Production of English connected speech processes: an assessment of Cantonese ESL learners’ difficulties obtaining native-like speech

Simpson W. L. Wong, Jessica Dealey, Vina W. H. Leung & Peggy P. K. Mok


To link to this article: https://doi.org/10.1080/09571736.2019.1642372

Published online: 28 Jul 2019.
Production of English connected speech processes: an assessment of Cantonese ESL learners’ difficulties obtaining native-like speech

Simpson W. L. Wong a, Jessica Dealeymb, Vina W. H. Leungc and Peggy P. K. Mokd

aDepartment of Education Studies, Hong Kong Baptist University, Kowloon Tong, Hong Kong; bDepartment of Theoretical and Applied Linguistics, University of Cambridge, Cambridge, UK; cDepartment of Psychology, The Education University of Hong Kong, Taipo, Hong Kong; dDepartment of Linguistics and Modern Languages, Chinese University of Hong Kong, New Territories, Hong Kong

ABSTRACT
Despite English being a core and compulsory part of the curriculum for Chinese English as a second language (ESL) learners, it is nevertheless often highly challenging for them. This is partly due to the discrepancies between English’s citation and spoken form and the lack of recognition this pronunciation receives within ESL classrooms. With this in mind, the current study focuses on a highly significant, yet understudied, feature of fluent English speech: connected speech phonological processes (CSPPs). Through analysing the oral recordings of 60 Cantonese ESL university students, the foundation of non-native-speech productions is examined. Results of error analysis highlight the multitude of interrelated sources of non-native-likeness and provide support for previously stipulated causes of non-native-likeness, e.g. L1/L2 phoneme inventory gaps, as well as previously unidentifiable influences, e.g. hypercorrection. The study presents invaluable data and implications regarding the successful assessment, teaching, and learning of English CSPPs and ultimately fluent connected English speech. (151 words)

KEYWORDS
English reduced forms; connected speech; Chinese learners of English as a second language; error analysis; pronunciation

During natural speech production, native English speakers frequently employ phonological processes that alter, blend or entirely remove sounds which are otherwise present within a word’s canonical form (Ernestus 2014; Ladefoged 2000). We refer to these features of natural speech production, occurring both word-internally and at word boundaries, as ‘connected speech phonological processes’ (CSPPs). Table 1 shows definitions and examples of CSPPs which occur in abundance in native English speech. Whilst the extent of modification is arguably partially constrained by contextual predictability (Bell et al. 2009) and word frequency (Bybee 2000), natural spoken output in fact rarely reflects unaltered canonical form, e.g. ‘do you have?’ is frequently reduced to /dʒav/. This feature of the English language can pose challenges for English-as-a-second language (ESL) learners. In this study, we set out to investigate more fully where difficulties lie in relation to CSPPs for ESL learners whose first language is Cantonese, a language typologically distant from English.

Current understanding of Cantonese ESL learners’ CSPP usage

To date, a large proportion of research investigating L2-English speech production appears limited to the assessment of canonical forms and/or focuses generally on L2 learner difficulties without considering the characteristics of CSPP. For example, Chan (2010) analysed Cantonese ESL learners’
difficulties in producing English vowels and consonants, as well as the strategies used to overcome these challenges. English consonants that are absent in Cantonese phonetic inventory (e.g. /ʃ/, /v/) typically pose greater difficulty for Cantonese ESL learners and are often substituted by similar consonants that exist in the native language (e.g. [w] for /ʃ/, [f] for /v/) (e.g. Bolton and Kwok 1990; Hung 2000). Other reported features of Cantonese ESL learners’ English speech production include deletion of final consonants, simplification of consonant clusters (Deterding, Wong, and Kirkpatrick 2008; Peng and Setter 2000), final consonant epenthesis (Setter and Deterding 2003) and vowel epenthesis (Hung 2000). Existing research provides valuable data with regard to the role of the L1, L1/L2 differences and the overall process of L2 phonological learning (see Chan 2006a, 2006b, 2014). However, findings often do not cover CSPP production and little is known on how learners produce natural, fluent, connected English speech. The importance of English CSPPs as a key obstacle in achieving native-like proficiency needs to be recognised, and a greater understanding of ways to overcome these difficulties needs to be developed.

Previous empirical research specifically examining the use of English CSPPs by Chinese ESL learners is particularly scarce. In Alameen’s (2014) study, although Chinese ESL learners were included, the analyses were performed on data from an aggregated group of university students that consisted of Koreans, Malaysians, Spanish, Indians, Portuguese and Arabians. Thus, the influence of Cantonese on English connected speech production was not revealed.

The studies by Mao and Chen (2013) as well as Liang (2015) revealed that the use of reduced forms by this population was limited. Liang’s (2015) analysis of 50 Chinese-speaking university English majors’ speech showed that students produced on average only 64 out of a possible 150 reduced forms. The author identified transfer from the first language as an important factor accounting for learners’ performance. Meanwhile, Mao and Chen (2013) compared the elision of schwa (the unstressed vowel form) by a native English speaker, a speaker of Uyghur and a speaker of Chinese Putonghua (the latter two were students majoring in English). They found that the Uyghur-speaker showed elision of schwa comparable to the native speaker, but the Chinese speaker did not. The absence of schwa elision in the Chinese participant’s speech was attributed again to the interference of first language and the difference in phonological systems of English and Chinese.

Kuo (2009) conducted a training study on primary school students in Taiwan and reported significant improvement in connected speech production after a 14-week training. While providing valuable insights into the production of CSPPs by Chinese ESL learners, these three aforementioned studies were limited in scale with the Chinese-speaking participants restricted to Putonghua-speaking English learners. Research on production of CSPPs by Cantonese Chinese speakers is absent in the literature. Moreover, the types of CSPP concerned in the three existing studies were limited; Liang addressed 4 types of CSPPs while Mao and Chen focused only on elision and Kuo (2009) merely trained features of linking. The current study, based on data from 60 students from a variety of majors, fulfills the need for a larger scale empirical study assessing a range of CSPPs with Cantonese ESL learners.

<table>
<thead>
<tr>
<th>Process</th>
<th>Description/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assimilation</td>
<td>A phoneme becomes, to varying degrees, more phonetically similar to a neighbouring segment. E.g. <em>I heard that speech</em> ... (<em>/ðæt/</em>)</td>
</tr>
<tr>
<td>Contraction</td>
<td>A type of elision involving fixed grammatical patterns and appearing in written forms. E.g. <em>He’ll need</em> ... (<em>/hæt/</em>)</td>
</tr>
<tr>
<td>Elision</td>
<td>Deletion of a phoneme. E.g. <em>... thanked them</em> ... (<em>/ðæŋk/</em>)</td>
</tr>
<tr>
<td>Glottalisation</td>
<td>Replacement of a phoneme with a partial or complete glottal closure. E.g. <em>... we hadn’t watched</em> ... (*/hæt/nt/)</td>
</tr>
<tr>
<td>Linking</td>
<td>Blending of a word-final consonant with the following articulatory event. E.g. <em>... need all his</em> ... (*/nɪːdəl/)</td>
</tr>
<tr>
<td>Vowel Weakening</td>
<td>Reduction of vowels in unstressed syllables to weaker vowels. E.g. <em>... trying to turn</em> ... (*/tɔl/)</td>
</tr>
<tr>
<td>Unreleased Stops</td>
<td>Articulation of plosives without audible release. E.g. <em>... why that child</em> ... (<em>/ðæt/</em>)</td>
</tr>
</tbody>
</table>

Table 1. The processes elicited by native English speakers within the production experiment.

Note. Each process is accompanied by a description and an example taken from the stimuli. The citation form is italicised, followed by native speakers’ productions in parenthesis.
Training of CSPP: difficulties, mechanisms and additional needs

The difficulties of perceiving English connected speech and producing native-like pronunciation have been acknowledged by researchers and educators for three decades in the field of second language learning (e.g. Henrichsen 1984; Practor and Robinett 1985). Shockey (2003) has given a comprehensive overview of the nature of CSPP in her book. Training approaches and materials have been reported in a handful of experimental studies (e.g. Brown and Hilferty 1986; Kuo 2009; Melenca 2001; Sardegna 2011) and summarised in Brown and Kondo-Brown’s (2006) edited book. A set of well-structured training materials is also available in Brown’s (2012) edited book. Despite the availability of these resources, recent studies consistently showed sub-standard performances of connected speech comprehension in ESL learners who have obtained high scores in internationally recognised English tests such as IELTS (Wong et al. 2017a; 2017b). Such inferior listening performances were attributed to the low quality of phonological representations as measured by a part-word recognition test and a spoken vocabulary test (Wong et al. 2017b). Given the low quality of phonological representations of English connected speech, non-native-like English pronunciation is expected in these learners.

Unlike meta-linguistic skills, of which learners can be conscious, phonological representations are operated at the unconscious level and therefore hard to be noticed by the learners and monitored by teachers. As postulated by the Exemplar-based models of connected speech acquisition (Bybee 2001; Hawkins 2003), sufficient exposure to English connected speech is needed for decoding English connected speech. However, given that many learners of English live in a non-immersive environment and rely heavily on visual aids (subtitles, transcription, etc) when learning English through multimedia materials, exposure to native spoken English is far from enough. Alternatively, as inspired by the Abstractionist models of connected speech acquisition (McClelland and Elman 1986; Norris and McQueen 2008), connected speech processing skills can be enhanced by introducing the phonological rules involved to learners. Whilst the aforementioned training materials have provided guidelines for introducing phonological rules and regularity of connected speech phonological patterns, learning about the nature of connected speech does not necessarily improve connected speech perception and production.

Moreover, without considering the influences of L1 phonology on the L2 phonological rules and L2 phonological representations, learners still do not have insights about their level of attainment in connected speech – either partial or even non-learning of connected speech. For example, the CSPP ‘linking’ in English connected speech is processed by a similar CSPP in Cantonese, resulting in listening difficulties and non-native like pronunciation. Only by identifying the interference caused by L1 phonology can ESL learners know where the difficulties lie and therefore allow teachers to develop appropriate strategies to overcome the challenges. The teaching of CSPP-related rules and regular phonological patterns (as implicated by the Abstractionist models) and provision of sufficient and targeted linguistic input (as implicated by the Exemplar-based models) are critical to the acquisition of L2 phonological representations of English connected speech. Extending from these two broad major principles for promoting connected speech acquisition, we conduct this research to identify the specific influences of L1 Cantonese on the acquisition of L2 English connected speech. The results will alert teachers the exact interferences caused by the students’ mother tongue so that teachers can devise respective strategies to help students overcome their listening and speaking challenges. Given the existence of many types of CSPP, we tried to include as many CSPP as possible in the current study. Our data will allow teachers to understand in greater detail the development of Cantonese-speaking students’ English connected speech production.

The influence of L1 on L2 speech learning

Theories regarding the way in which an L2 is learnt, as well as the role of the L1, have been evolving over several decades. One critical issue is whether the similarity between the L2 and L1 phonologies
promotes or hinders L2 phonological acquisition. It has been argued in the Contrastive Analysis Hypothesis (Lado 1957) that where the L1 and the L2 share linguistic and/or cognitive structures, positive transfer is likely to occur, resulting in native-like utterances. In contrast, negative transfer will occur when L1 and L2 phenomena are different, resulting in errors/ non-native-like speech. Thus L1/L2 phonetic inventory divergences are likely to be a significant – some have suggested the most significant – source of learner difficulty. Wode (1986), for example, contended that dissimilar phenomena should be easier to learn than similar ones because the latter are likely to lead to negative transfer and non-learning; for example, similar L1 and L2 phonemes will have low perceptual saliency which makes their differences hard to notice (Major 1987). Young-Scholten (1985) also argued that similar sounds reinforce negative transfer in the long run. It is important to note that contrastive analysis is not employed in the current study for predicting the transfer errors that our participants will make. Instead, contrastive analysis, in the usual way it is applied in studies of second language phonology, is used to explain learners’ errors that teachers may encounter in the English language classrooms (Ohata 2004).

More recently, Flege (1995) has grounded his Speech Learning Model (SLM) in the view that L1/L2 differences are indeed a source of pronunciation difficulties, but accepts that some L1/L2 similarities are also problematic. According to his idea of equivalence classification, similar L2 sounds may be difficult to acquire accurately because a speaker perceives and classifies them as equivalent to those in the L1 and a new phonetic category cannot be established. In contrast, the more dissimilar an L2 segment is to its closest L1 segment, the more likely it is to be successfully produced. The learnability of L2 production is thus dependent on the way in which the learner perceives the L2 in relation to the L1. Studies such as Flege, Takagi, and Mann (1996), Guion et al. (2000) and Lai (2010) support this notion. However, different patterns have been observed at different acquisition stages. For example, Major (1987) showed that beginning and advanced learners performed better on L2 sounds that were similar to their L1. Major’s (2001) Ontogeny Phylogeny Model argues that transfer has different roles to play at various stages of L2 phonological development; the influence of transfer on L2 phonological encoding is greatest at the beginning of the acquisition process, and decreases as L2 phonological competence develops. One particular observation in accordance with this model was that similar sounds between L1 and L2 became more difficult to acquire in the course of development. Major’s (1987) earlier Similarity Differential Rate Hypothesis (SDRH) had further suggested that dissimilar phenomena are acquired at faster rates than similar phenomena.

The aforementioned L2 speech learning models have typically been formulated to reflect, and advance, the study of the perception and production of consonants and vowels in their canonical/citation form. There has been little focus extended to CSPPs. Yet perceiving and producing L2 connected speech is arguably more challenging for learners because the speech signal in connected speech is in effect degraded, providing less salient features for differentiating phonetic contrasts. Assuming this, it seems likely that the L2 phonological knowledge typically acquired by EFL learners may not extend to English CSPP, leading to reliance on L1 knowledge to support connected speech production. We thus hypothesise that the L2 English connected speech produced by Cantonese speakers is likely to be characterised by L1 Cantonese features, particularly as there is a significant typological distance between the two languages.

The difference between Cantonese and English phonological systems

English has 24 basic consonants, 12 pure vowels and 8 diphthongs while Cantonese Chinese has 19 basic consonants, 8 pure vowels and 10 diphthongs (Chan and Li 2000). Certain contrasts in English, including the voicing contrast in consonants and tense-lax contrast in vowels, are absent in Cantonese. Certain English consonants and vowels are absent in Cantonese, such as /θ/, /l/, /r/, /æ/, /ɪ/, /ɒ/. These phonemes are known to cause difficulties for Cantonese ESL learners to produce, and consequently, similar Cantonese phonemes (e.g. /ʃ/, /s/, /w/, /ɛ/, /i/, /ɒ/) are often used as substitutes (Bolton and Kwok 1990; Chan 2006a, 2006b; Chan and Li 2000; Hung 2000; Kenworthy 1986).
English tense-lax vowels (e.g. /iː/ and /uː/; /ɔː/ and /ɒː/) are also often neutralised such that long and short cognate sounds are often conflated in Cantonese English L2 speech (Chan and Li 2000). This may result from similarities between the two languages: while the abovementioned tense-lax contrasts do not exist in the Cantonese phonological system, the corresponding Cantonese lax vowels (/i/, /u/ and /ɔ/) do all have tense-lax allophones realised in different phonological contexts (Bauer and Benedict 1997). Therefore, it is possible that learners recognise the difference between these sounds only at a position-dependent allophonic level (as in Cantonese) but not a phonemically-based level (as in English). Finding it hard to discriminate between the two L2 vowels, Cantonese learners tend to conflate them in speech (Chan 2010).

In terms of syllabic structure, a notable difference between English and Cantonese is that while consonant clusters appear in both onset and coda positions in English, usually playing a morphological role (e.g. as an inflectional morpheme indicating tense, plurality and persons), there is no final consonant cluster in Cantonese, and initial consonant clusters are restricted to combinations with the approximant /w/ (i.e. /gʷ/, /kʷ/, which are often considered as a single consonant). The possible configurations of consonant (C) and vowels (V) in Cantonese syllables are V, CV, VC, CVC, while the number of possible configurations in English syllables is larger: V, CV, VC, CVC, CCV, VCC, CCVC, CVCC, CVCVC.

A positive correlation has been found between perception and production (Chan 2014). Learners who showed difficulty in discriminating the target consonant pairs in production (/v, w/, /θ, f/, /ð, d/, /z, s/ and /r, w/) also showed confusion in perception for the same contrast pairs, and all the target sounds were misperceived as exactly their mispronounced versions. The smaller the perceived distance between an L2 sound and the closest L1 sound, the greater the difficulty in recognising the L2 sound (Chan 2012). While these studies have provided valuable insights on the influences of L1 Cantonese on L2 English pronunciation, the findings are limited to spoken words articulated in a non-connected manner. The nature of L2 English connected speech production in native Cantonese speakers has yet to be researched. The present study aims to identify the features of English pronunciation when Cantonese speakers attempt to speak in a fluent way.

The current study

The current study examines the CSPP in the English oral production of 60 native Cantonese ESL learners. In analysing this data, we seek explanations for non-native-like CSPP features, guided by the following research questions:

What are the key characteristics of Cantonese participants’ English CSPP as realised in their L2 English oral production?

Are the major types of CSPPs learnable?

Are new categories established?

What explanations can be given for non-native-like CSPP?

Are those consonants and vowels within the connected speech environment that diverge most from the L1 more likely to be learned?

Conversely, do segments that are dissimilar to the L1 pose greater difficulties for Cantonese ESL learners in achieving native-like CSPP?

Methods

Participants

Three groups of participants were recruited: a group of 60 Cantonese-speakers, a group of 10 native speakers of General American English and a group of 10 native speakers of Standard Southern English or Received Pronunciation (RP). The Cantonese speakers were all university students, recruited from four universities in Hong Kong and majoring in a variety of fields. Their mean age was 20 years and 11
months (SD = 3 months) and there were 21 men and 39 women. All had begun to learn English as a compulsory subject from 3 years old (Kindergarten Level 1). However, their proficiency levels, as indicated by their results in the Use of English examination in the Hong Kong Advanced Level Examination (HKALE) varied. Translating HKALE grade levels into International English Language Test (IELTS) band scores participants’ comprehension skills were shown to fall approximately between IELTS band scores 5.4 and 8.3 (mean score = 6.5 defined as intermediate to high-level proficiency). Participation in the experiment was rewarded with book tokens upon the completion of the experiment.

The production of English CSPP in native speakers was also found to differ across utterances, especially under different conversational contexts (Barry 1991). Therefore, it was deemed important to collect data from a number of native English speakers in order to establish a benchmark for comparison. The inclusion of native speakers of American and British Englishes further shed light on the consistency among native English speakers. The General American speakers were from metropolises in the United States and the Received Pronunciation speakers were from the southern region of England. The participants were either monolingual native English speakers or spoke 1–2 foreign languages with limited proficiency. They spoke with accents typical of the varieties they represented.

**Stimuli**

A total of 18 sentences were used for the current study. Sentences were adapted from the ‘Manual of American English Pronunciation’ (Prator and Robinett 1985). The sentences contained significant numbers of consonant clusters at word boundaries and, when spoken by native English speakers, elicited a range of connected speech processes (CSPPs). Using pre-prepared sentences allows for linguistic features such as vocabulary, grammar and phonetic segments to be controlled for. Reliable comparisons between, and generalisations across, participants’ productions can then be made. The original sentences had proven to be effective for English CSPP analysis through their use in a similar study examining Japanese ESL learners’ English CSPP production (Anderson-Hsieh, Riney, and Koehler 1994) but were slightly adapted for the present study.

Prior to recording the Cantonese participants, 2 native General American (GA) speakers and 2 native Received Pronunciation (RP) speakers were recorded reading the 18 sentences aloud to provide baseline data which formed the basis of the marking scheme for the participants’ recordings. The RP speakers had lived in England for 22 and 37 years. The GA speakers had been living in the United States for 39 and 28 years. The approximate speed of speech production was 195 and 224 syllables per minute (spm) for the RP speakers respectively, and 225 and 237 spm for the GA speakers respectively. CSPP that were elicited by three out of the four native speakers were deemed to be reliably present in native English speech, and formed the basis of the CSPP marking criteria. In this way, traits from both sets of native speakers were combined in order to avoid falsely attributing an accent difference to an erroneous production. The CSPP that were chosen as the focus of this study were: assimilation, contraction, elision, glottalisation, linking, vowel weakening and unreleased stops, as shown in Table 1. A total of 88 tokens of these 7 CSPP types were produced by at least 3 of the 4 native speakers, and thus included within the marking scheme.

**Procedure and materials**

Before data collection, all participants were asked to give written informed consent. Then, participants were asked to read the 18 English sentences aloud as quickly as possible, without reducing the intelligibility of their speech. Prior to recording their English speech, each participant was provided with the written form of the stimuli and was allowed to rehearse their productions for as much time as they needed. No participant required rehearsal time over 5 min. All participants remained unaware of the aim of the experiment. Their speech was recorded using a high quality, Roland R-09HR, recorder. The recorded speech productions were then marked according to the marking scheme by two raters, one English native speaker and one Cantonese native speaker, both formally trained in phonetics and phonology. The two raters did not know the participants,
which allowed for more reliable, unbiased judgments to be made. The rating was supervised by the fourth author who is a phonetician. One mark was given for each CSPP in the participant’s speech that was present on the marking scheme (i.e. produced by at least three of the four English native speakers). In the case of CSPPs that were absent from the participant’s production but present in the marking scheme, no marks were given. Where the raters’ judgments diverged, participants’ productions were discussed and a suitable scoring was agreed upon.

Errors were categorised and analysed based on comparison of linguistic traits when considering ‘the typological gap’, and in accordance with the language characteristics selected within Chan and Li’s (2000) extensive comparison of Cantonese and English. A list of the different categories of errors is given in Table 3. Errors (e)-(h) comprise forms of error that do not derive from L1/L2 phonemic inventory differences. E2, H1 and H2 are excluded from calculations regarding % of Overall Errors’ as these can also be attributed to other sources. Cases of Consonant Cluster Reduction (E2) can be attributed to alternative sources of error. For example, non-native-like productions such as /buk/ (participant reads ‘books’ and thus deletes the native word-final plural marker /s/) can equally be regarded as a case of consonant cluster reduction (avoiding the complex word-final cluster /ks/) and of suffix deletion. Similarly, perceptual errors (H1 and H2) are equally ascribed to error A2 (L2 consonant substituted with an L1 consonant) and B1 (Vowel substitution) respectively, e.g. the /sw/ alternation within /wisk/ (participant reads ‘risk’, pronounced by native English speakers as /usk/) can be categorised as both A2 and H1.

The total percentages of each error are calculated by considering their type, as opposed to token, frequency. For example, in calculating the percentage of errors attributable to A2 (Substitution of an L2 consonantal phoneme with an L1 consonantal phoneme), multiple instances of identical non-native-like productions, e.g. /spid/ (participant reads ‘speech’, thus substituting native-like word-final affricate /tʃ/ with plosive /d/) contributed one error. Therefore, when calculating the proportion of each source of error, the following equation was used: number of different types of error/total types of error across all sources x 100. This allowed for Table 3 to be fully representative of the different types of phonological error that participants produced, and avoided percentages being warped by repeated errors of the same form. Accumulatively, the following categories allow for all types of non-native-like divergences present within participants’ speech to be accounted for. Each proposed source of ‘error’ will be considered in greater depth in the discussion section.

**Results**

As shown in Table 2, a significant effect of group on the total score of CSPP production was found, $F(2, 77) = 17.97, p < .001$. Post-hoc comparisons using the Games-Howell test that took into account the unequal sample size across the three groups indicated that both the GA speakers and the RP speakers scored significantly higher than the Cantonese ESL university students ($p < .01$). No significant difference between the GA and RP groups was found.

Considering participants’ overall CSPP production scores, on average, the Cantonese ESL participants produced 64% of the CSPPs on the marking scheme (scoring 56/88), while GA and RP speakers produced 83.9% and 81.5% of the CSPPs (scoring 73.9/88 and 71.8/88). There was a 17.9–20.3% (15.8–17.9 marks) difference between the scores of the ESL and the two native-speaker groups.

Considering participants’ performance on the seven individual CSP processes, it should be noted that the frequency of each process within the stimuli differed, e.g. the marking scheme contains 27 cases of elision but only 3 cases of assimilation. The total number of cases of each process is provided in parenthesis beside the process label. For 4 out of the 7 processes, consistent with the general pattern, ESL participants scored significantly lower than both GA and RP speakers, while the performance of the 2 native speaker groups was comparable. The difference between the ESL and the two native groups was most robust for linking and vowel weakening ($p < .001$), followed by glottalisation ($p < .01$) and assimilation ($p < .05$). For contraction, significant difference was only found between the
Table 2. Participants’ scores organised according to process.

<table>
<thead>
<tr>
<th>Process (max)</th>
<th>Speakers of American English (A) (n = 10)</th>
<th>Speakers of British English (B) (n = 10)</th>
<th>Cantonese-speakers (C) (n = 60)</th>
<th>F-test, post-hoc test of Games–Howell</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Standard Deviation); Range</td>
<td>Mean (Standard Deviation); Range</td>
<td>Mean (Standard Deviation); Range</td>
<td></td>
</tr>
<tr>
<td>Assimilation</td>
<td>2(1.1); 0–3</td>
<td>1.9 (0.9); 1–3</td>
<td>1.0 (0.4); 0–2</td>
<td>20.3*** A=B; A&gt;C; B&gt;C</td>
</tr>
<tr>
<td>Contraction</td>
<td>4.2(0.4); 4–5</td>
<td>4.5 (0.5); 4–5</td>
<td>3.9 (1.1); 1–5</td>
<td>1.6 A=B; A=C; B&gt;C</td>
</tr>
<tr>
<td>Elision</td>
<td>22.5 (3.7); 16–27</td>
<td>22.8 (3.4); 18–27</td>
<td>23.2 (2.9); 16–27</td>
<td>0.2 A=B; A=C; B&gt;C</td>
</tr>
<tr>
<td>Glottalisation</td>
<td>0.7 (0.5); 0–1</td>
<td>0.8 (0.4); 0–1</td>
<td>0.06 (0.3); 0–1</td>
<td>36.5*** A=B; A=C; B&gt;C</td>
</tr>
<tr>
<td>Linking</td>
<td>26.3 (4.3); 17–34</td>
<td>27.8 (4.6); 19–33</td>
<td>16.8 (6.3); 6–34</td>
<td>22.3*** A=B; A&gt;C; B&gt;C</td>
</tr>
<tr>
<td>Vowel Weakening</td>
<td>7.8 (0.9); 6–9</td>
<td>8.1 (0.5); 7–9</td>
<td>3.6 (2.1); 0–9</td>
<td>38.6*** A=B; A&gt;C; B&gt;C</td>
</tr>
<tr>
<td>Unreleased Stops</td>
<td>8.3 (1.2); 6–9</td>
<td>8.0 (1.3); 6–9</td>
<td>7.8 (0.9); 6–9</td>
<td>0.7 A=B; A&gt;C</td>
</tr>
<tr>
<td>Total score</td>
<td>73.9 (10.2); 55–87</td>
<td>71.8 (10.1); 52–88</td>
<td>56.4 (10.7); 36–82</td>
<td>17.97*** A=B; A&gt;C; B&gt;C</td>
</tr>
</tbody>
</table>

Note. The mean, standard deviation, highest and lowest scores for each process are given to 1 decimal. *p<.05; **p<.01; ***p<.001

scores of RP and ESL speakers (p < .05). For two of the processes (elision and unreleased stops), performance of the ESL group followed the same trends as the native speakers’.

Inter-speaker variability

In the ESL group, the highest scoring participant achieved a mark of 82/88 (93%), whilst the lowest scored 36/88 (41%). This resulted in a range of 52% (46 marks) and a standard deviation between.

Table 3. Error analysis table categorising each type of non-native-like pronunciation produced by participants.

<table>
<thead>
<tr>
<th>Proposed source of error</th>
<th>Error type</th>
<th>% of Overall errors</th>
<th>Participant examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A2. Substitution with an L1 phoneme</td>
<td>A2. 17%</td>
<td>A2. Thanked /θæŋkt/ → /θæŋkt/</td>
</tr>
<tr>
<td></td>
<td>A4. Substitution of an L1 target phoneme with an alternative L1 phoneme</td>
<td>A4. 7.4%</td>
<td>A4. Desk /desk/ → /dɛst/</td>
</tr>
<tr>
<td>(c) Transfer of L1 articulatory habits</td>
<td>C1. Transfer of unreleased stops</td>
<td>C1.12.5%</td>
<td>C1. Lunched /lʌntʃt/ → /lʌntʃt/</td>
</tr>
<tr>
<td></td>
<td>C2. Transfer of aspiration</td>
<td>C2. 4.7%</td>
<td>C2.Instructor /ɪnstəkt/ → /ɪnʃtəkt/</td>
</tr>
<tr>
<td>(d) The influence of L1 phonemic distribution</td>
<td>D1. Syllable-final position constraints</td>
<td>D1. 6.8%</td>
<td>D1. Made /meid/ → /mei/</td>
</tr>
<tr>
<td>(e) Strategies</td>
<td>E1. Vowel Insertion</td>
<td>E1. 4.5%</td>
<td>E1. Request /skwɪst/ → /skwɪst/</td>
</tr>
<tr>
<td></td>
<td>E2. Consonant Cluster Reduction</td>
<td>E2. 2.5%</td>
<td>E2. Student /stjuːdənt/ → /stjuːdənt/</td>
</tr>
<tr>
<td></td>
<td>E3. Metathesis</td>
<td>E3. 2.5%</td>
<td>E3. Risk /rɪsk/ → /rɪks/</td>
</tr>
<tr>
<td>(f) The influence of L1 grammar</td>
<td>F1. Deletion of Articles and Suffixes</td>
<td>F1. 9.6%</td>
<td>F1. Watched /wɑːtʃt/ → /fæʃt/</td>
</tr>
<tr>
<td></td>
<td>F2. Superfluous use of Suffixes</td>
<td>F2. 4.1%</td>
<td>F2. Spring /spɛŋ/ → /spɛŋ/</td>
</tr>
<tr>
<td>(g) The Influence of the opaque nature of English orthography</td>
<td>G1. Adherence to Orthographic cues (grapheme-phoneme correspondence rules)</td>
<td>G1. 3.1%</td>
<td>G1. Pushed /pʊʃt/ → /pʊʃt/</td>
</tr>
<tr>
<td>(h) Perceptual differences</td>
<td>H1. Neutalised Consonantal Categories</td>
<td>H1. –</td>
<td>H1. Screen /skwin/ → /skwin/</td>
</tr>
<tr>
<td></td>
<td>H2. Neutalised Vocalic Categories</td>
<td>H2. –</td>
<td>H2. Graduate /ɡrædʒuət/ → /ɡrædʒuət/</td>
</tr>
</tbody>
</table>

Note. Altered phonemes, exemplifying the type of error, are emboldened. E2, H1 and H2 are excluded from calculations regarding ‘% of Overall Errors’ as this type of errors can also be attributed to other sources.
scores of 10.7. In the native groups, a range of 36–41% (32–36 marks) and a standard deviation of 10.1–10.2 were found. While all three groups showed inter-speaker variability in their production of CSPPs, the variability was much greater overall among the ESL learner group, but it should also be noted that the level of variability differed depending on the different processes involved.

The L2 speakers showed the greatest variability on vowel weakening (0–100%) and elision (18–100%). Conversely, production scores for ‘Unreleased Stops’ showed a smaller range (67–100%).

**CSPP productions specific to the ESL group**

Table 3 presents a summary of the different categories of deviant productions present within the participants’ responses.

**Discussion**

Building on previous studies on the perception and production of L2 segments, this study examines the performances of connected speech production by Chinese ESL learners. The highest scoring participant (scoring 82/88) from the Chinese ESL group was not able to obtain a score equivalent to native speakers’ highest score (87-88/88). Nevertheless, considering participants’ performance on the individual CSPPs, in six of the seven types, at least one ESL participant matched native usage (scoring 100%). It would thus seem that Cantonese ESL learners are able to access and produce native-like English CSPPs, but not without difficulties and variability across speakers. Whilst the overall CSPP score of the Cantonese speakers was significantly lower than both speakers of American English and British English, there were different patterns for the seven types of CSPPs. These findings suggest that CSPPs differ in terms of difficulties for Cantonese speakers, with similar performances found in elision and unreleased stop across the three groups as well as similar performances found in contraction between the Cantonese speakers and speakers of American English. The ESL learners’ more native-like performance on elision is possibly attributable to the absence of final consonant clusters in Chinese. Elision was used as a means of simplifying final consonant clusters, as we elaborate below. Likewise, the higher incidence of unreleased stops may be explained by a characteristic of Cantonese, which is that all final plosives are unreleased (i.e. /p/ /t/ /k/ in syllable-final position are pronounced as /p̚/, /t̚/, /k̚/). It is interesting that while the characteristics of Cantonese may lead to negative transfer and pronunciation errors, especially where the citation form of English is concerned (e.g. inappropriate omission and non-release of final consonants, such as omission of /t/ which marks past tense), they may also be a source of positive transfer in the context of CSPPs.

Another process that was found more frequently in the ESL participants’ speech was contraction, which is the only process that is represented in written form. Not only does this feature of the process provide extra visual cues that may facilitate ESL learners’ production, but it also increases L2 learners’ exposure to the form (through both written and auditory means) and increases familiarity with the process. It is thus plausible to explain ESL learners’ more native-like performance on weak forms by pointing to the fact that this process appears in the orthographical form. In the case of contraction, the beneficial effect brought about by orthography was not affected by the orthographical opacity of English, which can be a source of non-native-like pronunciation, as will be discussed below.

Frequent native-like cases of deletion were also produced by participants in the form of ‘consonant cluster simplification’. For example, in cases such as /pu:t/ (participant reads ‘pushed’, target pronunciation /put/), and /ˌimurɪ/ (participant reads ‘removed’, target pronunciation /imuvəd/), part of the consonant cluster is elided. Despite the difference in the selection of the phonemes to be elided within the consonant cluster, the elision process involved and the context in which it was used (consonant cluster) were similar across native speakers and L2 learners. The high frequency of use of elision (59.2–100%), which is comparable to the native speakers’ pronunciation, also reflects a positive transfer from L1, even if there is less learning taking place. A finer examination of consonant clusters in the connected speech of the ESL participants suggests that this process carries a lot of L1 Cantonese
characteristics, as evidenced in the cases of metathesis. For example, /k/ and /s/ in /aks/ and /deks/ are reversed respectively (participant reads ‘risk’ and ‘desk’), resulting in the consonant cluster becoming more alike to the Cantonese affricate /ts/. Additionally, metathesis occurs at word boundaries for the non-adjacent case / wifi/ (participant reads ‘if we’), supporting the avoidance of the consonant cluster /fw/ that consists of two labials, as well as altering the syllable structure to one that is accessible within the participants’ L1 (i.e. CV CV). This again may not be an effect of L1 influence alone, but possibly the result of a universal preference for CV syllable structure (Rogerson-Revell 2011), as predicted by the ONSET constraints of Optimality Theory (Prince and Smolensky 1993).

Examination of CSPP productions that diverge from native-like speech provide insights into processes specific to this group of ESL learners, addressing research question 1b. Although some native-like CSPP features are identified in the ESL participants’ productions, participants still clearly diverge from native-like phonology. In the following section, key features of our Cantonese participants’ L2 CSPP production are examined for the interplay between L1 and L2 as postulated by various models of L2 phonology such as SLM (Flege 1995), thus addressing research question 2.

Vowel insertion, or epenthesis, was identified as a distinctive non-native feature in the Cantonese ESL learners’ production of English connected speech. For example, the word-final consonant cluster in /agenst/ (participant reads ‘against’) was dismantled via insertion of an inter-consonantal /ə/ and spoken as /agenəs/ . This also occurs in the learner production /sunkwəstə/ (participant reads ‘request’) via insertion of an inter-consonantal /ə/ . The pronunciation /stənəθə/ (participant reads ‘strength’, native pronunciation /stənəθ/) similarly evidences vowel epenthesis occurring at word exteriors, with the non-native-like addition of schwa /ə/ appearing word-finally. As proposed by Young-Scholten and Archibald (2000), vowel epenthesis is more often used as a syllable simplification strategy in literate L2 learners who are not residing in an English speaking country and where contact with L2 English relies greatly on written material.

Further evidence of the role of L1 Cantonese in L2 English connected speech is shown in the case of contrastive quality of aspiration. This feature is found within English speech; the voicing contrast of aspirated English words such as /dʒər/ and /dʒɔr/ are reversed respectively (participant reads ‘speech you’). However, Cantonese participants instead produced deletion of the word-final /t/ of ‘speech you’. This exemplifies substitution of the word-final English affricate /dʒ/, whilst the latter presents a case of complete deletion. Substitution and deletion, especially when occurring at word boundaries, disrupt fluency and inhibit native-like CSPPs. For example, within the sentence ‘I heard that speech you made last night’ native English speakers exhibit linking, i.e. /spiːtʃjuː/ ‘speech you’. However, Cantonese participants instead produced deletion of the word-final /s/ of ‘speech’, thus impeding the native-like linking process. In this case, the native-like word-final L2 affricate /tʃ/ was not assimilated, and instead was substituted with alveolar plosive /d/.

In other cases, substitution and deletion were sometimes applied to the same phoneme. For example, when producing the nominal ‘George’ (native pronunciation /dʒɔr/), the ESL participants produced variants such as /dʒɔːk/ or /dʒɔː/. The former exemplifies substitution of the word-final English affricate /dʒ/, whilst the latter presents a case of complete deletion. Substitution and deletion, especially when occurring at word boundaries, disrupt fluency and inhibit native-like CSPPs. For example, within the sentence ‘I heard that speech you made last night’ native English speakers exhibit linking, i.e. /spiːtʃjuː/ ‘speech you’. However, Cantonese participants instead produced deletion or substitution, e.g. /spiː/ (participant reads ‘speech’), thus impeding the native-like linking process. In this case, the native-like word-final L2 affricate /tʃ/ was not assimilated, and instead was substituted with alveolar plosive /d/. Similar substitution has been observed in Hungarian, Russian and Thai learners of English (Lombardi 2003).
Syllable position also appeared to play a determining role in the realisation of L2 CSPPs. This factor is considered in relation to SLM (Flege 1995). Cantonese positional constraints are, when transferred, incompatible with native-English syllable structure. English plosives in syllable-final position which do not fit with the phonemic distribution of Cantonese are often deleted. Of the 51 cases of word-final plosives contained within the stimuli, 33 were deleted by at least one participant. As exemplified with /maan/ (participant reads ‘mind and’), the process of word-final /d/ deletion draws the participant closer to native English reduced speech /maan/. Similarly, word-final /t/ elision within the phrase ‘asked if’ (canonical form pronunciation /askɪt/) was produced by both native English speakers (i.e. /æskɪf/) and Cantonese ESL participants (e.g. /askɪʃf/).

Like English, nasals /m/, /n/ and /ŋ/ can appear in word-final position in Cantonese (Bauer and Benedict 1997). In accordance with positional constraints, participants displayed reduced difficulty in producing word-final nasals compared to word-final plosives. Only two lexical items, namely ‘thanksgiving’ and ‘long’ (with native-like pronunciations /θæŋskɪŋ/ and /lŋŋ/ respectively), were read aloud with non-native-like word-final nasal production.

L2 productions of affricates and fricatives also evidenced positional constraints L1 transfer. In the 9 non-native pronunciations of the nominal ‘George’ /dʒɔ:ɹɪg/, the word-initial affricate was successfully produced in all cases, while the word-final /dʒ/ was only produced in 2 instances. A similar avoidance of word-final consonants was evidenced in participants’ productions of fricatives, e.g. /θə/ ‘others’, /ʃeɪt/ ‘states’ (both of which diverge from native-like speech through deletion of word-final fricatives /z/ and /s/ respectively). With the most restricted position in spoken Cantonese being word-final, this type of transfer has transparent limiting effects on successful CSPP production in English.

The selection of phonemes used to substitute the ‘L2-exclusive’ consonants, were not restricted to those present within Cantonese. For example, in the non-native-like production /stæenʃʧ/ (participant reads ‘strength’) the word-final fricative /θ/, present within native speakers’ productions, i.e. /stæenθ/, was substituted with affricate /ʧ/. Both phonemes are in fact absent from Cantonese speech. Substitution, whilst arguably triggered by a gap in the L1 inventory, does not always result in an L2 phoneme being replaced by an L1 phoneme. One explanation for substituting an L2 segment with a more marked segment is that the L1 requires that the manner of articulation of the original segment be maintained in the phonetic variant (Lombardi 2003). In this case, affricate /ʧ/ which can be derived for stop-fricative clusters [tʃ]+[ʃ] is substituted by fricative /θ/. This is further emphasised when considering the posited role of L1 cues according to the SLM (Flege 1995). In SLM, it is proposed that perception precedes production, and therefore if a learner cannot perceive the phonetic cue they may well not be able to produce a native-like form.

Participants’ choice of substitute phonemes was also arguably influenced by a ‘devoicing’ preference. 25.7% of non-native-like productions of target L2 consonantal phonemes were cases in which a voiced phoneme was substituted with a voiceless counterpart. Participants’ tendency to devoice L2 consonants may be attributable to the limited saliency of the voicing contrast within Cantonese. It is plausible that, due to its lack of utility and frequency within their L1, learners fail to notice and distinguish this linguistic feature. L2 phoneme devoicing may or may not result in an L1 phoneme being used. For example, /ʌmʊfʊt/ (participant reads ‘removed’, native-like pronunciation /ʌmʊvʊd/) depicts the fricative /v/ being exchanged for its devoiced counterpart /f/; an L2 phoneme is replaced by one that is instead present in Cantonese. The tendency to devoice is consistent with the prediction made by the Marked Differential Hypothesis (MDH; Eckman 1977). Because voiceless phonemes are typologically more common, and therefore less marked, Chinese ESL learners would find devoicing voiced L2 English phonemes easier. This voicing characteristic, and its importance within English, may therefore require more explicit teaching in order to ensure its productivity within the L2.

Non-native-like vowel productions, in which an L2 vowel is directly replaced by a vowel present within the L1, were common in the ESL participants’ data. A highly frequent case was the /eə, e/ substitution, in which the former vowel (absent from the L1) was substituted by the latter (present in the L1) in non-native-like productions, since the two sounds were perceived as instances of the same L1 category. This process accounted for 20% of total errors dependent on L1 vowel inventory gaps.
Despite /e/ only appearing as part of a diphthong within Cantonese, its presence within the L1 inventory, as well as acoustic and articulatory similarity with /æ/, arguably encourages this substitution. For the lexemes ‘graduate’ (native English pronunciation /ɡrəˈɛdʒʊət/) and ‘thanked’ (/ˈθæŋkt/) alone, 41% and 33% of participants respectively substituted the native-like target form /æ/ with /e/. A previous study showed that inexperienced Mandarin-speaking learners of English failed to produce a significant height difference between the vowel pair /e-æ/. In assessing this error source, evidence points to the role of perceptual difficulties and the possible implementation of the ‘equivalence classification’ (perceiving L2 English /æ/ as /e/) of vowel inventory in Cantonese speakers (Flege 1987).

In addition, the process of vowel weakening is often avoided. ESL participants tended to substitute the English ‘schwa’ /a/ with a ‘stronger’ vowel form (e.g. /ɔʃ/ as opposed to native-like /æʊ/ ‘of’). Reduction to ‘schwa’ is extremely common in native English connected speech, and is a key characteristic in differentiating native speakers’ citation and connected speech forms (Kelly 2000). However, many of the acoustic variations of vowels cannot be attributed to language-universal coarticulatory constraints, but rather to language specific phonetic realisation patterns. The spectral and temporal structure of vowels in spoken continuous speech varies significantly even among speakers of homogeneous dialect (Strange et al. 2005). Whilst making speech more accessible to Cantonese learners, L2 vowel substitution disrupts the timing, fluency and native-like-ness of English connected speech (Deterding, Wong, and Kirkpatrick 2008).

The role of orthographical opacity must also be recognised when considering erroneous L2 CSPP productions. In the productions /ˈlʌntʃət/ and /ˈwʌntʃət/, (participant reads ‘lunched’ and ‘watched’), participants prioritised the orthographic form of the past tense suffix over production of a native-like CSPP. Participants diverged from the native-like pronunciations /ˈlʌntʃt/ and /ˈwʌntʃt/ respectively, instead explicitly pronouncing the ‘-ed’ affix as /tʃət/. This accented pronunciation could be considered as a form of hypercorrection; in a possible attempt to avoid affix omission, participants may instead have highlighted its orthographical presence (Setter and Deterding 2003). Of the 11 lexical items ending with the past tense ‘-ed’ suffix among the stimuli, 10 were ‘hypercorrected’ by at least one participant. However, instances of this error type are not isolated to lexical items containing the past tense suffix. In the production of ‘states’ as /ˈstɛɪtθəs/, for example, the participant also pronounced the orthographic form explicitly with word-final /əs/, deviating from the native-like pronunciation /ˈstɛɪtsəs/. These errors are more likely to be a result of the opaque nature of the English writing system (mismatch between spelling and pronunciation) rather than a hypercorrection. All the examples mentioned above are spelt with an ‘e’, which often suggests the presence of a schwa vowel, but does not match the actual pronunciation. These errors occurring at word-final position adversely affect CSPP production by inhibiting both reduction and linking processes.

Taking all the phenomena analysed here together, the English connected speech elicited from these Cantonese ESL learners clearly carries predominately L1 Cantonese features. According to the Ontogeny Phylogeny Model (OPM; Major 2001), the attainment of native-like English connected speech production is deemed to fall between the initial and medial interlanguage (IL) stages and far from reaching the final stage with a prevalence of L2 features. SLM holds that age of onset of L2 learning will determine the nativeness of production of sounds in the target language (Flege 1995). While native-like accent can typically emerge between the ages of 5 and 8 in L2 naturalistic settings, there appears to be no optimal age for the attainment of native-like (or near-native) L2 pronunciation for learners living in non-English-speaking countries (Flege and Liu 2001). Interpreted in the light of Flege’s (1991) ‘accented L2 input hypothesis’, non-native-like production could simply be a result of insufficient access to L2.

**Pedagogical implications**

With the knowledge about specific influences of L1 Cantonese on L2 English connected speech presented in the current study, teachers can assess students’ performances individually using the analytic method described in this study. Apart from introducing the topic of connected speech in a
traditional way by providing definitions and examples of various CSPPs, teachers can discuss the interference of L1 phonology on L2 connected speech acquisition. Moreover, students themselves can be encouraged to modify some of their learning and speaking behaviours so as to minimise the influence of L1 phonology. For instance, they may reduce code-switching when making a conversation, so that the L1 phonology is less automatically activated when speaking English.

Teachers can also make reference to our findings and optimise the linguistic input and improve classroom activities in the language classrooms. The application of electronic visual feedback system (Alameen 2014; Coniam 2002) is recommended, as it would allow students to visualise the manner in which their connected speech is articulated and better adjust their articulatory movement based on teachers’ feedback. Recent research has shown the effectiveness of shadowing (learners listening to sound recordings and speaking aloud simultaneously) in promoting connected speech production (Hamada 2011; Shiki et al. 2010). As repetitive shadowing itself is far from motivating, teachers may develop fun ways to encourage shadowing so as to ensure sufficient processing of English connected speech. As suggested by Alameen and Levis (2015), production of connected speech is hard to train, partly because students usually mix up the CSPPs. The information about the frequency occurrence of different errors provided in this study will help teachers to prioritise the training of different types of CSPPs. The sub-standard performances among university students also highlight the need for connected speech training. In agreement with Alameen and Levis’s (2015) suggestion, the amount of time allocated to pronunciation training in ESL classroom should be increased.

**Limitations and future research**

As with the majority of second language research, the extent of inter-speaker variability renders the study of CSPPs more complex. Considering Labov’s (1972) notion of the ‘observers paradox’, it must be ensured that participants speak naturally. The format of the current study, reading a list of sentences aloud, may encourage hypercorrection, adherence to grapheme-phoneme correspondence rules and more ‘formal’ speech, possibly reducing the instances of CSPPs, contributing to pronunciation errors or resulting in non-native-like speech processes. Therefore, future studies may consider how alike CSPPs can be elicited across participants in more conversational and natural settings, in which participants are less conscious of their speech and achieve full avoidance of the ‘observers paradox’.

**Conclusion**

In order to acquire fluent, native-like English speech, successful realisation of CSPPs is vital. It is evident from the results of the current study that ESL learners’ CSPP difficulties are heavily influenced by differences between the L1 and the L2. However, the variable ways in which the L1 is shown to be of influence (e.g. deletion of articles and suffixes due to L1/L2 differences, avoidance of consonant clusters that are absent from the L1) highlight the need for greater understanding of the source of errors by both educators and students, and implementation of targeted-error correction. One way forward might be to train educators to be able to diagnose phonological errors more effectively in order to address difficulties in learner-specific ‘individualized programmes’ (Chan 2010: 325). This may not only ensure progression in L2 oral production, but also maintain the student’s motivation by focusing on error types and solutions that are both relevant and beneficial to their learning of L2 speech.

**Notes**

1. HKALE is the territory-wide public examination administered for all the Secondary Form 7 students (aged 17-18) in Hong Kong. The Hong Kong Examinations and Assessment Authority (2015) provide an official conversion table for HKALE scores to IELTS scores.
Acknowledgement

We would like to thank all the participants and Dr Mark Harrison for coordinating the recording of native English speakers’ speech.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Early Career Scheme of the Research Grants Council (RGC) of Hong Kong [grant number ECS 846212].

ORCID

Simpson W. L. Wong  http://orcid.org/0000-0002-6606-6382

References


