USE OF PHONETIC CUES BY HONG KONG L1 CANTONESE LEARNERS IN L2 ENGLISH WORD STRESS PRODUCTION AND PERCEPTION

Wience Wing-sze Lai

College of Professional and Continuing Education, The Hong Kong Polytechnic University, Hong Kong
ccwience@hkcc-polyu.edu.hk

ABSTRACT

This study examines L2 English word stress production and perception by Hong Kong L1 Cantonese (CS).

Experiment 1 examined 22 CS's production of seven disyllabic English noun-verb pairs with an acoustic study and a listening evaluation test. Experiment 2 assesses the accuracies of 42 CS's perception of an English noun-verb pair of "contract" with F0, intensity or duration in either syllable manipulated.

A two-way ANOVA revealed significant main effects for nativeness/proficiency and for acoustic cues in the acoustic study of Experiment 1 and in Experiment 2.

CS's consistent F0 dependence but intensity negligence suggested a possible perception-production link and L1 influence. Highly proficient CS's F0 overuse, and better intensity and duration use suggested the effect of L2 proficiencies. These findings may inspire English word stress teaching and future studies on English prosody acquisition.

Keywords: L1 Cantonese, L2 English, word stress, perception, production, phonetic cues.

1. INTRODUCTION

This study extends from [9] which investigated the tonal assignment in Cantonese loanwords borrowed from English in investigating whether Hong Kong L1 Cantonese learners of L2 English rely mainly on F0 to denote English word stress in their production and whether such ability is related to their perception of English word stress.

While there appears to be a close connection between perception and production in L1 acquisition, previous studies such as [1, 3, 4-6, 11, 12] reported discrepant findings regarding the relationship between perception and production in second language acquisition. In addition, most perception studies focused on segmental features and most studies on Chinese learners of English targeted at L1 Mandarin learners [c.f. 2, 7, 8, 10, 13-15].

With extensive studies on the perception and production of English word stress by L1 Mandarin learners [c.f. 10, 14] but few on those by L1 Cantonese learners [c.f. 7, 8], there is a need for the current study which focused on suprasegmental features and targeted at L1 Cantonese learners, addressing previous findings on perception-production link, tone-stress relationship, L1 influence and L2 proficiencies.

1.1. Implication drawn from previous literature

Both as suprasegments, tone and stress use F0 as the common acoustic cue but differ in that the latter also uses intensity and duration (and vowel quality). It is thus expected that tone language users, such as L1 Cantonese and Mandarin learners of L2 English, differ from L1 English speakers in English word stress production and perception. L1 Mandarin learners' reliance on F0 as in [10, 14, 15] in production and/or perception of English word stress, and L1 Cantonese learners' experience with tone in English word stress perception as in [2] and their tone assignment in English stress production as in [13] suggest F0 as the dominant cue for stress production and perception among these tone language users. L1 Mandarin learners' negligence of intensity in English word stress perception as discovered in [15] also led us to question about Cantonese ESL learners' neglected cue(s) for comparison.

1.2. Research questions

This study thus aims to answer the following research questions:

- Do Hong Kong CS (highly proficient ones, HCS, and less proficient ones, LCS) and native English speakers (NS) produce English word stress differently in terms of fundamental frequency (F0), intensity and duration?
- Do Hong Kong CS (HCS and LCS) and NS perceive English word stress differently in terms of F0, intensity and duration?
- Is Hong Kong CS' production of L2 English word stress related to their perception?

2. METHODOLOGY

To answer the above research questions, two experiments were designed and implemented. Experiment 1 investigated the production of L2 English word stress by Hong Kong L1 Cantonese learners (CS), while Experiment 2 investigated the perception. All participants were ranged from 18 to 55 years of age. They all had normal speech and language ability by self-report. The target CS participants were recruited from Hong Kong Community College (HKCC) of The Hong Kong Polytechnic University (PolyU). They all gained prior English experience from classes in Hong Kong. The control L1 English (NS) participants were born and brought up in English-speaking countries such as the United Kingdom and the United States of America, and they were still living there or temporarily staying in Hong Kong.

2.1. Experiment 1: Production of disyllabic noun-verb pairs

Experiment 1 examined 22 CS participants (11 HCS and 11 LCS) and 14 NS participants' production. The speech samples produced by the participants were recorded by using audacity installed in a laptop computer. The recording was done in a quiet room where a high-quality unidirectional dynamic microphone was fixed at a distance of 10 cm from the participant's mouth to ensure the quality and consistency of the recording.

Experiment 1 was designed with reference to [10] and [15], both of which examined the perception and production of L2 English word stress by L1 Mandarin learners. All participants were asked to pronounce seven disyllabic word pairs for three times (see Table 1). Each word pair consists of a noun and a verb with identical spelling but different stress patterns, that is with the stress on the first syllable for nouns as in CONtract AND OBject, but with the stress on the second syllable for verbs as in conTRACT and obJECT. These stimulus pairs were formed from the following corpus of word forms: contract, desert, object, permit, rebel, record, and subject. Each target word was produced in three ways: (i) in isolation, (ii) in the semantically neutral frame sentence "I said __ this time", and (iii) in contextualised sentences created specifically for each word, as shown in the rightmost column of Table 1.

The produced tokens were first analysed with Praat to obtain F0 (Hz), intensity (dB) and duration (ms) measurements, which were then statistically analysed and compared.

Table 1: Stimuli of 7 pairs of disyllabic words and their corresponding contextualised sentences.

	Target Word	Noun/ Verb	Contextualised Sentence
1	contract	noun	Mr. Smith has finally agreed to sign the new *contract*.
		verb	Will steel *contract* when it is cooled?
2	desert	noun	They got lost in the *desert*.
		verb	Will he *desert* his team?
3	object	noun	What is the *object* on the table?
		verb	They won't *object* to your decision.
4	permit	noun	In order to park here, you need a *permit*.
		verb	Would you *permit* her request?
5	rebel	noun	The *rebel* army did this.
		verb	They *rebelled* at this unwelcome suggestion.
6	record	noun	Can I get a copy of my health *record*?
		verb	She *recorded* all songs her daughter sang yesterday.
7	subject	noun	What is the *subject* of this sentence?
		verb	Must you *subject* me to this boring twaddle?

Note. The disyllabic word pairs and contextualised sentences were adapted from [15].

A listening evaluation test was also conducted among 13 native English listeners with no previous phonetics training and 30 CS listeners (23 HCS, 4 LCS and 3 with moderate English proficiency) who did not participate in the acoustic studies.

During the experiment, all listeners were seated at a computer wearing headphones playing sounds at 60-65 dBA at adjustable volume in a sound-treated room. The sound samples were presented randomly to rule out any possible order effects.

For each token produced by CS speakers for the acoustic study, listeners first heard the word and then determined which word, for example, either conTRACT (verb) or CONtract (noun), they thought

had been said. Before the actual experiment, listeners were given a standard handout and explained by the same researcher the procedures, the acoustic cues for stress (pitch, loudness and duration) and the stress shift rule (the first syllable stressed in a disyllabic noun, e.g., CONtract, and the second syllable stressed in its disyllabic verb counterpart, e.g., conTRACT). A practice session identical to the actual experiment but composed of only eight questions using eight different tokens of the same word "contract", not used in the actual experiment, was then provided to familiarise listeners with the pace and format of experiment.

2.2 Experiment 2: Perception of a disyllabic noun-verb pair

In Experiment 2, participants underwent a stress identification test. The goal was to compare the relative importance of perceptual cues for English stress to the CS participants. The entire perceptual experiment was carried out in a sound-treated language lab. Participants were instructed to listen to 165 tokens of the English word "contract" in one block in a randomised order, including 150 different tokens {F0 [CON] (5) × F0 [tract] (5) + intensity [CON] (5) × intensity [tract] (5)] + duration [CON] (5) × duration [tract] (5) + F0 [con] (5) × F0 [TRACT] (5) + intensity [con] (5) × intensity [TRACT] (5)] + duration [con] (5) × duration [TRACT] (5)} plus 15 (ten percent) repeated for intra-reliability assessment. The tokens combined the syllables "con" and "tract" with natural and modified intensity, F0 and duration from both the noun "CONtract" and the verb "conTRACT" produced by a female native British English speaker. F0, duration, intensity values were first obtained from the natural production of the stressed "CON" in the noun "CONtract" and the unstressed "con" in the verb "conTRACT". Since acoustic cues, including F0, intensity and duration, for listeners' perception of English word stress were compared, other cues were held constant with values being averaged from the stressed and unstressed into five levels for concatenation with the five levels of the stressed "TRACT" and the unstressed "tract". Syntheses were executed to alter pitch or intensity, or lengthen syllables, only one parameter at a time. All of the stimulus manipulations were made with the use of Praat functions "To Manipulate - multiply pitch frequencies", "Modify - scale intensity" or "Convert - lengthen" respectively.

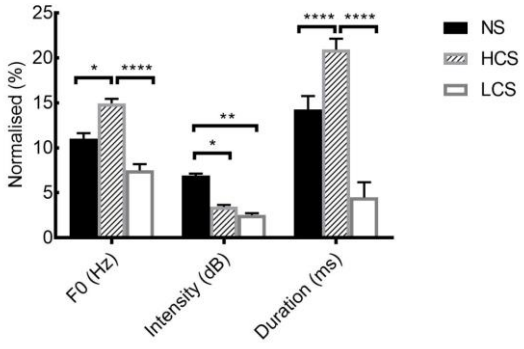
An answer sheet was provided on which the participants had to indicate whether each concatenated word token was a noun or a verb. Audio signals were presented via high-quality headphones to the participants at 65-70 dBSPL. There was a 1500-ms interval between two tokens to allow sufficient time for the participants to make a judgment, and the following token were played after three seconds. The experiment took 20-30 minutes per subject. Due to the demand for high concentration of the task, the participants were allowed to take breaks during the experiment. A brief practice period was provided such that the participants could familiarise themselves with the experimental procedure and environment. Five audio samples were randomly selected from the data corpus and used for practice.

3. RESULTS AND DISCUSSION

3.1 Experiment 1: Production of disyllabic noun-verb pairs

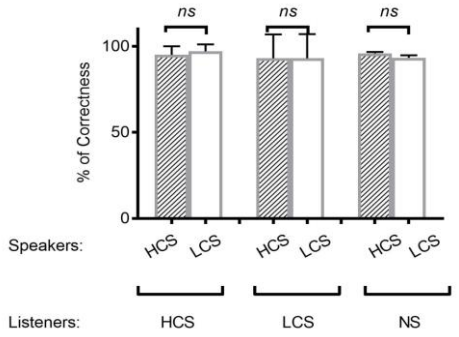
The acoustic results of Experiment 1 revealed duration as the most dominant cue while intensity as the least dominant for both NS and CS speakers in English word stress production. HCS speakers even appeared to overuse duration and F0 while LCS speakers tended to underuse these two cues when attempting to contrast different stress patterns. A two-way ANOVA also revealed a significant main effect for nativeness/proficiency [F (2, 99) = 60.85, $p < 0.0001$], and for acoustic cues [F (2, 99) = 73.04, $p < 0.0001$], and a significant interaction effect [F (4, 99) = 18.61, $p < 0.0001$] (see Figure 1).

Figure 1: Percentage difference between stressed and unstressed syllable within a disyllabic word in vowel F0, intensity and duration among NS, HCS and LCS speakers.



The listening evaluation test shows little difference in mean percentage of correct distinction between HCS and LCS speakers in all listeners' ratings. A one-way ANOVA revealed no statistically significant differences between group means [F (3, 82) = 0.5902, p = 0.6231] (see Figure 2).

Figure 2: Percentage of correct distinction associated with HCS, LCS and NS listeners between disyllabic noun-verb pairs produced by HCS and LCS speakers.

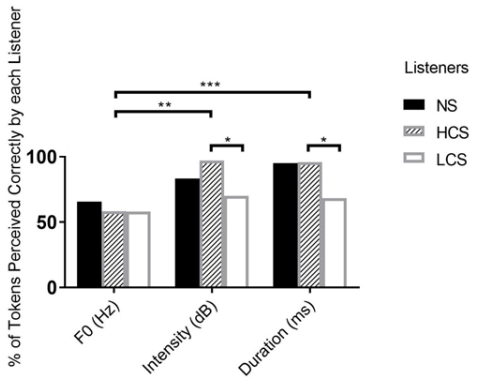


The acoustic study of Experiment 1 showed noticeable difference between HCS and LCS speakers' productions in terms of the use of vowel duration and F0 (but not intensity) within a disyllabic word to distinguish the stressed and unstressed syllables. On the contrary, the results of the listening evaluation task showed no significant difference between HCS and LCS speakers' production to all listener groups.

3.2 Experiment 2: Perception of a disyllabic noun-verb pair

Experiment 2 discovered that the accuracies for both CS and NS participants were the lowest in the abnormal case when only F0 was manipulated to be three levels lower in the stressed syllable than in the unstressed. This suggested F0 as the main acoustic cue for stress perception. This was particularly true for CS listeners as their accuracies were significantly lower with F0 manipulation than with intensity manipulation. A two-way ANOVA revealed significant main effects for nativeness/proficiency and for acoustic cues [F (2, 78) = 9.023, p = 0.0003]) (see Figure 3).

Figure 3: Percentage of tokens, with manipulated F0, intensity or duration values three levels higher in the unstressed syllable than in the stressed, perceived correctly by each listener.



What appeared to be common between stress production and perception as identified in the acoustic studies of Experiments 1 and 2 was that intensity seemed to be the least dominant cue.

3.3 Answers to research questions

CS speakers tend to use intensity less than NS speakers do, and CS speakers tend to overuse or underuse F0 and duration as compared to NS. With NS's production regarded as the standard, CS speakers in general fail to use all the three target acoustic cues properly, either overusing or neglecting them in English word stress production. CS's L1 influence thus serves to explain such difference in production. HCS speakers tend to overuse F0 and duration while LCS tend to underuse duration. Also, HCS speakers tend to use intensity better than LCS do. This shows that different English proficiencies do affect one's production.

CS listeners tend to rely on F0 more than NS listeners do. It could be because NS listeners spread their attention across the three acoustic cues for English word stress perception while CS listeners tend to be more sensitive to F0 for its common use in both lexical tone and word stress at the suprasegmental level. These suggest that L1 influence attributes to such difference in perception. HCS listeners resembled NS listeners in English word stress perception, but LCS listeners use intensity and duration less better than HCS and NS.

The results of the two experiments suggested that the stronger mastery of F0 and the weaker mastery of intensity established a link between CS participants' English word stress perception and production. Future studies should also include the vowel quality cue and investigate the relationship of vowel duration with F0 and speech rate.

4. CONCLUSION

CS participants' reliance on F0 in English word stress perception accounted for their capable use of F0 in the production. Their negligence of intensity in English word stress perception accounted for their less capable use of intensity in the production. In addition, HCS participants tended to overuse F0 in both perception and production, and use intensity and duration better than their LCS counterparts. Thus, both L1 influence and L2 proficiencies contribute to the perception-production link in CS learners' stress acquisition.

This study may inspire English teachers to develop a more effective English word stress teaching regimen and serve as a basis for future studies on CS's acquisition of English prosody such as speech rhythm, sentential stress and intonation.

5. REFERENCES

[1] Best, C. T. 1995. A direct realistic view of cross-language speech perception. In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross-language research*, Baltimore, MD: York Press.

[2] Chan, M. K. K. 2007. *The Perception and Production of Lexical Stress by Cantonese Speakers of English*. M.Phil Dissertation, The University of Hong Kong, Hong Kong.

[3] Diehl, R. L., & Kluender, K. R. 1989. On the objects of speech perception. *Ecol. Psychol.* 2, 121–44.

[4] Fowler, C. A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3-28.

[5] Kluender K. R. 1994. Speech perception as a tractable problem in cognitive science. In Gernsbacher, M. A. (ed), *Handbook of psycholinguistics* (pp. 173-217). San Diego, CA: Academic.

[6] Kuhl, P. K. 1988. Auditory perception and the evolution of speech. *Human Evolution*, *3*, 19-43.

[7] Lai, W. W. S. and Ng, M. 2016. "Different acoustic cues for emphasis in teaching English word stress to Hong Kong Cantonese ESL learners of different proficiencies," in *Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics (ExLing2016), St. Petersburg*, Russia, pp. 91-94.

[8] Lai, W. W. S., Ng, M. 2015. *An Acoustic Account of Hong Kong Cantonese ESL Learners' Production of English Word Stress*. Oral session presented at APCSLH2015 - the Asia Pacific Conference for Speech, Language and Hearing, October 9, Guangzhou, China.

[9] Lai, W.S. 2005. *Tone-stress Interaction: A study of English Loanwords in Cantonese.* M.Phil Dissertation. Hong Kong: The Chinese University of Hong Kong.

[10] Lai, Y. 2008. *Acoustic Realization and Perception of English Lexical Stress by Mandarin Learners*. PhD Dissertation. Kansas: University of Kansas.

[11] Liberman, A. M., & Mattingly, I. G. 1989. A specialisation for speech perception. *Science*, *245*, 489-494.

[12] Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review, 74*, 431-461.

[13] Luke, K. K. "Phonological Re-interpretation: The Assignment of Cantonese Tones to English Words," *ICCL-9 Conference Paper*. Singapore: National University of Singapore, 2000.

[14] Wang, Q. 2008. *Perception of English stress by Mandarin Chinese learners of English: An acoustic study*. PhD Dissertation. Canada: University of Victoria.

[15] Zhang, Y. 2008. *Production and Perception of English Lexical Stress by Native Mandarin Speakers.* PhD Dissertation. US: Purdue University.