Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Relative cue weighting in multilingual stop voicing production

Le Xuan Chan[1], *Annika Heuser*[1]

[1]Department of Linguistics, University of Pennsylvania, U.S.A.

lxchan@sas.upenn.edu, aheuser@sas.upenn.edu

Abstract

How does a multilingual speaker produce similar phonological contrasts across the different languages that they speak? Some theories predict crosslinguistic influence while others predict that multilinguals keep separate sound inventories for each language. In this paper, we present crosslinguistic data from early multilingual speakers in Malaysia. We investigate the interaction of a true voicing language (Malay), a variable voicing language (English), and an aspiration language (Mandarin). Using a random forest classification of nine acoustic correlates of stop voicing, we show that 1) all early multilinguals show language-specific productions of stop voicing, and 2) variation driven by dominance can still be observed despite this language-specificity. In addition, we present evidence that closure voicing is a salient correlate alongside aspiration in Malaysian English, and that English is more reliant on secondary correlates than Malay and Mandarin.

Index Terms: multilingual variation, language contact, stop voicing contrast, sociophonetics

1. Introduction

1.1. Crosslinguistic influence and language-specificity in multilingual speech

A major goal of multilingual phonetics is to determine whether and how speakers keep similar yet distinct crosslinguistic categories apart. Theories of L2 speech such as the revised Speech Learning Model [1], the Perceptual Assimilation Model [2], and the Native Language Magnet model [3] argue for the importance of similarity: the more similar a phonemic category in language A is to another sound in language B, the more likely we are to observe crosslinguistic influence. These theories advocate for a connected representation of phonemic categories in a multilingual speaker's grammar.

Yet, evidence also shows that multilingual speakers are able to, and even motivated to, keep crosslinguistic categories apart. The revised Speech Learning Model also emphasizes the importance of perceived dissimilarity – the idea that multilingual speakers will keep two phonemic categories apart as long as there is a perceptible difference between them. In other words, bilingual speakers will have two /p/ categories if the realization of /p/ is sufficiently different between their two languages. In the case of stop voicing, we have evidence from [4], [5], [6], and [7] that early multilinguals maintain language-specific production and perception of stop voicing categories across different languages.

1.2. Language contact and Malaysia

These theories are of particular relevance when examining phonetic variation in language contact settings, which are inherently multilingual in nature. One such community is Malaysia, a multiethnic and multilingual site of language contact, where the population consists of 58% Malays and *Bumiputeras* 'natives,' 30% Chinese, and 11% Indians [8].

The major languages spoken in Malaysia are Malay, English, Mandarin, and Tamil, while other Chinese (Cantonese, Hokkien, Hakka, Foochow, etc.) and Indic (Hindi, Malayalam, Telugu) languages are also spoken. These languages interact both at the individual and societal level in Malaysia, due to a combination of public and education policies, alongside Malaysia's history as a former British colony.

While Malay is the sole official language, as well as the main working language in the public sector, Malaysian English as a contact variety has undergone nativization [9, 10, 11] and is spoken by many as a native or dominant language. At the same time, mother tongue languages such as Mandarin and Tamil remain very salient languages within their respective communities [12, 13]. In fact, children can attend fully Mandarin- and Tamil-medium public schools, in addition to Malay-medium schools. All of this has resulted in a speech community where speakers are functionally bilingual, if not tri- or multilingual, with most speakers acquiring their non-dominant languages at an early age. All speakers whose data we discuss in this paper, for instance, reported acquiring their non-dominant languages no later than age 7.

1.3. Stop voicing in Malaysia

Our feature of interest is stop voicing. Malaysia presents an interesting case study of contact in stop voicing because of the interaction between three typologically different languages – Malay, Malaysian English, and Malaysian Mandarin. Malay is a "true voicing" language: the contrast between voiced /b,d,g/ and voiceless /p,t,k/ is in vocal fold vibration during the consonant closure (i.e. closure voicing). Aspiration has not been observed to be a salient cue for distinguishing the two categories [14, 15]. Mandarin contrasts with Malay in that it is an aspiration language, and has a phonemic aspiration contrast between /p,t,k/ and /p^h,t^h,k^h/, without voicing [16, 17]. [18] show that the aspiration contrast in Malaysian Mandarin patterns with that of other varieties.

English represents a middle ground between Malay and Mandarin, in that a phonological voicing contrast exists between /b,d,g/ and /p,t,k/, but aspiration is considered the primary cue for distinguishing between the two categories, at least phrase-initially [19]. In phrase-medial positions, however, [20, 21, 22] have shown that closure voicing does occur

10.21437/Interspeech.2025-2259

Table 1: *Coding and information about each participant group. F=female.*

Code	Group	N	Language Dominance
Mal-dom.bi	Malay-dominant bilinguals	12 (F=6)	**Malay** >English
Eng-dom.bi	English-dominant bilinguals	10 (F=5)	**English** >Malay
Eng-dom.tri	English-dominant trilinguals	14 (F=5)	**English** >Mandarin >Malay
Mand-dom.tri	Mandarin-dominant trilinguals	13 (F=5)	**Mandarin** >English >Malay

for /b,d,g/.

In Malaysian English, aspiration has been shown to be variable, with [23, 14] reporting short-lag voice onset time (VOT) for /p,t,k/, while other studies like [18, 24] report long-lag VOT values. These studies agree that closure voicing is also present in Malaysian English, though its importance relative to aspiration has not been ascertained. In this paper, we use a random forest classification model to do precisely that.

The goal of this paper is to characterize the multilingual variation present in a speech community where all three languages are spoken regularly among speakers and are in constant contact. In particular, we look at the crosslinguistic productions of four groups that differ in language dominance and linguistic repertoire, and ask two questions: 1) Do early multilinguals in Malaysia maintain language-specificity in their productions of stop voicing crosslinguistically, and if so, 2) Can we still observe variation within this speech community due to speakers' dominance and language backgrounds?

While focusing on the specific empirical values of each individual correlate is essential for an understanding of how stop voicing is implemented articulatorily and acoustically, this paper focuses on 1) the relative cue weighting of each correlate (i.e. how correlates "bundle" together to form a phonological contrast), and 2) how these weights compare across Malay, English, and Mandarin among different language backgrounds. In doing so, we show the relevant cue-trading strategies for producing voicing crosslinguistically by multilingual speakers in Malaysia, and how this varies by language background.

2. Methods

2.1. Data

Our data is a subset of a larger corpus of multilingual speech collected by the first author. It contains Malay, English, and Mandarin elicited from Malaysian speakers differing in dominance and language background. Our data subset consists of word list readings of target items containing word-initial stops in all three languages. All places of articulation (labial, alveolar, velar) were included for both voicing categories in each languages. Including repetitions, each speaker produced 60 Malay tokens (30 voiceless, 30 voiced), 70 English tokens (36 voiceless, 34 voiced), and 42 Mandarin tokens (26 unaspirated, 16 aspirated). All stimuli were inserted in carrier phrases in the respective language and were presented to speakers as slides with accompanying illustrations. Speakers produced all three languages in an identical context and received the same set of instructions: "Read each sentence in a way that feels the most natural to you."

We recruited four groups of Malaysian speakers based on *dominance*, i.e. which language they are dominant in, and *linguistic repertoire*, i.e. how many languages they speak (Table 1). All groups speak Malay and English, and only the trilingual groups speak Mandarin. Language dominance was determined using an adapted version of the Bilingual Language Profile [25]. Note that this questionnaire is a rubric that takes into account language history (e.g., age of acquisition, etc.), frequency of usage, self-rated proficiency, and language attitudes, and is not merely a measure of proficiency. In terms of ethnicity, the Malay-dominant group consisted of ethnic Malays, while the other groups were ethnic Chinese Malaysians. Malay, English and Mandarin data were elicited from both trilingual groups, while only Malay and English data were elicited from the bilingual groups. A total of 7,504 tokens were elicited.

2.2. Pre-processing

The data were hand-segmented and annotated in Praat for stop identity, closure duration, aspiration/VOT, burst intensity, following vowel duration, and following vowel identity. We extracted the proportion of voicing during oral closure (voicing hereafter). To calculate voicing, we detected periodic pulses during the oral closure interval and calculated the proportion of periodic frames. We also extracted the earliest defined onset f0 and F1 values. We checked at 5% vowel duration, then incrementally checked 10%, 15%, etc., until we found defined values for each of f0 and F1. Similarly, we extracted f0 and F1 at 25% (or the nearest defined vowel duration). We then calculated the difference between the early- and mid-vowel f0 and F1 values, respectively. We refer to this difference as the onset f0/F1 slope.

We z-score normalized closure duration and burst intensity by language and speaker. Remember that each speaker produced sentences in at least 2 languages. For Malay and English, we z-score normalized onset f0 and onset f0 slope by language and speaker. For Mandarin, we normalized these features by tone as well. Finally, we normalized vowel duration, onset F1, and onset F1 slope by language, speaker, and vowel. We did not normalize voicing nor aspiration/VOT. Voicing is a percentage so it is already constrained to a range of 0-100, regardless of the speaker. Whether a language is classified as a true voicing or aspiration language is based on its raw aspiration/VOT range. For this reason, aspiration/VOT is not usually normalized in the literature (e.g. [26, 27, 28, 5]). Doing so would likely mask the very language-specific effects that we are interested in.

For z-score normalized features, we replaced outliers with NaN values. We defined outliers as z-scored values above 3 or below -3. We did not delete data points with NaN values for any of the features. We simply ignored the NaNs and used the remaining available features. This again allowed us to maximize the amount of data available for our consequent random forest analysis.

2.3. Random forest analysis

We chose a random forest analysis to determine the relative cue weights of the voicing contrast for each participant group in each language. For the voicing contrast specifically, the feature importances of decision trees have reflected established perceptual results of American English speakers[29]. A random forest model is an ensemble model consisting of many decision trees.

Table 2: *Random forest voicing classification accuracy for each language and participant group permutation. Overall accuracies are based on all available features, while the next row of reported accuracies are excluding voicing and aspiration.*

Language	Malay (Ml)				English (En)				Mandarin (Mn)	
Group	Ml.bi	En.bi	En.tri	Mn.tri	Ml.bi	En.bi	En.tri	Mn.tri	En.tri	Mn.tri
Overall accuracy	0.98	0.98	0.94	0.89	0.97	0.96	0.99	0.96	0.99	0.93
Excluding voicing and aspiration features										
Accuracy	0.78	0.73	0.77	0.74	0.84	0.86	0.86	0.84	0.78	0.76
Accuracy decrease	0.20	0.25	0.17	0.15	0.13	0.10	0.13	0.12	0.21	0.17

Accordingly, a random forest's feature importances is the average of its decision trees' feature importances. Previous work have also had success in modeling cue weighting in other phenomena with random forests (e.g. Mandarin tone acquisition [30] and Korean laryngeal contrast [31]). We therefore believe that the random forest model is a good choice for our purpose of characterizing the variation in cue weights for the voicing contrast across different languages and different multilingual speaker profiles.

We used the Python scikit-learn RandomForestClassifier class and opted for all the defaults, except for split criterion and maximum tree depth. We conducted a grid search to find the optimal settings for these parameters for each participant group in each language. The constraint on the depth of the trees making up the random forest were one of a number of parameters available to prevent overfitting. The two split criterion options that we included were gini impurity and entropy.

First, we held out about 10% of the speakers for each language/speaker group combination. This prevents each model from being able to use any speaker-specific patterns to succeed on the test set. We held out 1 speaker for the smallest group (English-dominant bilinguals) and 2 speakers for every other group. The speakers held out from each group differed for the language-specific random forests (e.g. Malay/Mal-dom.bi vs. English/Mal-dom.bi). Overall, we held out between 10-16.7% from each language/speaker group combination.

The remaining data was split into 10 random 90/10 train/validation splits, without regard for the speaker identity, for each parameter configuration in the grid search (i.e. criterion and depth combination). We tested each maximum tree depth between 1 and 15 (inclusive). We determined the parameters with the best average validation performance and moved forward with these. We then used all the training data to train a 1000-tree random forest from which we extracted the feature importances. We also recorded this random forest's accuracy on the held-out-speakers test data. We repeated this entire process 10 times, and report the average test accuracy across these iterations in table 2. The best parameters as determined by the grid search changed with each iteration.

We also conducted this analysis without the voicing and aspiration/VOT (henceforth just aspiration) features. The goal of this was to determine whether the efficacy of secondary cues for classification changed between languages or participant groups. We conducted the same grid search but this time missing the voicing and aspiration features. The average accuracies for this analysis are also recorded in table 2.

3. Results

Overall accuracy across all the participant groups in each language is quite high. As shown in Table 2, the average across all the groups being about 96%. Overall, the relative weighting of voicing and aspiration is what we expect for each language: voicing is most important for Malay, aspiration is most important for Mandarin, and both are important for English. Despite each group maintaining language-specificity, we still observe variation between groups in the magnitude of importance assigned to each correlate. We predominantly discuss the multilingual variation of aspiration and voicing in this paper, leaving the discussion of other correlates for future work.

Looking first at Malay, we present the feature importances for each participant group in Figure 1. Unsurprisingly for a true voicing language, voicing is by far the most important feature for all the groups. We nonetheless see an effect of language dominance: voicing has a lower importance (under 0.5) for the Mandarin-dominant speakers than for the other groups (over 0.5). This indicates crosslinguistic influence from Mandarin, which does not utilize a voicing contrast in stops.

For English, we turn to the feature importances displayed in Figure 2. Here, both voicing and aspiration are strong pri-

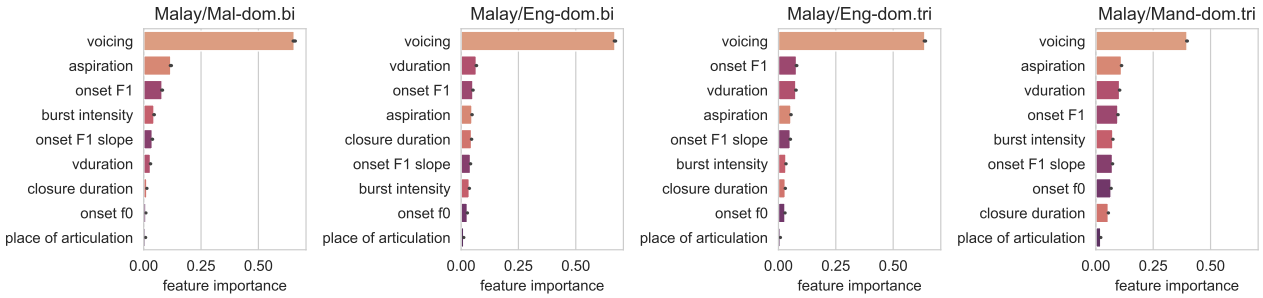

Figure 1: *Feature importances of each participant group in Malay.*

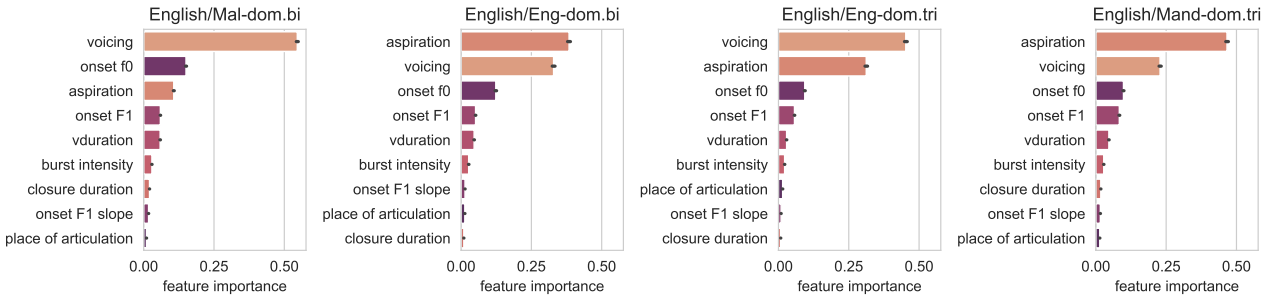

Figure 2: *Feature importances of each participant group in English.*

mary cues. Interestingly, which is the strongest cue appears to be a function of language dominance. For Malay-dominant bilinguals, voicing is much more important than aspiration. In fact, aspiration is not even the second-most important cue–onset f0 outranks it. For the English-dominant groups, both aspiration and voicing are important. For both the English-dominant *bilinguals*, voicing is the most important cue, though it is just barely more important than voicing. This is reversed for the English-dominant *trilinguals*, for whom voicing is more important than aspiration. We speculate that this discrepancy arises from wanting to differentiate English more from Mandarin, which is an aspiration language. Finally, for Mandarin-dominant trilinguals, aspiration is much more important than voicing. This group assigns more importance to aspiration for English than any other group, likely because of transfer from Mandarin.

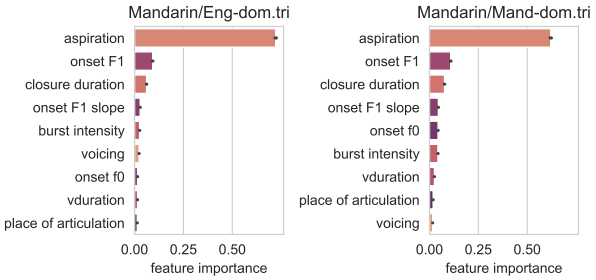

Figure 3: *Feature importances of each participant group in Mandarin.*

The Mandarin feature importances are displayed in Figure 3. Remember that only the trilingual groups produced Mandarin data. Unsurprisingly, aspiration is by far the most important feature for both groups. Surprisingly, it is more important for the English-dominant group than for the Mandarin-dominant group. This lends itself to our speculation for why voicing is more important for classifying the stops in these speakers' English. The English and Mandarin feature importances for English-dominant trilinguals place English and Mandarin further apart on the voicing to aspiration spectrum.

Integrating across the multilingual groups, we confirm that Malay is a voicing language, Mandarin is an aspiration language, and English is variably somewhere in between. We hypothesized that this variability might allow for secondary cues to be more predictive in English than for Malay and Mandarin. We tested this by training random forests with all features but voicing and aspiration. The accuracies are reported in Table 2. Notice that the average accuracy decrease (bottom row) is lower for English (average of 0.12) than for Malay (0.1925) or for Mandarin (0.19). This is evidence in support of our hypothesis. Additionally, onset f0 seems to be a consistently important secondary cue in English for all groups, only ranked behind aspiration and voicing.

4. Discussion

We conclude that early multilinguals in language contact settings are indeed good at keeping distinct categories apart. In the face of three typologically different stop voicing contrasts, multilingual speakers of Malay, English, and Mandarin do not apply wholesale the contrast from one language to another. These results corroborate previous findings of early multilinguals [5, 6, 16] as well as [1]'s theory of perceived dissimilarity.

Despite language-specific productions by all groups, we still see variation resulting from crosslinguistic influence across a highly multilingual society. In Malay, the importance of voicing as a correlate decreases for the Mandarin-dominant group compared to the other groups. In English, while English-dominant groups rely on both aspiration and voicing, the Malay-dominant group utilizes voicing more, while the Mandarin-dominant group utilizes aspiration more. Theories of socially-conditioned variation [32, 33, 34] would also then predict that in a contact setting like Malaysia, there exist phonetic cues that listeners can rely on to identify a speaker's language background. This also ties to further questions of phonetic adaptation in a contact setting.

Finally, through our random forest analysis, we show that voicing is a salient correlate in Malaysian English alongside aspiration. This accounts for discrepancies in reported aspiration and voicing patterns in Malaysian English [18, 14]. In addition, the availability of both correlates results in dominance-driven variation being more salient in English. This means that in natural settings, we would expect more variation in the production of /b,d,g/ and /p,t,k/ in Malaysian English over Malay and Mandarin. Accordingly, we predicted that there might also be a stronger reliance on secondary cues for perceiving stop voicing in Malaysian English. Although perceptual experiments are needed to confirm this, we find that random forest classification accuracy decreases less for English than for the other languages when voicing and aspiration are not available cues. This suggests that secondary correlates such as onset f0 play a salient role in the voicing contrast of Malaysian English.

5. Acknowledgments

Data collection for this project was supported by the National University of Singapore Graduate Research Support Scheme and the Student Project Fund from the NUS Department of English, Linguistics, and Theatre Studies' awarded to the first author. We would like to thank Rebecca Starr, Jianjing Kuang, Meredith Tamminga, and Marlyse Baptista, as well as the NUS Sociolinguistics Reading Group and the Penn Phonetics Lab for their insightful comments and suggestions. We also thank Andrew Zhu for helpful discussions about algorithm details.

6. References

[1] J. E. Flege and O.-S. Bohn, "The revised speech learning model (slm-r)," *Second language speech learning: Theoretical and empirical progress*, vol. 10, no. 9781108886901.002, 2021.

[2] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege*. John Benjamins Publishing Company, 2008, pp. 13–34.

[3] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson, "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (nlm-e)," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1493, pp. 979–1000, 2008.

[4] A. A. MacLeod and C. Stoel-Gammon, "Are bilinguals different? what vot tells us about simultaneous bilinguals," *Journal of Multilingual Communication Disorders*, vol. 3, no. 2, pp. 118–127, 2005.

[5] M. Geiss, S. Gumbsheimer, A. Lloyd-Smith, S. Schmid, and T. Kupisch, "Voice onset time in multilingual speakers: Italian heritage speakers in germany with l3 english," *Studies in Second Language Acquisition*, vol. 44, no. 2, pp. 435–459, 2022.

[6] J. Schertz, K. Carbonell, and A. J. Lotto, "Language specificity in phonetic cue weighting: Monolingual and bilingual perception of the stop voicing contrast in english and spanish," *Phonetica*, vol. 77, no. 3, pp. 186–208, 2020.

[7] C. Nagle, M. M. Baese-Berk, C. Diantoro, and H. Kim, "How good does this sound? examining listeners' second language proficiency and their perception of category goodness in their native language," *Languages*, vol. 8, no. 1, p. 43, 2023.

[8] Department of Statistics Malaysia, "The Population of Malaysia," https://open.dosm.gov.my/dashboard/population, accessed: 2025-02-19.

[9] L. Baskaran, "Malaysian english: Phonology," *A handbook of varieties of English*, vol. 1, pp. 1034–46, 2004.

[10] E. W. Schneider, "Evolutionary patterns of new englishes and the special case of malaysian english," *Asian Englishes*, vol. 6, no. 2, pp. 44–63, 2003.

[11] A. Hashim, "Malaysian english," *The handbook of asian englishes*, pp. 373–397, 2020.

[12] R. Vollmann and T. W. Soon, "Chinese identities in multilingual malaysia," *Grazer Linguistische Studien*, vol. 89, pp. 35–61, 2018.

[13] I. M. Naji and M. K. David, "Markers of ethnic identity: Focus on the malaysian tamil community," *International Journal of the Sociology of Language*, vol. 161, pp. 91–102, 2003.

[14] A. H. Shahidi and R. Aman, "An acoustical study of english plosives in word initial position produced by malays," *3L, Language, Linguistics, Literature*, vol. 17, no. 2, 2011.

[15] B. A. Hamid, C. J. Yee, and H. M. Ibrahim, "Acoustic analysis of voicing contrast in malay word-initial plosives produced by mandarin-speaking children." *Pertanika Journal of Social Sciences & Humanities*, vol. 30, no. 4, 2022.

[16] C. B. Chang, Y. Yao, E. F. Haynes, and R. Rhodes, "Production of phonetic and phonological contrast by heritage speakers of mandarin," *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 3964–3980, 2011.

[17] K.-Y. Chao, G. Khattab, and L.-m. Chen, "Comparison of vot patterns in mandarin chinese and in english," in *4th Annual Hawaii International Conference on Arts and Humanities*, 2006, pp. 840–859.

[18] B. K. Ng and P. S. Chiew, "L1 influence on stop consonant production: A case study of malaysian mandarin-english bilinguals," *Asian Englishes*, vol. 26, no. 2, pp. 310–336, 2024.

[19] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, vol. 20, no. 3, pp. 384–422, 1964.

[20] L. Davidson, "Variability in the implementation of voicing in american english obstruents," *Journal of Phonetics*, vol. 54, pp. 35–50, 2016.

[21] D. Deterding and F. Nolan, "Aspiration and voicing of chinese and english plosives," in *Proceedings of the 16th international congress of phonetic sciences*. Universität des Saarlandes Saarbrücken Germany, 2007, pp. 385–388.

[22] G. J. Docherty, *The timing of voicing in British English obstruents*. Walter de Gruyter, 1992, no. 9.

[23] H. S. Phoon, A. C. Abdullah, and M. Maclagan, "The consonant realizations of malay-, chinese-and indian-influenced malaysian english," *Australian Journal of Linguistics*, vol. 33, no. 1, pp. 3–30, 2013.

[24] T. Yamaguchi and M. Ptursson, "Voiceless stop consonants in malaysian english: Measuring the vot values," *Asian Englishes*, vol. 15, no. 2, pp. 60–79, 2012.

[25] L. M. Gertken, M. Amengual, and D. Birdsong, "Assessing language dominance with the bilingual language profile," *Measuring L2 proficiency: Perspectives from SLA*, vol. 208, p. 225, 2014.

[26] T. Cho and P. Ladefoged, "Variation and universals in vot: evidence from 18 languages," *Journal of phonetics*, vol. 27, no. 2, pp. 207–229, 1999.

[27] K. Shimizu, "A study on vot of initial stops in english produced by korean, thai and chinese speakers as l2 learners." in *ICPhS*, 2011, pp. 1818–1821.

[28] J. P. Kirby, D. R. Ladd *et al.*, "Stop voicing and f0 perturbations: Evidence from french and italian." in *ICPhS*, 2015.

[29] A. Heuser and J. Kuang, "Information-theoretic hypothesis generation of relative cue weighting for the voicing contrast," in *Interspeech 2024*, 2024, pp. 3585–3589.

[30] N. Rhee, A. Chen, and J. Kuang, "Going beyond f0: The acquisition of mandarin tones," *Journal of Child Language*, vol. 48, no. 2, pp. 387–398, 2021.

[31] J. Perkins, Y. Yan, S. J. Dahm Lee, and I. University, "Using machine learning to model the three-way laryngeal contrast in korean," in *Proceedings of the 20th International Congress of Phonetic Sciences*, 2023, pp. 783–787.

[32] K. Becker and E. L. Coggshall, "The sociolinguistics of ethnicity in new york city," *Language and Linguistics Compass*, vol. 3, no. 3, pp. 751–766, 2009.

[33] M. Newman and A. Wu, ""do you sound asian when you speak english?" racial identification and voice in chinese and korean americans' english," *American Speech*, vol. 86, no. 2, pp. 152–178, 2011.

[34] R. L. Starr and B. Balasubramaniam, "Variation and change in english/r/among tamil indian singaporeans," *World Englishes*, vol. 38, no. 4, pp. 630–643, 2019.