Language and Speech

Article

Language *and* Speech
56(2) 191–209

DOI: 10.1177/0023830912443948
las.sagepub.com
SAGE

Does Vowel Inventory Density Affect Vowel-to-Vowel Coarticulation?

Peggy P.K. Mok
Chinese University of Hong Kong, Hong Kong

Abstract
This study tests the output constraints hypothesis that languages with a crowded phonemic vowel space would allow less vowel-to-vowel coarticulation than languages with a sparser vowel space to avoid perceptual confusion. Mandarin has fewer vowel phonemes than Cantonese, but their allophonic vowel spaces are similarly crowded. The hypothesis predicts that Mandarin would allow more coarticulation than Cantonese. Eight native speakers of Cantonese and of Beijing Mandarin were recorded saying the target sequences /pV_1ˈpV_2pV_3/ (V = /i a u/) in carrier phrases. F1 and F2 frequencies were measured at vowel edge and midpoint, and were normalized for analyses. The results show that Cantonese and Mandarin do not differ in degree of vowel-to-vowel coarticulation in either F1 or F2. In addition, unstressed vowels exhibit more coarticulation than stressed vowels. Carryover coarticulation exceeds anticipatory coarticulation in both F1 and F2. Unstressed vowels in the carryover position are the most susceptible to coarticulation. The results show that vowel inventory does not predict vowel-to-vowel coarticulation. Fundamental assumptions of the output constraints hypothesis are evaluated to explain its failure in predicting language-specific patterns of vowel-to-vowel coarticulation. The importance of syntagmatic relationships in coarticulation is also discussed.

Keywords
Cantonese, Mandarin, vowel inventory, vowel-to-vowel coarticulation

1 Introduction

1.1 Vowel inventory density and vowel-to-vowel coarticulation

Vowel-to-vowel (V-to-V) coarticulation refers to the coarticulatory effects of one vowel on another across one or more intervening consonants. V-to-V coarticulation effects are known to vary from language to language. Öhman's (1966) seminal work showed that Russian exhibited the weakest V-to-V coarticulation compared with Swedish and American English. He suggested that the

Corresponding author:
Peggy P.K. Mok, Department of Linguistics and Modern Languages, Leung Kau Kui Building, Chinese University of Hong Kong, Shatin, Hong Kong.
Email: peggymok@cuhk.edu.hk

presence of secondary articulation of consonants in Russian (palatalization) reduced its freedom to coarticulate. Choi and Keating (1991) examined Öhman's suggestion by comparing plain and palatalized consonants in Russian, Bulgarian and Polish, and used American English as reference. They found that secondary articulation did not by itself block V-to-V coarticulation, but the four languages did exhibit different degrees of coarticulation, with American English allowing the most. They did not provide an explanation for the language differences, however.

Manuel (1987, 1990, 1999) was the only one who provided a clear hypothesis to account for language differences in V-to-V coarticulation. She suggested that linguistic contrast would affect coarticulation. She referred to such constraints as "output constraints". Languages differ in the number of distinctive sounds they have in their systems of contrast. This may have predictive value in how different vowels are allowed to coarticulate in particular languages. She suggested that articulatory or auditory proximity of phones affects coarticulation. How susceptible a phone is to the coarticulatory influence of neighboring phones can be predicted by how a language divides a particular phonetic space. If a language has only one phone in a particular region of the vowel space, then more variability in production (i.e., coarticulation) can be tolerated without causing perceptual confusion due to the blurring of phonetic contrasts. A crowded vowel space would limit the degree of V-to-V coarticulation allowed.

This hypothesis was based on the results of two studies. Manuel and Krakow (1984) compared V-to-V coarticulation in Swahili, Shona and American English. They found that Swahili and Shona, both with a five-vowel system, exhibited more V-to-V coarticulation than English, which has a much longer vowel inventory. But the results were only suggestive because there was only one speaker per language. Manuel (1987, 1990) compared three Bantu languages, Ndebele (five vowels), Shona (five vowels) and Sotho (seven vowels) with three speakers per language. She found that Sotho exhibited less V-to-V coarticulation of /ɑ/ in the F2 dimension. Given that the two studies were only based on a small number of speakers, more studies on different languages using a larger number of speakers are needed to test the hypothesis of output constraints.

In the intervening years, there have been only a few studies testing the hypothesis with conflicting results, although the idea of the output constraints hypothesis is often used to explain language differences in coarticulation (e.g., Cho, 2004). Okalidou and Koenig (1999) compared V-to-V coarticulation in Modern Greek (a five-vowel system) and American English using two speakers in each language. They found that Greek speakers showed more flanking vowel effects than American English speakers in the F2 dimension. But they only reported significant main effects of the flanking vowels for each speaker in the two languages individually, so it is unclear how reliable the language difference is. In contrast, also comparing a five-vowel language, Shona, with American English but using more speakers in each language (seven for Shona and five for English), Beddor, Harnsberger and Lindemann (2002) found that Shona exhibited less V-to-V coarticulation than American English for both stressed and unstressed vowels, contrary to the results of Manuel and Krakow (1984). Moreover, as mentioned earlier, based on three speakers in each language, Choi and Keating (1991) found that American English allowed the most V-to-V coarticulation compared with Russian, Bulgarian and Polish, although the last three languages have only five or six vowel phonemes. A more recent study on Korean (eight vowels) and Japanese (five vowels) using eight speakers in each language also found no difference in V-to-V coarticulation (Han, 2007).

None of the above few studies with contradictory results to the output constraints hypothesis provided a clear evaluation of the hypothesis and the patterns they found. One important issue that was not clearly distinguished by Manuel (1987, 1990) and subsequent studies testing the output constraints hypothesis is that of which type of contrast could limit coarticulation: surface or underlying. Although Manuel (1987, 1990) couched her idea largely in terms of segmental articulatory gestures, and she used the potentially ambiguous term "phones" in her work, her definition of vowel contrasts and the consistent use of // clearly indicate that she was referring to underlying phonemic contrasts. Subsequent studies were also based on the patterns of underlying phonemic vowel contrasts in the target languages. This may not be an issue in languages with a relatively simple mapping between phonemic and allophonic vowels, but as we will see in the next section, for some languages like Cantonese and Beijing Mandarin, vowel qualities are not simply determined by phonemic contrasts. The same phoneme can be realized quite differently depending on the neighboring segments and syllable structure. There are fewer vowel phonemes in Mandarin (five) than Cantonese (eight), but the allophonic vowel space can be similarly crowded in both languages. The complicated relationships in Cantonese and Mandarin highlight the fact that vowel variations are not always adequately captured by paradigmatic phonemic contrasts, and there can be substantial differences between phonemic and allophonic vowel variations. Issues like these were not addressed by previous studies on V-to-V coarticulation. This study evaluates the output constraints hypothesis by using data from these two widely spoken Chinese languages for which we have no relevant data on cross-linguistic comparisons of V-to-V coarticulation. Their special vowel systems also require us to take a wider perspective on linguistic contrasts and coarticulation.

1.2 Vowel phonemes in Cantonese and Beijing Mandarin

This section gives a brief overview of the vowel systems in the two languages. Interested readers should consult the references cited for detailed discussion.

In Chinese phonology, diphthongs or triphthongs ending in [-i] and [-u] are usually analyzed as 'vowel + glide' (for diphthongs) or 'glide + vowel + glide' (for triphthongs) combinations, with the final glide being in the coda position (e.g., Bauer & Benedict, 1997; Duanmu, 2000; Kao, 1971; Li, 1999; Shi, 2008b). This is not just a notational difference because such analysis is necessary for the analysis of Cantonese and Mandarin vowel systems, as vowel quality is dependent on segmental contexts.

1.2.1 Cantonese vowels. The main issue in Cantonese is the phonological status of vowel length. Acoustically, vowels have different durations in Cantonese, but their distributions complicate the assignment of phonemic status to vowel length (see discussion below). There are thirteen segmentally conditioned allophones in Cantonese: long [iː yː uː ɛː œː ɔː aː] and short [ɪ ʊ e o ɵ ɐ], with [e o] appearing only in diphthongs [ei] and [ou] (see Figure 1b). The formant frequency values for [ɪ] and [ɛː], [ʊ] and [ɔː] and [ɐ] and [aː] partly overlap, with [ɐ] and [aː] overlapping the least (Li, 1985; Shi, 2008a; Zee, 2003).

Using durational measurements from spectrograms, Kao (1971) suggested that duration is a relevant and primary feature of the Cantonese vowel system. She concluded that there are three factors affecting vowel duration in Cantonese: syllable structure (open or closed), vowel quantity (long or short) and coda endings (/j ([i]) w ([u]) m n ŋ/ vs. /p t k/). However, since vowel duration is dependent on three factors, vowel duration in Cantonese is not just a matter of length. Many of the phonemic analyses of Cantonese, e.g., Chao (1947), Hashimoto (1972), Yuan (1960), treated length as an allophonic feature except for the low vowel pair [a] and [ɐ], because other short vowels are in complementary distribution with their long counterparts. Only long vowels can appear in open syllables. The [a] and [ɐ] pair contrast before all possible consonants and glides.

Since only [a] and [ɐ] are contrastive, the most common analysis of Cantonese vowels consists of eight phonemes: /iː yː uː ɛː œː ɔː aː ɐ/ with only one short vowel /ɐ/ (see Figure 1a). Longer inventories of nine phonemic vowels (seven long /iː uː yː aː ɛː ɔː œː/ and two short /ə ɐ/) (Pulleyblank, 1997) and eleven phonemic vowels (seven long /iː yː ɛː œː aː uː ɔː/ and four short /e ɵ ɐ o/) (Bauer & Benedict, 1997) have also been proposed.

1.2.2 Mandarin vowels. There are many different allophonic vowel qualities in Mandarin. Figure 1e shows the vowel allophones in Mandarin, with [e o ɛ ɔ æ ɑ] occurring in diphthongs or triphthongs. The Mandarin vowels can be divided into three levels of height: high, mid and low, which gives a systematic entry-point to tackle the Mandarin vowel system.

The main debate about the high vowels is whether the two so-called apical vowels should be grouped under /i/ as one phoneme: [ɹ̩] occurs only after dental consonants *z c s* and [ɻ̩] occurs only after 'retroflex' consonants *zh ch sh r*. Different IPA symbols were used to

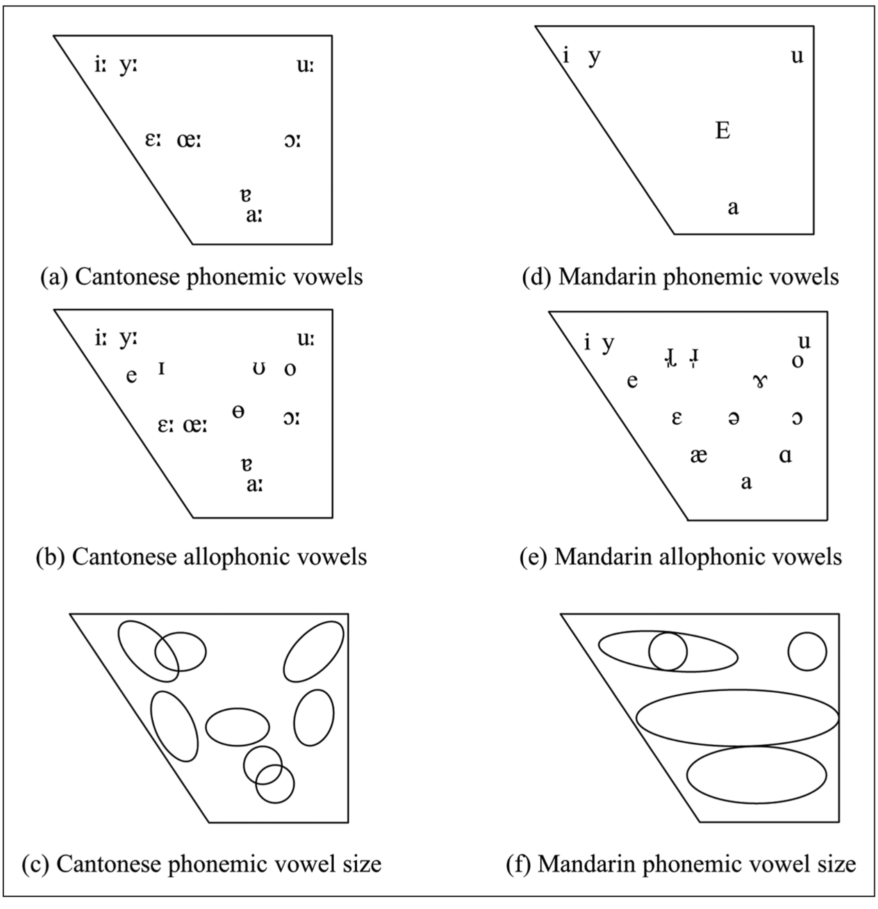

Figure 1. Phonemic and allophonic vowels in Cantonese and Mandarin in a schematic vowel space. 'E' in (d) represents an underspecified mid vowel phoneme. The symbols [ɻ̩] and [ɹ̩] in (e) represent two apical vowels. See the text for explanation.

transcribe these two vowels (see Lee & Zee, 2003; Li, 1999; Lin & Wang, 1992; Wiese, 1997; Wu & Lin, 1989). The two apical vowels are often viewed as syllabic consonants instead of vowels. Some argued that these segments are the syllabic prolongation of the preceding consonants (e.g., Chao, 1968; Duanmu, 2000; Hartman, 1944; Hockett, 1947; Lee & Zee, 2003; Li, 1999; Wiese, 1997). However, others believe that these segments are indeed vowels (e.g., C. C. Cheng, 1973; R. L. Cheng, 1966; Howie, 1976; Lin & Wang, 1992; Wu & Lin, 1989), but their formant frequencies resemble a mid vowel more than a high vowel (Wu, 1986). Ladefoged and Maddieson (1990, p. 117) called them "fricative vowels" and considered them as syllabic fricatives that are allophones of /i/. These two vocalic segments are in complementary distribution with [i] and the three can be analyzed as allophones of the same phoneme. It seems best to conclude that [ɹ̩] and [ɻ̩] are syllabic approximants in actual realization but are vowels phonologically, and they are allophones of /i/ (but see Cheng, 1973; Wang, 1991; Wong, 1953 for different analyses). The remaining two high vowels [y] and [u] are less controversial. Therefore, the most common analysis of Mandarin high vowels is that there are three vowel phonemes /i/, /y/ and /u/.

There are many different mid vowel qualities in Mandarin: [e ɛ ə ɤ ɔ o]. They differ along the front/back and round/spread dimensions. The realization of the mid vowel(s) depends heavily on segmental context, i.e., syllable structure. Only [ɤ] can occur in open syllables; all other realizations occur either before or after glides as diphthongs or triphthongs, e.g., [fe̲i^{55}] 'fly', [o̲u^{55}] 'Europe', [ɕiɛ̲^{51}] 'thank'. The distribution of the mid vowels is clearly complementary, so most phonemic analyses treat them as allophones of the same phoneme. But the wide variation of the mid vowels means that it is unclear what the canonical form of the mid vowel phoneme should be: suggestions include /e/ (Cheng, 1966; Hartman, 1944; Hockett, 1947), /ɤ/ (Cheng, 1973), /ə/ (Duanmu, 2000; Howie, 1976) and an underspecified mid vowel E (Wiese, 1997). In conclusion, based on the principle of complementary distribution, there is one underspecified mid vowel phoneme which can change its backness and rounding, but not its height.

The low vowels in Mandarin are the least contentious. All phonemic analyses agree that there is only one low vowel phoneme /a/ in Mandarin with several variants having complementary distribution. Like the mid vowel phoneme, the low vowel is also heavily influenced by its context and varies along the front/back dimension.

The most common analysis suggests that there are five phonemic vowels in Mandarin: /i y u E a/ where /E/ denotes an underspecified mid vowel as suggested by Wiese (1997) (see Figure 1d).

1.3 The present study

The section above briefly discusses the vowel systems in Cantonese and Mandarin. The evidence points to the conclusion that Cantonese has a more crowded phonemic vowel space than Mandarin, especially in the lower part of the vowel space around /a/. Based on the most common analyses, there are eight vowel phonemes in Cantonese and five in Mandarin (see Figure 1a, d). According to the output constraints hypothesis based on the density of phonemic contrasts, Mandarin should allow more V-to-V coarticulation than Cantonese. Although the phonemic vowel density is different in Cantonese and Mandarin, there are similar allophonic variations in both languages so their allophonic vowel space is similarly crowded (see Figure 1b, e). This intriguing situation was not considered in V-to-V coarticulation studies before. If V-to-V coarticulation depends on phonemic vowel contrasts, as postulated by the output constraints hypothesis, then Mandarin should allow more V-to-V coarticulation than Cantonese, especially for the low vowel /a/. However, the complicated mappings between phonemic and allophonic vowel qualities in the two languages may bring contrary results to this hypothesis.

The complicated mappings also lead to another interesting related situation. If we consider the 'size' of each vowel phoneme in the two languages by incorporating their allophonic variations, Cantonese would have more 'small' phonemic vowels with fewer variations while Mandarin would have fewer 'large' phonemic vowels with many variations (see Figure 1c, f). The Mandarin /y/ will completely overlap with /i/ which incorporates the two apical vowels [ɹ̩] and [ɻ̩] in the F1-F2 two-dimensional space, while the Cantonese /i/ and /y/ only overlap slightly (based on data in Zee, 2003). The Mandarin /E/ will take up most of the mid vowel space (Figure 1f). Using Manuel's own term (Manuel, 1990, p. 1287), the "range of productions" for each vowel should still be smaller in Cantonese than in Mandarin, but this time not because of the sparsity of the vowel space which allows more freedom to stray as such, but because of the intrinsic differences in allophonic variations. If V-to-V coarticulation is determined by the allowed range of productions as suggested by the output constraint hypothesis, we would also expect Mandarin to allow more V-to-V coarticulation than Cantonese, this time for both /i/ and /a/.

2 Methods

2.1 Speakers

Eight native speakers of Hong Kong Cantonese (four men, four women) and eight speakers of Beijing Mandarin (four men, four women) participated in the study. Except for the three Cantonese speakers who were visitors from Hong Kong at the time of recording, all speakers were graduate students in Cambridge, UK. The eight native speakers of Cantonese, all in their 20s, were born and brought up in Hong Kong. The eight native speakers of Mandarin were in their 20s and 30s. They were born in Beijing and spoke Beijing Mandarin natively. All subjects were fluent in English as a second language, but they mainly used Cantonese/Mandarin to communicate with their friends and relatives. They were paid for participating in the experiment.

2.2 Materials

The experimental design is the same as that of Beddor, Harnsberger and Lindemann (2002). Three vowels, /i/, /a/, /u/, were used as both target vowels and contexts. They were chosen in this study because of two reasons. First, these vowels are often used in previous V-to-V coarticulation studies. Second, they represent the most extreme points in the vowel quadrilateral for both Cantonese and Mandarin so as to maximize any potential effects on V-to-V coarticulation. The unaspirated /p/ was chosen because it exerts the least lingual constraint. The averaged VOT value for /p/ in both languages is less than 10 ms (Clumeck, Barton, Macken, & Huntington, 1981; Wu, 1987). The three syllables, /pi/, /pa/ and /pu/, carry a high level tone, a tone common to both languages. /pu^{55}/ is a marginal syllable in Cantonese. The rest of the syllables are all real words in both languages, but the resultant sequences are nonsense trisyllables. The experimental materials have the form of /pV_1ˈpV_2pV_3/ with three possibilities: $V_1 = V_2 = V_3$, $V_1 \neq V_2 = V_3$ or $V_1 = V_2 \neq V_3$. Since there is no lexical stress in Cantonese, only contrastive stress can be used to assess the effect of stress on V-to-V coarticulation. V_2 is always contrastively stressed, while V_1 and V_3 are not stressed but are also not reduced. Both target vowels and vowel context can be V_1, V_2 or V_3 depending on the conditions of different factors. Table 1 shows all of the trisyllables. The starting points and midpoints of V_2 and V_3 were used to assess carryover coarticulation from V_1 and V_2 respectively,

Table 1. Experimental trisyllables for Cantonese and Mandarin grouped according to target vowels, contexts and coarticulatory direction.

Target vowel conditions	Coarticulatory direction	Context /a/	Context /i/	Context /u/
Target /a/ stressed	Anticipatory	paˈ**pa**pa	paˈ**pa**pi	paˈ**pa**pu
	Carryover	paˈ**pa**pa	piˈ**pa**pa	puˈ**pa**pa
Target /a/ unstressed	Anticipatory	**pa**ˈpapa	**pa**ˈpipi	**pa**ˈpupu
	Carryover	paˈpa**pa**	piˈpi**pa**	puˈpu**pa**
Target /i/ stressed	Anticipatory	piˈ**pi**pa	piˈ**pi**pi	piˈ**pi**pu
	Carryover	paˈ**pi**pi	piˈ**pi**pi	puˈ**pi**pi
Target /i/ unstressed	Anticipatory	**pi**ˈpapa	**pi**ˈpipi	**pi**ˈpupu
	Carryover	paˈpa**pi**	piˈpi**pi**	puˈpu**pi**
Target /u/ stressed	Anticipatory	puˈ**pu**pa	puˈ**pu**pi	puˈ**pu**pu
	Carryover	paˈ**pu**pu	piˈ**pu**pu	puˈ**pu**pu
Target /u/ unstressed	Anticipatory	**pu**ˈpapa	**pu**ˈpipi	**pu**ˈpupu
	Carryover	paˈpa**pu**	piˈpi**pu**	puˈpu**pu**

Bold = target syllable.

while the midpoints and endpoints of V_1 and V_2 were used to assess anticipatory coarticulation from V_2 and V_3 respectively. For example, in the trisyllable /paˈpipi/, V_1 is an unstressed target /a/ with anticipatory coarticulation from the following /i/ context, while V_2 is a contrastively stressed target /i/ for assessing carryover coarticulation from the previous /a/ context. V_3 in this case is not analyzed. Target vowels in the context of themselves, e.g., /paˈpapa/, were all from the same stimulus, i.e., all three vowels were used.

The experimental trisyllables were embedded in short carrier phrases as shown below. The carrier phrases were designed so that the vowels of the whole sequence are as similar as possible in both languages in order to minimize unwanted variance on the target vowels.

Cantonese: [ŋɔ^{23} ka^{33} __________ka^{33} lɔ ^{55}]
(Gloss: It's my '_______' where [ka^{33} lɔ^{55}] are two sentence final particles.)
Beijing Mandarin: [wo^{214} tə __________ tə p^{u}o^{55}]
(Gloss: My '_______' wave.)

There are existing Chinese characters for the required syllables and tones in Mandarin. In Cantonese, there is also a character for /pa^{55}/. The syllable /pi^{55}/ originated from loanwords, but is now often used in colloquial expressions with no foreign origins, e.g., [tɐi^{55} pi^{55}] 'idiotic'; [ts^{h}ɛŋ^{55} pi^{55} pi^{55}] 'unpleasantly green'. The syllable was also included in the Cantonese syllabary by Bauer (1985), but there is no character for it. A character with similar pronunciation (/pɪt^{5}/) was used to represent /pi^{55}/. There is a syllable [pu^{21}] in Cantonese colloquial speech which has a marginal status, but it is also included in the Cantonese syllabary by Bauer (1985). A character with similar pronunciation (/pun^{55}/) was used to represent /pu^{55}/. Using characters with similar pronunciation to represent the two intended syllables is a practical solution. Since Cantonese is not a written language, Hong Kong Cantonese speakers often use non-standard characters, English words or even symbols to represent syllables without standard Chinese characters. Cases like this can be easily found in popular publications in Hong Kong. The speakers were told explicitly to pronounce the two syllables as [pi^{55}] and [pu^{55}]. After some practice, all speakers were able to pronounce them with no problem. Stressed syllables were bolded and underlined when presented to the subjects.

2.3 Procedures

All materials were presented to the speakers in Chinese characters. The speakers read ten randomized lists of all the trisyllables embedded in carrier phrases at a normal speaking rate. For the contrastive stress context, they were told to stress the second syllable in the experimental trisyllables in the way that they would normally stress the correct version of a misheard syllable (not stressing the entire trisyllable sequence). But they were not told what to do phonetically, e.g., say the syllable louder. All recordings were made in a sound-treated room in the phonetics laboratory at the University of Cambridge. Before the actual recording, the speakers practiced by reading a randomized list of the trisyllables several times. The speech was recorded directly onto a Silicon Graphics Indigo computer via a Sennheiser MKH 40 P48 microphone and a Symetrix SX 202 amplifier using *Xwaves* speech processing software with a sampling frequency of 16 kHz.

2.4 Acoustic measurements

The frequencies of the first two formants (F1 and F2) were measured from 18 pole 25 ms autocorrelation LPC spectra with a Hanning window, supplemented by wide band spectrogram and DFT spectra. F1 and F2 were measured at two temporal locations of the target vowels: vowel edge and midpoint. Vowel edge was either the onset or offset of the target vowels, depending on the directions of coarticulation: onset for assessing carryover coarticulation and offset for assessing anticipatory coarticulation. The temporal locations were identified from the waveform where the beginning and ending of periodic vocalic voicing was taken for onset and offset of the target vowels. The spectra windows were centered 12.5 ms inward from the onset and offset. The midpoint measures were taken at the point halfway between the two locations. For stressed target vowels in their own context, e.g., the middle syllable in /paˈpapa/, the F1 and F2 measurements at vowel midpoint are the same for both anticipatory and carryover directions.

2.5 Data normalization

In most previous studies, V-to-V coarticulation was investigated by comparing target vowels in the contexts of themselves (e.g., /papa/) with target vowels in other vowel contexts (e.g., /papi/) while keeping other things constant, e.g., the intervocalic consonant, so the difference in formant frequencies can be attributed to the vowel contexts (e.g., Choi & Keating, 1991; Fowler, 1981; Huffman, 1986; Hussein, 1990; Magen, 1997; Manuel, 1987, 1990; Öhman, 1966; Recasens, 1987; Recasens & Pallarès, 2000). The same method is adopted in this study, with only one difference. Previous studies used raw formant frequencies as the dependent variable, while this study used normalized formant frequencies for comparison.

The normalization procedure is described in Mok (2007b), which is essentially the same as the centroid method proposed by Watt and Fabricius (2002), but the two methods were developed independently. Information comparing the centroid method with other normalization procedures can be found in Watt, Fabricius and Kendall (2010). In the present study, each mean F1 (or F2) measurement averaged over ten repetitions for each target vowel in a given context was expressed as a proportion of the grand mean (F1 or F2) of all vowel tokens collected from a particular speaker. There are two steps involved: 1) Calculate the grand mean (F1 or F2) of all target vowels averaged across all conditions for a particular speaker. The grand mean represents the hypothetical center of the speaker's vowel space (e.g., 500 Hz for F1). 2) Calculate the F1 (or F2) proportion from that grand mean for each measurement of the target vowels ($current_{Target\ V}$ / grand mean), e.g., 750 Hz / 500 Hz = 1.5. A value bigger or smaller than 1 simply means that the vowel formant is higher or lower than the hypothetical center. The deviations of the proportions from the grand mean (i.e., 0.5 in the above example) only represent the normalized distance from the hypothetical center of each speaker's vowel space, but they do not show degree of coarticulation themselves. For example, the F2 of /i/ is always higher than 1, while the F2 of /u/ is always lower than 1. The proportions simply normalize formant frequencies across speakers. No expectation of any sort is built into the proportions because they only convert the raw formant frequencies onto a normalized scale (similar to converting formant frequencies onto a log or Bark scale), and the proportions are always positive. We can draw a traditional vowel plot using these normalized proportions, and the position of the vowels will correspond well to a plot drawn using raw formant frequencies, or formant frequencies on a Bark scale. The normalized proportions of each speaker were used for statistical analysis.

2.6 Statistical analysis

Four (2 formants × 2 temporal locations) 4-way repeated measures ANOVAs with factors Language (Cantonese, Mandarin) × Stress (contrastive stress, no stress) × Direction (anticipatory, carryover) × Context (/i, a, u/) were conducted for each target vowel (/i/ /a/ /u/) separately, so there were 12 ANOVAs (4 ANOVAs × 3 target vowels) altogether. A multi-factorial statistical design is necessary to ensure that any Language difference found in this study is not confounded by other factors. If the sphericity assumption of ANOVA is violated, the Huynh-Feldt correction is applied. A significant context effect is considered evidence of V-to-V coarticulation, which indicates that different contexts influenced the target vowels differently. An interaction of Context with other factors shows that the amount of V-to-V coarticulation is modified by these factors.

3 Results

Figure 2 shows the general effects of *Context* collapsed over the three target vowels under different conditions of Stress and Direction at both vowel edge and midpoint. The data points represent the average coarticulatory effects of each context vowel (*not* the positions of the target vowels in a vowel space), so they all cluster around 1, i.e., the hypothetical center of the vowel space. The triangles show the degree of V-to-V coarticulation. The bigger the triangles, the more the V-to-V coarticulation exhibited. At both temporal locations, unstressed vowels (Figure 2b, 2d) appear to have more coarticulation than contrastively stressed ones (Figure 2a, 2c). At vowel edge (Figure 2a, 2b), carryover coarticulation seems to have a larger effect than anticipatory coarticulation in both languages. At vowel midpoint (Figure 2c, 2d), the data points are clustered together and cannot be marked individually. It shows that V-to-V coarticulation has already diminished at midpoint. Finally, there does not seem to be any consistent difference between Cantonese and Mandarin. If anything, Cantonese unstressed vowels at vowel edge seem to have slightly less contextual variation in the F1 dimension than Mandarin (Figure 2b).

3.1 Language

The main focus of this study is whether languages with a sparser phonemic vowel space allow more V-to-V coarticulation than languages with a denser vowel space, that is, whether Mandarin allows more V-to-V coarticulation than Cantonese. Since significant Context effects show evidence of V-to-V coarticulation, a significant Context × Language interaction shows that there is Language

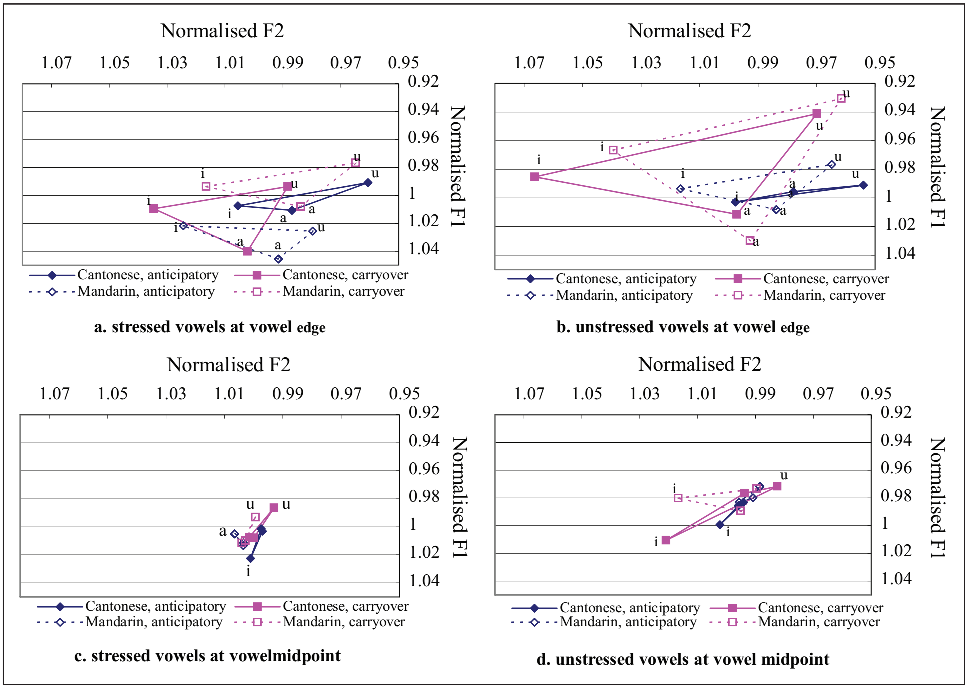

Figure 2. Coarticulatory effects of different contexts averaged across all three target vowels and speakers for contrastively stressed target vowels at (a) vowel edge and (c) midpoint, and for unstressed target vowels at (b) vowel edge and (d) midpoint in Cantonese and Mandarin.

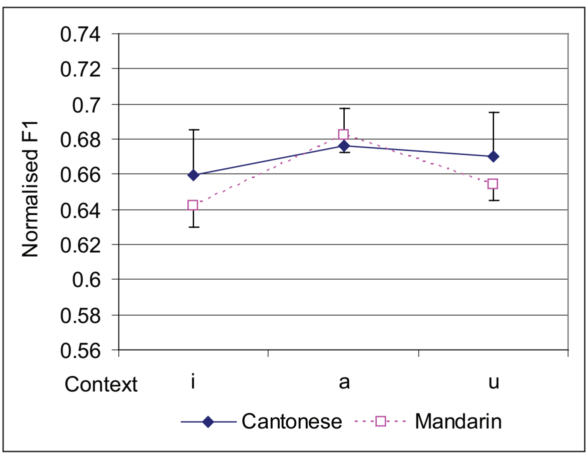

Figure 3. Context effects on F1 of target /i/ at vowel edge. The error bars show one standard error.

difference in V-to-V coarticulation collapsed across other factors (Stress and Direction). The results of all the ANOVAs show that there is only one significant Context × Language interaction for F1 of target /i/ at the vowel edge, $F(2, 28) = 4.266$, $p = 0.024$, partial $\eta^2 = 0.234$ (see Figure 3). There seems to be slightly more contextual variation in Mandarin than Cantonese in Figure 3. However, post hoc two-tailed independent samples t-tests confirm that there is no significant difference

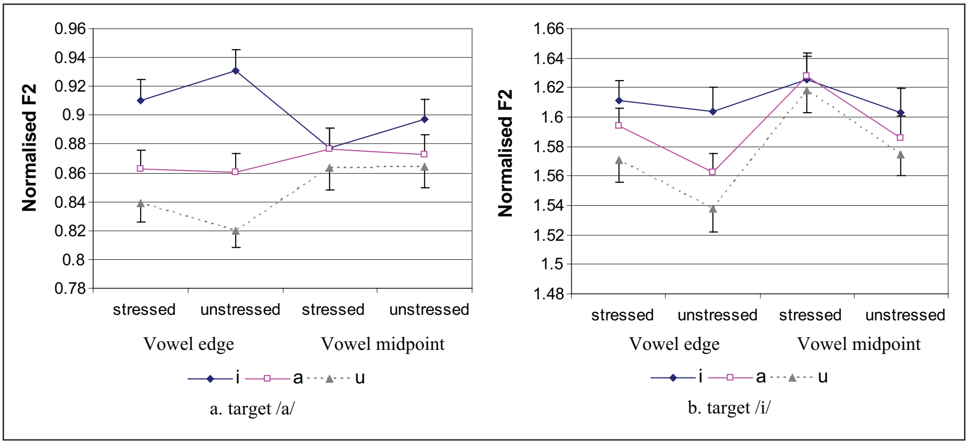

Figure 4. Context effects on F2 of stressed and unstressed (a) target /a/ and (b) target /i/ at both vowel edge and midpoint. The error bars show one standard error.

between Cantonese and Mandarin for all three contexts (/i/ $t(14) = -0.249$, $p = 0.807$; /a/ $t(14) = 0.636$, $p = 0.535$; /u/ $t(14) = 0.623$, $p = 0.544$).

For target /a/, a high order Language × Stress × Direction × Context interaction is significant for F1 at vowel midpoint, $F(2, 28) = 5.140$, $p = 0.013$. Nevertheless, the partial η^2 for this interaction is only 0.269, which indicates a small effect size. Post hoc two-tailed t-tests comparing Mandarin and Cantonese under different conditions of target /a/ are again all non-significant ($p > 0.05$). Moreover, there is no significant interaction involving Language and Context in the F2 dimension for all the ANOVAs. Therefore, the results point to the conclusion that Cantonese and Mandarin do not systematically differ in V-to-V coarticulation in either F1 or F2. This conclusion echoes quite well the observation we have based on averaged coarticulation presented in Figure 2 above.

3.2 Stress

The effect of stress on V-to-V coarticulation is consistent with findings in the literature. Unstressed vowels allow more V-to-V coarticulation than contrastively stressed ones. The Stress × Context interactions in F2 are significant for target /a/ and target /i/ at both vowel edge and midpoint (see Figure 4, target /a/: vowel edge $F(1.672, 23.414) = 9.559$, $p = 0.002$, midpoint $F(2, 28) = 5.02$, $p = 0.014$; target /i/: vowel edge $F(2, 28) = 6.853$, $p = 0.004$, midpoint $F(2, 28) = 12.661$, $p < 0.001$). For both target vowels, there are more contextual variations for unstressed than stressed vowels at both temporal locations (with a larger F2 difference induced by contexts /i/ and /u/). The same pattern is also found in F1 (target /a/ at vowel edge $F(2, 28) = 3.641$, $p = 0.039$ and target /i/ at midpoint $F(2, 28) = 6.765$, $p = 0.004$).

3.3 Direction

There is more carryover than anticipatory V-to-V coarticulation. The Direction × Context interactions are significant for F2 of target /i/ at both vowel edge and midpoint (see Figure 5, vowel edge $F(2, 28) = 7.578$, $p = 0.002$; midpoint $F(2, 28) = 4.124$, $p = 0.027$). The two contexts /i/ and /u/ induce a larger difference in normalized F2 in the carryover than the anticipatory direction at both temporal locations (vowel edge: carryover 0.064 vs. anticipatory 0.043; midpoint: carryover

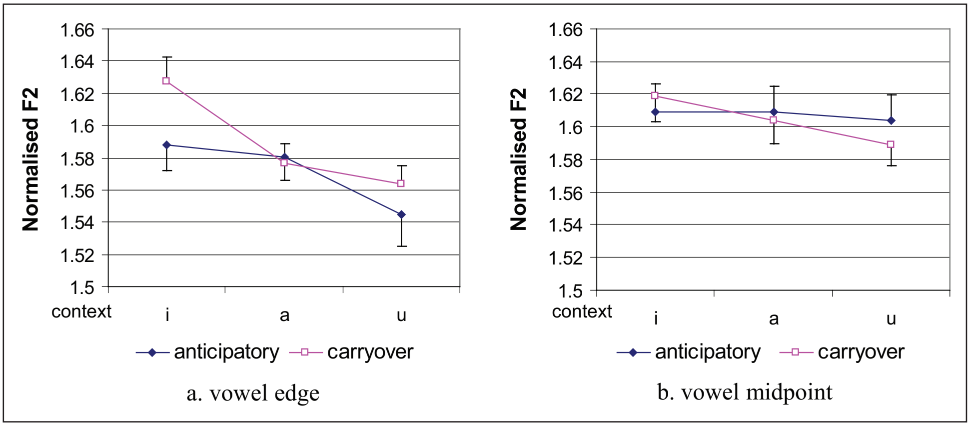

Figure 5. Contextual variations on F2 of target /i/ with anticipatory and carryover contexts at (a) vowel edge and (b) midpoint. The error bars show one standard error.

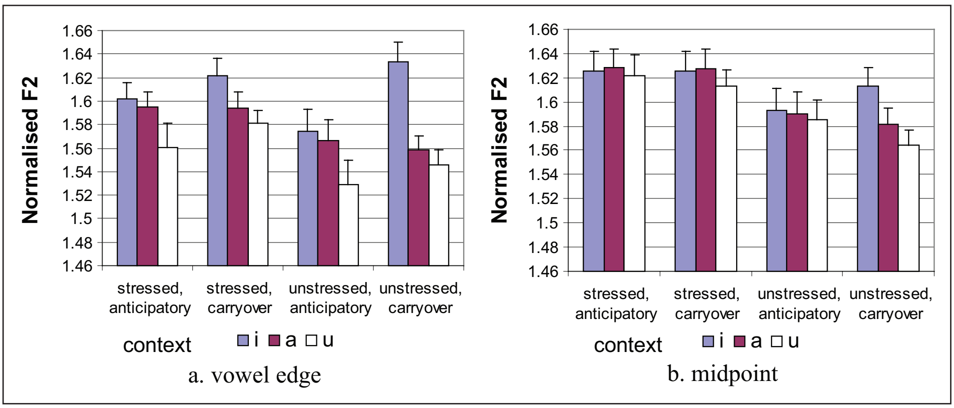

Figure 6. F2 context effects on target /i/ grouped according to Stress and Direction at (a) vowel edge and (b) midpoint. "Stressed, anticipatory" = data from stressed /i/ with anticipatory contexts; "unstressed, carryover" = data from unstressed /i/ with carryover contexts and so on. The error bars represent one standard error.

0.030 vs. anticipatory 0.006). The same pattern is found in F2 of target /a/ at midpoint $F(1.311, 18.347) = 4.420$, $p = 0.041$ (carryover 0.033 vs. anticipatory 0.0144), and in F1 of target /a/ at vowel edge $F(2, 28) = 13.257$, $p < 0.001$ (difference in normalized F1 between contexts /u/ and /a/: carryover 0.158 vs. anticipatory 0.019).

3.4 Stress and Direction

Finally, Stress and Direction jointly influence V-to-V coarticulation in F2, as is evident by the significant Stress × Direction × Context interactions at both temporal locations of target /i/ (vowel edge $F(2, 28) = 4.349$, $p = 0.023$; midpoint $F(1.7, 23.802) = 4.079$, $p = 0.036$, see Figure 6). Context /i/ and context /u/ exert the largest difference in F2 of unstressed /i/ with carryover context (vowel edge 0.087, midpoint 0.049) at both temporal locations. The combined effect of Stress and Direction renders unstressed vowels with carryover context as being the most susceptible to V-to-V coarticulation.

4 Discussion

Mainly target /i/ and target /a/ were sensitive to the effects of various factors on V-to-V coarticulation in this study. The results indicate that there is no Language difference in V-to-V coarticulation in both F1 and F2. As in many previous studies, unstressed vowels allow more V-to-V coarticulation than contrastively stressed vowels, and there is more carryover coarticulation than anticipatory coarticulation. Unstressed vowels in the carryover direction are the most susceptible to V-to-V coarticulation.

The main purpose of this study was to evaluate the output constraints hypothesis: whether linguistic contrast based on phonemic density limits V-to-V coarticulation. The results show that it does not. Beijing Mandarin, a language with a sparser phonemic vowel space, does not allow more coarticulation than Cantonese, a language with a more crowded phonemic vowel space. In addition to the differences in vowel space density, Mandarin phonemic vowels also allow a wider range of production (i.e., more allophonic variations) than Cantonese vowels, especially for the vowels /i/ and /a/. However, this intrinsic difference also does not affect the degree of V-to-V coarticulation exhibited. As mentioned in the Introduction, there is already some evidence in the literature showing that languages with fewer phonemes do not necessarily coarticulate more than languages with more phonemes. This study confirms these findings. Phonemic density by itself does not predict V-to-V coarticulation.

If phonemic density does not determine V-to-V coarticulation, would allophonic vowel density affect V-to-V coarticulation instead? It seems possible at first glance, since Mandarin and Cantonese have a similarly crowded allophonic vowel space and they do not differ in V-to-V coarticulation. Nevertheless, careful consideration reveals that allophonic vowel density is unlikely to determine V-to-V coarticulation as well. Previous independent studies testing the output constraints hypothesis quoted before (Beddor et al., 2002; Choi & Keating, 1991; Han, 2007) used languages with simpler mapping between vowel phonemes and allophonic vowel qualities, which means that their phonemic and allophonic vowel space should be similar in density. If allophonic contrast was at work, they should have results congruent with the hypothesis. However, these studies on diverse languages also found negative results. In order to fully assess the possible effects of allophonic vowel contrasts, further research should compare languages with a similar phonemic density but different allophonic inventory sizes. Nevertheless, the negative findings cast serious doubt on the possibility of allophonic vowel contrasts determining V-to-V coarticulation.

The studies mentioned above which found contradictory results to the output constraints hypothesis only suggested that some factors other than inventory size may be responsible for their results, without clearly discussing what those factors might be. Manuel herself (1990, 1999) pointed out that factors like speech style and individual differences could affect the role of phonological contrast in coarticulation. Nevertheless, the consistent negative results on diverse languages using the same speech style (read speech) do question the hypothesis of output constraints on V-to-V coarticulation based on phonemic vowel density and call for a thorough evaluation of the output constraints hypothesis itself.

In fact, the output constraints hypothesis is based on some implicit assumptions which need careful reconsideration. The first assumption of output constraints is that, in order to maintain phonological contrast (i.e., to facilitate perception), vowels should be less coarticulated with each other. More coarticulation would hamper perception. Recent data, however, cast much doubt on this assumption. For example, listeners compensate for coarticulation in perception (e.g., Beddor et al., 2002; Beddor & Krakow, 1999), and coarticulation can also enhance perception (e.g., Hawkins, 2003; Hawkins & Slater, 1994; Scarborough, 2004).

The second main assumption of output constraints is that languages fully utilize the available phonetic vowel space, and vowels are maximally dispersed within this space. More variability in production can be tolerated in a sparse vowel space. In a crowded vowel space, however, speakers only have limited freedom to stray away from the ideal patterns lest they intrude into the boundaries of surrounding vowels which would blur phonemic distinctions and hamper perception. This assumption is intuitively appealing, but the literature is not all supportive. Languages allow extensive overlap of vowel formant frequencies in running speech (e.g., Keating & Huffman, 1984). Moreover, a number of studies which compared languages with small inventories (e.g., Spanish and Modern Greek) with languages with large inventories (e.g., American English and German) showed that vowels in these languages have comparable variability, i.e., have similar acceptable area of allophonic variations (e.g., Bradlow, 1995; Flege, 1989; Hawks & Fourakis, 1995; Johnson & Martin, 2001; Jongman, Fourakis, & Sereno, 1989; Mendez, 1982; Meunier, French-Mestre, Lelekov-Boissard, & Le Besnerais, 2003).

The assumptions of output constraints on V-to-V coarticulation are based on the view of maximal perceptual contrast of the vowel inventory (Liljencrants & Lindblom, 1972), which was also stated in Manuel (1987, pp. 33–34, 42–43). Nevertheless, the notion of maximal perceptual contrast of vowels is inadequate to account for the vowel inventories of the world's languages (Lindblom, 1986, 1990a, 1990b, 1991), and there are other important factors at work (De Boer, 2000; Diehl, Lindblom, & Creeger, 2003; Schwartz, Boë, Valleé, & Abry, 1997). In addition, the size of the acoustic vowel space itself can vary between languages: a more expanded vowel space for larger inventories (e.g., Al-Tamimi & Ferragne, 2005; Bradlow, 1995; Butcher, 1994; Fletcher & Butcher, 2003; Jongman et al., 1989). Therefore, vowels in smaller inventories do not necessarily have more freedom to coarticulate than vowels in larger inventories.

The third implicit assumption of the output constraints hypothesis is that vowels have similar susceptibility to coarticulation. However, open vowels like /a/ are more susceptible than closed vowels like /i/ and /u/ in many languages (e.g., Beddor & Yavuz, 1995; Kondo & Arai, 1998; Recasens, 2002; Recasens, Pallarès, & Fontdevila, 1997). Open vowels with more surrounding phonemes could still show more coarticulation than closed vowels with fewer surrounding phonemes simply because of different intrinsic susceptibilities. Mok (2011) examined the susceptibility of different vowels to V-to-V coarticulation. She found that generally the lower the vowel is, the more susceptible to coarticulation it is. Therefore, intrinsic susceptibility of vowels is a more important factor in limiting V-to-V coarticulation than phonemic density.

The above discussion reveals that the implicit assumptions of the output constraints hypothesis on V-to-V coarticulation are actually questionable, but the previous few studies comparing language differences have not examined them at all. Despite these limitations, Manuel (1987, 1990, 1999) was the only one to propose a simple, clear and testable explanation to account for how languages may differ in V-to-V coarticulation. Her hypothesis represents an important first step in our understanding of language-specific V-to-V coarticulation, although it needs revision. Manuel was also cautious in suggesting that the role of contrast in limiting V-to-V coarticulation can be modulated by other factors, and more studies are needed to assess the validity of the hypothesis. The present study has fulfilled this role.

If vowel inventory does not affect coarticulation, what can account for language differences in V-to-V coarticulation? Moreover, what further insights can we obtain from the present study, in addition to the critical evaluation of the rationales of the output constraints hypothesis with concrete experimental data? The complicated relationships between vowel phonemes and allophones in Cantonese and Mandarin prompt us to re-examine language-specific V-to-V coarticulation from a new perspective and to consider other unexplored factors which may affect V-to-V coarticulation. It is clear from the Introduction that the vowel variation in Cantonese and Mandarin is dependent on syntagmatic relationships (syllable structure). Phonemic theories mainly concern minimal contrast between different phones in identical environments, emphasizing the paradigmatic aspect of phonological relationships at the expense of the syntagmatic aspect (Robins, 1957). Phonetic differences not capable of differentiating words were treated as functionally and phonologically unimportant, but recent studies show that phonetic fine detail is important in understanding speech (e.g., Hawkins, 2003). The present results show that paradigmatic phonemic contrast alone does not determine language-specific V-to-V coarticulation, whereas syntagmatic phonological relationships receive very little attention in coarticulation studies.

Beckman (1999) rightly pointed out that since coarticulation originated as a problem for the linear segmental model of speech, it is important to re-assess the status of coarticulation in non-alphabetic frameworks. All this points us to consider the role of phonological units larger than phonemes and to consider syntagmatic relationships in V-to-V coarticulation, since V-to-V coarticulation is essentially about the syntagmatic relationship between vowels. The insights we get from the Cantonese and Mandarin vowel systems suggest that syllable structure of a language is worth considering. Syllable structure complexity of a language often correlates with other important typological phonological differences between languages. This line of inquiry has not been explored in the coarticulation literature before. Nevertheless, previous studies on cross-language differences in V-to-V coarticulation already gave us some hints in this respect. For example, Beddor et al. (2002) found more V-to-V coarticulation in English than Shona, an African language with a very simple syllable structure (mainly CV). Japanese and Korean, both with a very simple syllable structure, do not differ in V-to-V coarticulation (Han, 2007). In the present study, Cantonese and Mandarin also do not differ in V-to-V coarticulation in view of their simple syllable structure. These results suggest that there may be a link between syllable structure complexity and V-to-V coarticulation. Inspired by the results of the present study, Mok (2007a, 2010) found that languages with more complicated syllable structure (e.g., English) allow more V-to-V coarticulation than languages with simpler syllable structure (e.g., Thai). Her results strongly suggest that syllable structure complexity of a language does have a role in determining V-to-V coarticulation. Of course, more studies comparing languages with complex and simple syllable structure are needed to substantiate this idea and to investigate the principles behind it.

This study disconfirms the hypothesis of phonemic contrast posing output constraints on V-to-V coarticulation in languages based on their phonemic vowel inventories. Vowel inventory does not constrain coarticulation. Manuel herself also pointed out that "there probably is a tendency for coarticulation to be constrained so as to maintain contrast, but it is just that – a tendency. If there truly is only a tendency for coarticulation patterns to be derivative from patterns of contrast, then we would expect to find examples in which the principle of contrast maintenance makes the wrong prediction about coarticulation" (Manuel, 1999, p. 196). We have found such examples, and have discussed in depth the reasons why the hypothesis made wrong predictions. The data and the special features of Cantonese and Mandarin vowel systems show us clearly that a segmental approach is inadequate in predicting language differences in V-to-V coarticulation, and lead us to consider a wider perspective for V-to-V coarticulation. It is hoped that the various insights from the present study can bring us closer to identifying the factors which contribute to language-specific patterns of V-to-V coarticulation.

Acknowledgements

The author is grateful to Sarah Hawkins for help and guidance throughout the project. Thanks also go to the Associate Editor Keith Johnson and the reviewers Robert Hagiwara and Yao Yao for their helpful suggestions on the manuscript. This research was supported by the Sir Edward Youde Memorial Fellowship for Overseas Studies from Hong Kong and an Overseas Research Studentship from the United Kingdom.

References

Al-Tamimi, J. E., & Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech-Eurospeech 2005)*, 2465–2468. Lisbon.

Bauer, R. S. (1985). The expanding syllabary of Hong Kong Cantonese. *Cahiers de Linguistique Asie Orientale, 14*, 99–111.

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

Beckman, M. E. (1999). Implications for phonological theory. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 199–225). Cambridge: Cambridge University Press.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 30*, 591–627.

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalisation by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America, 106*, 2868–2887.

Beddor, P. S., & Yavuz, H. K. (1995). The relationship between vowel-to-vowel coarticulation and vowel harmony in Turkish. *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS)*, 44–51. Stockholm, Sweden.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*, 1916–1924.

Butcher, A. (1994). On the phonetics of small vowel systems: Evidence from Australian languages. *Proceedings of the 5th Australian International Conference on Speech Science and Technology*, 28–33. Canberra.

Chao, Y. R. (1947). *Cantonese primer*. New York: Greenwood Press.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Cheng, C. C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics, 2*, 135–262.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics, 32*, 141–176.

Choi, J. D., & Keating, P. (1991). Vowel-to-vowel coarticulation in three Slavic languages. *UCLA Working Papers in Phonetics, 78*, 78–86.

Clumeck, H., Barton, D., Macken, M. A., & Huntington, D. A. (1981). The aspiration contrast in Cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics, 9*, 210–224.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics, 28*, 441–465.

Diehl, R., Lindblom, B., & Creeger, C. (2003). Increasing realism of auditory representations yields further insights into vowel phonetics. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1381–1384. Barcelona.

Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Flege, J. E. (1989). Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Language and Speech, 32*, 123–147.

Fletcher, J., & Butcher, A. (2003). Local and global influences on vowel formants in three Australian languages. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 905–908. Barcelona.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research, 46*, 127–139.

Han, J. I. (2007). The role of vowel contrast in language-specific patterns of vowel-to-vowel coarticulation: Evidence from Korean and Japanese. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, 509–512. Saarbrücken, Germany.

Hartman, L. M. (1944). The segmental phonemes of the Peiping dialect. *Language, 20*, 28–43.

Hashimoto, A. (1972). *Cantonese phonology*. Cambridge: Cambridge University Press.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31*, 373–405.

Hawkins, S., & Slater, A. (1994). Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 57–60.

Hawks, J. W., & Fourakis, M. (1995). The perceptual vowel spaces of American English and Modern Greek: A comparison. *Language and Speech, 38*, 237–252.

Hockett, C. F. (1947). Peiping phonology. *Journal of the American Oriental Society, 67*, 253–267.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Huffman, M. K. (1986). Patterns of coarticulation in English. *UCLA Working Papers in Phonetics, 63*, 26–47.

Hussein, L. (1990). VCV coarticulation in Arabic. *Ohio State University Working Papers in Linguistics, 38*, 88–104.

Johnson, K., & Martin, J. (2001). Acoustic vowel reduction in Creek: Effects of distinctive length and position in the word. *Phonetica, 58*, 81–102.

Jongman, A., Fourakis, M., & Sereno, J., A. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech, 32*, 221–248.

Kao, D. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.

Keating, P., & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica, 41*, 191–207.

Kondo, Y., & Arai, Y. (1998). Prosodic constraint on v-to-v coarticulation in Japanese. *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP)*, 1755–1758. Sydney, Australia.

Ladefoged, P., & Maddieson, I. (1990). Vowels of the world's languages. *Journal of Phonetics, 18*, 93–122.

Lee, W. S., & Zee, E. (2003). Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association, 33*, 109–112.

Li, W. C. (1999). *A diachronically-motivated segmental phonology of Mandarin Chinese*. New York: Peter Lang Publishing.

Li, X. D. (1985). Guangzhouhua yuanyin de yinzhi ji changduan duili [Opposition of vowel quality and length in Cantonese]. *Fangyan, 1*, 28–38.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839–862.

Lin, T., & Wang, L. J. (1992). *Yuyinxue Jiaocheng* [A course in phonetics]. Beijing: Beijing University Press.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). London: Academic Press.

Lindblom, B. (1990a). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht: Kluwer.

Lindblom, B. (1990b). On the notion of "possible speech sound". *Journal of Phonetics, 18*, 135–152.

Lindblom, B. (1991). The status of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 7–24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics, 25*, 187–205.

Manuel, S. (1987). *Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages* (Unpublished PhD thesis). Yale University.

Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America, 88*, 1286–1298.

Manuel, S. (1999). Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 179–198). Cambridge: Cambridge University Press.

Manuel, S., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research, SR-77/78*, 69–78.

Mendez, A. (1982). Production of American English and Spanish vowels. *Language and Speech, 25*, 191–197.

Meunier, C., French-Mestre, C., Lelekov-Boissard, T., & Le Besnerais, M. (2003). Production and perception of vowels: Does the density of the system play a role? *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 723–726. Barcelona.

Mok, P. (2007a). Effects of syllable structures on v-to-v coarticulation (Thai vs English). *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, 421–424. Saarbrücken, Germany.

Mok, P. (2007b). *Influences on vowel-to-vowel coarticulation* (Unpublished PhD thesis). Department of Linguistics, University of Cambridge.

Mok, P. (2010). Language-specific realizations of syllable structure and vowel-to-vowel coarticulation. *Journal of the Acoustical Society of America, 128*, 1346–1356.

Mok, P. (2011). Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation. *Language and Speech, 54*, 527–545.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151–168.

Okalidou, A., & Koenig, L. L. (1999). Patterns of vowel-to-vowel coarticulation in Greek and English. *Proceedings of Berlin 99: The 137th regular meeting of the Acoustical Society of America and 2nd Convention of the EAA. Forum Acusticum*. Deutsche Gesellschaft für Akustik e.v. Oldenburg.

Pulleyblank, E. G. (1997). The Cantonese vowel system in historical perspective. In J. Wang & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 185–217). Berlin: Mouton de Gruyter.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics, 15*, 299–312.

Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *Journal of the Acoustical Society of America, 111*, 2828–2841.

Recasens, D., & Pallarès, M. D. (2000). A study of F1 coarticulation in VCV sequences. *Journal of Speech, Language and Hearing Research, 43*, 501–512.

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America, 102*, 544–561.

Robins, R. H. (1957). Aspects of prosodic analysis. In F. R. Palmer (Ed.), *Prosodic analysis* (pp. 188–200). London: Oxford University Press.

Scarborough, R. A. (2004). *Coarticulation and the structure of the lexicon* (Unpublished PhD thesis). Department of Linguistics, University of California Los Angeles.

Schwartz, J. L., Boë, L. J., Valleé, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics, 25*, 255–286.

Shi, F. (2008a). Guangzhouhua de yuanyin geju [Cantonese vowel patterns]. In *Yuyin Geju: Yuyinxue Yu Yinxixue De Jiaohuidian* [Sound patterns: Interface between phonetics and phonology] (pp. 72–81). Beijing: The Commercial Press.

Shi, F. (2008b). *Yuyin Geju: Yuyinxue Yu Yinxixue De Jiaohuidian* [Sound patterns: Interface between phonetics and phonology]. Beijing: The Commercial Press.

Wang, L. J. (1991). *Yinxixue Jichu* [Basics of phonology]. Beijing: Yuwen Chubenshe.

Watt, D., & Fabricius, A. (2002). Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1~F2 plane. *Leeds Working Papers in Linguistics and Phonetics, 9*, 159–173.

Watt, D., Fabricius, A., & Kendall, T. (2010). More on vowels: Plotting and normalization. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 107–118). London: Routledge.

Wiese, R. (1997). Underspecification and the description of Chinese vowels. In J. L. Wang & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 219–249). Berlin: Mouton de Gruyter.

Wong, H. (1953). Outline of the Mandarin phonemic system. *Word, 9*, 268–276.

Wu, Z. J. (1986). *Hanyu Putonghua Danyinjie Yutuce* [The spectrographic album of mono-syllables of Standard Chinese]. Beijing: Chinese Social Science.

Wu, Z. J. (1987). Aspirated vs non-aspirated stops and affricates in Standard Chinese. *Proceedings of the 11th International Congress of Phonetic Sciences (ICPhS)*. Tallinn, Estonia.

Wu, Z. J., & Lin, M. C. (Eds.) (1989). *Shiyan Yuyinxue Gaiyao* [An outline of experimental phonetics]. Beijing: Gaodeng Jiaoyu Chubanshe.

Yuan, J. H. (1960). *Hanyu Fangyan Gaiyao* [A survey of Chinese dialects]. Beijing: Wenzi Gaige Chubanshe.

Zee, E. (2003). Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 1117–1120. Barcelona.