

Role of regional language background and speech styles on the production of Voice Onset Time (VOT) in English among Indonesian multilinguals

Ewynurul Laily Zen

Department of English, Faculty of Letters, Universitas Negeri Malang, Jalan Semarang 05, Malang, East Java, Indonesia

ABSTRACT

This paper seeks to contribute to the nature of cross-linguistic transfer in the production of English Voice Onset Time (VOT) by adult multilingual speakers in Indonesia in view of how different regional home languages and speech settings shape the phonetic realizations. Three adult multilinguals participated in this pilot project. They are all learners of English as the third language (L3) at the Department of English of a state university in Malang, Indonesia who acquire different regional home languages – Javanese, Sundanese, and Madurese – as the first language (L1) and speak Indonesian as the second language (L2). The participants' production of bilabial stop consonants of English /p/ and /b/ were elicited from two different speech settings; a *careful speech* via text readings (monologue and dialogue) and wordlist reading, and a *spontaneous speech* through natural conversation among participants. Twenty-one tokens from each participant were then analyzed acoustically in Praat. The findings show that the bilingual speaker with L1 Sundanese consistently produced the shortest VOT values of both /p/ and /b/. The Javanese speaker produced the intermediate lag, whereas the Madurese speaker produced the longest aspiration interval. It is shown that the Sundanese language provides the strongest transfer effect, while Madurese gives the least effect. In light of cross-linguistic transfer, however, the overall VOT productions clearly put forth evidence of L1 phonological transfer. The production of non-native bilabial stop VOTs of English is largely due to the absence of this phonetic property in Javanese and Sundanese, while Madurese shows marginal similarities. The findings also demonstrate that speech styles play only a marginal role in determining the production of VOTs that the VOTs of /p/ and /b/ in careful speech is found to be slightly longer than in the spontaneous settings. This study makes an original contribution to the area of phonological acquisition in adult speakers by giving attention to the understudied languages of Indonesia in order to more fully understand the interaction of different language systems in multilingual language acquisition and development.

Keywords: Bilabial stops of English; cross-linguistic transfer; regional home languages; Voice Onset Time

First Received:

25 May 2020

Revised:

28 August 2020

Accepted:

10 September 2020

Final Proof Received:

25 September 2020

Published:

30 September 2020

How to cite (in APA style):

Zen, E. L. (2020). Role of Regional Language Background and Speech Styles on the Production of Voice Onset Time (VOT) in English among Indonesian Multilinguals. *Indonesian Journal of Applied Linguistics*, 10(2), 359-368. <https://doi.org/10.17509/ijal.v10i2.28604>

INTRODUCTION

This paper explores how the production of English stop consonants among adult multilinguals may vary depending on different regional home languages spoken and speech settings. Central to the entire

discipline of multilingual acquisition is the concept of cross-linguistic transfer. With particular respect to the source of linguistic transfer, the multilingual acquisition is seen to be distinct from bilingual acquisition due to the possibility for learners to have

more than one language systems prior to the learning of the third language (Rothman, 2015). Multilingual transfer is, therefore, unique as it embodies multidirectional interaction of three language systems (Cenoz, 2001; Clyne, 1997; Herdina & Jessner, 2002; Sanz et al., 2015). Recent developments in this area have led to the proliferation of studies that favor multilingual acquisition in adulthood with an increased interest in morphology and syntax (for further discussion, see Antonova-Ünlü & Sağın-Şimşek, 2015; Bardel & Falk, 2007; Flynn et al., 2004; García-Mayo & Slabakova, 2012; Sereno & Jongman, 1997) and with a lack of research on phonology (Fallah et al., 2016; Jaensch, 2011). Besides, previous major works have relied heavily on Western language pairings (see Gut, 2010; Llana et al., 2010; Mayo & Slabakova, 2015; Mayr & Montanari, 2015; Missaglia, 2010; Rah, 2010; Sanchez, 2015). This indicates that investigations involving non-Western languages are inadequate. Thus more works in the area are needed. This current study, therefore, sets out to address these research gaps by investigating phonological production of English as the third language (L3) among adult multilinguals acquiring Indonesian as the second language (L2) and either Javanese, Madurese, or Sundanese as the first language (L1). Participants of this study have experienced complex linguistic processing since they speak three languages on a daily basis – a regional home language in private and inter-ethnic communication domains, Indonesian in public and intra-ethnic settings, and English in the classroom and other academic settings due to their engagement and professional status as students of English Department.

This paper intends to determine the extent to which L1 and L2 come to influence L3 within the underlying analytical framework of transfer by mainly following Smith and Kellerman's (1986) definition of transfer as the incorporation of elements from one language to another. English and the other languages under investigation are typologically unrelated languages so much so that their phonological structures vary considerably. Among a variety of different properties, aspiration becomes a key feature in stop consonant production (Davenport & Hannahs, 1998; Ladefoged & Johnson, 2011). In English, factors of consonant distribution and environment take a significant impact in determining the degree of aspiration in which word-initial stops are clearly aspirated while their word-final counterparts are not (Clark et al., 2007). In Indonesian, voiceless stops are obviously unaspirated (Muslich, 2014; Sneddon et al., 2010). Stops in Javanese, Madurese, and Sundanese exhibit similar phonetic features as Indonesian with Madurese considering voiced aspirated stop (Horne, 1961; Nothofer, 2006b; Poedjosoedarmo, 1993). The presence of aspiration has marked a delay in

voice onset time following a voiceless stop that is crucial in the phonological system of English. VOT, according to Ladefoged and Johnson (2011), is a period between the stop burst after the release of the closure to the start of the voicing that it is divided into two categories; short-lag (less than 30ms) and long-lag (above 30ms). The aspiration interval or VOT is considered to be long in English and other Germanic languages; 50-60ms or even longer at 60-80ms (Kang & Guion, 2006; Ladefoged & Johnson, 2011). With the absence of aspiration, the VOT of voiceless stops in Indonesian, Javanese, Madurese, and Sundanese can be predicted to be very small or even negative. Drawing upon two different phonetic contrasts, the learning of L3 English will presumably be more demanding as the other two languages do not share similar phonetic realizations. It can also be said that the phonological knowledge of Indonesian and regional home languages is accumulated altogether to bring non-facilitative effect during the acquisition of English VOT. In this account, Kehoe et al. (2004) propose that L2 learners may never acquire target-like VOT values when L1 and L2 do not share the same VOT qualities. Following their argument, this present study assumes that phonological transfer from other previously learned languages will be anticipated and also that the VOT production will vary across participants with different regional language backgrounds. The findings of this study should make an important contribution to the area of phonological transfer by involving more than one understudied language – Javanese, Madurese, and Sundanese – in order to be able to explore the degree of transfer effects. This study will serve as a baseline for further research looking at how regional languages in multilingual Indonesia may determine the learning of English as a foreign language. In a wider context of language pedagogy, such findings are critically important in shaping the direction of teaching and learning.

In the area of phonological acquisition, one of the most extensive explorations is the acquisition of VOT; Kehoe et al. (2004), for example, measured the VOT production of word-initial stop consonants of German by four German-Spanish bilingual children and compared them to the three monolingual German peers using naturalistic speech recordings. Similarly, Fabiano-Smith and Goldstein (2010) examined the VOT of voiceless bilabial [p] and velar stops [k] of Spanish and English produced by eight monolingual Spanish, eight monolingual English, and eight Spanish-English bilinguals. Also, Netelenbos et al. (2015) studied the VOT of French stop consonants produced by 56 French-English children enrolling in an early French immersion program in Alberta, Canada, by experimenting on the production of 54 words beginning with [p], [t], [k], [b], [d], and [g] and compared them to the 45 age-matched monolingual English-speaking

children. Departing from the different patterns of VOT in the two languages, children in these previous studies were found to demonstrate the followings: in the German-English data, they indicated (1) delay in the phonetic realization of voicing, (2) transfer of voicing features, and (3) no cross-language influence in the phonetic realization (Kehoe et al., 2004), in the Spanish-English, (1) monolingual and bilingual children generally differed on VOT in English, but not in Spanish and (2) no statistically significant differences were found between the Spanish and the English VOT of the bilingual children, but the VOT values did differ significantly for monolingual Spanish- versus monolingual English-speaking participants (Fabiano-Smith & Goldstein, 2010), whereas in the French-English data, (1) for the French voiceless stops, French immersion students display non-native-like VOT values in the intermediate range between monolingual English voiced and voiceless stops, (2) their English voiceless stops exhibit higher VOT values than the monolinguals' and are separate from those of their French, (3) for voiced stops, their English and French are indistinguishable, located within the range of voiced stops for monolingual English speakers (Netelenbos et al., 2015). These findings have provided important insights into how cross-linguistic interaction takes place during the acquisition of non-native language VOTs. However, previous studies have not dealt in much detail with how adult speakers construct target-like VOTs.

Pertaining to the idea of the length of L2 learning, Flege (1991) studied the production of Spanish and English [t] to test whether early and late L2 learning would affect the VOT of English [t] and whether their learning experience affects their production of Spanish [t]. The findings illustrated that the cross-linguistic transfer was mainly performed by the late L2 learners with an intermediate to short-lag VOT value of English [t] which is in contrast to the early learner who could produce the target-like English [t].

Extensive works have also been devoted to exploring the role of language settings or speech styles in the instances of VOT production. In a given conversation, monolingual and bilingual speech pattern as reflected primarily in code-switching practices are found to bring a fundamental effect on segmental phonetic production as well as the degree of phonetic transfer (Olson, 2013). Antoniou et al.'s (2011) empirical study, for instance, examined the VOT of Greek-English bilinguals' productions of word-initial and word-medial [b], [d], [p], [t] in both monolingual and bilingual mode. They found that all English stops were produced as code-switches from Greek, regardless of context, had more Greek-like VOTs, which is in contrast to Greek stops that showed no shift toward English VOTs, with the exception of

medial voiced stops. Their study highlighted the pervasive influence of L1 even in L2-dominant individuals as they aim to contrast the opposite argument. There has been, however, little analysis conducted to investigate the role of speech settings in determining the production of VOT. The speech settings here are operationalized as to whether the targeted sounds are produced in spontaneous speech or controlled settings. In such naturally occurring speech, particular sounds are generally produced with less cognitive control and in a continuous phonetic environment. In contrast to careful speech, when in isolation the targeted consonants are generally produced with relatively high cognitive control.

This study takes the form of a case study of three adult trilingual speakers speaking L1 Javanese, Madurese, and Sundanese with each speaker acquired Indonesian and English as an L2 and L3 respectively. An underlying concern of this present study is how these different L1s provide transfer effect toward the production of voiceless stop [p] and voiced stop [b] of English. In particular, the research questions are of two folds: (1) how does the VOT of word-initial bilabial stop consonants of English differ among speakers of different regional home languages? and (2) how do the speech settings provide effects on the VOT production?

METHOD

In this study, the production of targeted stop consonant [p] and [b] of English was collected from three multilingual participants; RR, EM, and AI. They are all seventh-semester students at the Department of English of a state university in Malang, Indonesia who acquire different regional home languages as a mother tongue and speak Indonesian as the national language. Nurtured in a comparatively similar linguistic environment, these participants were exposed to their regional language primarily in the family and ethnic community from birth. It is important to also note that Javanese, Sundanese, and Madurese are the languages with the largest number of speakers in the island of Java and Madura as reported in the 2010 national census (Ananta et al., 2015). These ethnic groups as well as their languages are widely spoken in its home provinces – Javanese in Central Java, Yogyakarta, and East Java, Sundanese in West Java, and Madurese in the island of Madura – even though the massive growth of Indonesian is said to gradually take over the role of regional home language in private domains at the expense of modernization, urbanization, and inter-racial marriage (Steinhauer, 1994). In the context of language acquisition and use, the three bilingual participants are proficient speakers of their own regional language who did not start learning Indonesian until the school age. As Indonesian is the sole official language of education

and other formal circumstances, they have also developed an advanced competence in both Standard and Colloquial variety of the language. In addition to being balanced bilinguals in regional language and Indonesian, these speakers have started learning English in a tutored setting since lower secondary level. Furthermore, these speakers have taken English as a major at the university level, meaning that English has been used intensively and extensively ensuring their level of L3 competence.

The participants' production of bilabial stop consonants of English [p] and [b] were elicited from two speech settings; a *careful speech* via text readings (monologue and dialogue) and wordlist reading, and a *spontaneous speech* through natural conversation among participants. They were also asked to do a self-introduction task using regional home language to obtain data on VOTs of their own home language. Fifteen tokens from each participant were generated from the corpus (see Table 1). The VOT values of these elicited words were then acoustically measured in Praat through several steps of analysis; (1) measuring VOT value of /p/ and /b/ in different sound sequences; consonant–vowel (CV) such as in [POssible] and consonant–vowel–consonant (CCV) such as in [PLAy], (2) quantifying mean VOT value of [p] and [b] in all sound sequences, (3) identifying mean VOT value of [p] and [b] from the two speech settings; careful speech and spontaneous speech, (4) comparing the result of VOT measurement across different L1 backgrounds and speech settings. In short, these VOT measurements and quantifications were then attempted to discover the VOT realizations across different regional home language speakers and different speech settings.

Table 1
Tokens for Praat Analysis

Targeted sounds	Targeted words
[p]	Possible People Parents Play Prom Problem Project
[b]	But Boy Big Because Been Blocker Brought Brother

FINDINGS AND DISCUSSION

In the section that follows, the data on how multilingual speakers under observation produce

word-initial bilabial stops of English will be presented with regard to how Javanese, Madurese, and Sundanese speakers differ in the extent of VOT productions. Further discussion will also be made in light of cross-linguistic transfer. The following subsection will then be a presentation of finding with respect to how the VOT of bilabial stops of English is realized in different speech settings.

The role of regional languages in the VOT productions

As Figure 1 shows, the results of the mean VOT value of [p] in both sequences indicate that the VOTs are realized shorter than the average of native speakers. As widely reported, the VOTs of average native speakers of English fall within the range of 50–60ms in voiceless bilabial [p] and 15–18ms in voiced bilabial [b]. Sundanese L1 speaker consistently produced a short lag VOT [p] within the range of 17ms and 31ms respectively. The Javanese speaker exhibits a longer VOT of [p] (32ms – 35ms) than his Sundanese peer but shorter than his Madurese peer (32ms – 46ms). The Madurese speaker produced the longest VOT in CV sequence (46ms), yet not long enough to reach the average value of native speakers. The aspiration interval is also determined by phonetic environments in which it tends to obviously longer in word-initial positions followed directly by vowels; in such cases where word-initial voiceless stops are followed by another consonants, the degree of aspiration is said to be shorter as a result of assimilation (Ladefoged, 2001). To this end, this study has established that the VOT values of [p] in a CCV sequence produced by the Sundanese and Javanese speakers is particularly longer than those in the CV sequence leading to an opposite direction from the target phonetic feature of L3 English.

Another intriguing finding is shown in the realization of voiced stop [b] in which the Sundanese speaker demonstrated the longest VOT (41ms) compared to the Javanese (33ms) and Madurese speaker (38ms) in the CCV sequence. While in the CV sequence, the result is consistent with the production of [p] in which Sundanese speaker (17ms) exhibited the shortest VOT compared to the other regional language speakers; 29ms and 25ms accordingly. An important point to highlight is that the English voiced stops are phonetically realized as a short lag at around 15–18ms among native speakers (Ladefoged & Johnson, 2011), whereas it is realized longer within the range of 17ms – 41ms in this study. It seems suggestive that these adult L3 learners of English have established a unique VOT realization in an attempt to compromise the target production. Figure 2 provides a clear illustration on how the VOT production can notably differ across regional language speakers with a similar tendency of moving away from target-like VOT realization. The

inability of adult bilingual speakers in my data to produce native-like VOT of English is a key finding, even though this result might have been expected as bilinguals would naturally experience such kind of cross-linguistic influence during the acquisition

process. Yet, the fact that the voiceless stop [p] is realized shorter while the voiced [b] is longer than the average of native speakers of English is particularly exceptional.

Figure 1
Mean VOT Value of [p] in Both Sound Sequences

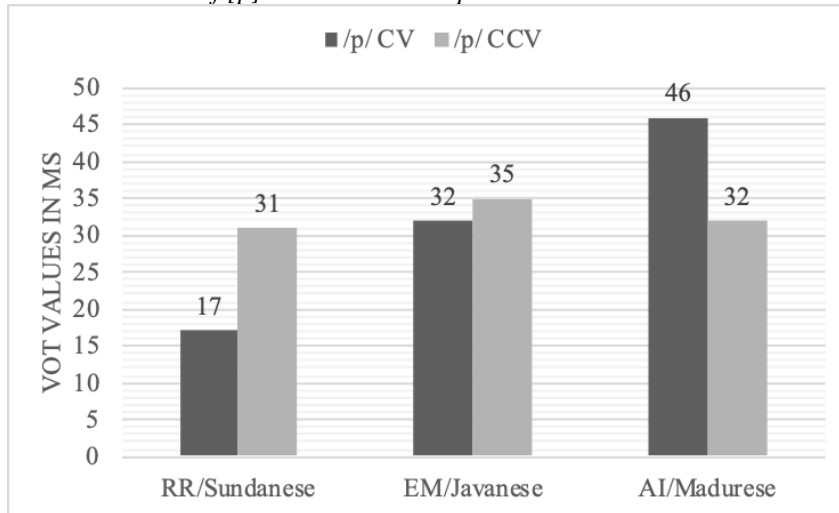
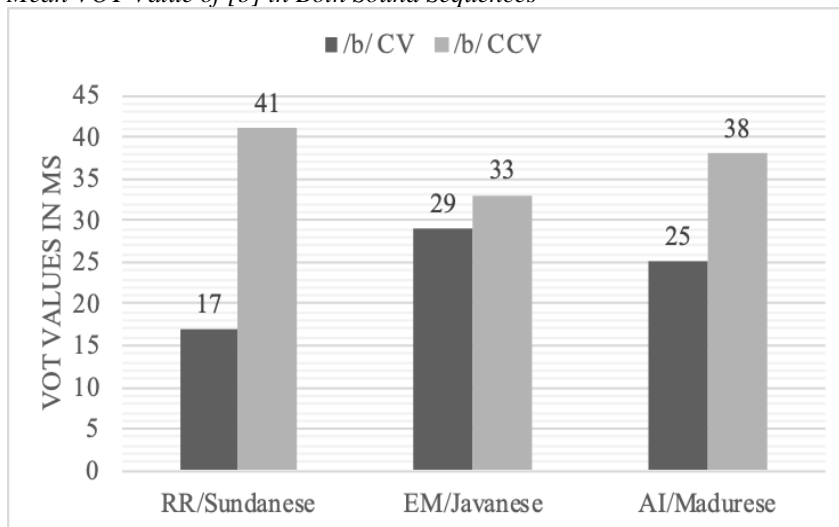


Figure 2
Mean VOT Value of [b] in Both Sound Sequences



These findings are in support of Paradis and Genesee's (1996) hypothesis on a segmental transfer underlining the concept that consonants and/or vowels along with their properties in one language will transfer to the productions of the other language(s). It is now convincing to put forward an assumption that these multilingual speakers' VOT productions of L3 English have undergone phonological transfer from both their regional language and Indonesian because there is no sharp contrast between voiceless and voiced stop consonants in these languages. In other words, voiceless consonants are unaspirated, the same way as the voiced counterparts except in Madurese, so

that these speakers have incorporated this phonological knowledge and use when learning a language whose voiceless consonants are significantly aspirated. In this regard, the VOT values in both voiceless and voiced pairs suggested in this study are typically non target-like resulting in the so-called accented speech. This is however predictable as the acquired languages whose VOTs stand at a different continuum would most possibly give significant influence toward the VOTs of non-native language(s) (Simonet, 2014). The unique feature of accented speech has also been apparent in Flege's (1991) study projecting to compare early and late learners. His analysis reveals that the late

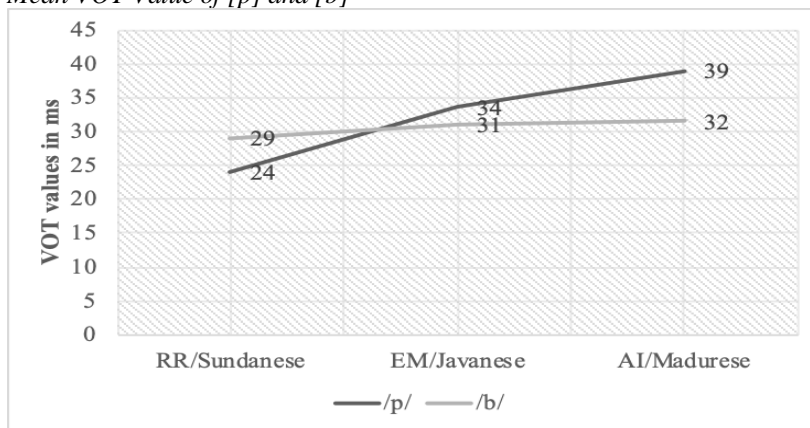
learners perceived L2 phonemes on the basis of L1 phonemic categories, unlike those found in early learners who were more successful in establishing phonetic independence.

With respect to how mother tongue provides transfer effects, the overall results of mean VOT values in Figure 3 show us that Sundanese speaker consistently produced short VOT of [p] and [b]. Her VOT for [b] is interestingly longer than her voiceless [p] where it is supposed to be in the opposite direction. It is suggestive that Sundanese provides the strongest transfer effect compared to Javanese and Madurese. The Javanese speaker's VOT realization is at the intermediate level producing voiceless [p] accurately longer than the voiced [b], yet not close enough to the average native speakers of English. The VOT production of Madurese speaker is the longest among other speakers with the voiceless [p] takes in a longer shape than the voiced [b]. Beyond this comparative result, a focus should be given to not only that all speakers produce a shorter VOT value of voiceless bilabial [p] compared to standard English, but also that they all produce VOT in their voiced bilabial [b]. The latter can be said to provide stronger evidence of cross-linguistic transfer because English voiced bilabial stop [b] should basically be realized

with very small or even negative VOT values. The word-initial voiced stops themselves are already voiced so the airstream closure is released together with the voicing of the following vowel, as Ladefoged and Disner (2012) elaborate. This feature seems to be very distinctive particularly in Javanese, where the voiced consonants are pronounced like voiceless stops with breathy voice (Nothofer, 2006a). Madurese, on the other hand, has a voiceless (tense stop) [p], voiced (lax stop) [b], and voiced aspirated (voiced stop with indifferent tension followed by strong aspiration) /b^h/ (Nothofer, 2006b). This phonetic property explains why the Madurese speaker tended to produce the longest VOT for both [p] and [b] compared to the other regional language speakers. In this way, Madurese has been found to give the least effect of transfer in the course of English VOT acquisition. This particular finding hints at the expected nature of cross-linguistic interaction where the possible outcome of it is that the regional language (L1 Javanese/Sundanese/Madurese) and the lingua franca (L2 Indonesian) have created cumulative effects in influencing the phonological production of a foreign language (L3 English) and that the L1 effects remain strong even with the intensive uses of L3 (Antoniou et al., 2011).

Figure 3

Mean VOT Value of [p] and [b]



On the basis of the current finding, the outcome of L3 phonological learning seems to depend largely on internal linguistic features of background languages as well as complex multilingual environments. In addition to the significant mother tongue (L1) influence as clearly suggested in this study, the absence of native environment of English from which learners can get primary exposures becomes a contributing factor in the appearance of non-target like outcomes here. Mayr and Montanari (2015) point out that this input conditioning factor is crucial during the acquisition process. They argue that when learners receive non-native speech from the environment, it will be difficult for them to extract specific phonological

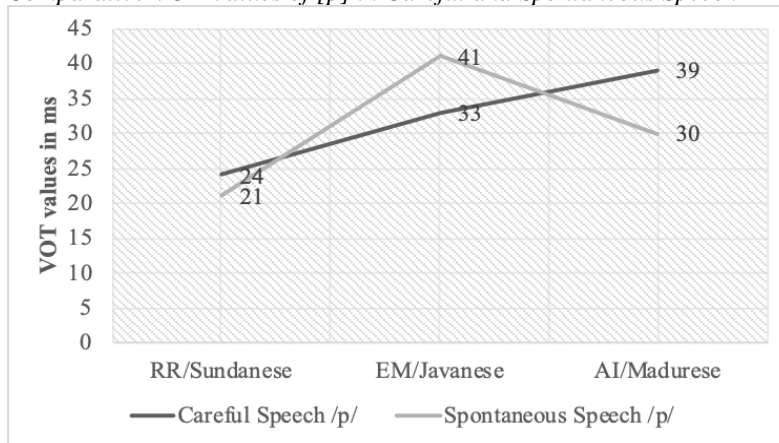
properties of the new language they learn. Place and Hoff's (2015) research provides supports to this premise that non-native input has become a negative predictor of language skills among their Spanish-English bilingual participants growing up in the US. Additional support for native environment as a prerequisite of foreign language learning comes from Lin and Johnson's (2010) study toward Mandarin-English bilingual children in English immersion class in China. The ability of their participants to acquire target-like L2 English phonology even when exposure to English was limited in school was mainly due to the appearance of native-speakers of English in classrooms.

The role of speech settings in the VOT productions

In addition to the degree of transfer effects provided by each of the regional languages, a careful observation on how speech settings influence the production of English VOTs is also carried out. From a brief comparative analysis on the VOT value of [p] in careful speech and spontaneous speech, the following result in Figure 4 is identified. The Javanese speaker (41ms) demonstrates a longer

aspiration interval of voiceless stop [p] in the spontaneous speech setting, while the Sundanese (24ms) and Madurese (39ms) speakers exhibit longer VOTs in the careful speech. With respect to regional language, bilingual speaker with Sundanese language background consistently produced the shortest VOTs of [p] in both controlled and natural speech. Regarding the settings, the interval value is longer when produced in careful speech as evident from Sundanese and Madurese speakers.

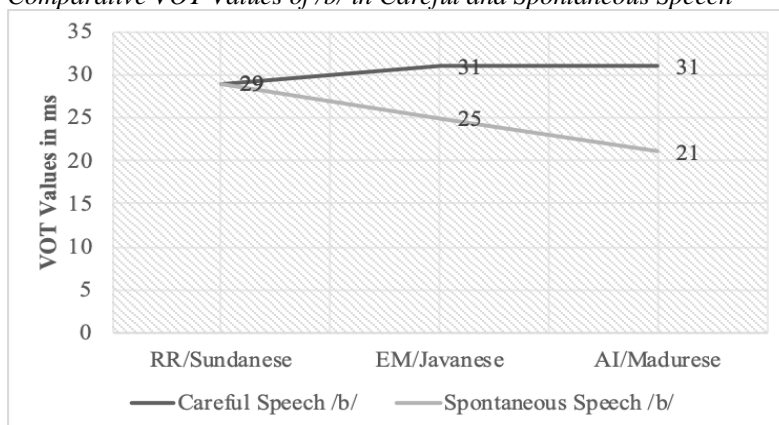
Figure 4
Comparative VOT Values of [p] in Careful and Spontaneous Speech



In the production of VOT value for voiced stop [b], Figure 5 shows that the same VOT value in both speech settings is maintained by the Sundanese speaker data. In this way, the context of speech production does not significantly determine the VOT realization of this speaker. It is consistent with her [p] production in which the minor gap between careful [p] (24ms) and spontaneous [p] (21ms) has clearly been established. The Javanese and

Madurese speakers, on the other hand, retain longer VOT of [b] in careful speech (31ms) and relatively shorter values (25ms and 21ms) in spontaneous setting. As reflected from Figure 4 and 5, however, the aspiration interval between the two speech settings marginally show a consistent difference in that the VOTs are more likely to be longer in careful speech rather than the spontaneous counterpart.

Figure 5
Comparative VOT Values of /b/ in Careful and Spontaneous Speech



However, if we look at the overall result of mean VOT value across different speech settings and speakers in Figure 6, the findings suggest that spontaneous speech only slightly determines the longer value of [p] while careful speech seems to

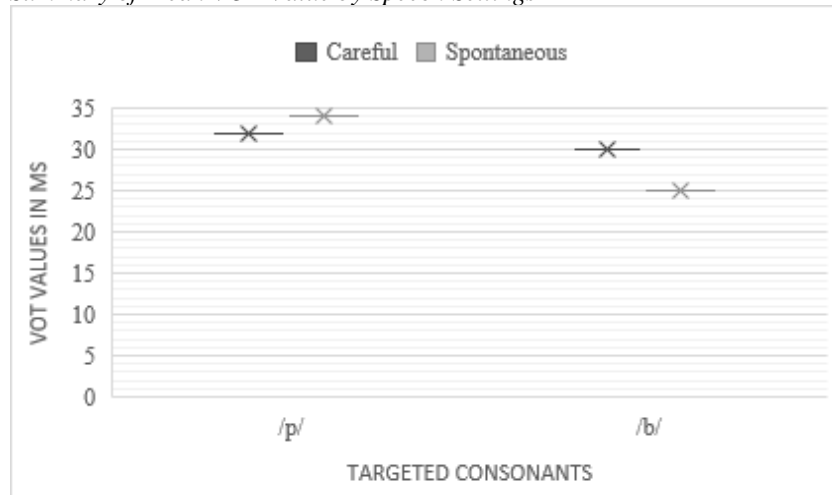
influence longer values of [b]. Marginal gaps between voiced and voiceless stops as controlled by speech settings, however, reflect insignificant role played by the settings. This is particularly in contrast to Gosy's (2001) study looking at the

behavior of three Hungarian voiceless stops when they are in isolation and in spontaneous speech showing a high tendency of the sounds to carry different VOT values. Bilabials and velars are

considerably shorter in spontaneous than in careful speech, in addition to the influence of vowels following the stops sounds in careful than in spontaneous speech.

Figure 6

Summary of Mean VOT Value by Speech Settings



CONCLUSIONS, LIMITATION AND FUTURE DIRECTION

To conclude, the acoustic measurements of word initial bilabial VOTs of English as produced by three adult bilinguals with different regional language backgrounds have shown that bilingual speakers with L1 Sundanese consistently produced the shortest VOT values of [p] and [b] with the range of 24ms and 29ms respectively. The Javanese speakers produced the intermediate lag of [p] (34ms) and [b] (31ms), whereas the Madurese produced the longest aspiration interval of [p] (39ms) and [b] (32ms). As these findings show, Sundanese language can be said to provide the strongest transfer effect, while Madurese gives the least effects when learning English. In light of cross-linguistic transfer framework, however, the overall VOT productions clearly suggest evidence of L1 phonological transfer. The realization of non-native bilabial stop VOTs of English here is considerably due to the absence of this phonetic property in Javanese and Sundanese with Madurese showing marginal similarities. With regard to the role of speech settings, empirical evidence in this present study indicates that speech settings take insignificant part in determining the production of VOTs. In this case, however, the VOTs of [p] and [b] in careful speech is found to be marginally longer than in the spontaneous speech.

Taken together, the findings provide empirical support for Kehoe’s et al., (2004) line of research highlighting the idea that non-native language learners may never acquire target-like VOT values as a result of cross-linguistic dissimilarities. It is in addition to minimum native input that the learners

might have during the acquisition process as well as the extent of L1 dominance. However, the readers should bear in mind that this study is based solely on the limited production of bilabial stops of English. This paper, therefore, cannot provide a comprehensive review of phonetic aspects of other articulation places such as velar and alveolar stop consonants. In the future, this current study should incorporate the whole stop consonant members with more varied and larger number of multilingual participants. Such studies would be critically important in helping language pedagogists, particularly English, in mapping out relevant learning needs to better assist the acquisition and development of English in a complex multilingual setting of Indonesia.

ACKNOWLEDGMENT

This paper was a preliminary report of an individual project in *Language Variation and Change* class at the National University of Singapore. I thank Dr Rebecca Lurie Starr for the fruitful ideas and comments. Of course, all remaining errors are my own.

REFERENCES

Ananta, A., Arifin, E. N., Hasbullah, M. S., Handayani, N. B., & Pramono, A. (2015). *Demography of Indonesia’s ethnicity*. ISEAS Publishing.
 Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals:

- Asymmetries in phonetic code-switching. *Journal of Phonetics*, 39(4), 558–570. <https://doi.org/10.1016/j.wocn.2011.03.001>
- Antonova-Ünlü, E., & Sağın-Şimşek, Ç. (2015). The use of verbal morphology in Turkish as a third language: The case of Russian–English–Turkish trilinguals. *International Journal of Bilingualism*, 19(3), 347–362. <https://doi.org/10.1177/1367006914526953>
- Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, 23(4), 459–484. <http://eric.ed.gov/?id=EJ778032>
- Cenoz, J. (2001). The effect of linguistic Distance, L2 status and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 8–19). Multilingual Matters Ltd.
- Clark, J., Yallop, C., & Fletcher, J. (2007). *An introduction to phonetics and phonology* (3rd Ed.). Blackwell Publishing.
- Clyne, M. G. (1997). Some of the things trilinguals do. *International Journal of Bilingualism*, 1(2), 95–116. <https://doi.org/10.1177/136700699700100201>
- Davenport, M., & Hannahs, S. J. (1998). *Introducing phonetics and phonology* (2nd ed.). Oxford University Press.
- Fabiano-Smith, L., & Goldstein, B. A. (2010). Phonological acquisition in bilingual Spanish–English speaking children. *Journal of Speech, Language and Hearing Research*, 53, 160–178. [https://doi.org/10.1044/1092-4388\(2009/07-0064\)](https://doi.org/10.1044/1092-4388(2009/07-0064))
- Fallah, N., Jabbari, A. A., & Fazilatfar, A. M. (2016). Source (s) of syntactic cross-linguistic influence (CLI): The case of L3 acquisition of English possessives by Mazandarani – Persian bilinguals. *Second Language Research*, 32(2), 225–245. <https://doi.org/10.1177/0267658315618009>
- Flege, J. E. (1991). Age of learning affects the authenticity of Voice-Onset Time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, 89(1), 395–411. <https://doi.org/10.1121/1.400473>
- Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for language acquisition: Comparing adults’ and children’s patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, 1(1), 3–16. <https://doi.org/10.1080/14790710408668175>
- García-Mayo, M. D. P., & Rothman, J. (2012). L3 morphosyntax in the generative tradition. In J. Cabrelli, S. Flynn, & J. Rothman (Eds.), *Third Language Acquisition in Adulthood* (pp. 9–32). John Benjamins Publishing Company. <https://doi.org/10.1075/sibill.46.04pil>
- Gosy, M. (2001). The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. *International Journal of Speech Technology*, 4(1), 75–85. <https://doi.org/10.1023/A:1009608900453>
- Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, 7(1), 19–38. <https://doi.org/10.1080/14790710902972248>
- Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics*. Multilingual Matters Ltd.
- Horne, E. C. (1961). *Beginning Javanese*. Yale University Press.
- Jaensch, C. (2011). L3 acquisition of German adjectival inflection: A generative account. *Second Language Research*. <https://doi.org/10.1177/0267658310386646>
- Kang, K.-H., & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean–English bilinguals. *The Journal of the Acoustical Society of America*, 119(3), 1672–1683. <https://doi.org/10.1121/1.2166607>
- Kehoe, M. M., Le, 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>
- Ladefoged, P. (2001). *A course in phonetics* (4th ed.). Harcourt College Publishers.
- Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants* (3rd ed.). Wiley–Blackwell.
- Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Wadsworth.
- Lin, L. C., & Johnson, C. J. (2010). Phonological patterns in Mandarin–English bilingual children. *Clinical Linguistics and Phonetics*, 24(4–5), 369–386. <https://doi.org/10.3109/02699200903532482>
- Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, 7(1), 39–57. <https://doi.org/10.1080/14790710902972255>
- Mayo, M. del P. G., & Slabakova, R. (2015). Object drop in L3 acquisition. *International Journal of Bilingualism*, 19(5), 483–498. <https://doi.org/10.1177/1367006914524643>
- Mayr, R., & Montanari, S. (2015). Cross-linguistic interaction in trilingual phonological development: The role of the input in the acquisition of the voicing contrast. *Journal of Child Language*, 42(5), 1006–1035. <https://doi.org/10.1017/S0305000914000592>

- Missaglia, F. (2010). The acquisition of L3 English vowels by infant German–Italian bilinguals. *International Journal of Multilingualism*. <https://doi.org/10.1080/14790710902972289>
- Muslich, M. (2014). *Fonologi Bahasa Indonesia*. Bumi Aksara.
- Netelenbos, N., Li, F., & Rosen, N. (2015). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism, 20*(3), 346–357. <https://doi.org/10.1177/1367006914564566>
- Nothofer, B. (2006a). Javanese. In *Encyclopedia of Language & Linguistics* (Vol. 6, pp. 113–115). Elsevier Ltd.
- Nothofer, B. (2006b). Madurese. In *Encyclopedia of Language & Linguistics* (Vol. 6, pp. 433–435). Elsevier Ltd.
- Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics, 41*(6), 407–420. <https://doi.org/10.1016/j.wocn.2013.07.005>
- Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children. *Studies in Second Language Acquisition, 18*(1), 1–25. <https://doi.org/10.1017/S0272263100014662>
- Place, S., & Hoff, E. (2015). Effects and noneffects of input in bilingual environments on dual language skills in 2½-year-olds. *Bilingualism: Language and Cognition, 19*(July), 1–19. <https://doi.org/10.1017/S1366728915000322>
- Poedjosoedarmo, G. R. (1993). Uses of phonation type in Javanese. *Oceanic Linguistics Special Publications, 24*(1), 123–131.
- Rah, A. (2010). Transfer in L3 sentence processing: Evidence from relative clause attachment ambiguities. *International Journal of Multilingualism, 7*(2), 147–161. <https://doi.org/10.1080/14790710903276342>
- Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism, 18*(2), 179–190. <https://doi.org/10.1017/S136672891300059X>
- Sanchez, L. (2015). L2 activation and blending in third language acquisition: Evidence of crosslinguistic influence from the L2 in a longitudinal study on the acquisition of L3 English. *Bilingualism: Language and Cognition, 18*(2), 252–269. <https://doi.org/10.1017/S1366728914000091>
- Sanz, C., Park, H. in, & Lado, B. (2015). A functional approach to cross-linguistic influence in ab initio L3 acquisition. *Bilingualism: Language and Cognition, 18*(02), 236–251. <https://doi.org/10.1017/S1366728914000285>
- Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition, 25*(4), 425–437.
- Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics, 43*(1), 26–37. <https://doi.org/10.1016/j.wocn.2014.01.004>
- Smith, M. S., & Kellerman, E. (1986). Crosslinguistic influence in second language acquisition: An introduction. In M. S. Smith & E. Kellerman (Eds.), *Crosslinguistic Influence in Second Language Acquisition* (pp. 1–9). Pergamon Press.
- Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian reference grammar* (2nd ed.). Allen & Unwin.
- Steinhauer, H. (1994). The Indonesian language situation and linguistics: Prospects and possibilities. *Bijdragen Tot de Taal-, Land- En Volkenkunde, 150*(4), 755–784. Retrieved from <http://www.jstor.org/stable/27864615>