Project Title	**:**	Phonetic and phonological characteristics of workplace English produced by university graduates in Hong Kong
Grantee	**:**	City University of Hong Kong
Principal Investigator	**:**	LI Bin Department of Linguistics and Translation City University of Hong Kong
Co-investigator	**:**	YAN Xiu Department of Linguistics and Translation City University of Hong Kong

Final Report

by

Principal Investigator

(a) Title: Phonetic and phonological characteristics of workplace English produced by university graduates in Hong Kong

(b) Abstract

Local university graduates in Hong Kong who are second language learners of English (ESL learners) have been reported to face challenges in achieving satisfactory levels in oral proficiency to cope with professional communication at workplaces. In response to an immense concern to maintain the competitiveness of Hong Kong in the international market, the government has endeavoured to improve the knowledge of English among the young working population. This project aimed to address the needs for effective ESL for professional settings by investigating phonetic and phonological characteristics in workplace English by local university graduates.

We conducted studies and experiments to first identify local ESL professional needs and then to examine ESL speech in local settings. We analysed phonetic patterns in production of segments and suprasegments, and examined phonological processes to identify influences from ESL learners' first languages on the development of oral skills in English. Our survey with nursing and pharmacy majors prompted follow-up investigation on ESL Lombard speech under adverse conditions, which revealed imminent issues and challenges in ESL professional training targeting at authentic workplace interactions.

Our project and findings contribute to enrich the literature on theoretical and empirical issues relating workplace English in Hong Kong. The data and analysis from this project enable us a comprehensive understanding of phonetic and phonological (dis)similarities between ESL speakers and first-language speakers of English. We also propose pedagogical and strategic recommendations for more effective design and development of ESL English training that is applicable to professional needs in Hong Kong and beyond. (249 words)

(c) Keywords

Phonetic cues, L2 phonology, workplace English, English-as-a-second-language (ESL) in Hong Kong, Lombard speech, English for special purposes, phonetic adjustment.

(d) Introduction

The last stage of the project recorded our continued investigation on phonetic and phonological analysis of the workplace English, in four inter-related studies which were designed and modified to cope with the persistent pandemic situation that has challenged feasibility of inviting participants to campus for research activities. All proposed changes and studies are completed successfully in the final stage of the project (Summary of changes and outcome can be found in Table d-1.).

In previous stages, our surveys and preliminary studies involving nurses and nursing interns confirmed that accuracy and intelligibility of workplace English by local graduates is affected by both linguistic and non-linguistic factors including the first language and affective states correlated with work environment. For example, workplaces in clinics and emergence rooms involve alarms, babbles and other types of noises, which all add degrees of stresses to nurses and other service providers, who are overwhelmed with tasks and noises at the same time. This leads to speaking faster and louder unconsciously than normal, to compensate for signal degradation, but at the sacrifice of sufficient attention to accuracy in English-as-a-second-language (ESL) speech production. In order to assess linguistic strategies ESL users adopt when speaking under noises and in different styles, we designed four studies to examine linguistic modifications in ESL speech under adverse conditions and their effects on communication. We examine production and perception of English-for-special-purpose by ESL nurses and nursing interns. Their phonetic realization and perceptual patterns were assessed and compared across various situations relating to common use in professional communication and under various affective states. Specifically, **Study 1** contained the phonetic analysis of speech by Cantonese ESL speakers, results from which provided baseline and reference for analysis of speech under adverse conditions to follow. **Study 2** examined Cantonese ESL speakers' speech production in noise that simulated the work environment for nurses and caretakers in emergency rooms and hospitals. By assessing effects of babble noise on speech production and perception, our study showed that Cantonese ESL speakers at high proficiency levels have limited experience in adapting to non-L1 speaking and listening in noise. **Study 3** assessed and recorded effects of acoustic modifications to speech features on speech intelligibility in communication under noises. **Study 4** evaluated and showed effects of speaking styles on ESL perception of speakers' emotions.

Our results reveal unique phonetic and phonological patterns in workplace English by local university graduates. Their English use in workplaces is significantly affected by subject and nature of jobs which may put different requirement on accuracy and fluency in their English use, but also request awareness and proficiency in coping with unfavourable yet unavoidable conditions in workplaces. The university graduates we investigated showed certain knowledge in adopting verbal strategies to overcome adverse influence from environment noises. Their phonetic modifications however show distinct patterns from those of L1 speakers of English, which is attributed to the phonological influence of their L1, Cantonese, as a tonal language. Moreover, when identifiable linguistic information is present in the background noise such as multi-talker babbles, ESL speakers' attention is significantly distracted, which may lead to stresses in speaking. Consequently, their fluency and accuracy in speech communication are adversely affected. Such phonetic modifications to speaking styles under different conditions do not only lead to changes in perception of contents in speech but also in interpretation of emotions. Therefore, we suggest that training on perception and production of speech under noises and in emotional status should be introduced to advanced levels of English pedagogy especially in English for special purposes. Last but not least, our project was extended to special groups at workplaces: the visually impaired ESL speakers who aspire to hone their phonetic skills in

English (please see Appendix 4).

Table d-1. Summary of research activities in previous and current reporting periods.

a	Studies completed and reported in previous three progress reports	Summary of results and findings
1	Phonetic analysis of 'so' as a discourse marker in Hong Kong English	Grammatical complexity requires more time and cognitive load in planning, which may add to the sources of difficulty in speech production and perception for ESL (English-as-a-second-language) speakers in Hong Kong.
2	Online resources to showcase research collaboration and findings arising from the project	Online practices and tips for pronunciation and verbal communication targeting ESL speakers in Hong Kong.
3	Summary of training pedagogy relating workplace English in two universities of Hong Kong	Implications for curriculum design and planning at tertiary levels involving teaching ESL workplace communication
4	Survey of workplace English in Hong Kong among nursing students and interns	- Linguistic background and preparedness of nursing majors in Hong Kong. - Need assessment of ESL pedagogical materials and students' expectation for career preparation.
5	Speech production in noise by HK ESL speakers	- ESL production of lexical stresses was comparable as that of L1 speakers in quiet, but varied significantly under noise conditions. - Output was accepted in two international conferences: one in HK and the other to be held in Spain.
b	**Proposed changes and outcome of the last stage**	
1	Phonetic analysis of general speech in quiet.	
2	Emotional ESL pronunciation: speech production and perception in workplaces.	
3	Cantonese ESL pronunciation in quiet and in noise.	
4	Intelligibility of Cantonese ESL pronunciation in quiet and in noise.	
5	Outputs: - Dataset of English and Cantonese Lombard Speech - Publication: IPA guide for groups with special needs	

(e) Review of literature of the project

Our studies in previous stages, as presented in progress reports, revealed that nurses in emergency rooms and hospitals are constantly working under noises including alarms and babbles. Such adverse conditions trigger anxiety and degrade speech signals, which are detrimental to oral communication especially that involving non-L1 languages. Literature has established that human vocal system remains dynamic and adaptive under noise, e.g. amplitude involuntarily increases to compensate signal degradation. This phenomenon is called the Lombard effects (Lombard, 1911). Previous studies have confirmed the Lombard effects across languages in first-language (L1) communication. However, relatively fewer studies have devoted to non-L1 speakers' adjustment of speech in response to background noises. The current study examined effects of noise on speech production in L1 and second languages (L2) by focusing on production of segments and suprasegments in English by Cantonese ESL speakers.

People tend to involuntarily increase vocal efforts when speaking in adverse conditions such as under noise. The Lombard effect as such defined in early 20^{th} century has been examined and found in languages such as English (Junqua, 1996), French (Garnier, Henrich, & Dubois, 2010), Spanish (Castellanos, Jose-Miguel, & Casacuberta, 1996), Japanese (Takizawa & Hamada, 1990) and so on. Speech under noises showing Lombard effects include an increase in vocal amplitude in response to an increase in background noise, which was originally interpreted as an automatic increase of voice intensity for the benefit of auditory feedback of speakers' own voice. However, speakers are not always aware of this change. Recent research reports higher cortical areas are involved in Lombard effects and that the essential circuits responsible are located in the brain stem, particularly, in the region for the integration of vocal production and auditory perception (Zollinger & Brumm, 2011).

Phonetic modifications triggered by background noise are results of attempts to facilitate speech communication in adverse conditions, where speakers employ measures available to enhance clarity and speech intelligibility. Much less is known, however, whether and how non-L1 speech shows similar effect and modifications, L2 speakers who are limited by their linguistic experience may have more difficulty in identifying L2 speech targets in noisy conditions than L1 listeners (Li, Wayland, & Zhang, 2010). Among the limited studies in this aspect, Marcoux and Ernestus (2019) reported the non-L1 Lombard speech evidenced by changes to F0 range of L1 speakers of Dutch when speaking English in noisy conditions. So, previous studies warrant an investigation of the impact of noise on speech production and coping strategies if any by Chinese ESL speakers.

Acoustically, Lombard effect is accompanied not only by an increase in vocal amplitude, but also by modifications to other phonetic cues. For instance, fundamental frequency and vowel duration present in Lombard speech tend to increase. In addition, Energy shift from low frequency bands to higher bands. For the first formant (F1) of vowel, the frequency centre of the first formant shifts to higher frequency. However, the degree to which these acoustic cues are modified depends on the phonetic contexts. Previous studies show that acoustic adaptation in noisy conditions favours vowels rather than consonants in general. Junqua (1996) found that vowels are more likely to be elongated, while consonants tend to be slightly shortened. For intensity and spectral energy, even both vowel and consonant were increased, vowel intensity increased more in noisy conditions than did consonant intensity. But spectral energy did not significantly increase more in noise for vowels than for consonants (Garnier, Bailly, Dohen, Welby, & Lœvenbruck, 2006). In addition, Junqua (1996) found that the first formant frequency (F1) for vowels, glides, liquids, and nasals is more likely to increase. This tendency is more important for female speakers than for male speakers. However, second formant frequency (F2) tends to increase for the female speakers only. The author also noticed that voiceless consonants (i.e. /p, t, f/) are more likely to be deleted in coda position. In addition, experimental condition/setup could affect the Lombard speech. Garnier et al. (2010) found that phonetic modifications (intensity, F0, vowel duration, and centroid (Hz)) in Lombard speech were greater when the noise was played in headphones than over loudspeakers. While speech with interaction (e.g. conversation) was more modified in noise than that without interaction (e.g. speakers describe their actions aloud in the noisy environment).

All these acoustic modifications can be interpreted (1) as the consequences of automatic regulation of vocal intensity from the attenuated auditory feedback that a speaker gets from his/her own voice (Egan, 1972; Lombard, 1911; Tonkinson, 1994; Tourville, Reilly, & Guenther, 2008) and/or (2) as the expression of a listener-oriented adaptation that aims to maintain an acceptable level of speech intelligibility in the acoustic domain (Junqua, Fincke, & Field, 1999; Lane & Tranel, 1971).

Lombard speech has been examined and found in a few languages, such as English (Junqua, 1996; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988), French (Garnier et al., 2006, 2010), (Spanish (Castellanos et al., 1996), Japanese (Suzuki, Nakqjima, & Abe, 1994; Takizawa & Hamada, 1990), Cantonese (Zhao & Jurafsky, 2009), etc.. For all the languages, there are global adjustments in Lombard speech, including fundamental frequency, amplitude, duration, frequency formants and spectral tilt. However, a few extended tendencies are found since there are many degrees of freedom in these studies, due to, for instance, gender of participants, degree of noise, type of noise (white noise vs. babble noise), etc. For instance, Castellanos and colleagues (1996) show that significant differences between male and female speakers for the F0 and the spectral energy distribution. Garnier et al. (2006) found that the speech adjustment was greater in white noise than in cocktail party noise.

Acoustic characteristics of Lombard speech are language specific. Zhao and Jurafsky (2009) investigated Cantonese lexical tones in the presence of noise and found all tones in Cantonese are produced with higher F0. Arciuli and colleagues (2014) examined English lexical stresses in noise and found that English speakers did not increase vocal intensity equally across syllables but altered degree of contrast between strong and weak syllables in Lombard speech. In the production of lexical stresses, both intensity and fundamental frequency showed greater differences between strong and weak syllables in noise than in quiet.

The phonetic modifications triggered by background noise are interpreted as attempts to facilitate speech communication in adverse conditions, where speakers employ measures available to enhance the intelligibility of speech in their L1 language. Much less is known, however, whether and how non-L1 speech shows such effect and modification, as second-language (L2) speakers who are limited by their linguistic experience may have more difficulty in identifying L2 speech targets in noisy conditions than L1 listeners (B. Li et al., 2010; Mayo, Florentine, & Buus, 1997; Rogers, Lister, Febo, Besing, & Abrams, 2006). Cooke and Lecumberri (2012) examined the Lombard speech in noise and in quiet for Spanish learners of English and found that non-L1 listeners gained a large benefit from Lombard speech in noise, even though less than that displayed by L1 listeners. More recently, Marcoux and Ernestus (2019) reported the non-L1 Lombard speech in increased F0 ranges among Dutch-English bilinguals.

English and Dutch differ in their median pitch, where English has a wider pitch range than Dutch. By comparing the median pitch of the two languages in quiet and in noise, the authors noticed Dutch speakers modulated their median pitch in noise in L1 and in their ESL production in tasks involving question-answer sentence pairs in quiet and under noise (through headphones with babble noise at 83 dB SPL): an early-focus pair and a late-focus pair, with reference to production by L1 speakers of American English. Results showed pitch increase in late-focus sentences in noise for Dutch speakers' Lombard speech in both L1 and L2 similar as in AE production. But for the early-focus sentences, only Dutch speakers showed an increase in pitch in noise in both L1 and L2, but AE speakers did not. The difference was interpreted as L1 effect on L2 Lombard speech.

Another speech style induced by noisy conditions is "clear speech" in which people manage to increase speech intelligibility by altering global phonetic cues resulting in slower speech rate, higher F0, larger F0 range. Clear speech may also be achieved by enhancing the acoustic distance between contrast categories, namely, "linguistic enhancement" based on phonological features, for instance, less vowel reduction to enhance contrast between individual vowels (Johnson, Flemming, & Wright, 1993; Moon & Lindblom, 1994; Picheney, Durlach, & Braida, 1986). Bradlow and Bent (2002) found that non-L1 speakers who have less experience with the sound structure of the target language gained a very limited clear speech benefit from L1 clear speech linguistic enhancements. However, Smiljanic & Bradlow (2010) showed that fluent non-L1 talkers benefited from clear speech modifications by L1 speakers. Meanwhile, fluent non-L1 talkers can alter phonetic cues to increase speech intelligibility for listeners. These two studies suggest less fluent ESL speakers can benefit from training on improving in L2 processing on global acoustics modifications and language-specific enhancements.

(f) Theoretical and/or conceptual framework of the project

Our study involves ESL speakers from Hong Kong who speak Cantonese as their first language, typologically different from English. Cantonese is tonal that employs the fundamental frequency (F0) as the primary acoustic cue to tonal contrast. There are six distinct lexical tones based on F0 contrast alone: T1 (55, high–level), T2 (25, high–rising), T3 (33, mid–level), T4 (21, low–falling), T5 (23, low–rising), and T6 (22, low–level) (Bauer & Benedict, 1997; Fok-Chan, 1974). The numbers such as 55, represent the relative values of F0 at onset and offset of each tone, with 5 being the highest and 1 being the lowest F0 of the pitch range (Chao, 1930, 1947). Based on their mean F0 range, these tones are classified into two registers: T1, T2, and T3 in high register, while T4, T5, and T6 in low register. In world tonal languages, in addition to F0, other phonetic parameters may cue tonal differences, for example duration and intensity. However, these cues are often secondary compared with the role of F0 modulation. English uses F0 modulation too, but for lexical stresses, the realization of which employs also variations to duration and intensity (Fry, 1958). F0 is found the strongest cue in perceiving English lexical stress, followed by syllable duration and intensity (Bolinger, 1965; Fry, 1958). Both lexical tones and lexical stress may contribute to distinguishing minimal pairs. For instance, in Cantonese, /ma/ in a high-level tone T1 means ‘mother,’ while in a high rising tone T2 /ma/ means ‘horse.’ In English, the words “CONtent” (strong-weak syllables) and “conTENT” (weak-strong syllables) varying by stress placement results in different parts of speech and lexical meanings.

Much research has been devoted on ESL acquisition of suprasegments by learners whose L1 rhythmic mechanisms differ. Cantonese learners are known to have difficulties in mastering English lexical stresses. It is worth noting that L1 speakers of English comprehended better if the speech was produced in L1-like prominent stress patterns compared to those lacing such stress patterns (Hahn, 2004 as cited in Theodoropulos, 2014). Though there is few research on the suprasegmental characteristics of Hong Kong English at workplace, a majority of research on general speech production has demonstrated the influence of Cantonese tones in the perception and production of English lexical stresses (Choi, Tong, & Singh, 2017; Hung, 2012; Luke, 2008; Tong, Lee, Lee, & Burnham, 2015). Tong et al. (2015) tested the perception of Cantonese tones and English stress patterns by Cantonese-English bilingual children, adults, and English monolingual adults. They claimed that tone and stress are processed in comparable patterns by Cantonese-English bilingual children and adults, at least within the stage of tone/stress shared acoustic cues (i.e. average F0 and F0 onset). In a later study, Choi, Tong and Singh (2017) found that Cantonese lexical tone sensitivity contributed to English lexical stress sensitivity among Cantonese children who learned English as a second language, where underlying L1 lexical tone perception was recruited in service of L2 English lexical stress perception. They also tested the roles of general auditory sensitivity and working memory in L2 sensitivity. Results revealed that the general auditory sensitivity, which is a lower auditory level of perceptual sensitivity to acoustic-phonetic variation, have an impact on the perception of L2 English lexical stress, but not for the work memory. These results are consistent with the previous findings of lexical prosodic transfer from L1 lexical tone to L2 lexical stress (Nguyễn, Ingram, & Pensalfini, 2008; Q. Wang, 2008; Yu & Andruski, 2010; Zhang et al., 2008). For production, there is a ”tonalization” of English lexical stress, in which Luke (2008) showed three pitch profiles of HK English: L, H and L!, where L is a Low level tone, H is a High level tone and L! is a Mid-Low Falling tone. Luke argued that perceptually prominent syllables are assigned a high tone by Cantonese speakers. While the preceding syllables are assigned low tones, and all following syllables assigned mid-low falling tones. For instance, *consider* is LHL!, *physical* is HL!L!, and *encyclopaedia* is LLLHL!L! in Hong Kong English.

(g) Methodology

With many studies on Lombard effects in L1 communication and general ESL characteristics, much fewer studies focus on L2 Lombard speech, especially by speakers of tonal languages such as Cantonese. Both theoretical and practical significance of such studies is discussed in previous sections. Therefore, the current research aimed to investigate universality and embodiment of the Lombard effects by comparing L1 and non-L1 speech production under noise. To achieve the research goals, we examined the Lombard effects in Cantonese ESL speech under noise with a reference to their L1, in terms of phonetic realizations of lexical stresses. These studies were designed to answer research questions below.

- Do Cantonese ESL speakers modify their English speech, especially lexical stresses, in noise condition? If yes, how?
- Do they employ strategies, as similar as those by first-language (L1) speakers of English?
- Are there other factors, such as L1 and task type, influencing Cantonese ESL Lombard speech?

In-depth understanding of Lombard speech in ESL communication requires references to pattern variability in speech production under normal situations. As reported in the application for changes, in addition to our own recording of ESL speakers in local workplaces under different environment conditions, we identified an online corpus, ALLSSTAR (Bradlow, n.d.), which contains ESL speech in various styles including read passages and storytelling by ESL speakers from different language backgrounds, which added significantly to the variability of speech data for analysis of phonetic and phonological patterns in ESL speech. Extracting speech recording and transcription from ALLSSTAR, we first carried out analysis of segments and suprasegments in non-L1 speech production with a comparison with those in L1 speech production (Study 1). Study 1 provided empirical evidence and warranted the follow-up experiments in Study 2 on production of ESL Lombard speech, Study 3 on perception of ESL Lombard speech, and Study 4 on affective analysis of ESL speech.

Study 1. Phonetic characteristics of ESL speech in quiet.

<u>Part 1: Production of vowels</u>

This part examined production of speech sounds in English with a focus on tense and lax vowels produced by L1 and non-L1 speakers. The vowel inventories in English differ from those in Cantonese and Mandarin in number and category, which provides a good scenario to assess the role of L1 language on ESL speech production. The speech data were based on recordings of “The North Wind and the Sun” retrieved from the ALLSTAR corpus (Bradlow, n.d.).

<u>Method</u>

<u>*Participants and speech materials*</u>

The recordings were speech production from 17 L1 speakers of English, 14 Cantonese ESL speakers and 14 Mandarin ESL speakers. Four tense vowels /a:, i:, u:, ɔ:/ and six lax vowels /æ, ʌ, ə, e, ɪ, ʊ/ were selected for analysis (Table 1). In total, 1555 tokens were transcribed and measured. Phonetic data such as correlates of formants were extracted from these tokens. Vowel spaces were plotted using values of the first and second formants (F1 and F2) of a vowel. Duration of vowels was also measured to compare temporal features of lax and tense vowels in L1 and ESL production.

Table g-1. List of target vowels and their phonetic contexts. (The targets are in bold fonts.)

Tense Vowels	Words	Syllable structure	Lax Vowels	Words	Syllable structure
/i/	b**e** agr**ee**d succ**ee**ded imm**e**diately	CV CCVC CVC CV	/ɪ/	w**i**nd wh**i**ch cons**i**dered	CVCC CVC CV
/ɑ/	h**a**rd	CVCC	/e/	wh**e**n th**e**n att**e**mpt conf**e**ss	CVC CVC CVCCC CVC
/ɔ/	n**o**rth w**a**rm w**a**rmly m**o**re al**o**ng str**o**nger	CVCC CVC CVC CV CVC CCCVC	/æ/	tr**a**veller wr**a**pped th**a**t l**a**st	CCV CVCC CVC CVCC
/u/	wh**o** bl**ew** tw**o**	CV CCV CV	/ʊ/	sh**ou**ld c**ou**ld t**oo**k	CVC CVC CVC
			/ə/	s**u**cceeded c**o**nsidered c**o**nfess	CVC CVC CVC
			/ʌ/	s**u**n	CVC

Part 2: Production of consonants

Chinese phonotactics is different from that in English, for example, consonant cluster is not allowed at syllable-final positions in Chinese and coda consonants are not as variated either. Due to such influence of L1 phonotactics, it is often found that Chinese ESL speakers may delete coda or insert vowels following coda consonant. In this part, we analysed production of English consonants by Chinese ESL speakers, with a focus on coda consonants. The data is from recordings of ALLSSTAR corpus (Bradlow, n.d.). Our analysis reveals that Mandarin ESL speakers modified syllable structures by deleting the coda consonant or by adding an epenthetic vowel after the consonant, the former of which is more frequently used. Moreover, phonetic contexts (i.e. followed by a vowel, or a consonant, or at word boundary) play an important role in phonological processes involving coda consonant in ESL speech production. Cantonese ESL speakers on the other hand seemed to benefit from their L1 phonotactics which allows non-nasal coda. Cantonese speakers did not resort to vowel epenthesis when faced with coda consonants. Instead, they often missed the coda consonants especially the plosives.

Mandarin Chinese does not allow non-nasal at syllable-final positions, or consonant clusters at syllable-initial or final positions. The maximal syllable structure allowed in Mandarin is CGVX, where C is a consonant, G a prenucleus glide, V a vowel, and X is either a nasal (N) or an off-glide (the off-glide is considered as the second vowel of a diphthong in (Duanmu, 2007). Thus, if L1 phonotactic knowledge is affecting production of English coda consonant, such as *should*, *take*, then Mandarin ESL speakers are expected to modify a seemingly illicit syllable structure by deleting a coda consonant or by inserting a vowel after the consonant.

Method

Participants

We selected from the ALLSTAR corpus data from 14 Mandarin ESL speakers whose average age was 23 years old (21-25 years) and mean Versant scores were 59 (B1-C2 range of The Common European Framework of Reference for Languages (CEFR)). Data from fourteen Cantonese ESL speakers were selected. Their average age was comparable as the Mandarin peers, at 22 years (19-27 years), but no Versant scores were reported for Cantonese group. For the control group, recordings from 15 L1 speakers of English were analysed. The L1 speakers' average age was 29 years old (18-26 yrs).

Speech materials

We selected one of the speech recordings, which is a widely used passage in phonetic analysis, i.e. "The North Wind and the Sun". The full list of target words contained coda stops /d, k, t/ and onsets of the following syllables listed in Table 2. The preceding phonemes were vowels, including /i, ɪ, ʊ, eɪ, aʊ, oʊ/.

Table g-2. Target elements and structures. (#: word boundary; -0, -1, -2: trial repetitions).

Coda	Onsets of the following syllable	Words
t	C (w)	Out
d	C (b)	should
	C (ð)	agreed
	V (i)	succeeded
	#	could
k	C (h)	take
	V (ɑ)	cloak-1
	V (ə)	cloak-2
	V (ɑ)	took
	#	cloak-0
	#	cloak-3

Part 3: Production of lexical stresses

This part examines ESL production with a focus on suprasegmental features such as lexical stresses. The examination included phonetic realization of key correlates of stresses by ESL speakers and L1 speakers of English.

In English, fundamental frequency (F0), duration and intensity are primary cues to the realization of lexical stresses. For tonal languages such as Cantonese and Mandarin, correlates of F0 such as its contour and height are cues to differentiate lexical tones. Mandarin uses durational difference as secondary cues to neutral tones. Previous studies show that Mandarin speakers outperform Cantonese speakers and Taiwan Mandarin speakers in perception and production of English lexical stresses (Li & Grigos, 2018; Qin, Chien, & Tremblay, 2017). This advantage has been attributed to neutral tone in Mandarin. In our study, we analysed production of disyllabic words varying stress assignment. Our findings indicate a greater difference between stressed and unstressed syllables produced by L1 speakers than among ESL speakers.

Study 2. ESL speech production in simulated adverse conditions at workplaces

Results from Study 1 provided further warrant and evidence to our study design targeting at ESL Lombard speech (Studies 2, 3, 4). Study 2 containing two experiments. Experiment 1 uses a reading task in English to explore whether and to what extend ESL speakers modify their speech in noise to maintain a speech-to-noise ratio favourable for their auditory feedback. Experiment 2 adopts a reading task in Cantonese to investigate whether and to what extend Cantonese speakers modify their L1 in noise. Experiment 3 employs a conversation task to test the influence of spontaneous speech in L2 Lombard speech. ESL speakers need maintain a speech-to-noise ration favourable for auditory feedback of their own voice, and also for intelligible communication with other interlocutors.

Methods:

Experiment 1 Reading in English in quiet and noisy conditions

Participants

Eighteen L1 speakers of Hong Kong Cantonese participated in the current study (8 females and 10 males, average age: 25.6). All participants are fluent ESL speakers. Nine were undergraduate students who majored in nursing, business, engineering or liberal arts. They all had internship experiences and prior working experiences. The rest participants are full-time employees in healthcare sectors. No participant reported having known neurological, speech or language disorders. In order to further characterize the L2 English participants' experience and proficiency in Mandarin and English, each person completed a language experience and proficiency questionnaire (Appendix 1). The institutional ethics committee approved all experimental procedures and all participants gave informed consent before taking part in experiments. They were remunerated for their time in participating in the study. Six L1 speakers of English were recruited as control group in Experiment 1 (1 female and 5 male, average age: 28).

Speech materials

Trisyllabic English words were selected where the first two syllables vary in the location of stress assignment: strong-weak vs. weak-strong. Complex syllable structure with consonant clusters is not included in the stimuli due to potential phonological processes such as consonant deletion or substitution commonly found among Cantonese ESL speech. Our choice of speech materials for recording took consideration of common words in the workplace for medical and financial professionals, which had been chosen from our previous survey of students and pedagogical materials. Twelve fillers with one or multiple syllables were added to the list totalling thirty-three words. Target words in strong-weak stress patterns were: Salary, Memory, Medicine, Surgery, Manager, Pharmacy, Summary, Vitamin, Therapy, Minimal, Physical, Revenue, Agency. The words in weak-strong patterns were: solution, location, relation, reporter, emotion, infection, referral, insurance, revival, addition. Fillers were: fee, ill, sale, sick, immunity, information, average, recovery, association, application. All targets and fillers were produced twice in quiet and noise conditions, yielding 152 trials for each participant and 3,648 tokens in total for phonetic transcription and analysis.

Procedure

Participants were instructed to wear earphones and to read aloud the English words in the following target conditions: in multi-talker babble noise vs. in a quiet environment. The two conditions were randomly presented to the participants. A pretest was conducted to verify the participants' knowledge of the words and their ability to place the lexical stress correctly. In this block, participants were asked to read the word list and to report if they have any doubt or questions about the words.

In the reading blocks, a multi-talker babble noise was played through headphones. Target participants were asked to sit comfortably and try to stay still in front of a computer screen in a sound-attenuated booth. Participants were introduced to read aloud the target words, which were embedded in a carrier sentence as follows: *Let me show you ______ slowly*. All sentences were presented on a computer screen one at a time using Psychopy 3. Each single sentence appeared randomly twice in total for each condition (quiet and noise) over the entire recording session, with an inter-stimulus interval of one second. The participants' reading was directly recorded onto Adobe Audition with a sampling rate of 44.1 kHz and 16 bits using a high-quality microphone connected to a laptop. The distance between the microphone and the corner of a speaker's mouth is about 3 cm.

Experiment condition: Noise settings

The multi-talker noise came from the Freesound database (https://freesound.org/) and was composed of unintelligible mixed voices, which involved background sounds in hospitals and restaurants in Hong Kong, Barcelona, England. The spectral energy concentrated below 1000Hz, see Figure 14.

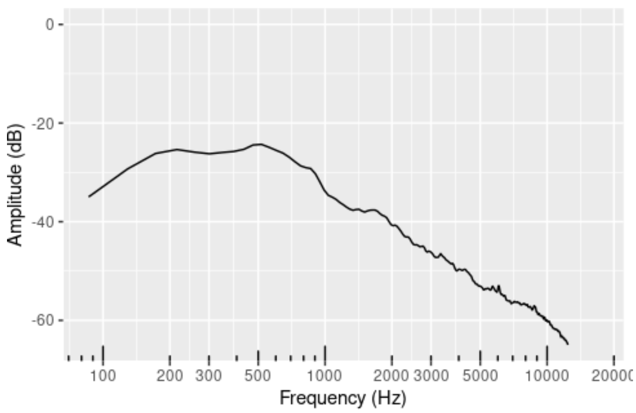


Figure g-1: Spectrum profiles of the multi-babble noise used in the experiments.

As closed headphones can affect internal and external hearing, independently from any noise played into them (Garnier et al., 2010), the level of the participant's own voice played into his headphones was calibrated so that their perception of talking with and without the headphones would be comparable (Alghamdi et al. 2018). During the calibration procedure, headphone over one ear was wore while the other one remained uncovered. A speaker was requested to speak, and her voice was presented through headphones. The voice loudness level was adjusted gradually to balance auditory feedback into both ears. The noise output levels (in dB SPL) at the speaker's ears through headphones were adjusted to 78 dB SPL by using an artificial ear (AWA6163 Ear Simulator artificial ear). The microphone signal was measured by a sound level meter (RS PRO Digital Sound Level Meter 146-4651 8k Hz).

<u>Experiment 2 Reading in Cantonese in quiet and in noisy conditions</u>

The second experiment aims to investigate the effect of L1 language knowledge on the production of English lexical stress in quiet and in noise. To this end, twenty trisyllablic Cantonese words varying in lexical tones were used to compare with the production of lexical stresses in both conditions. As mentioned above, Cantonese tonal system has an influence on the production of English words. Previous studies (Cheung, 2009; Gussenhoven, 2012, 2014; Luke, 2008; Wee, 2016) argue that perceptually prominent syllables are assigned T1 (55) by Cantonese speakers, while the preceding syllables are assigned T6 (22), and all following syllables assigned T4 (21).

<u>Participants</u>

The same group of L1 speakers of Cantonese in Experiment 1 were invited to Experiment 2.

<u>Speech materials</u>

In the present study, the target words consisted of ten trisyllabic words with T1-T4 tones in the beginning and the other ten with T6-T1 tones to compare the target words with strong-weak and weak-strong contrast stresses in Experiment 1, as listed in Table g-3. Additionally, ten trisyllable Cantonese words with three T1 were included as fillers. Altogether, 120 productions per participant and a total of 2160 productions were obtained.

Table g-3. Cantonese target words varying tones in the first two syllables.

T1-T4 (55-21 tones)		T6-T1 (22-55 tones)	
faa1 kei4 sam1	花旗參	hau6 saang1 zai2	後生仔
saa1 tin4 jau2	沙田柚	ngoi6 sang1 neoi5*2	外甥女
si1 man4 jan4	斯文人	sau6 sing1 gung1	壽星公
so1 fu4 lei4	梳乎厘	se6 lei1 ngaan5	射喱眼
hoi1 fung4 ce1	開篷車	si6 baa1 naa2	士巴拿
ziu1 paai4 fo3	招牌貨	si6 do1 fong4*2	士多房
maau1 tau4 jing1	貓頭鷹	fung6 sin1 faa1	鳳仙花
ziu1 tau4 zou2	朝頭早	zoeng6 gan1 gwu1	橡筋箍
sin1 jan4 zoeng2	仙人掌	je6 maa1 maa1	夜麻麻
cung1 loeng4 fong4*2	沖涼房	long6 saam1 gaa3	晾衫架

<u>Procedure</u>

This experiment was conducted after Experiment 1 and the same procedure was used. The target words were embedded in a carrier sentence in Cantonese as follows: "我講_____ 三次" (/ŋɔ: kɔ:ŋ tsɔ: ... sa:m ts^{h}i:/, *I say*_____*three times*).

<u>Prediction</u>

It is predicted that L1 speakers of Cantonese will modify the production of Cantonese sentences in noise with higher F0 on each syllable.

<u>Experiment 3 Verbal interaction in English in quiet and noisy conditions</u>

Participants

Same groups of ENG and Cantonese ESL speakers as in previous experiments.

Speech materials

Eight target words were embedded in scenarios and are listed in Table g-4. Altogether, 16 trials per participant, totalling 288 utterances were obtained. More details about the face-to-face conversation are described in Appendix III.

Table g-4. List of target words by stress patterns. Primary stresses are capitalized.

Strong-Weak	Weak-Strong
THErapy	reCEPtion
PHARmarcy	reFErral
PHYsical	inFECtion
SURgery	
VItamin	

Procedure

Participants were instructed to make conversations on the following topics: give directions and explain a referral letter with a partner seated 2 meters in front of them. After the experiment, participants were asked to evaluate on their perception in noise by answering several questions shown in Appendix III (see also, Garnier et al., 2010).

In addition to the multi-talker noise, the auditory feedback of participant's own voice and the voice of his/her partner were played into the headphones using Adobe Audition. Similarly, the partner can hear her own voice and the voice of the participant through headphones. The voice levels were calibrated as mentioned in Experiment 1. It is predicted that Lombard effect, which has been found in reading tasks, will be amplified in interactive conversation.

Study 3: Awareness of speech modifications in noise.

In adverse conditions such as multi-talker babble noise, speakers tend to modify their speech. Such modification is more often interpreted as an automatic increase of voice intensity for the benefit of auditory feedback of speakers' own voice as well as attempts to facilitate speech communication in noise (Bosker & Cooke, 2020). Previous studies show that language learning experience plays a role in identifying non-L1 speech targets in noise. In particular, it is more difficult to identify non-L1 speech in noise for non-L1 listeners than L1 listeners (Van Engen, 2010). Study 3 intended to examine whether modification strategies by non-L1 speakers of English are effective in enhancing speech intelligibility in noise and whether the perception by L1 speakers of English and Cantonese ESL speakers differ.

Method

Participants

The same group of participants who participated in Experiment 1 of the study 2.

Speech materials

Two sets of questions were made to collect responses from the participants in the face-to-face conversation tasks. The first set of questions was about asking directions in a hospital while the second set was asking information on a referral letter (see Appendix III). The second question in the feedback questionnaire was designed to collect participants' perception of Lombard speech produced by his or her partner and examine whether the partner's modification strategies in noise are effective. The question is provided below (see also Appendix III).

> Question (2): Was your partner's voice audible enough in this
> condition? Did your partner change her voice in this condition?
> If yes, at what moment?

Study 4: Perception of affective states in ESL speech.

Speech communication exchanges messages and emotions. In workplaces, interaction needs be accurate and clear, which requires proper control and expression of emotion. However, expression or cues of emotions may be attenuated or misperceived, especially in non-L1 communication (Bond & Lai, 1986). The objective of this experiment is to investigate the correlation between phonetic characteristics within Cantonese-accented English and emotion perceived by L1 speakers of English.

<u>Method</u>

<u>Participants</u>

A L1 speaker of French (hereafter, French listener) and a L1 speaker of Mandarin (hereafter, Mandarin listener) were asked to listen to the stories produced by the L1 speakers of Cantonese and detect the emotion based on the following three categories: positive emotion, negative emotion, and neutral emotion. They are advanced speakers of English. Speech speeds and proficiency levels are decided by auditory judgement of two experimenters of this study who are also advanced speakers of English.

<u>Speech materials</u>

Utterances were English spontaneous speech recordings of story-telling tasks retrieved from the ALLSTAR corpus (Bradlow, n.d.). In the story-telling task, the participants (13 Cantonese L1 speakers and 24 English L1 speakers) were asked to tell a story based on pictures of cartoon stories. Utterances from two story-telling tasks, Story 1 and Story 2, were used. There are some examples of Story 1 produced by L1 speakers of English.

Speaker 51: *One day, a bird was walking along, and he found a nice big, round top hat…*

Speaker 55: *Olly the odd bird was walking along one day when he found a top hat in his path.*

Speaker 56: *This is a story about Barry the bird. One day Barry's walking along and he comes upon a hat that's lying on the ground.*

Examples of Story 1 from L1 speakers of Cantonese are listed below.

Speaker 74: *There is a bird called Tommy. And he have a hats…*

Speaker 75: *A bird, uh, with the black hair uh, called Peter uh, when he is walking on the uh pathway, he sought a hat…*

Speaker 76: *One day, a bright burrow, uh, tips. He f- he found a has on the grounds…*

We predicted that the different emotions would be detected from these utterances. Next, if the recognition of emotion is universal, French listener's perception would be similar to that of Mandarin listeners. If the recognition of emotion is language-specific, their recognitions would be different.

(h) Data collection and analysis

Study 1. Phonetic characteristics of ESL speech in quiet.

Part 1: Production of vowels

To compare duration across vowel and speaker group, linear mixed-effects models were computed using the lme4 package (Bates, Maechler, & Walker, 2015) in R (R Core Team, 2020). The independent variables were language (English, Mandarin and Cantonese) and vowel (four tense vowels and six lax vowels). For random effect, we included random intercepts for participants and words. In the process of model selection, we used backwards elimination of non-significant effects. The full model included language and vowel as fixed factors, as well as the interaction between them. The second model excluded the interaction language * vowel from the full model. Both Bayesian Information Criterion (BIC) values (Schwarz, 1978) and Akaike Information Criterion (AIC) values (Akaike, 1987) were used to assess model fit. *P*-values were generated using likelihood ratio tests and Tukey's HSD post-hoc analyses using the emmeans package (Lenth, 2016).

For F1 and F2, multivariate regression models were performed using the MCMCglmm package (Hadfield, 2010) in R. The independent variables in this model were language (English, Mandarin and Cantonese) and gender (male and female). Moreover, the interaction between language and gender was included as an independent variable. Gender was included as an independent variable, since gender differences in the formant frequency values of adults are well established (Fant, 1960; Peterson & Barney, 1952). For contrast coding of categorical variables, the 3-level factor, language, was coded so that each level of the factor (Mandarin and Cantonese) was compared to the reference level (English). The binary variable, gender, was coded using sum-to-zero contrast coding such that each factor had a mean of zero. We included random intercepts for participants and words. The critical independent variable in question was language, and for this variable, we included correlated random slopes for participants and words (this quantifies by-participant and by-word variability in the effect of language). We reported 95% highest posterior density (HPD) interval for each coefficient and related *p*-values.

Part 2. Production of consonants.

All selected recordings were manually inspected with careful auditory judgement and visual examination of waveforms and spectrograms in PRAAT (Boersma & Weenink, 2019). Modifications of a coda consonant were coded according to the phonological process, as shown in Table 3. If multiple errors occurred, each error was labelled accordingly. A token was coded as Epenthesis if the post-coda production (after frication or release of a stop) had vocalic part containing visible formants. A token was coded as Deletion if the final consonant was missing. A token was coded as Correct if no error listed in Table h-1 was found. We examined all data for manner and place of articulation as well as voice specifications of the input based on auditory judgement and visual inspection of spectrogram. In total, 473 items were annotated and segmented by three trained phoneticians. Ten percent of the target words were removed prior to further analysis, because they were produced mistakenly as another word (e.g. *coat* for *cloak*), or because the voice quality was poor.

Table h-1. Coding phonological processes in speech data, with 'agreed' ([əˈgriːd]) as examples.

Transcription	**Codes**	**Description**
[əˈgriːd]	Correct	Target is produced fully and correctly.
[əˈgriː]	Deletion	A coda consonant is deleted.
[əˈgriːdV]	Epenthesis	Vocalic part (V) is added after a coda consonant.

Part 3. Production of lexical stresses

The data for this study were based on recordings of "The North Wind and the Sun", taken from the ALLSTAR corpus (Bradlow, n.d.). Recordings were collected from sixteen L1 speakers of English, fourteen L1 speakers of Cantonese and fourteen L1 speakers of Mandarin. Three target words "other", "along", "agreed" were used in this study, where "other" has an initial stress, while "along" and "agreed" are stressed at the second syllable. In total, 260 tokens were analysed for vowel duration, F0 peak, F0 mean, and intensity peak.

The percentage of stressed and unstressed vowels was calculated for duration, intensity and F0 of each pair of vowels within a word using the formula: Percentage of $vowel_1$= $vowel_1$/ ($vowel_1$+ $vowel_2$). If the percentage is higher than 50%, it means the vowel has a larger proportion than the other vowel within a word in terms a particular acoustic correlate.

Study 2. ESL speech production in simulated adverse conditions at workplaces

Experiment 1 Reading in English in quiet and noisy conditions

The recorded data was manually segmented into utterances, words, syllables and phonemes, by auditory judgement and visual examination their waveforms and spectrograms in Praat (Boersma & Weenink, 2019). The dynamic range for spectrogram display in Praat was set to the default value of 70 dB. Mean fundamental frequency (F0) peak, peak F0, peak intensity and duration of the target syllables were measured. The pitch range is set to 75-300 Hz for male and 100-500 Hz for female. Also, one tense vowel tense vowel /ɑ/ (AA1) and four lax vowels /ɪ, e, æ, ə/ were analysed in this study, shown in Table h-2. Formant frequencies (F1 and F2) of vowels were extracted at the midpoint of the vocalic intervals using LPC (Linear Predictive Coding) analysis. The maximum formant frequency is 5500HZ for female speakers and 5000 Hz for male speakers.

Long Term Average Speech Spectra (LTASS) provides a means of viewing the average frequency distribution of the sound energy in a continuous speech sample. The average LTASS on the entire recordings across conditions (noisy vs. quiet) was calculated automatically using a script in Matlab. Pairwise variability index (PVI) was a ratio of phonetic parameters between chosen syllables and has been commonly used in typological analysis of speech prosody (Ballard, Djaja, Arciuli, James, & van Doorn, 2012; Ballard, Robin, McCabe, & McDonald, 2010; Low, Grabe, & Nolan, 2000). We calculated PVI between syllables with primary stress and those that are unstressed, using the formula (1), PVI = 100 × ((dk – dk+1) / ((dk + dk+1)/2)), where d is value of a phonetic parameter (duration, intensity or pitch) of the kth syllable in a word. A higher PVI means the greater contrastivity between chosen stressed and unstressed syllables (Arciuli et al., 2014; Nolan & Asu, 2009). The phonetic parameters measured are syllable duration, peak F0 and peak intensity. Speaking rate was also measured using the formula (2): Speaking rate = number of syllables / total duration of a clause.

Table h-2. List of target vowels and words in Experiment 1.

	Vowels	Code	Words	Syllable structure
Tense	/ɑ/	AA	ph**a**rmacy	CVV
Lax	/ɪ/	IH	ph**y**sical v**i**tamin **i**nfection r**e**ception r**e**ferral vitam**i**n	CV CV VC CV CV CVC
	/e/	EH	rec**e**ption ther**a**py	CVC CV
	/æ/	AE	fin**a**ncial	CVC
	/ə/	AH	pharm**a**cy physic**a**l referr**a**l ther**a**py vit**a**min	CV CV CVC CV CV

Based on the acoustic modifications of speech in noise and similarity of acoustic correlates shared by Cantonese lexical tones and English lexical stresses, the following predictions are made on Cantonese ESL learners' modifications of L2 speech in response to noise condition in Experiment 1.

Prediction 1: There are L2 Lombard effects. Cantonese ESL learners will increase first formant frequency in noise. They will speak slower, louder, with higher F0 and more vocal effort in noise than quiet condition.

Prediction 2: Learners will not enhance the contrastivity of stress-unstressed syllables. As L2 learners may lack experience in dealing with degraded speech signals or self-monitoring under adverse conditions, the patterns of interlanguage or L1 transfer they show in their L2 speech in a quiet condition may not be altered freely under noise

Prediction 3: Learners' phonetic performance will be different from that of L1 speakers in quiet, and would remain different if not more distant in babble noise, because of the acoustic and verbal interference.

Prediction 4: Learners will increase F0 on each syllable constantly from quiet to noise condition due to the influence of their L1 phonological knowledge and phonetic correlates.

<u>Experiment 2 Reading in Cantonese in quiet and in noisy conditions</u>

Participants were instructed to make conversations on the following topics: give directions and explain a referral letter with a partner seated 2 meters in front of them. The partners are L1 speakers of Mandarin and advanced speakers of English. After completing the interactional conversation tasks in noise, the participants were asked to evaluate on their perception in noise by answering the feedback questionnaire.

In addition to the multi-talker noise, the auditory feedback of participant's own voice and the voice of his/her partner were played into the headphones using Adobe Audition. Similarly, the partner can hear her own voice and the voice of the participant through headphones. The voice levels were calibrated as mentioned in Experiment 1 of the study 2.

<u>Experiment 3 Verbal interaction in English in quiet and noisy conditions</u>

Participants were instructed to make conversations on the following topics: give directions and explain a referral letter with a partner seated 2 meters in front of them.

In addition to the multi-talker noise, the auditory feedback of participant's own voice and the voice of his/her partner were played into the headphones using Adobe Audition. Similarly, the partner can hear her own voice and the voice of the participant through headphones. The voice levels were calibrated as mentioned in Experiment 1. It is predicted that Lombard effect, which has been found in reading tasks, will be amplified in interactive conversation.

Study 3: Awareness of speech modifications on perception in noise.

Participants were instructed to listen to conversations conducted under noise. The procedure was the same as in Study 2, on the following topics: give directions and explain a referral letter with a partner. The voice levels were calibrated as mentioned in Experiment 1 of the study 2. The participants were asked to evaluate on their perception in noise by answering questions in a questionnaire, which is provided in Appendix 2 (see also, Garnier et al., 2010).

Study 4: Perception of affective states in ESL speech.

Phonetic correlates of intonation were extracted and measured, as an index to emotional speech. Intonation, as an important cue to emotion in speech, is relatively clear and accessible to speakers with proper training. In this experiment, we conducted a preliminary emotion perception experiment by two listeners to detect the emotion of utterances produced by Cantonese ESL learners. Next, we provided an acoustic analysis of the utterances in ALLSSTAR corpus (Bradlow, n.d.) to examine whether intonation is a cue to emotion in speech.

(i) Results and Discussion

Study 1. Phonetic characteristics of ESL speech in quiet.

<u>Part 1: Production of vowels</u>

Tense vowels: formant frequencies and vowel normalization

After the measurement of F1 and F2, we performed vowel normalization to minimize cross-speaker variations in phonetic analysis (Thomas, 2002). We selected bark scale for this method is a psychoacoustic calibre that fits human sensitivity and perception of frequencies (Adank, Smits, & van Hout, 2004). Converting formant frequencies (Hz) into bark scales minimizes speaker effects and also reduces arithmetic values from above 1000 to lower than 16, which simplifies calculation and comparison. The formula we used for bark conversion is: Bi = [26.81/(1+1960/Fi)] - 0.53 (Bi = B1 or B2; Fi = F1 or F2, where Fi is formant frequency and Bi is bark value (Traunmüller, 1990, 1997). Vowel spaces were then plotted for each vowel in L1 and ESL speaker groups based on the bark values generated in vowel normalization (Figure i-1).

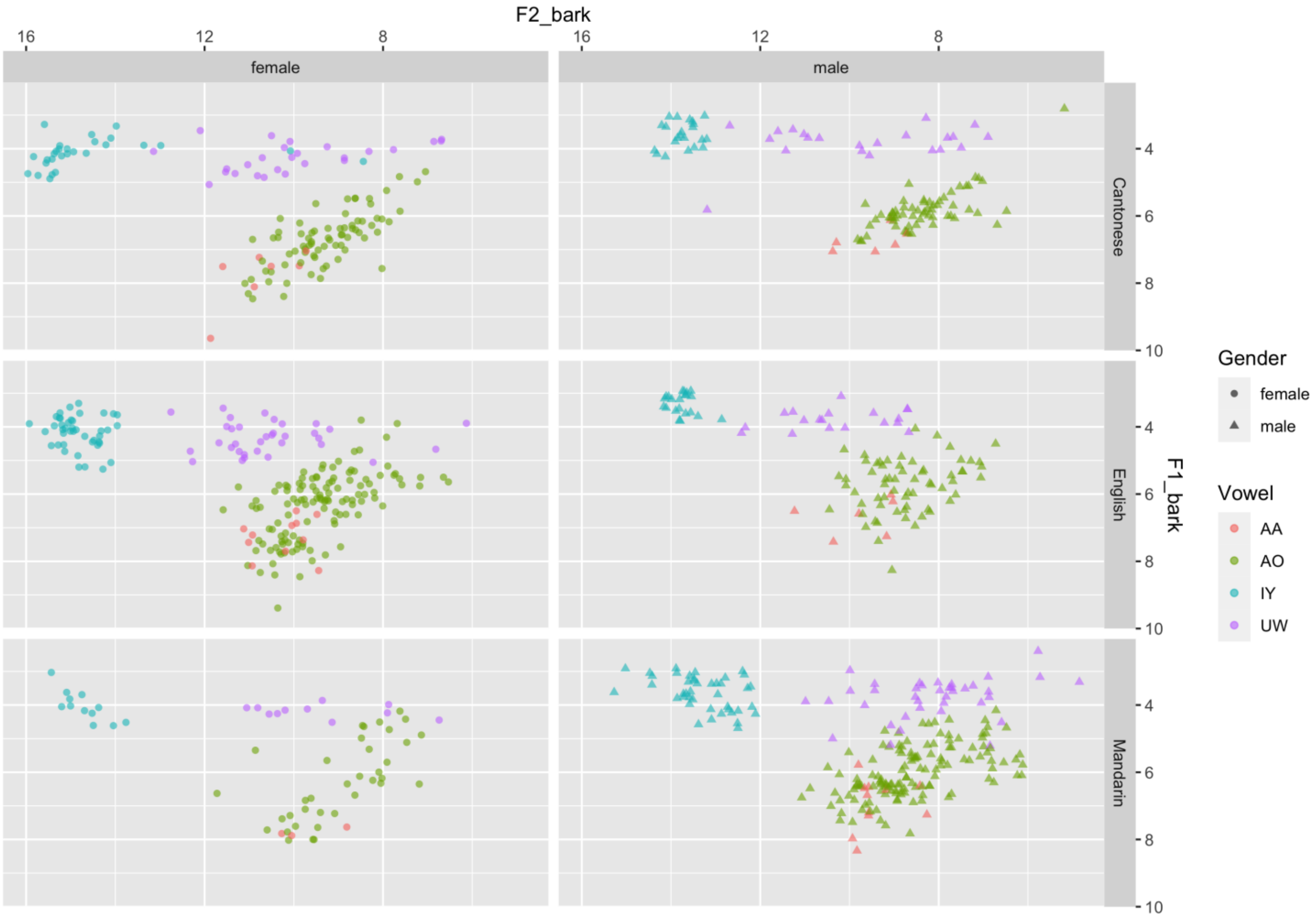


Figure i-1. Vowel spaces of /ɑ, i, u, ɔ/ (AA, IY, UW, AO) produced by L1 and ESL speakers.

The multivariate regression model shows that Mandarin and Cantonese ESL speakers produced tense vowels in a similar way as L1 speakers of English (ENG) did. Only the F2 values of /ɔ:/ produced by male Cantonese ESL speakers is found significantly lower than those by male NSEs (F1*Cantonese*male = -1.51, 95% HPD [-2.89, -0.35], p = 0.027). As expected, there was gender effect on some formant values, where male speakers had lower formants in general than female speakers. That is, male speakers' F1 of /a:/ was lower (F1* male = -0.62, 95% HPD[-1.19 -0.05], p = 0.044), and so were their F1 and F2 of /i:/ (F1* male = -0.84, 95% HPD [-1.22 -0.46], p < 0.001, F2* male = -1.05, 95% HPD[-1.67 -0.49], p < 0.001, respectively).

Tense vowels: Duration

Generally speaking, duration of the four tense vowels /a:, i:, u:, ɔ:/ by L1 speakers of English (ENG) is significantly shorter than that by ESL group, especially that by Cantonese ESL speakers (Figure i-2). The statistic model with L1 language, vowel type and their interaction as fixed factors proved the best model to interpret the duration data with a lower AIC value. This is indicated both by a chi-squared test of log-likelihood ratios (($\chi^2(20) = 25.466$, $p <0.001$). There was effect of the interaction between L1 language and vowel type on the duration of these vowels. Tukey's HSD post-hoc tests further showed that the duration of /a:, ɔ:/ for Cantonese ESL speakers is significantly different from that by ENG speakers (both $p < 0.0001$), the latter of which produced significantly shorter /a:, ɔ:/. However, there is no difference both between Mandarin and Cantonese speakers, or between Mandarin and ENG Speakers (/a:/: *p*-values of Mandarin - Cantonese and Mandarin - English are > 0.0001; /ɔ:/: *p*-values of Mandarin - Cantonese and Mandarin - English are > 0.0001). As for /u:/, ESL speakers produced it with longer duration than ENG speakers did. The statistical analysis shows that ESL production is in general different from ENG (*p*-values both of Cantonese - English and Mandarin - English are < 0.0001). The difference between Cantonese and Mandarin speakers is not significant ($p > 0.0001$). For the duration of /i:/, production by ENG and ESL groups was comparable (all *p*-values are > 0.0001).

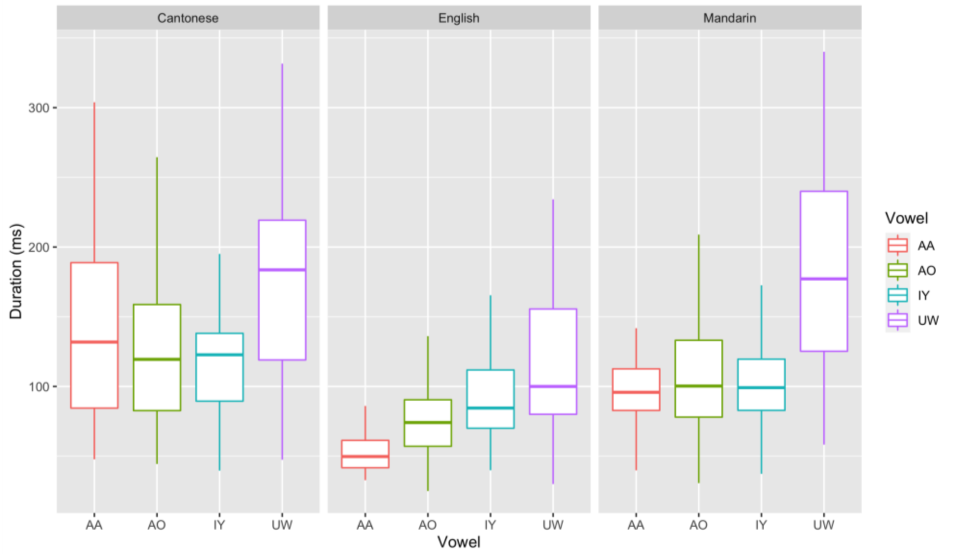


Figure i-2. Duration of /a, i, u, ɔ/ (AA, IY, UW, AO) produced by ENG and ESL groups.

Lax vowels: Formant frequencies and vowel normalization.

The same procedure was performance as for the tense vowels. Vowel spaces were then plotted based on bark values yielded from the normalization as shown in Figure i-3.

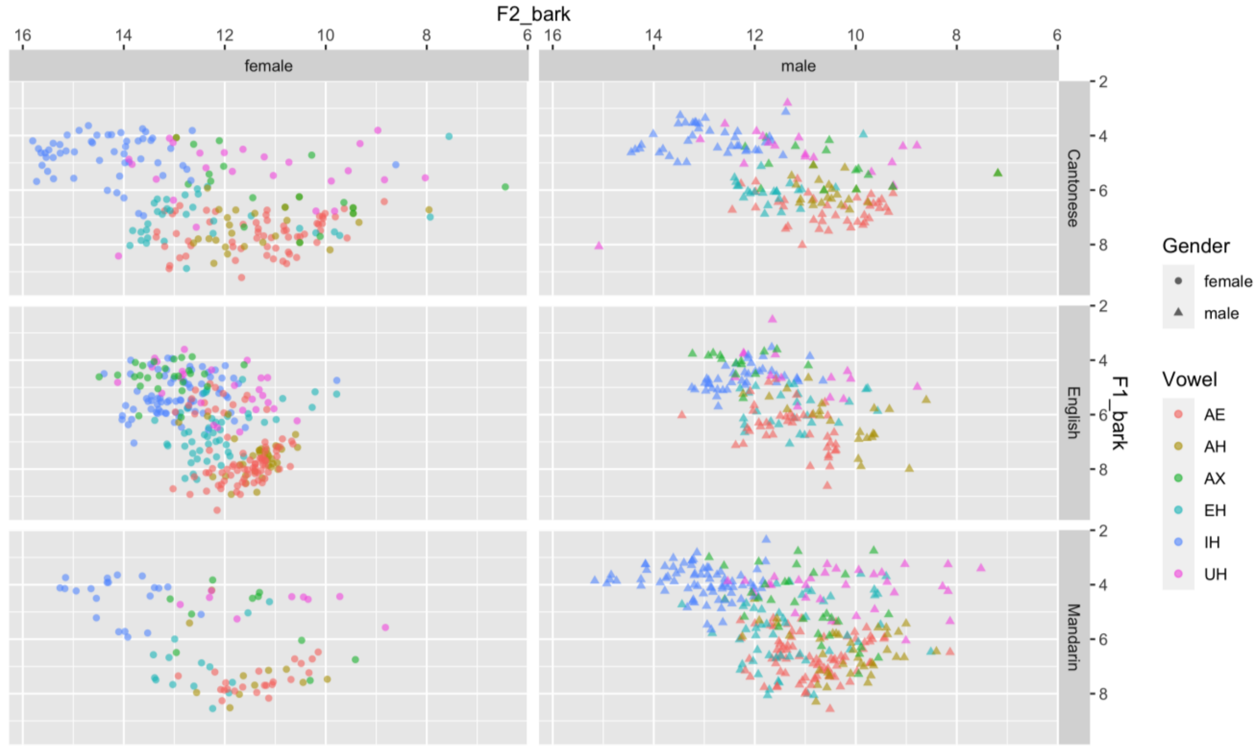
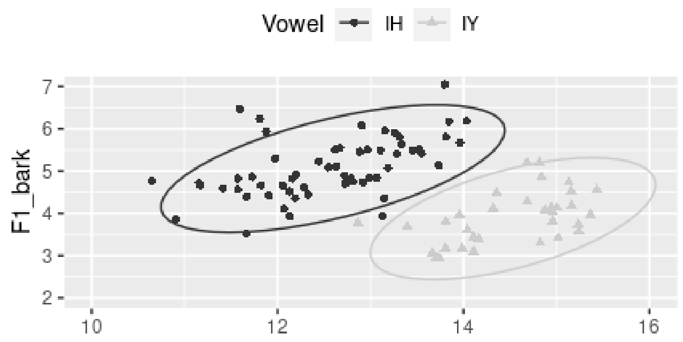


Figure i-3. Vowel spaces of /æ, ʌ, ə, e, ɪ, ʊ/ (AE, AH, AX, EH, IH, UH) produced by ENG and ESL groups.

The multivariate regression model shows that Mandarin and Cantonese ESL speakers produced the lax vowels /æ, ʌ, e, ʊ/ in similar ways as ENG group did. However, for /ɪ, ə/, the two groups differed from each other (Figure i-4). The F2 of /ə/ by Cantonese and Mandarin ESL speakers were significantly lower than that by ENG group (F2*Cantonese = -2.39, 95% HPD [-3.90, -0.76], p = 0.007; F2*Mandarin = -1.9, 95% HPD [-3.37, -0.65], p = 0.02). For the vowel /ɪ/, its F1 and F2 by ESL speakers were both significantly different from those by ENG group, where F1 was lower but F2 was higher in ESL production (for F1: F1*Cantonese = -0.62, 95% HPD [-1.27, -0.04], p = 0.04; F1*Mandarin = -0.85, 95% HPD[-1.62, -0.15], p = 0.029; for F2: F2*Cantonese = 1.37, 95% HPD [0.61, 2.02], p < 0. 001; F2*Mandarin = 1.21, 95% HPD [0.31, 2.09], p = 0.009). For closer inspection, Figure 4 presents a comparison of the lax and tense pair of the front high vowels by ESL speakers. It is clear that ENG group shows distinct patterns of lax and tense vowels. But ESL speakers' lax vowel contained lower F1 and higher F2, which overlapped with the tense counterpart /i/, especially among Mandarin ESL speakers.



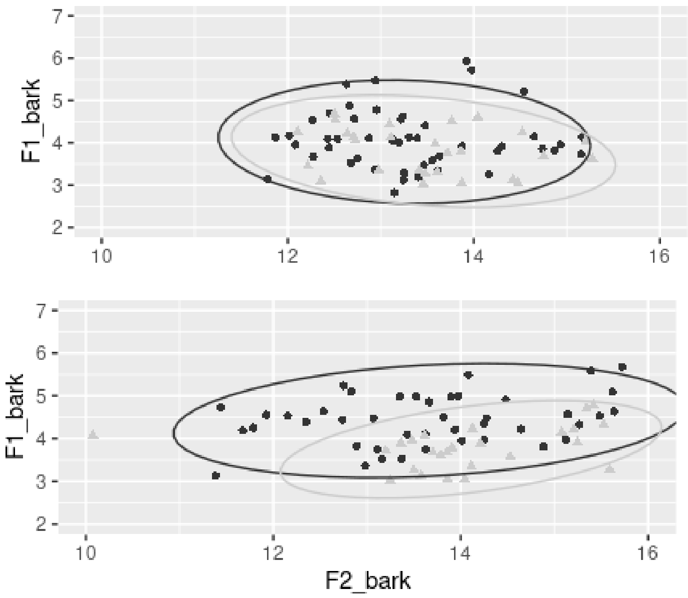


Figure i-4. Vowel spaces of /i/ (IY) and /ɪ/ (IH) produced by ENG speakers (top), and Mandarin (mid) and Cantonese (bottom) ESL speakers.

Secondly, gender plays a role in vowel production, where male speakers' formants are lower than those of female speakers, except for /ə/. Specifically, F1 and F2 of /ʌ, e, ɪ/ were lower in the production by male speakers than that by females (for /ʌ/, F1* male = -1.21, 95% HPD[-1.74 -0.51], p < 0.001, F2* male = -1.28, 95% HPD[-1.87 -0.73], p < 0.001; for /e/, F1* male = -0.83, 95% HPD[-1.38, -0.26], p = 0.01, F2* male = -0.92, 95% HPD[-1.55 -0.35], p = 0.004; for /ɪ/, F1* male = -0.67, 95% HPD[-1.07, -0.25], p < 0.01, F2* male = -0.65, 95% HPD[-1.25, -0.08], p = 0.004, respectively). For /æ/, F1 was lower (F1* male = -1.13, 95% HPD[-1.6 -0.65], p = 0.002). For /ʊ/, only F2 was lower in the male speech (F2* male = -0.97, 95% HPD[-1.78, -0.19], p = 0.03).

Lax vowels: duration

Duration was measured in the steady part of a vowel. Data were shown in Figure i-5.

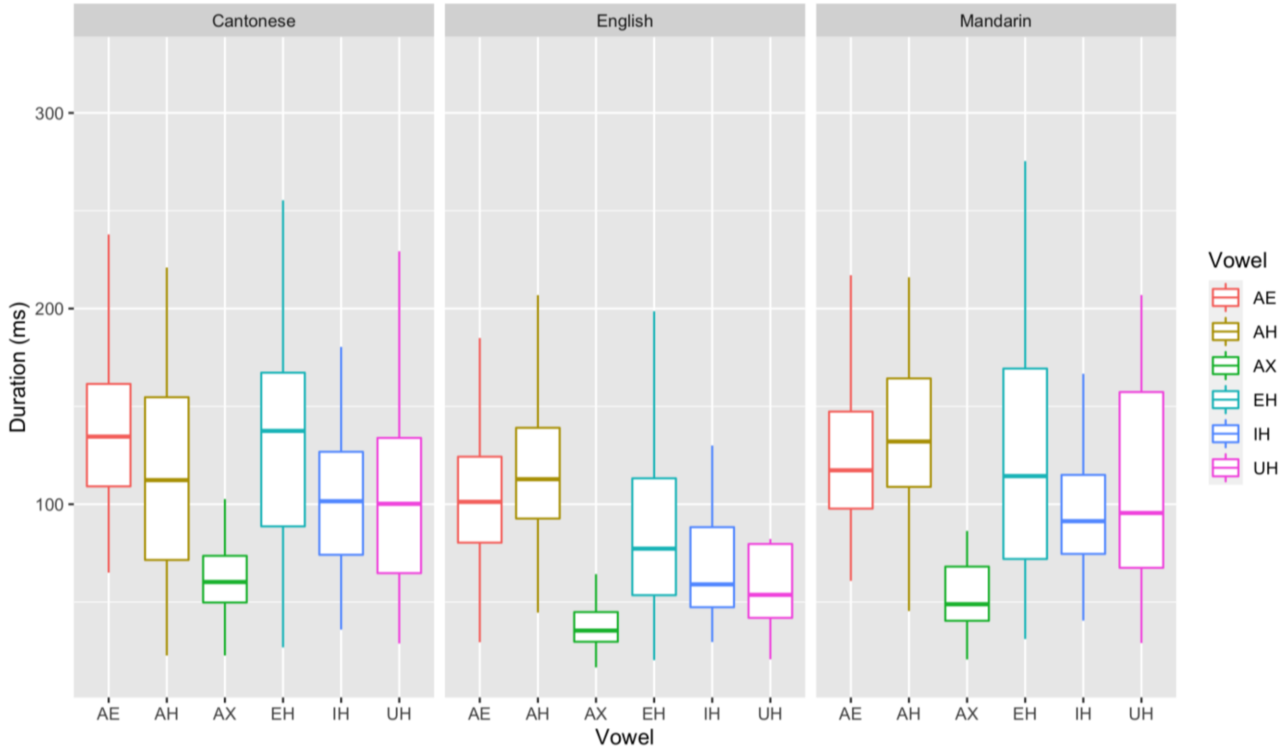


Figure i-5. Duration of /æ, ʌ, ə, e, ɪ, ʊ/ (AE, AH, AX, EH, IH, UH) produced by ENG and ESL groups.

As shown, ENG group produced shorter lax vowels than ESL speakers, especially in /e/. The model with L1 language, vowel, stress location, and the interaction between language and vowel as fixed factors proved the best model to interpret the duration data. This is indicated both by a chi-squared test of log-likelihood ratios ($\chi^2(10) = 63.181$, $p < 0.001$), and by the lowest AIC value. There was effect of the interaction between L1 language and vowel type on the duration of /æ, ʌ, ə, e, ɪ, ʊ/. Tukey's HSD post-hoc tests further showed that ENG and ESL groups produced different duration of /e/ (both $p < 0.0001$). However, there is no significant difference between the duration of /e/ among Cantonese and Mandarin ESL speakers. For the duration of /æ/ and /ɪ/, the production between ENG and Cantonese ESL groups is significantly different (both $p < 0.0001$), but no difference is found between Mandarin and Cantonese groups or between Mandarin and ENG groups (/æ/: *p*-values of Mandarin - Cantonese and Mandarin - English are > 0.0001; /ɪ/: *p*-values of Mandarin - Cantonese and Mandarin - English are > 0.0001). For /ʌ, ə, ʊ/, the duration in ENG production and that in ESL are comparable (all *p*-values are > 0.0001).

Discussion

This part of Study 1 investigated monophthong vowels produced by Cantonese and Mandarin ESL speakers with a comparison with those by L1 speakers of English. Cantonese and Mandarin ESL groups produced most vowels in similar ways as L1 speakers would do. However, the lax vowel /ɪ/ of ESL speakers has significantly lower F1 and higher F2 than that by L1 speakers; and the lax vowel /ə/ by ESL speakers shows significantly lower F1. Moreover, ESL speakers appeared not able to distinguish the tense /i/ and the lax /ɪ/, which were shown overlapping in the charts of vowel spaces. For temporal feature of vowels, ESL groups produced longer vowels than

L1 speakers of English did. The results are in line with previous findings of English vowels produced by Chinese ESL learners (Chen, Robb, Gilbert, & Lerman, 2001; Jia, Strange, Wu, Collado, & Guan, 2006; Munro & Derwing, 2008). Mandarin uses the vowel /i/ and Cantonese the vowel /i:/, both of which have /ɪ/ as an allophone (Luo, Li, & Mok, 2020). This phonological status and distribution of the two sounds in their L1 languages pose difficulties for Chinese ESL speakers when they are faced with the distinct pair of /i/ and /ɪ/ in English. L1 allophony may have reduced their sensitivity to the phonemic contrast in English, even though both involve temporal differences. However, perceptually, ESL speakers may still rely on durational differences to discriminate between the pair in English (Wang & Munro, 1999). Lastly, for the central vowel /ə/ which involves less specified features as compared to the cardinal vowels in a vowel space, more variation in F1 and F2 were found in ESL production than in ENG group.

<u>Part. 2. Production of consonants</u>

The average accuracy rate in production of coda consonant by Mandarin ESL speakers was 74.15%. As hypothesized, Mandarin speakers changed the illicit syllable structure by deleting the coda consonant (17.69%) or by adding an epenthetic vowel after the coda consonant (7.5%). Cantonese ESL speakers, on the other hand, produced the coda consonants with an accuracy rate at 56%, which is not higher than that of Mandarin group and contradicted our prediction. Cantonese ESL speakers also made modification to codas by deleting a consonant in 44% of all words. They did not add any epenthetic vowel as Mandarin speakers did. Similarly, ENG group also deleted the coda consonant sometimes (21% of all words), and showed no epenthesis following a coda consonant either.

Production of coda consonants by Mandarin ESL speakers

Accuracy rates by Mandarin and ENG groups were calculated and compared, as shown in Table 4.

Table i-1. Modification of coda consonant by Mandarin ESL speakers and L1 speakers of English.

Coda	Words	Onset of following words	Accuracy rates		Deletion		Epenthesis
			Mandarin ESL	English L1	Mandarin ESL	English L1	Mandarin ESL
d	should	C (b)	15%	80%	69%	20%	16%
	agreed	C (ð)	23%	41%	69%	59%	8%
	succeeded	V (i)	57%	100%	22%	0	21%
	could	#	92%	100%	0	0	8%
k	take	C (h)	78%	100%	7%	0	14%
	cloak-1	V (ɑ)	93%	88%	7%	12%	0
	cloak-2	V (ə)	100%	100%	0	0	0
	took	V (ɑ)	93%	100%	7%	0	0
	cloak-0	#	100%	100%	0	0	0
	cloak-3	#	77%	71%	15%	29%	8%
t	out	C (w)	93%	6%	0	94%	7%

In general, Mandarin ESL speakers tended to delete the coda consonant to repair the illicit syllable structure in Mandarin than to add an epenthetic vowel, except for *could* and *take*. They also preferred to modify coda-/d/ than coda-/k/ or /t/, though deletion of a coronal consonant is a very common phenomenon in English (Labov, 1989; Neu, 1980). It should be noticed that results of coda-/t/ is very limited in our study due to the availability of data in the sample texts. Phonetic contexts of the following words can also exert strong influence on coda modification. We examined three types of phonetic contexts, where the onsets of the following syllables contained a consonant (C), a vowel (V) and nothing (#, i.e. word boundary). Coda consonants in target words followed by a consonant were more often deleted than when it was followed by a vowel or by a word boundary. For instance, the coda-/d/ in both *should* and *agreed* were deleted in 69% of all tokens, when the two words were followed by a labial plosive /b/ or a dental fricative /ð/. In summary, production of Mandarin ESL speakers shows a pattern of phonological process involving modification to coda. The phonetic context to the right of a target that may trigger consonant deletion is: consonant > vowel > pause, with pauses the most probable in such trigger effect.

Word frequency is found to play a role as well in modification to coda consonants. For instance, Neu (1980) found that deletion was more common in high frequency words for the speakers of American English while Gregory et al. (1999) found that high frequency words underwent deletion 11 times greater than low frequency words. In our study, all target words are frequently-used words, except for *cloak*. The frequencies for *other*, *along* and *cloak* are 1454307, 255917, 4009 respectively at COCA corpus (Davies, 2008). Our results echo previous studies in that the coda /k/ in *cloak* was least modified by Mandarin ESL speakers.

L1 speakers of English modified the coda too but not as frequently or as varied in type as the Mandarin ESL speakers did. The latter group sometimes inserted a vowel after the coda and created a new CV syllable at the end. This may be due to the canonical CV structure in Mandarin Chinese. Modification of coda consonants or the tendency of keeping CV structures by Mandarin learners may be influenced by either universal phonology or their knowledge in Mandarin, which may require further analysis on more varieties of phonetic contexts and greater number of words and participants.

Coda deletion was found in production by all Mandarin speakers, with rates ranging from 9% to 30%. However, only six speakers out of all fourteen participants adopted a strategy of vowel epenthesis. Different speakers seemed to use different strategies in phonological processes in order to maintain a CV structure (Table i-2). It is hypothesized that such preference or performance may correlate with language proficiency or experience with the target languages. Individual variation to a much smaller extent is also noted in L1 speakers of English, who would occasionally omit a final consonant, which is common in connected speech in English.

Table i-2. Overall accuracy in consonant production and distribution of phonological processes.

a

Mandarin ESL (Speaker ID)	**11**	**18**	**16**	**32**	**12**	**20**	**21**	**30**	**43**	**37**	**39**	**33**	**35**
Accuracy (%)	91	91	90	90	82	73	73	73	73	70	67	60	60
Deletion (%)	9	9	10	10	18	18	27	18	27	30	11	20	30

Epenthesis (%)	0	0	0	0	0	9	0	9	0	0	22	20	10

b

L1 English (Speaker ID)	**57**	**62**	**64**	**63**	**65**	**66**	**131**	**69**	**70**	**49**	**56**	**58**	**68**	**133**	**51**	**72**	**61**
Accuracy (%)	91	89	89	80	80	80	80	78	78	75	75	75	75	75	73	73	71
Deletion (%)	9	11	11	20	20	20	20	22	22	25	25	25	25	25	27	27	29

c

Cantonese ESL (Speaker ID)	**88**	**74**	**77**	**83**	**87**	**86**	**84**	**85**	**75**	**73**	**76**	**79**	**78**	**82**
Accuracy (%)	89	82	82	80	73	67	64	56	44	33	33	22	18	17
Deletion (%)	11	18	18	20	27	33	36	44	56	67	67	78	82	83

Production of coda consonants by Cantonese ESL speakers

Accuracy rates by Cantonese and ENG groups were calculated and compared, as shown in Table i-3. Overall, Cantonese group omitted the coda consonants more often than ENG who also deleted codas sometimes but never in *succeeded*, *could*, *take*. This confirms our prediction that Cantonese ESL speakers produced coda stops as unreleased at word final positions, as what they would in their L1 language. However, the prediction requires further verification with analysis of phonetic cues of unreleased stops including F1, F2 , and F3 transitions during the preceding vowel into the stop coda constriction (Ciocca, Wong, & So, 1994; Khouw & Ciocca, 2006) and normalized duration of final syllables.

Table i-3. Modifications of coda consonant by Cantonese ESL speakers and English as L1 speakers.

Coda	Words	Onset of following words	Accuracy rates		Deletion	
			Cantonese ESL	English L1	Cantonese ESL	English L1
d	should	C (b)	7%	80%	93%	20%
	agreed	C (ð)	31%	41%	69%	59%
	succeeded	V (i)	44%	100%	56%	0
	could	#	79%	100%	21%	0
k	take	C (h)	57%	100%	43%	0
	cloak-1	V (ɑ)	89%	88%	11%	12%
	cloak-2	V (ə)	64%	100%	36%	0
	took	V (ɑ)	71%	100%	29%	0
	cloak-0	#	82%	100%	18%	0
	cloak-3	#	82%	71%	18%	29%
t	out	C (w)	29%	6%	71%	94%

Speaker variation in production is also noticed among Cantonese ESL speakers, as shown in Table 5-c above. There is greater difference among individual speakers among the Cantonese group than in Mandarin or English as L1 groups. It is suspected that greater differences in language proficiency may be true in Cantonese group, which has a great impact on their varied production. But this speculation requires verification from the corpus producers.

Our examination of Chinese ESL production of coda consonants reveals shared but also distinct patterns in consonant modification by Mandarin and Cantonese speakers. The former group preferred a combination of strategies while the latter one method only. The differences may be attributed to influence of L1 phonological rules and also universal preference of a CV structure in world languages. Experience in the target language also warrants attention in the account of ESL speakers' performance.

Part. 3. Production of lexical stresses.

First, vowel duration was measured as the steady part of a vowel. In general, as shown in Figure i-6, vowel duration by non-L1 speakers is longer than that of L1 speakers of English. Within the two non-L1 groups, no difference is found between Cantonese speakers (CAN) and Mandarin speakers (CMN). Examination of individual words as shown in Figure i-7 reveals the percentage of stressed vowels is larger than 50% for all three speaker groups. It suggests that stressed vowels are realized longer than unstressed vowel, except for 'other' in the production by L1 speakers of English (ENG) group.

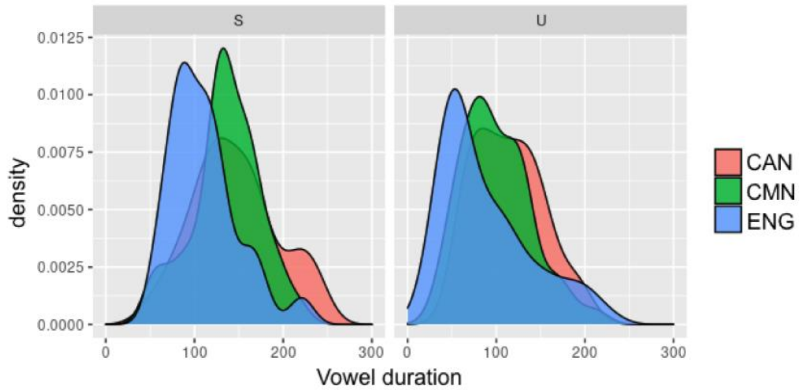


Figure i-6. Duration (ms) of stressed (S) and unstressed (U) vowel produced by speaker groups.

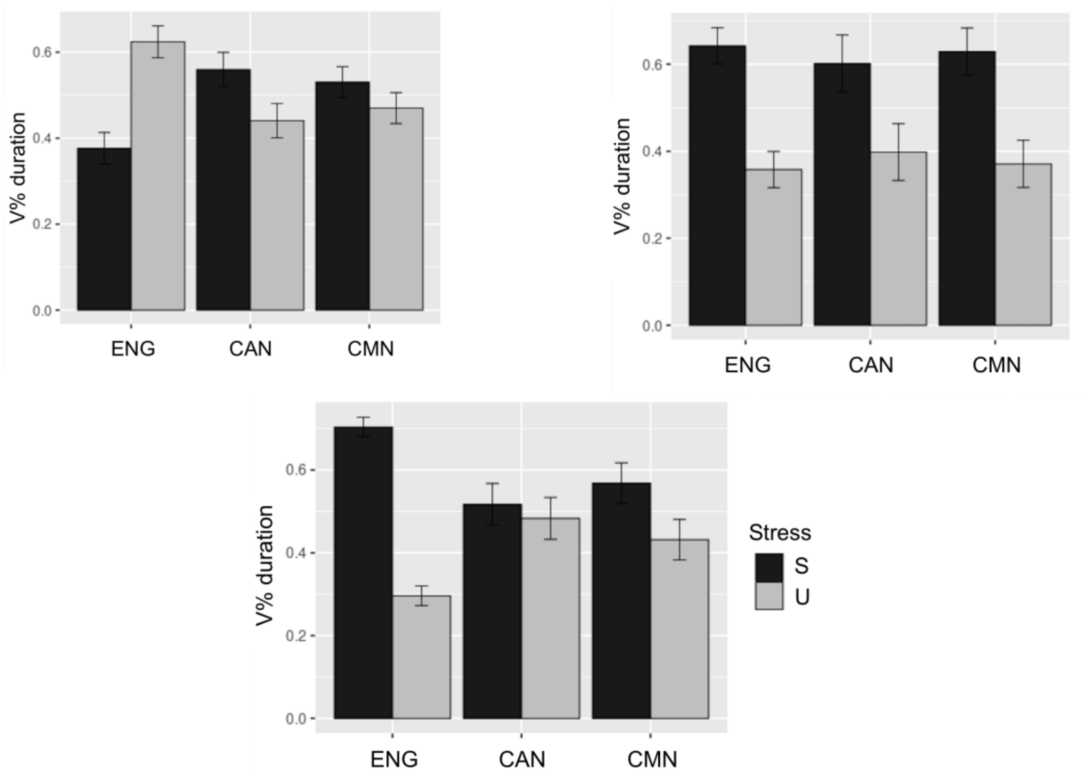


Figure i-7. Percentage of vowel duration in three words, "other, along, agree" (top left, right, bottom) by speaker groups.

Secondly, for F0 peak, female speakers have higher F0 peak and more variations than male speakers, as shown in Figure i-8. No difference has been found between L1 and non-L1 speakers in the production of F0 peak. Figure i-9 shows that the percentage of stressed vowels is larger than 50%, except for the word “along”, where unstressed vowels have higher percentage, but the difference is not very big. Moreover, stressed vowels have higher F0 mean than unstressed vowels for the word “other” by non-L1 speakers. Overall, stressed vowels have higher F0 peak than unstressed vowels.

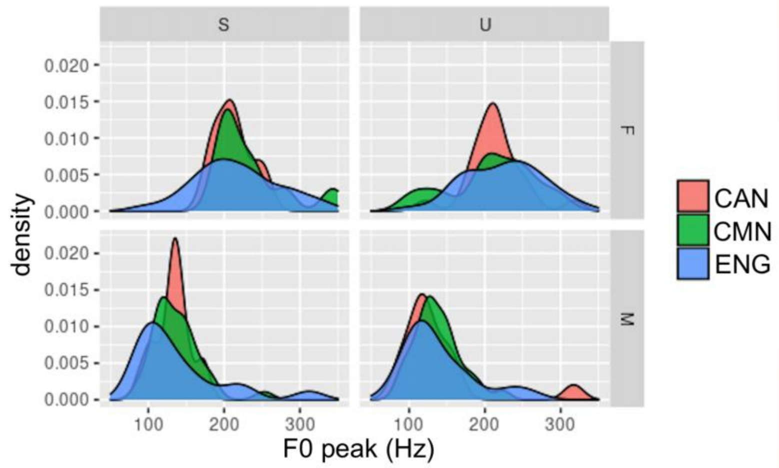


Figure i-8. F0 peak (Hz) in stressed (S) and unstressed (U) vowels by three speaker groups.

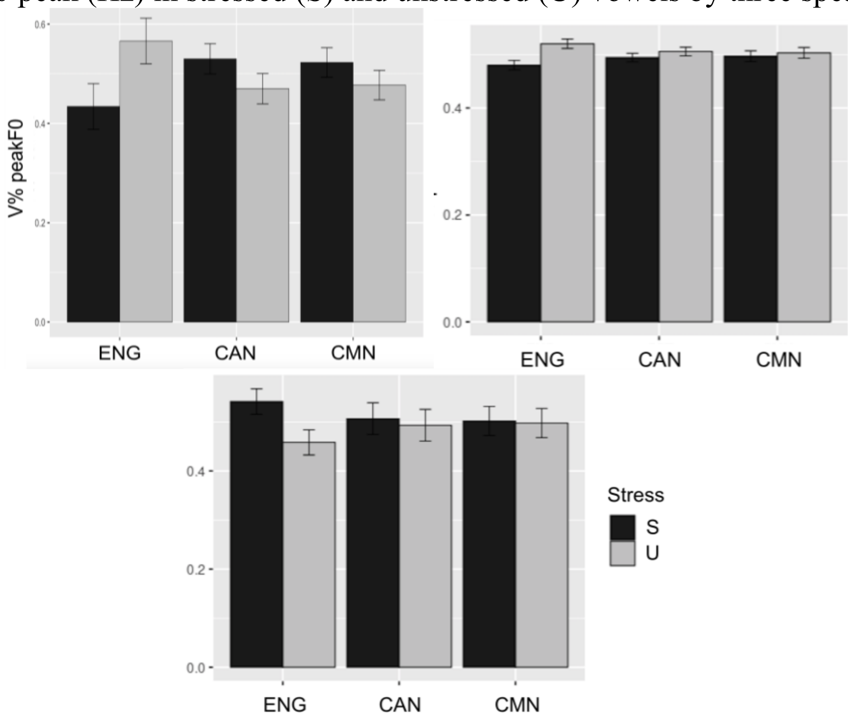


Figure i-9. Percentage of F0 peak for stressed and unstressed syllables in “other, along, agree” (left, mid, right) by three speaker groups.

Thirdly, for F0 mean, Figure i-10 shows that female speakers have higher F0 peak and more variations (pitch range) than male speakers. In Figure i-11, stressed vowels always have lower F0 mean than unstressed vowel in the production by L1 speakers. For non-L1 speakers, stressed vowels have higher F0 mean than unstressed vowels for the word “other” while all the three groups produced stressed vowels with higher mean F0 than unstressed ones for the word "agreed". But for "along", there is no difference between stressed and unstressed vowels.

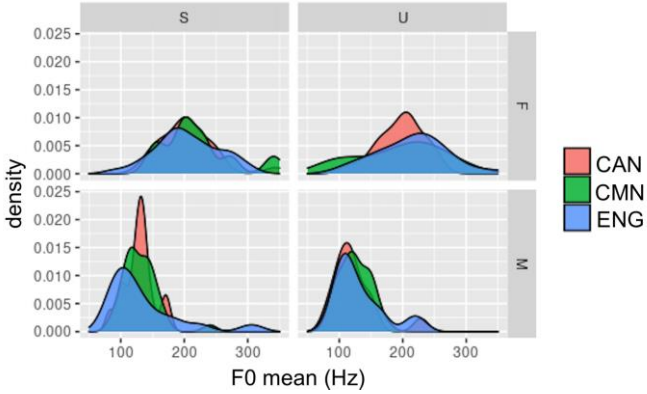


Figure i-10. Distribution of F0 mean (Hz) in the production of stressed (S) and unstressed (U) by three speaker groups.

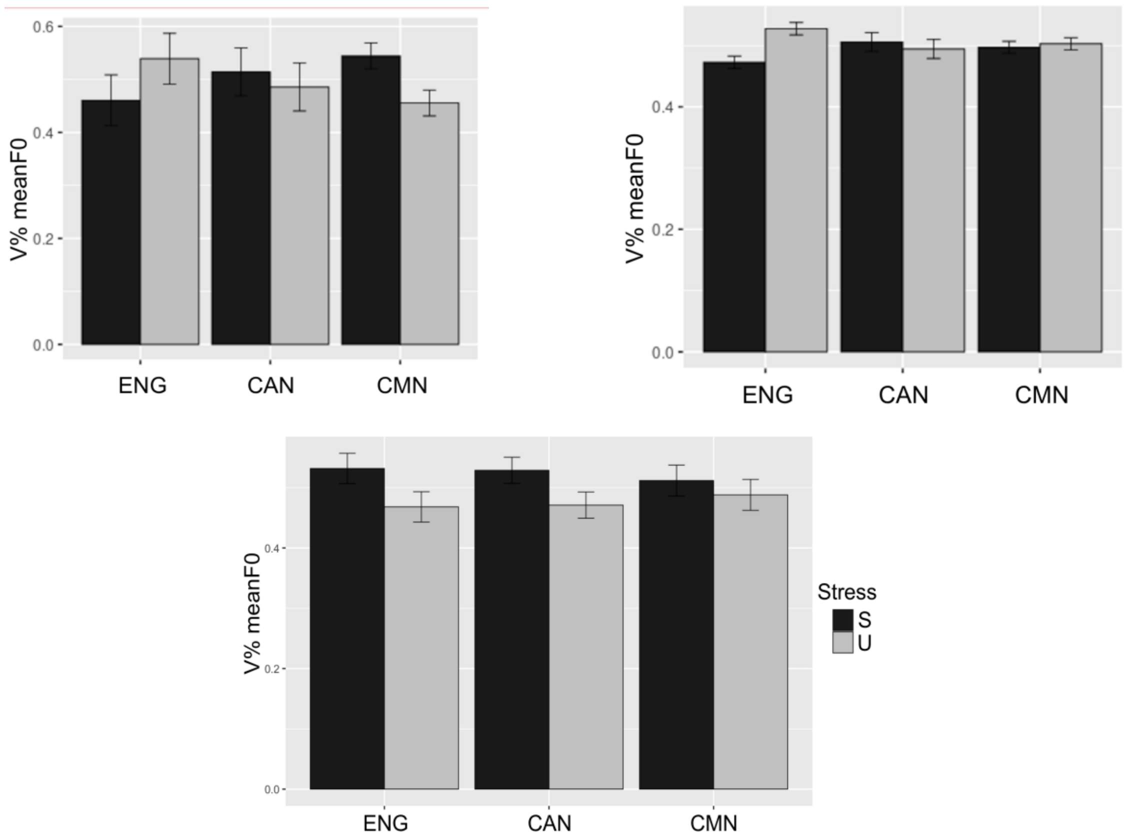


Figure i-11. Percentage of F0 mean for stressed (S) and unstressed syllables (U) in three words, "other, along, agree" (top left, right, bottom) by three speaker groups.

Fourthly, for intensity peak, there is no significant difference between L1 and non-L1 speakers (Figure i-12). Percentage of peaks falling on stressed vowels is much higher than that on unstressed vowels (Figure i-13).

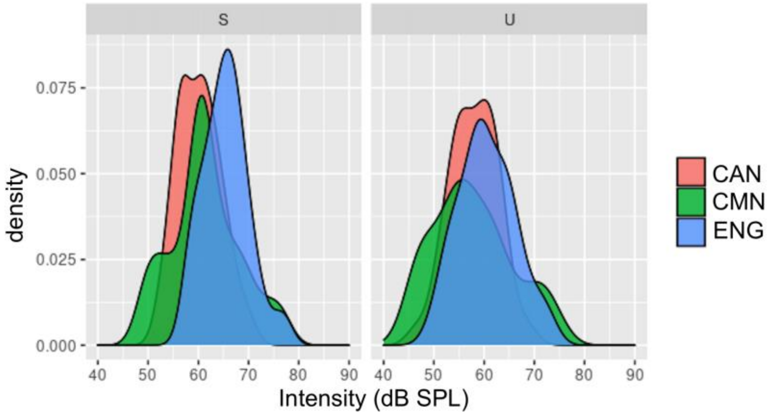


Figure i-12. Intensity peak (dB SPL) for stressed (S) and unstressed (U) vowels by three speaker groups.

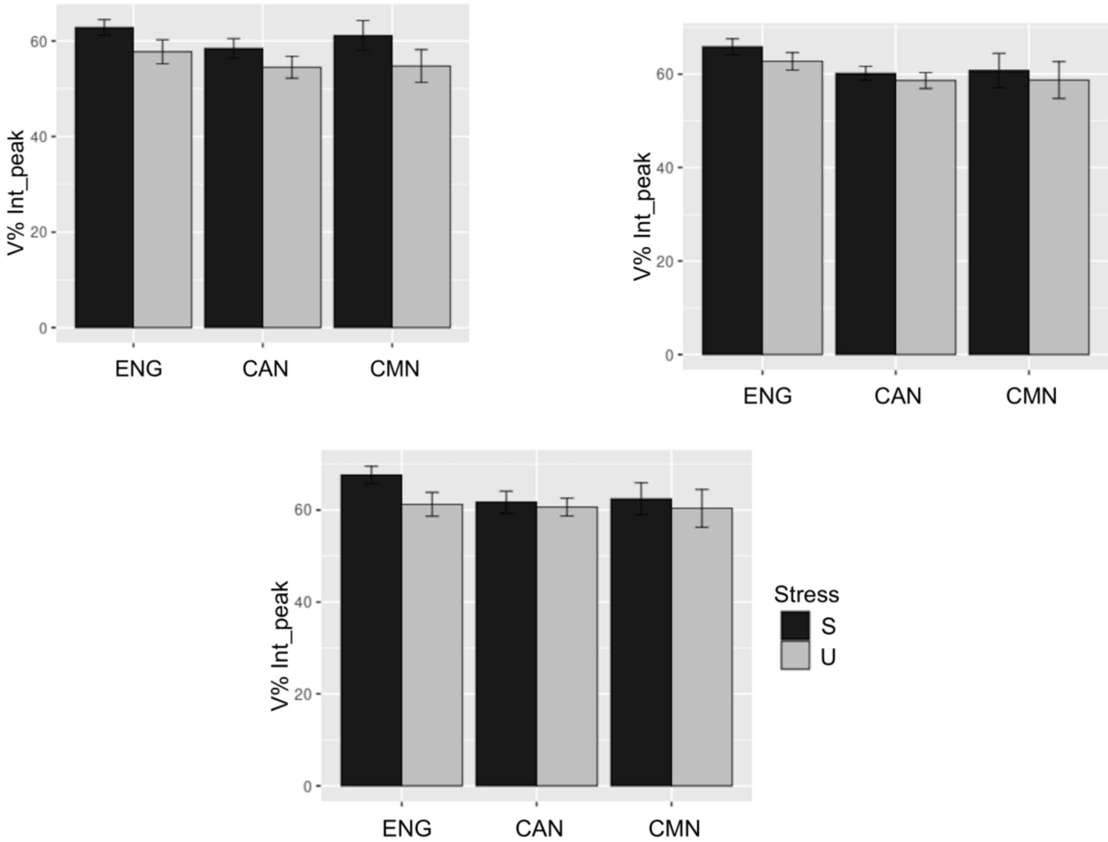


Figure i-13. Percentage of intensity peak for stressed (S) and unstressed (U) syllables in three words, "other, along, agree" (top left, right, bottom) by three speaker groups.

Discussion

In general, greater differences between stressed and unstressed vowels are found in production by L1 speakers of English (ENG) than that by ESL speakers. This could be explained by the typological difference in prosodic systems between English and Chinese. The strong-weak pattern comprises "foot" of the rhythms in English. However, Chinese does not assign such varied weights on its monosyllabic characters which show less distinct contrast phonetically as syllables bearing lexical stresses do in English. L1 speakers of Chinese may not be sensitive to prosodic contrastivity in English due to their familiarity with L1 systems. Therefore, they encounter difficulties in effectively utilizing all phonetic cues to distinguish stressed and unstressed syllables in English words. They could, however, improve in attending to duration after training (Lai, 2009; Zhang, Nissen, & Francis, 2008).

Not only phonetic knowledge in correlates of lexical stresses, but also phonological knowledge of syllabification seemed to affect NSL speakers' realization of stress contrastivity. In the word "other", the vowel /ʌ/ in the stressed syllable is actually shorter on average as produced by ENG speakers, but with higher F0 peak and intensity than the vowel in the unstressed syllable. For ESL speakers, vowels in the stressed syllable have greater values in all parameters than the unstressed one. ENG tended to attach the consonant to the stressed vowel while ESL speakers are less sensitive to the relation between syllable weights and stress placement. This is in conformity of Burzio's loose-requirement approach (Burzio, 1994). The approach aims to account for the main stress placement and prerequisites that "syllabification is sensitive to stress (or foot structure). In addition, syllabification and stress are not carried out sequentially but are checked simultaneously. Specifically, a CVCV sequence is not always syllabified as CV.CV, but can be CVC.CV if stress is on the first syllable" (Duanmu, Kim, & Stiennon, 2003, p. 14).

Summary

Study 1 investigated ESL speech production of vowels, consonants and lexical stresses, with a reference to that by ENG speakers. We analysed phonetic realization of segmentals and prosodic correlates such as duration, F0 and intensity. General phonetic characteristics are identified for Chinese ESL speakers. There are deviations in ESL speech from that by ENG speakers. Moreover, as Mandarin and Cantonese ESL speakers showed systematic variations from the ENG and also among the two learner groups themselves, we propose that phonetic characteristics of ESL production could be due to the influence of L1 phonological systems and also to ESL speakers' to-be-honed knowledge in L2 phonology. The results provide baselines for further examination of ESL pronunciation at workplace where interfering factors such as noises to speech communication are common. This leads to our Study 2 on ESL production in multi-talker babbles and on linguistic strategies that ESL speakers may adopt when speaking under noises.

Study 2. ESL speech production in simulated adverse condition at workplaces

This study examined Cantonese ESL learners' overall phonetic performance in noise with a reference to that in quiet and that by L1 speakers of English. A total of 759 tokens of ten target words were selected out of the word list participants read, for acoustic and statistical analysis.

Vowel spaces

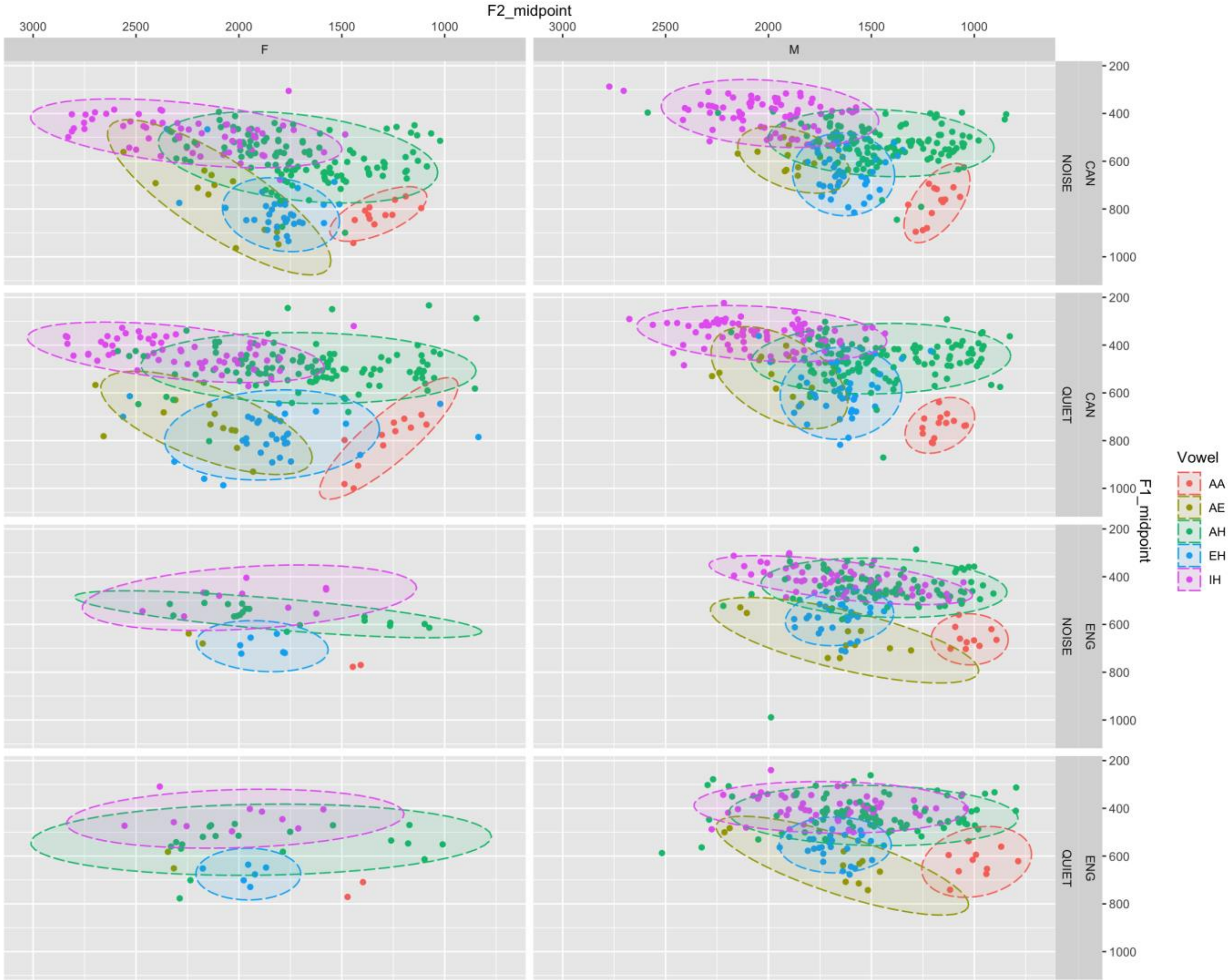


Figure i-15. F1 (y-axis) and F2 (x-axis) (Hz) of /ɑ, ɪ, e, æ, ə/ (AA, IH, EH, AE, AH) produced by Cantonese ESL (CAN) and L1 English (ENG) groups in quiet and under noise.

Figure i-15 shows the F1 and F2 values of the vowels /ɑ, ɪ, e, æ, ə/ (AA, IH, EH, AE, AH) produced by Cantonese ESL learners (Cantonese group) and L1 speakers of English (English group). The multivariate regression model shows that Cantonese learners produced English vowels different from L1 speakers of English, and the significant difference is from the production of F2 (F2 * Cantonese = 3.980e-01, 95% HPD [1.746e-02, 7.963e-01], p = 0.05). Cantonese ESL learners produced /ɪ/ with higher F2 than the L1 speakers of English. That is, the lax /ɪ/ produced by the Cantonese group with higher F2 was similar to a tense vowel /i/. This result is consistent with our previous findings in section 3.3 of Study 1. However, there was no difference in the production of F1. The sex also influenced the F1 and F2 values, where male speakers produced significantly lower F1 and F2 than female speakers (F1*male = -7.521e-01, 95% HPD [1.063e+00 -4.378e-01], p < 0.0001; F2*male = -8.385e-01, 95% HPD [-1.232e+00, -4.135e-01], p < 0.0001). The noisy environment did not affect the production of vowels for both Cantonese and English speakers.

Long-term average speech spectra (LTASS)

As shown in Figure i-16, LTASS increased in noise for both two groups. Nevertheless, there is greater difference for the Cantonese group than the English group in terms of the magnitude of increase in noise. This shows that the Cantonese group produces speech with more vocal effort when confronted with noise interference.

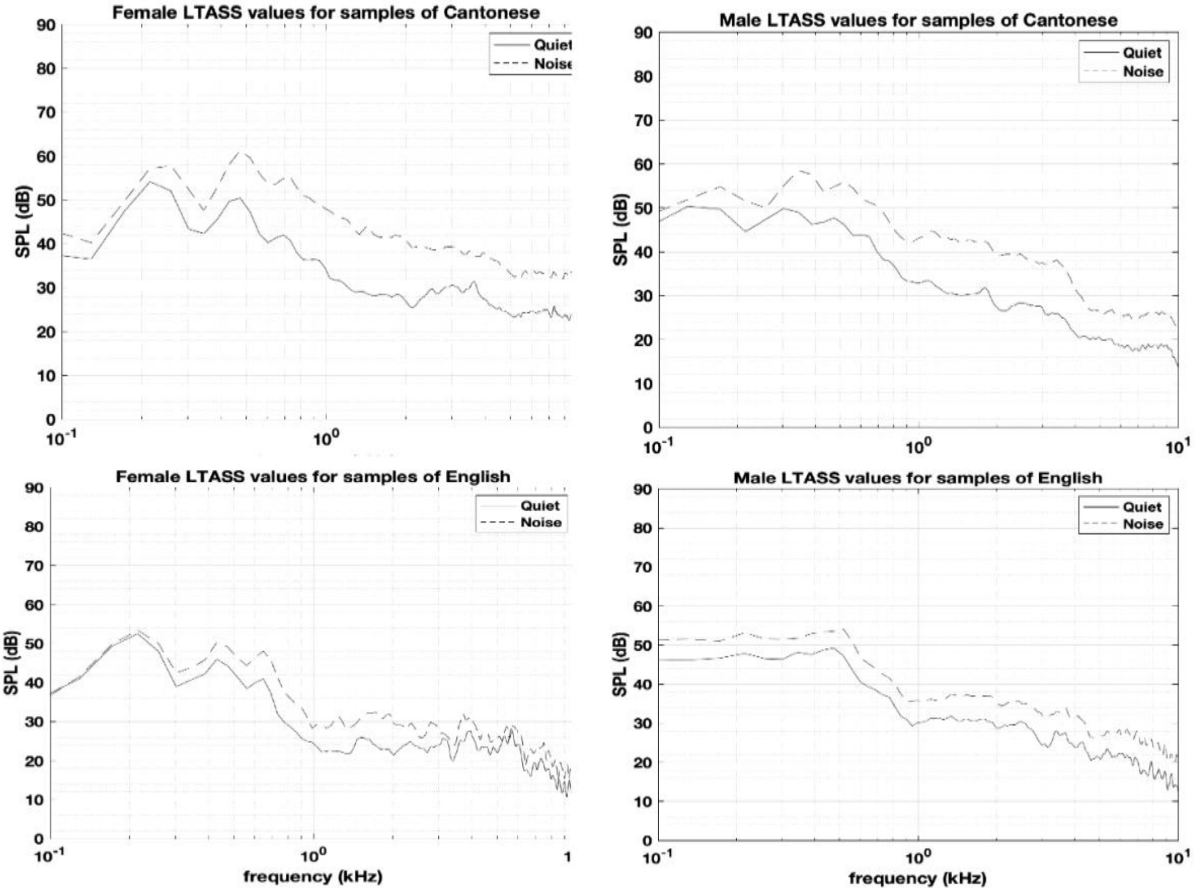


Figure i-16. LTASS for Cantonese ESL (CAN) and L1 English (ENG) groups in quiet and under noise.

Speaking rate

Speaking rates are comparable between the CAN and ENG groups in quiet condition, and also between quiet and noise conditions for the ENG group (Figure i-17). Nevertheless, the speaking rate is significantly slower in noise than quiet condition for CAN group ($p < 0.0001$), which indicates that Cantonese ESL speakers spoke much more slowly in noise than they did in quiet.

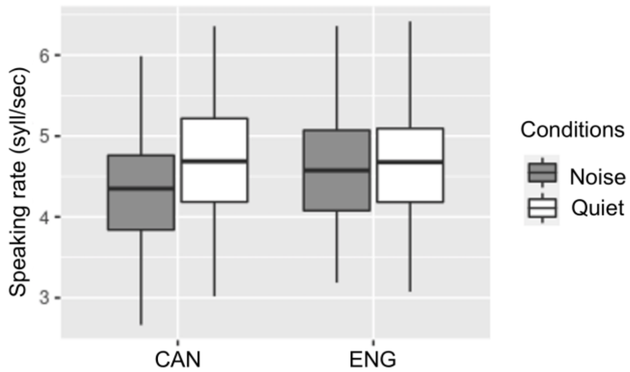


Figure i-17. Speaking rates of Cantonese ESL (CAN) and L1 English (ENG) groups in quiet and under noise.

Peak F0

The CAN group produced stressed syllables with significantly higher F0 peak in quiet than the ENG group did, and more so in noise condition (p's < 0.0001 for male speakers), shown in Figure i-18. Meanwhile, peak F0 is comparable between quiet and noise condition for English group, whereas the Cantonese group used significantly higher peak F0 in noise than quiet condition (p's < 0.0001).

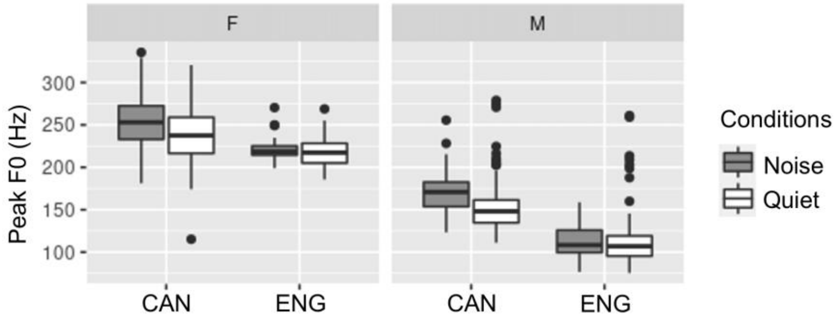


Figure i-18. Peak F0 of Cantonese ESL (CAN) and L1 English (ENG) groups in quiet and under noise.

Mean F0

Similar to the results of peak F0, the CAN group produced stressed syllables with significantly higher mean F0 in quiet than the ENG group did and more so in noise condition (p's < 0.001 for male speakers), shown in Figure i-19. Meanwhile, mean F0 is comparable between quiet and noise conditions for male ENG speakers, whereas the CAN speakers used significantly higher mean F0 in noise than in quiet condition (p's < 0.01).

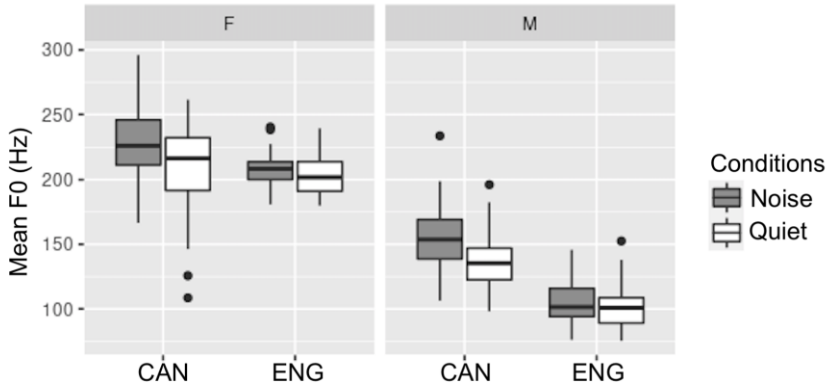


Figure i-19. Mean F0 of Cantonese ESL (CAN) and L1 English (ENG) groups in quiet and under noise.

Peak intensity

There is significantly difference between quiet and noise condition for both Cantonese ESL speakers and ENG group (p < 0.0001) (Figure 20).

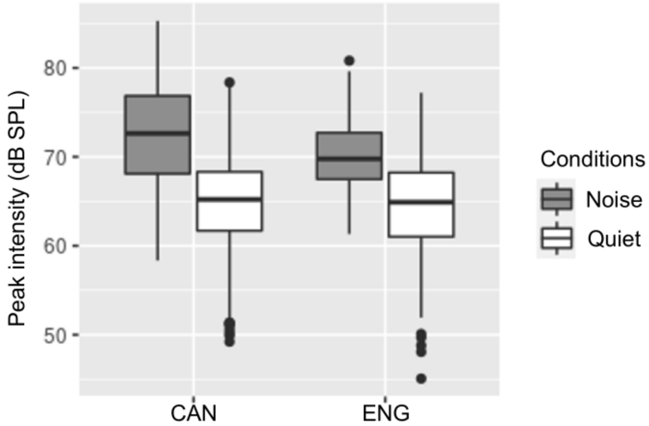


Figure i-20. Peak intensity of Cantonese ESL (CAN) and L1 English (ENG) groups in quiet and under noise.

Pairwise variability index

The overall PVI is larger than 0 for both two groups and both conditions, which means that primary stressed syllables have larger value than unstressed syllables in terms of duration, peak F0 and peak intensity.

There is no significant difference in PVIs of duration and peak F0 for both two groups as well as the quiet and noise conditions (Figure i-21). In terms of peak intensity, the English group produced larger stress contrast (as indicated by higher positive PVI) than the CAN group in quiet ($p = 0.0141$). Moreover, ENG group produced larger stress contrast in quiet than in noise ($p < 0.0001$).

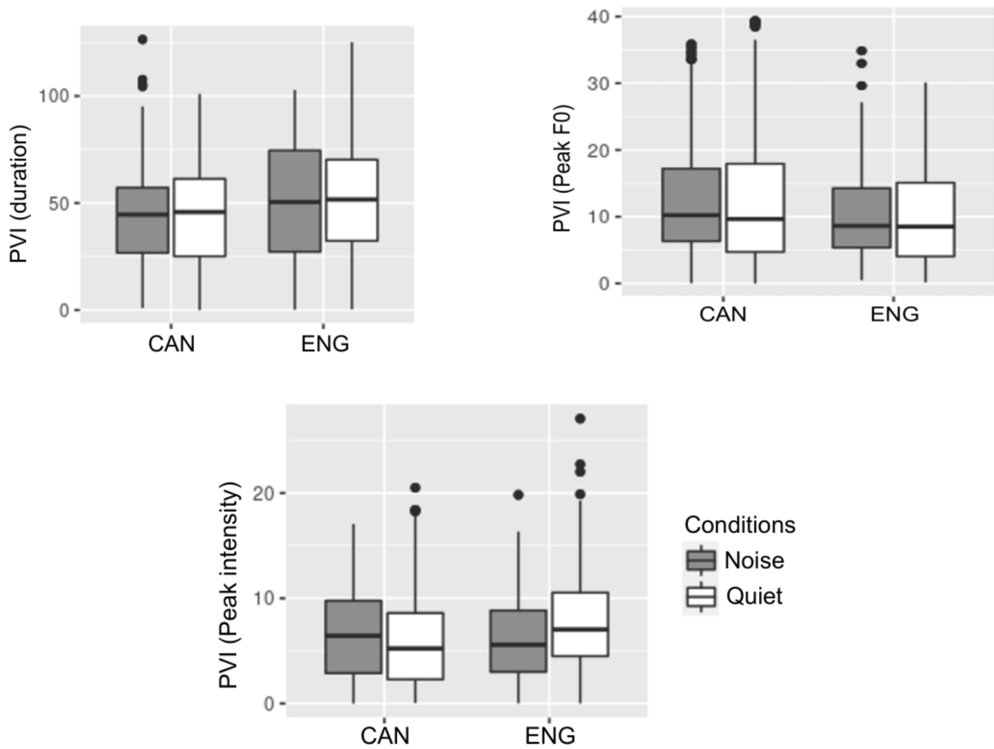


Figure i-21. PVIs of duration, peak F0 and peak intensity (top left, top right, bottom) among Cantonese ESL (CAN) and L1 English (ENG) groups in quiet and under noise.

Discussion

Summing up the results from Study 2, we confirmed that the Cantonese ESL speakers spoke louder, more slowly and with higher peak F0 and more vocal effort in noise than L1 speakers did (as summed in Table i-3). Compared to speech in quiet, Cantonese speakers maintained the contrast between primary stressed and unstressed syllables under noise, whereas English speakers reduced the peak intensity contrast under noise. We also found that formant frequencies of vowels were comparable in quiet and noisy conditions in both groups. Nevertheless, Cantonese speakers produced the lax vowel /ɪ/ with higher F2 than English speakers. In terms of primary stresses in disyllabic words, Cantonese ESL speakers adopted a slower speaking rate, higher peak fundamental frequency (F0), higher intensity and more vocal effort in noise than in quiet condition, whereas English speakers did not change their speaking rate or peak F0 when speaking with background noise. Furthermore, the contrast between primary stressed and unstressed syllables was achieved as indicated by positive PVI of duration, F0 and intensity. Our findings confirmed that Cantonese ESL speakers used duration, F0, and intensity as major means to signal stresses in noise, but with a different strategy from L1 speakers of English. Patterns of phonetic modifications in noise appear language-specific, which may be attributed to ESL speakers' limited experience in non-L1 communication under adverse conditions. They may, instead, tend to compensate degradation of speech signals by resorting to their L1.

Table i-3. Phonetic differences between speaking conditions by speaker groups.

	Cantonese ESL	English L1
F1 and F2 (Hz) of vowels	Noise = Quiet	Noise = Quiet
LTASS	Noise >> Quiet	Noise > Quiet
Peak intensity (dB SPL)	Noise > Quiet	Noise > Quiet
Peak F0 and mean F0	Noise > Quiet	Noise = Quiet
Speaking rate (syllable/sec)	Noise < Quiet	Noise = Quiet
Contrastivity: stressed vs. unstressed	Noise = Quiet (dur/pitch/intensity)	Noise = Quiet (dur/pitch) Noise < Quiet (intensity)

Effects of noise

First of all, Lombard effects were confirmed in Cantonese ESL speech (Prediction 1). Under the noise condition, both two groups display an increase in vocal amplitude reflected by higher LTASS and peak intensity. This confirmed the conclusion from previous studies that using more vocal effort is one of the globally acoustic modifications of Lombard (Garnier et al., 2010; Junqua, 1996). However, the F1 and F2 of the vowels are comparable for quiet and noisy conditions across the two groups, which indicate that it may be easier to modify the suprasegmental properties than the segmental ones.

The Cantonese ESL group amplified the speech modifications more than the L1 speakers of English (ENG speakers) (Prediction 3). Firstly, there is higher enhancement in LTASS for the Cantonese group than the English group. Moreover, the learners spoke slower in noise than quiet condition while the English group did not. It is expected that non-L1 speakers have a slower speaking rate compared to L1 speakers. In the current study, we also found such relatively slower speaking rate in L2 Lombard speech. Furthermore, this shows that ESL speakers are not only struggling in limited L2 linguistic knowledge but also confronted with limited knowledge of realizing and perceiving acoustic modifications under noise. Additionally, L2 speakers have more cognitive burden when they must process linguistic information in an L2 despite adverse conditions that harm speech communication in general.

Positive PVIs for duration, intensity and F0 indicate that both two groups could use higher F0, longer duration and higher intensity for primary stressed syllables in noise. However, English group reduced the degree of contrastivity in noise whereas Cantonese group did not (Prediction 2). This is different from the results in Arciuli and colleagues (2014). They found larger degree of contrast between the primary stressed and unstressed syllables in terms of intensity and fundamental frequency. The asymmetry performance in the English group could be explained by speakers' desire of communication. Junqua (1996) points out that "speech production may also be altered by a combination of human factors involving speech perception but also, for example, the desire of the speaker to communicate" (p.14). In Experiment1, participants read sentences from a computer screen without interactive communication with a speech partner. Garnier et al. (2010) show that "communicative interaction on speech production […] influences the studied phenomenon of speech adaptation from quiet to noisy conditions" (p. 599).

Effects of L1

The ENG group's speech did not show significant difference between quiet and noise condition in terms of peak F0. This is in line with Arciuli and colleagues (2014) that L1 speakers of English did not increase acoustic correlates of lexical stresses equally across syllables, but altered degree of contrast between strong and weak syllables in Lombard speech. However, Cantonese ESL group produced speech with higher peak F0 than English group in quiet and more so in noise (Prediction 4). The constant higher peak F0 may be associated with speakers' L1. In Cantonese, each syllable carries a lexical tone with correlates of F0 as primary cues. Cantonese ESL speakers tended to assign a High tone as equivalent to primary stresses in English. They also preferred modification of F0 correlates as accommodating measures to speech in noise. This is similar as their strategies in speaking Cantonese under noises, and echoes findings from previous studies such as Zhao and Jurafsky (2009) who found all tones in L1 Cantonese speech carried higher F0s under noisy conditions.

Study 3: Effectiveness of speech modifications on perception in noise.

We assessed effects of Lombard speech in L1 and L2 speech with a post-experiment questionnaire. All L1 speakers of English (ENG) reported reasonable audibility of their partners, but less than a quarter of our Cantonese ESL speakers did not hear clearly from theirs (Table i-4). The results show that Lombard speech is effective for enhancing speech intelligibility in noise to a certain extent, but such facilitative effect is not significant among ESL speakers who encountered great difficulties in capturing speech in noise. Around 61% of Cantonese ESL speakers reported that they did not think their partners changed their voices in noise, whereas over 80% of ENG speakers detected changes in voices of their non-L1 partners.

Table i-4. Responses from Cantonese ESL speakers and L1 speakers of English.

Questions	Responses	
	Cantonese ESL	English L1
Was your partner's voice audible enough in the noise condition?	○ **Yes 77.78%** ○ Not hear clearly 16.67% ○ No 0.00% ○ *No response* 5.56%	○ **Yes (6/6) 100%** ○ Not clearly 0.00% ○ No 0.00% ○ *No response* 0.00%
Did your partner change his/her voice in noise? If yes, did you notice when that happened?	○ Yes 27.78% ○ **No 61.11%** ○ *No response* 11.11%	○ **Yes 83.33%** ○ No 16.67% ○ *No response* 0.00%

Discussion

The results show that Lombard speech could increase intelligibility. Similarly, Bosker & Cooke (2020) find that the adjustment of speech in noise would benefit intelligibility and discuss the influence of language background on speech-in-noise perception.

Furthermore, nearly two thirds of Cantonese ESL speakers did not detect the change in voice in Lombard speech while most of the L1 speakers of English reported the change in voice, which indicated that the modification strategies could be affected by speakers' L1 language. Since the Lombard speech was produced by L1 speakers of Mandarin and Mandarin and Cantonese are both tone languages, the adjustment in acoustic correlates in noise might be more familiar to Cantonese speakers than to L1 speakers of English. Next, more analysis will be conducted regarding the association of acoustic correlate of Lombard speech by non-L1 speakers and the results of speech-in-noise perception.

Study 4. Perception of affective states in ESL speech.

We conducted a small-scaled perceptual test on affective judgement of ESL speech. The test was completed by two naïve bilingual speakers: French-English and Mandarin-English. The Mandarin listener detected 46% of the speech tokens containing positive emotions and regarded 54% as neutral. The French listener however judged only 15.4% of the tokens as containing positive emotions, 7.7% as negative and 76.9 % as neutral. There seems an effect of shared backgrounds between speaker and listener: Mandarin speakers are more familiar with features in Cantonese ESL speech and tended to transfer their shared linguistic knowledge into judging affects in ESL. Results are also analysed by correlating with phonetic parameters including speech rates and ESL proficiency levels (Figures i-22 and i-23). Utterances with normal speech speed were recognized as positive by the Mandarin listener more often than by the French listener. Meanwhile, utterances by intermediate speakers were more likely recognized as positive by the Mandarin listener than by the French listener.

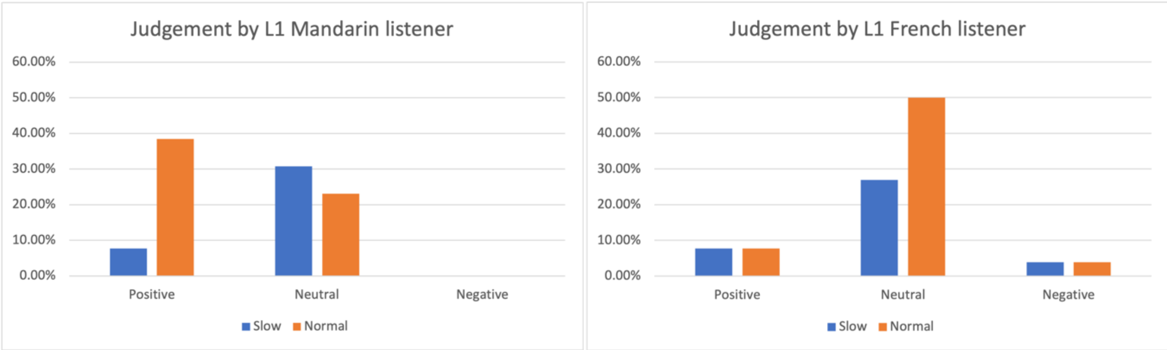


Figure i-22. Results of affective judgements by listeners on ESL speech in different speaking rates.

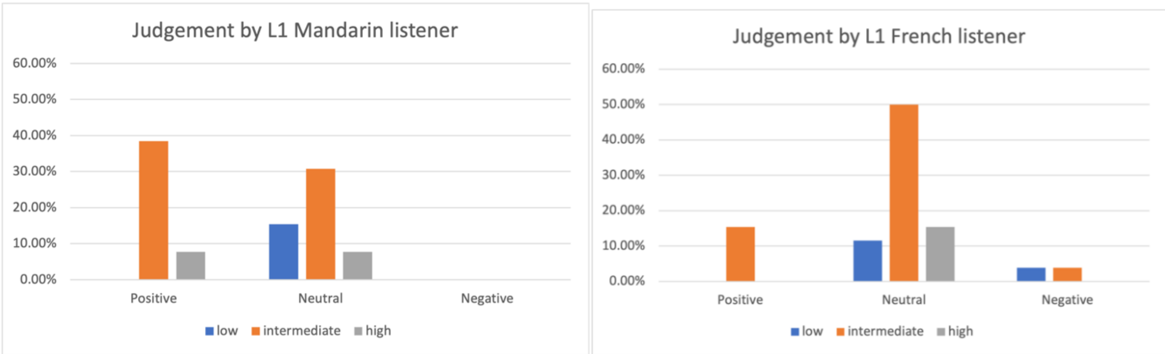


Figure i-23. Results of affective judgements by listeners on ESL speech at different proficiency levels.

Our listeners had similar judgements for utterances in Story 1. They both judged Speaker 78 and Speaker 84' production as containing positive emotions. Additionally, the Mandarin listener found that Speaker 88's production was positive. However, in Story 2, the results from the two listeners were very different. Although they agreed that both Speaker 78 and Speaker 84 expressed positive emotions in storytelling, the Mandarin listener detected greater positivity than the French listener did.

(j) Conclusions and Recommendations

To identify phonetic and phonological characteristics of local workplace English pertaining public health care, we conducted a series of studies to investigated ESL speech production in normal and adverse situations, with a reference to that by L1 speakers of English. We examined phonetic realization of segmentals and prosodic correlates such as duration, F0 and intensity. General phonetic characteristics are identified for Chinese ESL speakers. There are deviations in ESL speech from that by L1 speakers of English. Moreover, as Mandarin and Cantonese ESL speakers each showed systematic variations from the L1 group, the L1 groups themselves also differ in terms of phonetic patterns among major cues in both normal and special situations. Specifically, the two ESL groups' performance in general speaking reveal different strategies in speech production, such as deletion or epenthesis to deal with challenges with final consonant, keeping or withdrawing distances between formant frequencies in diphthong production, and adjusting PVIs to realize contrastivity in lexical stresses.

In noisy conditions, Lombard speech is found in both ESL and L1 groups who seemed to adopt a similar strategy, to primarily modify one major phonetic cue, speech intensity, as a means of strengthening signal transmission in communication. But Cantonese ESL speech production showed differences from that of L1 speakers of English. ESL production displayed traits of their L1 phonology and phonetics, i.e. Cantonese lexical tones and F0 as the major acoustic correlate. They also produced Lombard speech with higher F0 and with more vocal efforts, indicating that they were not familiar with modification strategies involving non-L1 speech under noise. Moreover, our post-study questionnaire reveals that ESL speakers were in general unaware of acoustic changes to voices under noise. Last but not least, ESL speakers' variations on phonetic correlates may lead to different judgements of their affective states. Based on these findings, we propose that phonetic characteristics of ESL production could be due to the influence of L1 phonological systems, to ESL speakers' to-be-honed knowledge in L2 phonology, as well as their strategies in speech accommodation in special conditions. Our results provide baseline data for further examination of ESL pronunciation at workplaces where interfering factors such as noises to speech communication are common.

Recommendations

Our studies identified background noise such as babbles as a common aspect of our daily lives and also in workplaces such as emergency rooms and hospitals. The noises especially multi-talker babble noise which contains speech information can be a grave challenge for ESL speakers. This type of noise contains speech information, and thus can be distracting in communication and also interfere verbal interaction more significantly than other types of background noises. We propose that linguistic and vocational trainings should be introduced to local ESL programmes targeting students who will be working in linguistically-adverse situations or be associated with noisy interactional situations. ESL students need be aware of potential interfering external factors and of effects on their speech communication. They could also benefit from courses and practices on strategies in L1 and L2 speech accommodation, and also on controlling affects in speech communication in adverse conditions.

When demands in workplace training increase, alterL1 efforts can be made to provide graduates with self-learning resources tailored for the specific professional communication. Background noise such as babble noise is often neglected in professional language teaching and training, but is inevitable for most of the actual situations at workplaces where quietness is a luxury but effective and accurate communication is a key to job efficiency. For example, nurses working under the noise in emergency rooms may feel more burdens on achieving effective communication when accomplishing nursing responsibility. Our research indicate that both internal and external factors such as background noise and affective states affect the quality of speech communication even for high intermediate and advanced L2 speakers of English, due to their lacking experience in speech accommodation or in modulating acoustic cues under special situations. Therefore, we recommend that linguistic trainings on effective adjustment and adaptation in speech could be introduced into advanced ESL speaking and listening to facilitate more efficient communication in workplaces. Such training could also benefit from a combination of various speaking environments and speaking styles that increase the variability of spoken English.

(k) Bibliography

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, (52), 317–332.

Arciuli, J., Simpson, B. S., Vogel, A. P., & Ballard, K. J. (2014). Acoustic Changes in the Production of Lexical Stress during Lombard Speech. *Language and Speech*, *57*(2), 149–162.

Ballard, K. J., Djaja, D., Arciuli, J., James, D., & van Doorn, J. (2012). Developmental trajectory for pro- duction of prosody: Lexical stress contrastivity in children 3 to 7 years and adults. *Journal of Speech, Language, and Hearing Research*, *55*, 1822–1835.

Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, *53*, 1227–1245.

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

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program] Version 6.1.08. Retrieved December 5, 2019, from http://www.praat.org/

Bolinger, D. L. (1965). Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (Eds.), *Forms of English: Accent, Morpheme, Order* (pp. 139–180). Cambridge: Harvard University Press.

Bond, M., & Lai, T. M. (1986). Embarrassment and code-switching into a second language. *Journal of Social Psychology*, *126*, 179–186.

Bosker, H. R., & Cooke, M. (2020). Enhanced amplitude modulations contribute to the Lombard intelligibility benefit: Evidence from the Nijmegen Corpus of Lombard Speech. *The Journal of the Acoustical Society of America*, *147*(2), 721–730.

Bradlow, A. R. (n.d.). ALLSSTAR: Archive of L1 and L2 Scripted and Spontaneous Transcripts And Recordings. Retrieved from http://groups.linguistics.northwestern.edu/speech_comm_group/allsstar2/#!/materials.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-L1 listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272–284.

Burzio, L. (1994). *Principles of English stress*. Cambridge University Press.

Castellanos, A., Jose-Miguel, B., & Casacuberta, F. (1996). An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication*, *20*, 23–35.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, *45*, 24–27.

Chao, Y. R. (1947). *Cantonese Primer*. Cambridge, Massachusetts: Harvard University Press.

Chen, Y., Robb, M., Gilbert, H., & Lerman, J. (2001). Vowel production by Mandarin speakers of English. *Clinical Linguistics and Phonetics*, *15*(6), 427–440.

Cheung, W. H. Y. (2009). Span of High Tones in Hong Kong English. *Annual Meeting of the Berkeley Linguistics Society*, *35*(1), 72.

Choi, W., Tong, X. L., & Singh, L. (2017). From lexical tone to lexical stress: A cross-language mediation model for Cantonese children learning English as a second language. *Frontiers in Psychology*, *8*(MAR), 1–9.

Ciocca, V., Wong, L. L. N., & So, L. K. H. (1994). An acoustic analysis of unreleased stop consonants in word-final position. In *ICPhS*.

Cooke, M., & Lecumberri, M. L. G. (2012). The intelligibility of Lombard speech for non-L1 listeners. *The Journal of the Acoustical Society of America*, *132*(2), 1120–1129.

Davies, M. (2008). The Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. *Language* (Second). Oxford University Press.

Duanmu, S., Kim, H.-Y., & Stiennon, N. (2003). Stress and Syllable Structure in English: Approaches to Phonological Variations.

Egan, J. J. (1972). Psychoacoustics of the Lombard voice response. *J Aud Res*, *12*, 318–324.

Fant, G. (1960). *Acoustic Theory of Speech Production* (2nd ed.). 1970 version. Hague.

Fok-Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: Hong Kong University Press.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*, 126–152.

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Lœvenbruck, H. (2006). An acoustic and articulatory study of lombard speech: Global effects on the utterance. In *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP* (Vol. 5, pp. 2246–2249).

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of Sound Immersion and Communicative Interaction on the Lombard Effect. *Journal of Speech, Language, and Hearing Research*, *53*(June 2010), 588–608.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, (35), 151 – 166.

Gussenhoven, C. (2012). Tone and intonation in Cantonese English. *Proceedings of The 3rd International Symposium on Tonal Aspects of Languages*, *Nanjing*, *C*, 26–29.

Gussenhoven, C. (2014). On the intonation of tonal varieties of English. In M. Filppula, J.

Klemola, & D. Sharma (Eds.), *The Oxford handbook ofworld Englishes*. Oxford: Oxford University Press.

Hadfield, J. D. (2010). MCMC methods for Multi–response Generalised Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, *33*(2), 1–22.

Hahn, L. D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*, 201–223.

Hung, T. T. N. (2012). Hong Kong English. In *English in Southeast Asia: Features, Policy and Language in Use* (pp. 113–134).

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, *119*(2), 1118.

Johnson, K., Flemming, E., & Wright, R. (1993). The Hyperspace Effect: Phonetic Targets Are Hyperarticulated. *Language*, *69*, 505–528.

Junqua, J. C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, *20*(1–2), 13–22.

Junqua, J. C., Fincke, S., & Field, K. (1999). The Lombard effect: A reflex to better communicate with others in noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 2083– 2086).

Khouw, E., & Ciocca, V. (2006). An Acoustic and Perceptual Study of Final Stops Produced by Profoundly Hearing Impaired Adolescents. *Journal of Speech, Language, and Hearing Research*, *49*(1), 172–185.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, *1*, 85–97.

Lai, Y. W. (2009). *Acoustic Realization and Perception of English Lexical Stress by Mandarin Learners*. Ph.D. dissertation, University of Kansas.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *J Speech Hear Res*, *14*, 677–709.

Lenth, R. V. (2016). lsmeans: Least-Squares Means. R package version 2.22. https: //CRAN.R-project.org/package=lsmeans.

Li, B., Wayland, R., & Zhang, C. (2010). Contextual information in (non)L1 speech perception in noise. In *Proceedings of the Third ISCA Tutorial and Research Workshop on Experimental Linguistics* (pp. 25–27).

Li, J. J., & Grigos, M. (2018). The Effect of L1 Prosody in the Perception and Production of Non-L1 Lexical Stress. In *9th International Conference on Speech Prosody 2018* (pp. 898–902). ISCA.

Lombard, E. (1911). Le signe de l'élévation de la voix [The sign of the rise of the voice]. *Maladies Oreille, Larynx, Nez, Pharynx*, (27), 101–119.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*, 377–401.

Luke, K. K. (2008). Stress and intonation in Hong Kong English. In *14th IAWE Conference*.

Hong Kong.

Luo, J., Li, V. G., & Mok, P. P. K. (2020). The Perception of Cantonese Vowel Length Contrast by Mandarin Speakers. *Language and Speech*, *63*(3), 635–659.

Marcoux, K., & Ernestus, M. (2019). Differences between L1 and non-L1 Lombard speech in terms of pitch range. In *PROCEEDINGS of the 23rd International Congress on Acoustics* (pp. 5713–5720).

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686–693.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and specking style in English stressed vowels. *Journal of the Acoustical Society of America*, *96*, 40–55.

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*(3), 479–502.

Neu, H. (1980). Ranking of constraints on /t,d/ deletion in American English. In W. Labov (Ed.), *Locating language in time and space* (pp. 37-54)). New York: Academic Press.

Nguyễn, T., Ingram, C. L. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158–190.

Nolan, F., & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica*, *66*(1–2), 64–77.

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America*, *24*(2), 175–184.

Picheney, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434–446.

Qin, Z., Chien, Y. F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics*, *38*(3), 541–570.

R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*(3), 465–485.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Smiljanic, R., & Bradlow, A. (2010). Teaching and Learning Guide for: Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, *4*(3), 182–186.

Summers, W. Van, Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the*

Acoustical Society of America, *84*(3), 917–928.

Suzuki, T., Nakqjima, K., & Abe, Y. (1994). Isolated word recognition using models for acoustic phonetic variability by lombard effect. In *ICSLP* (pp. 999–1002).

Takizawa, Y., & Hamada, M. (1990). Lombard speech recognition by formant-frequency-shifted LPC cepstrum. In *Proc. Int. Conf. Spoken Language Proc.* (pp. 293–296).

Theodoropulos, C. (2014). Prosody in the production and processing of L2 spoken language and implications for assessment. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, *14*(2), 1–20.

Thomas, E. R. (2002). Instrumental phonetics. In J. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 168–200). Oxford, MA: Blackwell.

Tong, X. L., Lee, S. M. K., Lee, M. M. L., & Burnham, D. (2015). A Tale of Two Features: Perception of Cantonese Lexical Tone and English Lexical Stress in Cantonese-English Bilinguals. *PLoS ONE*, *10*(11), 1–33.

Tonkinson, S. (1994). The Lombard effect in choral singing. *Journal of Voice*, *8*(1), 24–29.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*, 1429–1443.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, *88*, 97–100.

Traunmüller, H. (1997). Auditory scales of frequency representation. [Available online: http://www.ling.su.se/staff/hartmut/bark.htm].

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11–12), 943–953.

Wang, Q. (2008). *Perception of English stress by Mandarin Chinese learners of English: An acoustic study*. Ph.D. dissertation, University of Victoria.

Wang, X. C., & Munro, M. J. (1999). The perception of English tense-lax vowel by L1 Mandarin speakers: The effect of training on attention to temporal and spectral cues. In *ICPhS99* (pp. 125–128).

Wee, L. H. (2016). Tone assignment in Hong Kong English. *Language*, *92*(2), e67–e87.

Yu, V. Y., & Andruski, J. E. (2010). A Cross-Language Study of Perception of Lexical Stress in English. *Journal of Psycholinguistic Research*, *39*(4), 323–344.

Zhang, Y. H., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by L1 Mandarin speakers. *The Journal of the Acoustical Society of America*, *123*(6), 4498–4513.

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, *37*(2), 231–247.

Zollinger, S. A., & Brumm, H. (2011). The Lombard effect. *Current Biology*, *21*(16), R614–R615.

(l) Appendices, if any

Appendix 1: Language background questionnaire in Study 2 and summary of responses.

- Age, gender, self-rated proficiency:

First language	Age (years)	Female/Male	English Proficiency
Cantonese (n=18)	25.6 (SD=5.8)	8/10	12/6 (advanced/intermediate)
English (n=6)	28.3 (SD=11)	1/5	N.A.(1 American, 2 British, 1 Canadian, 2 South African)

- Profession:
  Twelve Cantonese ESL speakers were students, six employees. All L1 speakers of English were students.
- Degree program:

	Nursing/pharmacy	Languages	Engineering	Business	other
L1CAN (n=18)	3	3	3	2	7
L1ENG (n=6)	0	0	2	1	3

- Work experience or nature of jobs:
  - Cantonese ESL speakers: TUNS (Hospital Authority), 4 weeks in rehabilitation ward, Assistant Education Consultant, Film making, Research intern, Student helper, Bright Funture ltd, Vitagreen, The Hong Kong and China Gas Company Limited (Purchasing Officer), Research Assistant, Clerk Assistant, Sales Assistant, Educational Consultant, Translator, rugby coach, insurance , IT project management, Project Assistant at University, Clerk, part time writer, Project Assistant, Clerical Officer, administration.
  - English L1 speakers: site inspector, bank employee, NGO clerk, Fire Compliance Auditor, retailing sales.
- Language status at workplace for Cantonese ESL speakers

Language	major	secondary	not frequently used
Cantonese	12	6	0
English	6	10	1
Mandarin	0	1	9

- Other languages self-reported as known or in use: CAN_13 speaks French and CAN_18 speaks Korean, both at intermediate level. ENG_1 is an advanced speaker of Vietnamese. ENG_4 is an advanced speaker of German. ENG_6 is an advanced speaker of Japanese. ENG_3 speaks Spanish, Cantonese, Korean and Japanese, all at intermediate level.
- Length of living in English-speaking region: CAN_4 spent 5 years in US and CAN_17 spent 8 years in Canada.
- First languages: Among the 18 L1 speakers of Cantonese, three recognised themselves as bilinguals of Cantonese and Mandarin (CAN_4, CAN_9 and CAN_12).
- English Proficiency (intermediate/advanced/beginning): Twelve recognised themselves as advanced speakers in English and six as intermediate speakers. Three recognised themselves as advanced speakers in Mandarin and nine as intermediate speakers.
- Scores in English language tests:

DSE-5	DSE-4	DSE-3	A-level E	No report
7	5	3	1	2

- Medium of Instruction (primary, secondary and tertiary levels): Mostly Cantonese in primary schools; either English or Cantonese in secondary schools; mostly English in college and universities.
- HKEAA benchmarking studies on international examinations

IELTS (2004)

HKEAA conducted a research to equate performance levels on the HKCEE and HKALE to that of the IELTS. The two tables below express the grade levels of these two examinations in respect to the IELTS band scores. They are based on equi-percentile equating for stratified samples of candidates taking the 2004 HKCEE and HKALE.

Grade	Equivalent Range of Overall IELTS Band Score for HKCEE English Language (Syllabus B)	Equivalent Range of Overall IELTS Band Score for HKALE Use of English (AS-Level)
A	6.85 - 8.10	7.41 - 8.30
B	6.41 - 6.84	6.92 - 7.40
C	5.92 - 6.40	6.51 - 6.91
D	5.32 - 5.91	6.03 - 6.50
E	4.50 - 5.31	5.40 - 6.02

Home > Recognition of Exam > Benchmarking Studies on International Examinations > HKDSE

International English Language Testing System (IELTS) (2012)

The HKEAA has conducted a benchmarking study to compare the standards between the International English Language Testing System (IELTS) and the 2012 HKDSE English Language Examination. The tables below show the equivalent of level 2 to level 5** in HKDSE with respect to IELTS band scores.The results were developed based on the stratified samples of candidates taking the 2012 HKDSE.

HKDSE English Language Subject Level	Overall IELTS Band Score Equivalent Range
5**	7.51 – 7.77
5*	7.16 – 7.32
5	6.81 – 6.99
4	6.31 – 6.51
3	5.48 – 5.68
2	4.79 – 5.07

https://www.hkeaa.edu.hk/en/recognition/benchmarking/ce_al/ielts/
https://www.hkeaa.edu.hk/en/recognition/benchmarking/hkdse/ielts/
https://www.ielts.org/-/media/pdfs/comparing-ielts-and-cefr.ashx
https://ibo.org/news/news-about-the-ib/benchmarking-diploma-programme-language-courses-to-the-cefr/
“International Baccalaureate Diploma Programme” (IBDP)

Appendix 2: Interactive tasks used in Experiment 3 of Study 2

(1) First task : Giving directions

Participants were instructed to give directions to a partner (an experimenter) using directory signs provided (as shown at bottom left). They needed use a complete sentence. For example,

- *Experimenter: Excuse me. Could you tell me the way to the Canteen?*
- *Participant: Sure, the canteen is in B Block, on LG1.*

(2) Second conversation: to explain a letter

Participants were asked to explain the content of a referral letter (as shown at bottom right) to his/her partner (an experimenter) and answer questions. They needed provide all details of this letter. The experimenter elicited uses of target words with the following questions:

- *Could you please tell me what is this letter?*
- *From which department did I receive this letter?*
- *Where am I supposed to bring this letter to?*
- *What kind of problems do I have according to the doctor?*

A Block	FLOOR	B Block
	5	Eye Centre
Renal Centre	4	
Obstetrics & Gynaecology	3	
	2	Physical Therapy
Operating Theatre	1	
Emergency & Reception	G	Pharmacy & Cashier
	LG1	Canteen
MRI Centre	LG2	

醫院管理局

Referral letter
XXX Hospital
Surgery
Tel: XXXXXXX
06/06/2020
Case Nº: XXXXXXX
[Referral Nº: XXXXX]

To: Department of Physical therapy

Dear Consultant in-charge,

Re: XXX Sex: F Age: 43 y

Reason for referral: Chronic pain

This above-named patient had unremarkable past health.
She was admitted for chronic pain in knees and back with suspected infection and vitamin D deficiency.
...

Please follow-up and offer your expert opinion. Thank you!

Signature: ____________________

Name in Block letters: XXXXXXX
RESIDENT
Department of Surgery

Examples of directory signs (Left) and referral letters (Right) used in Experiment 3.

Appendix 3: Survey questions in Study 3 and summary of response.

Questions:

1) Could you hear your own voice in the noise condition? Did you ever consciously try to change your voice in this condition? If yes, do you remember when it happened? At what moment?
2) Was your partner's voice audible enough in this condition? Did your partner change her voice in this condition? If yes, do you remember when it happened? At what moment?
3) Is the noise you heard in the experiment similar to that of your workplace environment? (Yes / No, the noise used in this experiment is louder. / No, the noise used in this experiment is quieter).
4) Please indicate how uncomfortable it was to speak in this condition using the following scale: from 1 = very uncomfortable to 5 = not uncomfortable at all.
5) Please indicate how fatigued your voice was in this condition using the following scale: from 1 = very fatigued to 5 = not fatigued at all.

	ESL Speakers	**NL speakers**
Could you hear your own voice in the noise condition?	**Yes : 83.33%** No: 11.11%	**Yes: 100%** No: 0%
Did you ever consciously try to change your voice in this condition?	**Yes: 66.67 %** No: 33.33%	Yes: 33.33% **No: 66.67%**
If yes, at what moment?	**when I can't hear my own voice** (17%) for the first few words (8%) no comment (75%)	when need to hear myself clearly (33%) No comment (67%)
If yes, how?	**louder** (50%) **louder and slower** (17%) no comment (33%)	**increase volume (17%)** No comment (83%)
Please indicate how fatigued your voice was in the noise condition.	Very uncomfortable (0) Fatigued (27.78%) **Fair (50%)** Less fatigued (22.22%) Not uncomfortable (0%)	Very fatigued (0%) Fatigued (33%) **Fair (50%)** Less fatigued (17%) Not fatigued (0%)

19

Appendix 4: Project Deliverables and Outcome in the reported period

- Dataset of English and Cantonese Lombard Speech

The Dataset of English and Cantonese Lombard Speech corpus contains recordings in quiet environment and noisy environments from 18 L1 speakers of Cantonese and 6 L1 speakers of English. L1 speakers of Cantonese produced Lombard speech in Cantonese and English while L1 speakers of English produced English Lombard speech.

Details of Recordings

Talkers: 18 L1 speakers of Cantonese and 6 L1 speakers of English

Languages: L1 Cantonese, L1 English and L2 English

Items and condition:

(a) 1368 L2 English sentences and 456 L1 English sentences in noise condition;

(b) 1368 L2 English sentences and 456 L1 English sentences in quiet condition;

(c) 1080 L1 Cantonese sentences in noise and 1080 in quiet;

(d) 24 spontaneous conversations in noise and 24 in quiet.

Annotation

Word and phoneme aligned TextGrids. The materials contain English and Cantonese simple sentences with similar syntactic structure. The details could be found in the Materials section of Study 2. Regarding the spontaneous conversations, dyads of ESL talkers with HK Cantonese as L1 were invited to conduct scripted conversation simulating inquiries and answers in a hospital.

- Publication and Presentation

a) IPA Guide for Visually Impaired Learners

Yan, X. J., Glynn, D., **Li, B.**, Stevenson, S., Jing, X. (2020). *IPA guide for visually impaired learners*. Hong Kong: City University of Hong Kong & The Hong Kong Society for the Blind. https://julac.hosted.exlibrisgroup.com/permalink/f/iffgrl/CUH_IZ21574917070003408

We collaborated with the Hong Kong Society for the Blind to compile an International Phonetic Alphabet (IPA) guide for visually impaired (VI) learners of English. The IPA guide provides accessible information of IPA symbols and phonetics for English speech sounds. There are examples of vowels and consonants and more importantly easily-understood description for VI learners to grasp shapes and meanings of IPA symbols. For instance, the /ə/ sound, known as the schwa as in words like 'sofa', is described in the guide as "a 180-degree rotated 'e'". There is also tactile presentation of each IPA symbol for the VI ESL learners. They can hone their pronunciation in English and proceed with more in-depth knowledge in phonetics if they desire such learning.

b) Empirical verification of alterL1 approaches to experiments in phonetic laboratories.

Guan, Y. and **LI, B.** (2021). Usability and practicality of speech recording by mobile phones for phonetic analysis. Full paper in *Proceedings of the 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 24-26 Jan. 2021. Hong Kong.

As a situational response to inaccessibility of campus facilities due to Covid-19 and also strategic development in experimental paradigms in post-pandemic situation, we assessed reliability and validity of speech recordings collected by using non-professional devices and by non-linguistically trained persons. Non-professional devices such as mobile phones and laptops were examined with a reference to professional recording equipment. Results in acoustics parameters confirm feasibility and reliability of our proposed alterL1 approach. The outcome also provides baseline information for future phonetic analysis of general purposes.

c) Professional English training

LI. B., Guan Q., Fan Z., & He, M. (2021). Phonetic modification in English under noise by Chinese ESL learners. Presentation at *The HKCPD International Conference 2020 for English teaching professionals worldwide*. January 8-10, 2021, Hong Kong, China.

Our study and findings on ESL speech production and perception under noise were presented to teachers and researchers in English for academic and special purposes. We discussed the pedagogical implications from our research to training of workplace English at tertiary levels in local contexts. The experiment design and results provide reference to development of pedagogical materials and training paradigms that target at introducing strategies and methods in effective adjustment and adaptation in speaking under noise among advanced ESL learners.