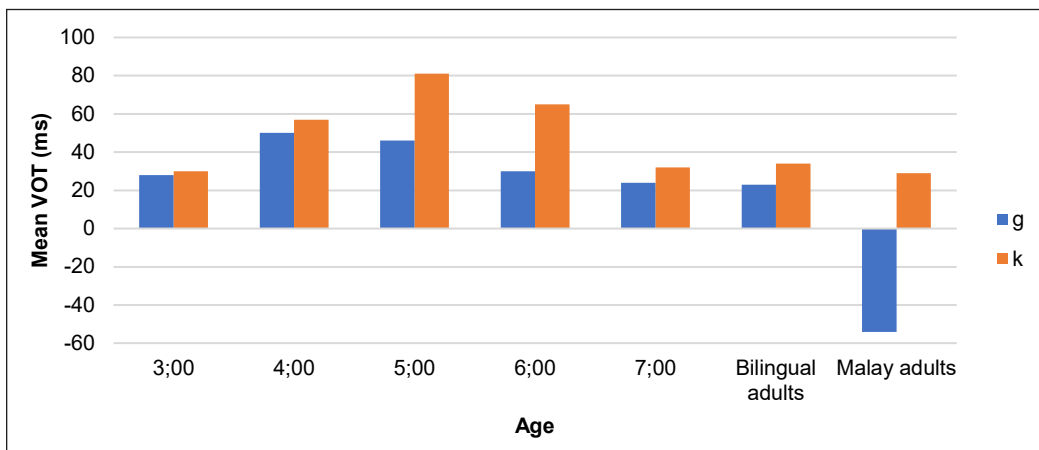


Figure 5. Mean VOT values for velar plosives /k/ and /g/ across age groups

DISCUSSION

This study investigates the VOT pattern of Malay word-initial plosives produced by Mandarin-speaking children by measuring the VOT values and determining the age of acquisition voicing contrast.

VOT of Malay Word-Initial Plosives Produced by Mandarin-Speaking Children

Three-stage models were proposed by Macken and Barton (1980) on the acquisition of English syllable initial plosives. However, Macken and Barton (1980) were conducted on L1 English (short lag vs. long lag), while the current study was conducted on L2 Malay (pre-voiced vs. long lag). Therefore, it is noted that children in the present study achieve the stages at different ages than subjects in the previous study.

Stage I. In the present study, Mandarin-speaking children aged 3;00 only produced a short lag for VOICED and VOICELESS plosives. It is consistent with past studies of word-initial stops in other languages (Shahidi et al., (2016) on Malay monolingual children, Yang (2018) on Mandarin monolingual children, and Shahidi and Aman (2011) on Malay-English bilinguals). According to Abdul Hamid (2016), children are expected to produce only a short lag in the early stage of acquiring voicing contrast. According to the mentalist theory of language learning, an individual's innate predisposition toward language acquisition is universal. Based on articulatory and aerodynamic considerations, Kewley-Port

and Preston (1974) hypothesized that short-lag VOT is acquired first because it is easier to produce short-lag VOT than long-lag VOT productions. Short-leg VOT production does not exceed the abilities of the immature motor system of a young child (Green et al., 2000).

Stage II. Before age 7, Mandarin-speaking children produced short and long lag, which differed from the expected VOT pattern. According to Macken and Barton (1980), children initially produced longer VOT for plosives than adult speakers, then experienced a drop in VOT before achieving adult-like VOT. In line with that, Tanaka (2018) reported that Japanese-Mandarin bilinguals aged 4;00 and 5;00 produced relatively longer VOT values than the children aged 3;00 and 6;00 across all stop categories. The same phenomenon could be observed in the VOT means produced by Mandarin-speaking children in the present study.

Stage III. The expected pre-voiced VOT in VOICED plosives emerged when the child was 7 years old for phonemes /b/ and /d/. In line with current findings, Deuchar and Clark (1996) hypothesized that a fairly adult-like model would develop at around seven. However, as age increased, the VOT values approached adult-like VOT values. Although children had acquired voicing contrast of L2, it was believed that the fine-tuning continued and extended into school-age years. It was particularly evident in languages with pre-voiced and

short-leg VOT contrasts (MacLeod, 2016). Meanwhile, the phoneme /g/ was produced with voicing lag rather than pre-voiced. Kewley-Port and Preston (1974) reported the same phenomenon where children aged 6;00 produced voicing lag for the phoneme /g/, which differed from the expected VOT pattern.

Other than this, Zlatin and Koenigsnecht (1976) also proposed that the acquisition of voicing contrast in English-speaking children changed from a unimodal (in favor of short lag VOT) to an emerging bimodal to a typical bimodal pattern. A similar developmental path could be observed in the present study. In the present study, the VOT distribution demonstrated a unimodal pattern (short lag VOT) in the 3-year-old, which developed into an emerging bimodal pattern (short lag-long lag contrast) in the 4 to 6-year-old and gradually approximated the bimodal distribution of the adult speakers (pre-voiced- short lag VOT).

The VOT of plosives produced by Mandarin-speaking children aged 3;00 was consistent with previous findings (Cho & Ladefoged, 1999; Shahidi et al., 2012). The further the back the place of articulation, the shorter the duration of pre-voiced, and the longer the duration of voicing lag (short or long). However, the Malay VOICELESS plosives produced by Mandarin-speaking children aged 4;00 to 7;00 did not conform to the general agreement. Phoneme /t/ had shorter mean VOT than phonemes /p/ and /k/, similar to VOT produced by Malay-Mandarin bilingual adults (Abdul Hamid et al., 2020). It is also interesting to note that

the finding was consistent with previous studies on Mandarin monolinguals (Chao & Chen, 2008; Chen et al., 2007; Liao, 2005). The VOT duration for Mandarin aspirated /t^h/ had shorter VOT than /p^h/ and /k^h/ (Rochet & Yanmei, 1991). Researchers hypothesized that the place effect may be language-specific and may not be so robust in Mandarin. Therefore, CLI was suggested as it may be a sign of influence from L1 (Mandarin) in the realization of the L2 (Malay) VOT pattern in the present study.

Acquisition of Voicing Contrast

According to MacLeod (2016), task demands may influence a child's ability to manipulate acoustic-phonetic cues. Prior studies of languages that included a pre-voiced VOT contrast focused on three types of tasks: repetition (Eilers et al., 1984), picture naming (present study; MacLeod, 2016), and spontaneous speech (Bóna & Auszmann, 2014). Thus, variability in the age of acquisition of voicing contrast in previous studies may be partly explained by the different tasks used and the languages examined. A study by Abdul Hamid in 2016 used a similar procedure as the present one. Abdul Hamid (2016) reported that Malay monolingual children acquired voicing contrast (pre-voiced vs. short lag) at 4-years-old. However, the present study revealed that Mandarin-speaking children acquired the voicing contrast in Malay word-initial plosives (pre-voiced vs. short lag) at 7 years old, except for the plosive /g/.

The possible explanations for the later acquisition of voicing contrast and

pre-voiced in Mandarin-speaking children may be based on: (i) Motor development and coordination, (ii) Cross-Linguistic Influence (CLI), and (iii) Perceptual, social and environmental factors.

Motor Development and Coordination.

Compared to the apparent simplicity of voicing lag production in L1, pre-voiced in L2 requires a high degree of motor coordination (MacLeod, 2016). It requires coordination between vocal fold adduction and plosives closure, leading to increased intraoral pressure (Cho & Ladefoged, 1999). During the closure phase, when the oral cavity is sealed off, it is difficult for speakers to maintain the requisite aerodynamic condition (difference in air pressure) for vocal fold vibration. In the present study, the later acquisition of pre-voiced in Malay VOICED plosives supported the idea that pre-voiced is more difficult to produce than voicing lag. In line with studies done on Malay monolingual children (Abdul Hamid, 2016) and Malay-Mandarin bilingual adults (Abdul Hamid et al., 2020), Mandarin-speaking children (subjects) undergo the same process where pre-voiced in /b/ is acquired first, followed by /d/ and /g/. The shorter the length of the pre-voiced, the easier the pre-voiced is produced (Abdul Hamid et al., 2020). Plosive /g/ has the shortest pre-voiced and requires the highest motor coordination. Kent (1976) reported that as age increases, the accuracy of motor control increases. Children aged 7; 00 producing pre-voiced VOT proved the maturation of motor skills and coordination.

Cross-Linguistic Influence (CLI). The use of the L1 VOT pattern in acquiring voicing contrast of L2 was often interpreted as CLI. The acoustic analysis of the VOT of initial Malay-accented English plosives confirmed the presence of CLI (Shahidi et al., 2012). In the present study, the absence of pre-voiced VOT production in Malay VOICED plosives by children aged 3;00 to 6;00 could be a sign of influence from L1 (Mandarin) in the VOT pattern of L2 (Malay) and yet to acquire the phonetic system of L2 (Abdul Hamid et al., 2020). It was also noted that children aged 4;00 to 6;00 produced VOT that fell within the short and long lag regions, contrary to the expected VOT pattern. They used the L1 (Mandarin) VOT pattern (short lag vs. long lag) in the process of acquiring voicing contrast (pre-voiced vs. short lag) in L2 (Malay). There is a ‘carry-over’ of L1 phonetic features of voicing contrast onto L2 plosives production, which causes the late acquisition of voicing contrast in L2 by Mandarin-speaking children.

Perceptual, Social, and Environmental Factors. Flege’s Speech Learning Model (Flege, 1995) was developed to account for L2 sound acquisition and pronunciation. The model claimed that different phonetic categories were likely to be established for each language if bilinguals perceived some of the two languages’ phonetic dissimilarity. Thus, it was suggested that bilingual children could produce native-like VOT values accurately if they perceived the phonetic dissimilarity between the sounds of their two languages (Tanaka, 2018). In agreement

with the model, Davis (1995) suggested that pre-voiced acquisition may be tied to the acoustic salience of this contrast relative to short-leg and long-lag VOT. Aslin et al. (1981) studied the VOT perception by infants. The results showed that pre-voiced VOT was less salient than the short-leg or long-lag VOT. Thus, in the present study, the later acquisition of pre-voiced at the age of 7;00 could be due to pre-voiced in L2 (Malay) being less salient than short lag-long lag contrast in L1 (Mandarin).

According to the normal development pattern for VOT production, all plosives are produced with short lag VOT initially (Zlatin & Koenigsnecht, 1976). Therefore, Khattab (2002) proposed that a bilingual's use of short lag instead of pre-voiced might be developmental rather than CLI. However, bilingualism might hinder the acquisition of certain features that require early and extensive exposure if bilinguals receive inadequate input on these features. A study by Khattab (2002) on English-Arabic children revealed that adequate Arabic pre-voiced input at an early age was necessary for children to master the complex articulatory features required for pre-voiced production.

In the present study, Mandarin-speaking children acquiring voicing contrast at 7;00 may be due to the Malay language's inadequate input at a younger age. It is similar to the factor suggested by Abdul Hamid et al. (2020); the period being exposed to L2 causes VOT variation among subjects. Children also only began attending primary school at the age of 7, which might, in turn,

have increased the Malay input in their environment. Subsequently, this increase could have led to pre-voiced becoming more salient in their environment, thus explaining Mandarin-speaking children acquiring pre-voiced VOT at 7. It is hypothesized that children with maids/ friends/neighbors who spoke Malay may acquire Malay's voicing contrast earlier. It is due to the frequency of using Malay in communicating increases, which then ease acquiring L2 phonetic (Abdul Hamid et al., 2020).

Khattab (2002) also mentioned that if parents or caregivers did not consistently produce pre-voiced Malay voiced plosives, it might affect the child's acquisition of this feature. In the current study, it is worth noting that Mandarin-Malay bilingual adults did not produce pre-voiced VOT for phoneme /g/, similar to Mandarin-speaking children. A study by Deuchar and Clark (1996) on English-Spanish subjects found similar results; children produced VOT patterns like their parents, which differed from the expected pattern. Further research on this topic could consider factors such as types of bilingualism (e.g., simultaneous/ sequential bilinguals) and the amount of exposure to L2.

Multiple Bursts and Aspiration

Although it was not one of this study's objectives, some children were noted to produce aspiration and multiple bursts in their spectrogram. Aspiration serves as a longer noise interval after the stop is released and before the vowel. It is a phonetic cue that differentiates Mandarin plosives. Therefore,

the presence of aspiration suggests CLI occurs when Mandarin-speaking children produce Malay plosives. It is consistent with Abdul Hamid et al. (2020) study on Malay-Mandarin bilingual adults. Yang (2018) also reported that children tended to produce more multiple bursts than adults. According to Yang (2018), children produce multiple bursts while producing plosives because they have smaller articulator sizes that cause higher subglottal pressure. High intraoral pressure followed by negative pressure caused constriction formation for subsequent bursts. Thus, multiple bursts in plosives are observed. Future studies could further examine the relationship between VOT and the number of bursts.

CONCLUSION

The study investigates the age of acquisition of voicing contrast in Malay word-initial plosives produced by Mandarin-speaking children through acoustic analysis of VOT. The findings from the study indicate that Mandarin-speaking children acquire voicing contrast (pre-voiced vs. short lag) in Malay word-initial plosives at the age of 7;00. VOT distributions follow the developmental path proposed by Macken and Barton (1980) and Zlatin and Koenigsknecht (1976). The further the back the place of articulation, the shorter the duration of pre-voiced, and the longer the duration of voicing lag (short or long), except for the phoneme /t/. The voicing contrast's developmental acquisition findings also give insight into the motor control mechanism and coordination, CLI, and social and environmental exposure

influence VOT acquisition. As age increases, the VOT values approach adult-like VOT values. However, interesting findings are seen in younger children in this study. They did not follow the typical VOT pattern and produced multiple bursts of aspiration, denoting the carry-over influence of L1 phonetic features of voicing contrast onto L2 plosives production. This information helps us to describe the developmental trend of phonetic categories in bilingual children. The present study had a relatively small sample size; further studies using a larger sample size and controlling for the duration of exposure to L2 are recommended to obtain more representative data.

ACKNOWLEDGEMENT

The authors want to thank all the participants who participated in this study and the reviewers of this paper for providing constructive comments.

REFERENCES

- Abdul Hamid, B. (2016). *Analisis kontras penyuraan bunyi plosif kanak-kanak Melayu berasaskan pendekatan akustik fonetik* [An acoustics phonetics approach of voicing contrast of plosives sound amongst Malay children; Unpublished doctoral thesis]. Universiti Kebangsaan Malaysia.
- Abdul Hamid, B., Mohd Ibrahim, H., & Keria, N. F. (2020). Analisis pola masa mula suara konsonan plosif Bahasa Melayu yang dihasilkan oleh penutur Mandarin dewasa [Analysis of voice onset time on Malay plosives by Mandarin adult speakers]. *GEMA Online® Journal of Language Studies*, 20(4), 55-73. <https://doi.org/10.17576/gema-2020-2004-04>

- Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52*(4), 1135-1145. <https://doi.org/10.2307/1129499>
- Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism, 19*(4), 423-443. <https://doi.org/10.1177/1367006913516035>
- Bóna, J., & Auszmann, A. (2014). Voice onset time in language acquisition: Data from Hungarian. In S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (Eds.), *Proceedings of the 10th International Seminar on Speech Production, ISSP* (pp. 40-44). http://www.issp2014.uni-koeln.de/wp-content/uploads/2014/Proceedings_ISSP_revised.pdf
- Chao, K. Y., & Chen, L. M. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *International Journal of Computational Linguistics & Chinese Language Processing, 13*(2), 215-232.
- Chen, L. M., Chao, K. Y., & Peng, J. F. (2007, September 6-7). *VOT productions of word-initial stops in Mandarin and English: A cross-language study* [Paper presentation]. 19th Conference on Computational Linguistics and Speech Processing, ROCLING 2007, Taipei, Taiwan.
- 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>
- Davis, K. (1995). Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. *Journal of Child Language, 22*(2), 275-305. <https://doi.org/10.1017/s030500090000979x>
- Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics, 24*(3), 351-365. <https://doi.org/10.1006/jpho.1996.0019>
- Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.
- Eilers, R. E., Oller, D. K., & Benito-Garcia, C. R. (1984). The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study. *Journal of Child Language, 11*(2), 313-336. <https://doi.org/10.1017/s0305000900005791>
- Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics, 8*(2), 162-177. <https://doi.org/10.1093/applin/8.2.162>
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). York Press.
- Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America, 76*(3), 708-721. <https://doi.org/10.1121/1.391257>
- Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research, 43*(1), 239-255. <https://doi.org/10.1044/jslhr.4301.239>
- Kehoe, M. M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition, 7*(1), 71-88. <https://doi.org/10.1017/S1366728904001282>
- Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of speech and*

- hearing Research*, 19(3), 421-447. <https://doi.org/10.1044/jshr.1903.421>
- Kewley-Port, D., & Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, 2(3), 195-210. [https://doi.org/10.1016/s0095-4470\(19\)31270-7](https://doi.org/10.1016/s0095-4470(19)31270-7)
- Khattab, G. (2002). VOT production in English and Arabic bilingual and monolingual children. In D. B. Parkinson & E. Benmamoun (Eds.), *Perspectives on Arabic Linguistics Papers from the Annual Symposium on Arabic Linguistics* (Vol. XIII-XIV, pp. 1-37). John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.230>
- Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America*, 133(2), EL142-EL147. <https://doi.org/10.1121/1.4778281>
- Liao, S. J. (2005). Interlanguage production of English stop consonants: A VOT analysis [Unpublished master's thesis]. National Kaohsiung Normal University.
- Lim, H., & Lim, S. (2018). Sebutan Bahasa Melayu oleh penutur Cina (ChinMalay) berdasarkan latar belakang persekolahan [Chinese speakers' pronunciation of Malay (ChinMalay) based on school background]. *Jurnal Bahasa*, 17(2), 334-358.
- Lin, Y. H. (2007). *The sounds of Chinese*. Cambridge University Press
- 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>
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7(1), 41-74. <https://doi.org/10.1017/s0305000900007029>
- MacLeod, A. A. (2016). Phonetic and phonological perspectives on the acquisition of voice onset time by French-speaking children. *Clinical Linguistics & Phonetics*, 30(8), 584-598. <https://doi.org/10.3109/02699206.2016.1152509>
- Melo, R. M., Mota, H. B., Mezzomo, C. L., & Brasil, B. D. C. (2015). Production and discrimination of voicing contrast of stops in cases of phonological disorder. *Revista CEFAC*, 17, 135-144. <https://doi.org/10.1590/1982-0216201517s14213>
- Ogasawara, N. (2011). Acoustic analysis of voice-onset time in Taiwan Mandarin and Japanese. *Concentric: Studies in Linguistics*, 37(2), 155-178.
- Park, J., & Byeon, H. (2015). Variability and duration of voice onset time in childhood apraxia of speech in Korea: Results of the spectrum study. *International Journal of Bio-Science and Bio-Technology*, 7(4), 67-76. <https://doi.org/10.14257/ijbsbt.2015.7.4.08>
- Rochet, B., & Yanmei, F. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, 19(4), 105-106.
- Shahidi, A. H. (2010). *An Acoustic and perceptual analysis of the phonetic properties of Malay English in comparison to those of Malay* [Unpublished doctoral dissertation]. University of Newcastle Upon Tyne.
- Shahidi, A. H., & Aman, R. (2011). An acoustical study of English plosives in word initial position produced by Malays. *3L: Language, Linguistics, Literature®*, 17(2), 23-33.
- Shahidi, A. H., Abdul Hamid, B., Mokhtar, S., & Kechot, A. S. (2016). Kontras penyuaran bunyi plosive awal kata oleh kanak-kanak Melayu: Satu penelitian awal [Voicing contrast of plosives in initial words by Malay children: A preliminary study]. *Journal of Social Sciences and Humanities*, 11(2), 217-231.

- Shahidi, A. H., Aman, R., & Hamid, Z. (2012). Kajian akustik realisasi kontras penyuaran bunyi plosif bahasa Melayu [Acoustic study of the realisation of the plosives voicing contrast in Malay]. *GEMA Online® Journal of Language Studies*, 12(2), 745-760.
- Shimizu, K. (1989). A cross-language study of voicing contrasts of stops. *Studia Phonologica*, 23, 1-12.
- Shimizu, K. (2011, August 17-21). A study on VOT of initial stops in English produced by Korean, Thai and Chinese speakers as L2 learners. *Online Proceedings of the International Congress of Phonetic Sciences (ICPhS) XVII 2011*, 1818-1821. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Shimizu/Shimizu.pdf>
- Tanaka, M. (2018). VOT of stops by Mandarin-Japanese bilingual children. *Concentric: Studies in Linguistics*, 44(1), 1-32.
- Umat, C., Hamid, B. A., Baharudin, A. (2015). Voice onset time (VOT) In prelingually deaf malay-speaking children with cochlear implants. *Journal of Phonetics & Audiology*, 1(1), 1-5. <https://doi.org/10.4172/jpay.1000106>
- Yang, J. (2018). Development of stop consonants in three- to six-year-old Mandarin-speaking children. *Journal of Child Language*, 45(5), 1091-1115. <https://doi.org/10.1017/S0305000918000090>
- Yu, V. Y., De Nil, L. F., & Pang, E. W. (2015). Effects of age, sex and syllable number on voice onset time: Evidence from children's voiceless aspirated stops. *Language and Speech*, 58(2), 152-167. <https://doi.org/10.1177/0023830914522994>
- Zlatin, M. A., & Koenigsnecht, R. A. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research*, 19(1), 93-111. <https://doi.org/10.1017/S0305000918000090>

