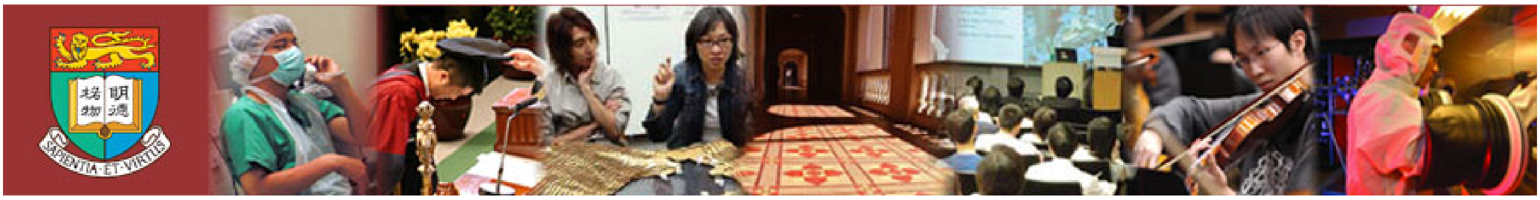

Relative Contributions of Vowels and Consonants in Recognizing Isolated Mandarin Words

Wong Lok Yi

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2014.

Abstract

This study investigated the relative contributions of vowels and consonants in recognizing isolated Mandarin words. Nineteen normal-hearing native speakers of Mandarin were recruited and were asked to recognize isolated Mandarin words with different proportions of consonant or vowel segment preserved. The accuracy in recognizing the isolated Mandarin words, phonemes, and tones were scored. It is found that there is a greater contribution of vowels than consonants to isolated word recognition in Mandarin, which is different from previous outcomes in English. Possible reasons for this language difference in isolated word recognition were discussed. Contribution of consonant-vowel transitional boundary to isolated word recognition in Mandarin was also examined. It is found that the word recognition performance improves with increased amount of consonant-vowel boundary information presented.

Introduction

Vowels and consonants are two traditional categories of speech sounds, which are used to produce words in languages throughout the world (Ladefoged & Disner, 2012). Vowels are produced with a more open vocal tract and their duration is longer. In contrast, production of consonants involves constriction of vocal tract and their duration is shorter (Abercrombie, 1967; Wright, 2004). Due to the differences in the manner of production, vowels and consonants carry different acoustic information (Kent & Read, 2002). Formant frequencies are the acoustic cues that are important for identifying vowels. For consonants, the acoustic cue differs in terms of place and manner of articulation. In general, spectral pattern is the acoustic cue for the manner of articulation, whereas burst (including energy level and spectral center of gravity and variance) and formant transition are the acoustic cues for the place of articulation (Pickett, 1999).

Many researchers suggested that vowels and consonants have different roles in speech perception. For example, Nespor, Pena, and Mehler (2003) and Toro, Nespor, Mehler and Bonatti (2008) suggested that vowels contain information about prosody which is used to interpret the syntactic structure whereas consonants' higher distinctive power within a word provides more cues for lexical identification than vowels. There were a number of studies investigating the relative contributions of vowels and consonants to speech (word and sentence) intelligibility (Chen, Wong & Wong, 2013; Cole, Yan, Mak, Fanty & Bailey, 1996; Fogerty & Humes, 2010; Fogerty & Humes, 2012; Kewley-Port, Burkle & Lee, 2007; Owren & Cardillo, 2006).

The study of Cole, Yan, Mak, Fanty and Bailey (1996) showed a two-to-one advantage of vowels over consonants in word recognition at sentence level using the noise-replaced stimuli (i.e., vowel-only sentences with consonants replaced by noise, and vice versa). Similar results were also found in the studies of Kewley-Port, Burkle and Lee (2007) and Forgerty and Humes (2012). A recent study concerning the segmental contribution to sentence intelligibility in Mandarin revealed a three-to-one advantage for perception of vowel-only sentences over perception of consonant-only sentences (Chen, Wong & Wong, 2013). This ratio was found to be higher than that in English, suggesting that vowels may play a more important role in sentence intelligibility for Mandarin.

Speech perception of sentences is different from that of isolated words in two aspects. One is the presence of linguistic context in sentences, and the other is the change in acoustic features in sentences. The linguistic context in sentences provides semantic and syntactic information, which allows listeners to predict the words by a top-down processing in addition to the acoustic cues. On the other hand, perception of isolated words is lack of context for prediction of words, and listeners mainly recognize isolated words by their acoustic features using a bottom-up processing (Denes & Pinson, 1993; Kewley-Port, Burkle & Lee, 2007). In terms of change in acoustic feature, there are prosodic changes in sentences (Shattuck-Hufnagel & Tuck, 1996). Since vowels are found to be the main carrier of prosodic information and convey information about syntax (Nespor, Pena & Mehler, 2003), the vowel advantage in word recognition at sentence level may be related to the syntactic content itself. Second, the acoustic information of phonemes spreads out to their adjacent words in sentences, which means that adjacent words also contain acoustic cues for word recognition in sentences (Janse & Ernestus, 2011). Due to the above-mentioned factors, the relative contributions of vowels and consonants to sentence intelligibility may not be the same as those to isolated word intelligibility.

Regarding the relative contributions of vowels and consonants to the perception of isolated word, there were a few studies conducted in English (Forgerty & Humes, 2010; Owren & Cardillo, 2006). Owren and Cardillo (2006) showed that listeners discriminated talker identity better when they were presented with vowel-only isolated words, and discriminated word meaning better when they were presented with consonant-only isolated words using a forced-choice judgment task. Forgerty and Humes (2010) used a word identification task to investigate the segmental contribution to recognize isolated English words. Although no significant difference was found between consonant-only and vowel-only words in word recognition performance, the performance with consonant-only words was more influenced by lexical difficulty than the performance with vowel-only words, suggesting that consonant has a more important role in lexical access. The possible reasons for greater contribution of consonants to lexical access include larger number of consonants and more distinctive nature of consonants compared to vowels in English (Forgerty & Humes, 2010).

Although there are studies concerning the segmental contribution to isolated word intelligibility in English, to date there is no study investigating the relative contributions of vowels and consonants to recognize isolated Mandarin words. Mandarin phonology is different from English. Mandarin is a tonal language which consists of four lexical tones (i.e., high level, rising, falling and rising, and falling) while English is not a tonal language (Howie, 1976). Moreover, Mandarin has a simpler syllable structure than English. Consonant cluster can appear in onset and coda positions for English syllables (McMahon, 2002) but not for Mandarin syllables. The number of vowels and consonants in Mandarin is also different from that in English. There are 35 vowels and 21 consonants in Mandarin (Chen, Wong & Wong, 2013). The number of vowels is greater than the number of consonants in Mandarin. In contrast, there are 20 vowels and 24 consonants in English according to the classification in English Pronouncing Dictionary (Jones, 2006). The number of vowels is fewer than the number of consonants in English. Therefore, the segmental contribution to isolated word intelligibility in English may be different from that in Mandarin.

Apart from the segmental contribution to isolated word intelligibility, it is also demonstrated that increasing the amount of consonant-vowel (C-V) boundary transitional information improved the isolated word intelligibility in English (Fogerty & Humes, 2010). Similar findings were also discovered in the study of Mandarin sentences (Chen, Wong & Wong, 2013). It is believed that C-V boundary consists of co-articulatory information which helps the identification of adjacent phonemes (Kent & Minifie, 1977; Recasens, 1999). No study has investigated the contribution of transitional boundary to isolated word intelligibility in Mandarin yet.

The purpose of the present study is (a) to determine the relative contributions of vowels and consonants in recognizing isolated Mandarin words; (b) to compare the relative contributions of vowels and consonants in isolated word recognition in Mandarin and that in English; and (c) to determine the contribution of C-V transitional boundary in recognizing isolated Mandarin words. In order to find out possible reasons for the relative contributions of vowels and consonants in recognizing isolated Mandarin words, the relative contributions of vowels and consonants in identification of consonants, vowels, and tones were also examined.

It was hypothesized that there would be a greater contribution of vowels in recognizing isolated words in Mandarin than that in English. And it was predicted that increasing C-V transitional boundary information would improve the isolated word recognition performance in Mandarin.

Method

Participants

Nineteen normal-hearing native speakers of Mandarin were recruited from the University of Hong Kong, including eight males and 11 females. The age range of the participants was 18 to 27 (mean age = 20.84). All of them were undergraduates or postgraduates studying at the University of Hong Kong and came from Mainland China. They all passed the hearing screening and their bilateral pure tone thresholds were all below 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 1996).

Materials

The isolated Mandarin words used in the test were taken from a corpus of 1128 isolated Mandarin (monosyllabic) words, covering almost all daily-used words in Mandarin Chinese. All the isolated words were produced by a female native Mandarin talker at a normal speaking rate and with broadcaster's voice quality. The fundamental frequency of recorded words ranged from 130 to 330 Hz. The C-V boundaries, defined based on traditional segmental boundaries, were labeled manually by an experienced phonetician, and later verified by another experienced phonetician. All final nasal consonants were counted as part of their preceding vowels. The average duration of the words was 468 ms (consonants: 118 ms, and vowels: 350 ms). The words include 21 consonants (/b/, /d/, /g/, /p/, /t/, /k/, /z/, /c/, /zh/, /ch/, /j/, /q/, /f/, /s/, /sh/, /r/, /x/, /h/, /l/, /m/, /n/), 35 vowels (/a/, /o/, /e/, /i/, /u/, /ü/, /ai/, /ei/, /ao/, /ou/, /ia/, /ie/, /iao/, /iou/, /ua/, /uo/, /uai/, /uei/, /üe/, /an/, /en/, /ang/, /eng/, /ong/, /ian/, /in/, /iang/, /ing/, /ion/, /uan/, /uen/, /uang/, /ueng/, /üan/, /ün/), and four tones (high level, rising, falling and rising, and falling).

Five types of stimuli were created. The first type was consonant-only (C-only) stimuli in which only the consonant segment was preserved (i.e., vowel replaced by noise). The second type was vowel-only (V-only) stimuli in which only the vowel segment was preserved (i.e., consonant replaced by noise). The third type preserved the consonant segment and some proportion of vowel transition at onset (C + *p*V), and similarly the forth type preserved the vowel segment and some proportion of consonant transition at offset (*p*C + V). The fifth type preserved some proportion of consonant and vowel transitions across C-V boundary (*p*C + *p*V). The proportion factor *p* included three values of 0.1, 0.2, and 0.5, representing 10%, 20%, and 50% of the consonant/vowel segment centered at the C-V boundary to be retained. There were a total of 11 speech processing conditions, as illustrated in Figure 1. Speech-shaped noise was used to replace the vowel/consonant segment to create stimuli in the 11 conditions, with signal-to-noise ratio -16 dB (Chen, Wong & Wong, 2013).

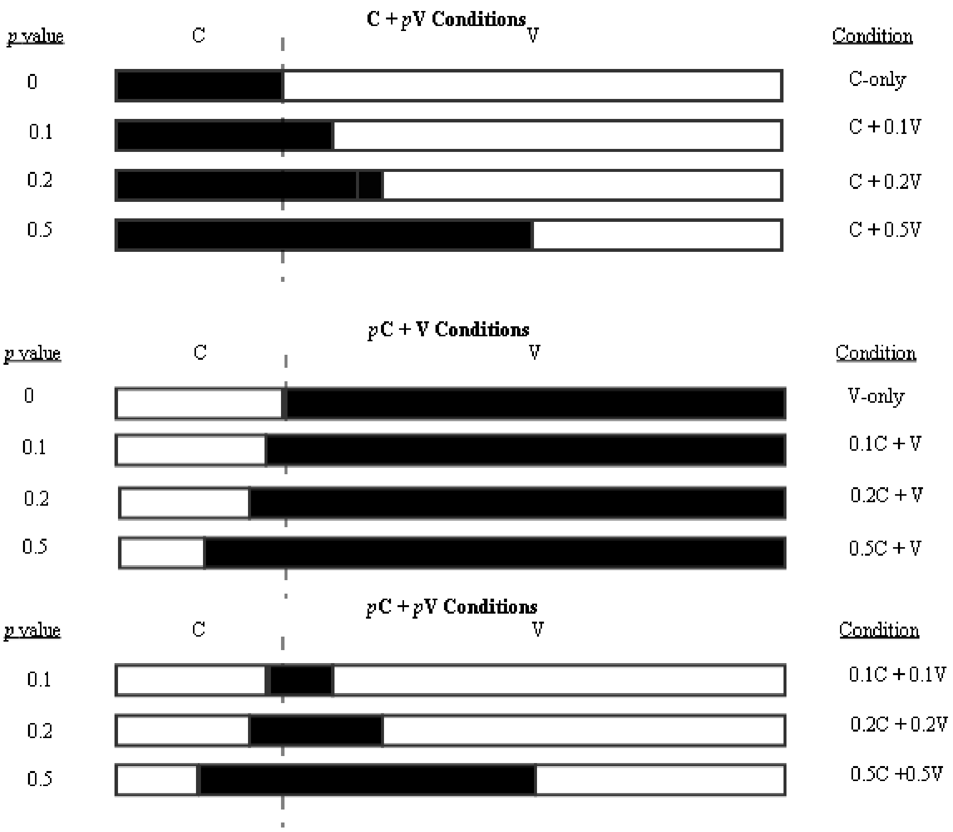

Figure 1. Schematic of the 11 speech processing conditions created. Dashed line represents the C-V boundary. Black bars represent the speech segments preserved. White bars represent the speech segments replaced by noise.

Procedure

The experiment was conducted in a sound-proof booth located at the Division of Speech and Hearing Sciences, the University of Hong Kong. Participants listened to the stimuli through a circumaural headphone, and the stimuli were played at a comfortable listening level. Each participant listened to all the 11 conditions in the experimental trials. The order of the conditions was randomized to minimize the practice effect. For each condition, 40 words were randomly taken from the word corpus for signal processing. Participants were allowed to listen to each stimulus at most three times. They were asked to (a) repeat the recognized word after listening to the stimulus presented, and (b) choose the corresponding HanYu PinYin (a standard phonetic writing system in Mainland China) from a MATLAB interface showing the 21 consonants, 35 vowels, and four tones in Mandarin. Before the experimental trials, all participants listened to 40 words in C + 0.4V condition, 40 words in 0.4C + V condition, and 40 words in 0.4C + 0.4V condition as practice trials in order to familiarize them with the noise-replaced stimuli and experimental procedure, including the use of the MATLAB interface for choosing the HanYu PinYin. Feedback was given in the practice trials. Their responses were scored based on four aspects: words, consonants, vowels, and tones. A five-minute break was given in every 30 minutes. The score for each condition was computed as the ratio between the number of the correctly recognized words/phonemes/tones and the total number (i.e., 40) of words/phonemes/tones contained in each condition.

Results

A. Contribution of Full Vowel versus Full Consonant

The scores of word recognition, consonant identification, vowel identification, and tone identification in the 11 conditions were converted into rational arcsine unit (RAU) for statistical analysis. Figure 2 shows the mean scores (in percentage accuracy) of consonant identification, vowel identification, tone identification, and word recognition in C-only condition and V-only condition. The scores of word recognition, consonant identification, vowel identification, and tone identification in C-only condition and V-only condition were compared using paired *t*-tests. It is demonstrated that the score of word recognition in C-only condition was significantly lower than that in V-only condition [$t(18) = -14.17$, $p < .001$]. Similarly, the scores in C-only condition were significantly lower than those in V-only condition in identifying vowels [$t(18) = -27.17$, $p < .001$] and tones [$t(18) = -37.39$, $p < .001$]. In contrast, the score of consonant identification in C-only condition was significantly higher than that in V-only condition [$t(18) = 7.82$, $p < .001$].

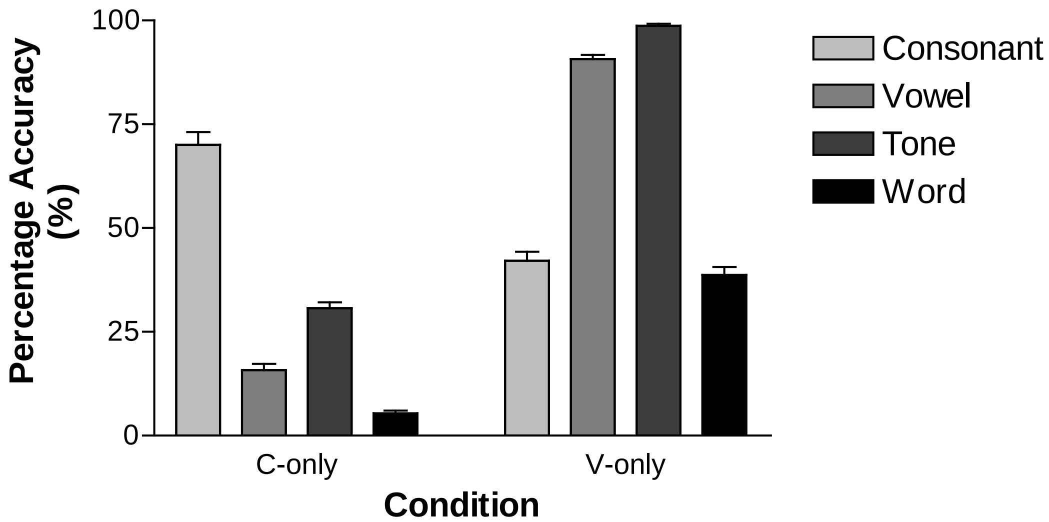

Figure 2. Mean scores of consonant identification, vowel identification, tone identification, and word recognition in consonant-only and vowel-only conditions. The error bars show the standard deviation of the mean.

B. Contribution of Proportion Factor *p*

The contribution of proportion factor *p* to the scores of word recognition, consonant identification, vowel identification, and tone identification were investigated. Figure 3 (a) shows the mean recognition scores (in percentage accuracy) of words in C + *p*V conditions, *p*C + V conditions, and *p*C + *p*V conditions. One-way repeated measure analysis of variance (ANOVA) was used to investigate the effect of altering proportion factor *p* in different signal processing conditions on word recognition. Significant main effect was found for C + *p*V conditions [$F(3, 54) = 185.21$, $p < .001$] and *p*C + *p*V conditions [$F(2, 36) = 142.65$, $p < .001$]. However, no significant main effect was found for *p*C + V conditions ($p = .37$). Pairwise comparison adjusted by Bonferroni method was used to compare word recognition score in C-only condition and word recognition score in C + 0.1V condition. It was found that the word recognition score in C + 0.1V condition was significantly higher than that in C-only condition ($p < .001$), that is, 16.7% versus 5.4%.

Figure 3 (b) shows the mean identification scores (in percentage accuracy) of consonants in C + *p*V conditions, *p*C + V conditions, and *p*C + *p*V conditions. One-way repeated measure ANOVA was used to investigate the effect of altering proportion factor *p* in different signal processing conditions on consonant identification. Significant main effect was found for C + *p*V conditions [$F(3, 54) = 37.00$, $p < .001$] and *p*C + *p*V conditions [$F(2, 36) = 66.49$, $p < .001$]. However, no significant main effect was found for *p*C + V conditions ($p = .45$). Pairwise comparison adjusted by Bonferroni method was used to compare consonant identification score in C-only condition and consonant identification score in C + 0.1V condition. It was found that the consonant identification score in C + 0.1V condition was significantly higher than that in C-only condition ($p < .001$), that is, 90.3% versus 70.0%.

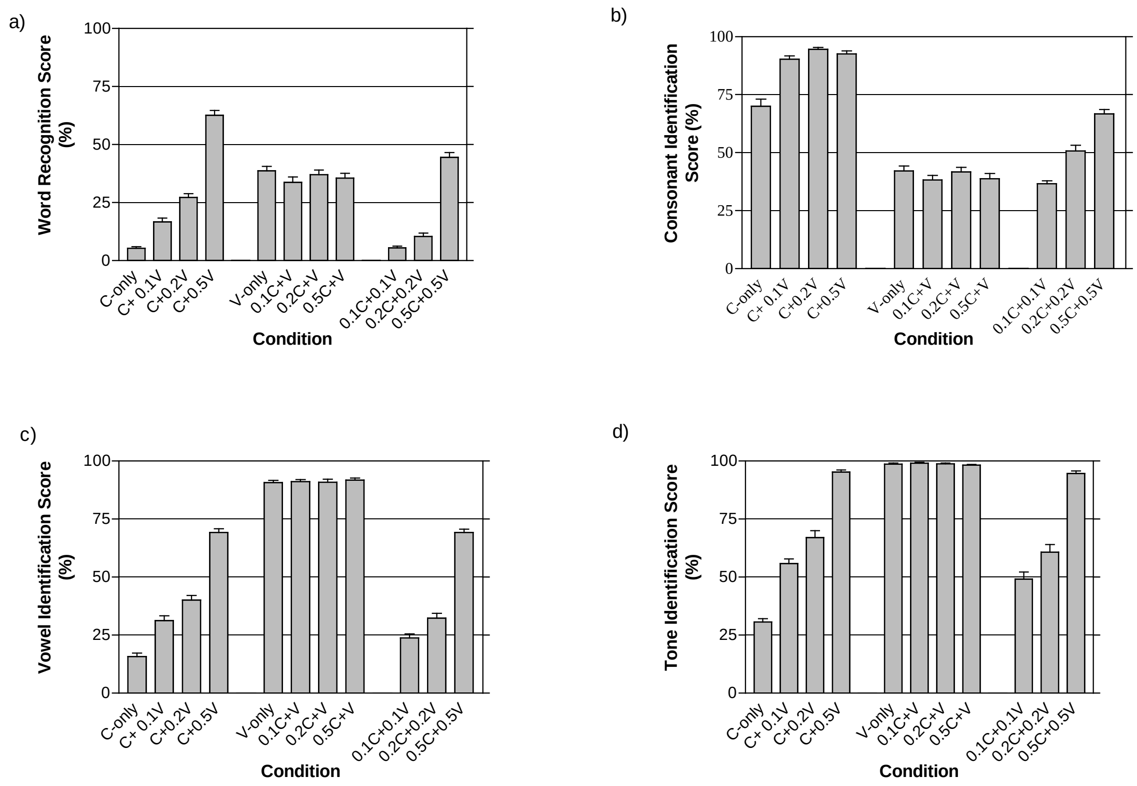

Figure 3. Mean scores in C + *p*V, *p*C + V, and *p*C + *p*V conditions in (a) word recognition, (b) consonant identification, (c) vowel identification, and (d) tone identification. The error bars show the standard deviation of the mean.

Figure 3 (c) shows the mean identification scores (in percentage accuracy) of vowels in C + *p*V conditions, *p*C + V conditions, and *p*C + *p*V conditions. One-way repeated measure ANOVA was used to investigate the effect of altering proportion factor *p* in different signal processing conditions on vowel identification. Significant main effect was found for C + *p*V conditions [$F(3, 54) = 123.03$, $p < .001$] and *p*C + *p*V conditions [$F(2, 36) = 211.75$, $p < .001$].

For *p*C + V conditions, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2 = 13.22$, $p = .022$, therefore multivariate tests are reported ($\varepsilon = .67$). The results show that the vowel identification score was not significantly affected by *p*C + V conditions, $V = .047$, $F(3, 16) = .26$, $p = .85$. Pairwise comparison adjusted by Bonferroni method was used to compare vowel identification score in C-only condition and vowel identification score in C + 0.1V condition. It was found that the vowel identification score in C + 0.1V condition was significantly higher than that in C-only condition ($p = .001$), that is, 31.3% versus 15.8%.

Figure 3 (d) shows the mean identification scores (in percentage accuracy) of tones in C + *p*V conditions, *p*C + V conditions, and *p*C + *p*V conditions. One-way repeated measure ANOVA was used to investigate the effect of altering proportion factor *p* in different signal processing conditions on tone identification. Significant main effect was found for C + *p*V conditions [$F(3, 54) = 232.54$, $p < .001$] and pC + pV conditions [$F(2, 36) = 161.27$, $p < .001$]. However, no significant main effect was found for *p*C + V conditions ($p = .22$). Pairwise comparison adjusted by Bonferroni method was used to compare tone identification score in C-only condition and tone identification score in C + 0.1V condition. It was found that the tone identification score in C + 0.1V condition was significantly higher than that in C-only condition ($p < .001$), that is, 55.9% versus 30.7%.

C. Comparison between Tone Identification Scores of C + *p*V Conditions and *p*C + *p*V Conditions

The tone identification scores in C + *p*V conditions and *p*C + *p*V conditions with proportion factor 0.1, 0.2, and 0.5 were compared using paired *t*-tests. The result in Figure 3 (d) demonstrated that the tone identification score in C + 0.2V condition (M = 66.87, SE = 3.11) was significantly higher than that in 0.2C + 0.2V condition (M = 60.39, SE = 3.17), $t(18) = 2.737$, $p = .014$. The tone identification score in C + 0.1V condition (M = 55.60, SE = 1.91) was marginally significantly higher than that in 0.1C + 0.1V condition (M = 49.06, SE = 3.00), $t(18) = 2.10$, $p = .05$. No significance difference was found in tone identification score between C + 0.5V condition (M = 105.02, SE = 2.32) and 0.5C + 0.5V condition (M = 105.16, SE = 2.90), $p = .94$.

D. Correlation between Consonant/Vowel/Tone/Word Scores and Duration

Note that the average duration of vowels and consonants were found to be 350 ms and 118 ms respectively in this study. The ratio of average duration of vowels to that of consonants was around 3:1. The duration of presented speech segments was calculated using this ratio (e.g., 25% for C-only condition; 32.5% for C + 0.1V condition).

The correlation between the duration of the stimuli and the scores of word recognition, consonant identification, vowel identification, or tone identification was analyzed. Figure 4(a)-(d) are four scatter plots of the recognition/identification scores of words, consonants, vowels, and tones against the duration of stimuli respectively. Correlational analysis demonstrated that the duration of the presented speech segments was significantly correlated with word recognition score, $r = .71$, consonant identification score, $r = -.26$, vowel identification score, $r = .93$, and tone identification score, $r = .88$ (all ps < .001). Though the duration of the presented speech segments was significantly correlated with all four recognition/identification scores, it is noted that its correlation with vowel and tone identification scores were much larger than that with consonant identification score (i.e., .93 and .88 vs. -.26).

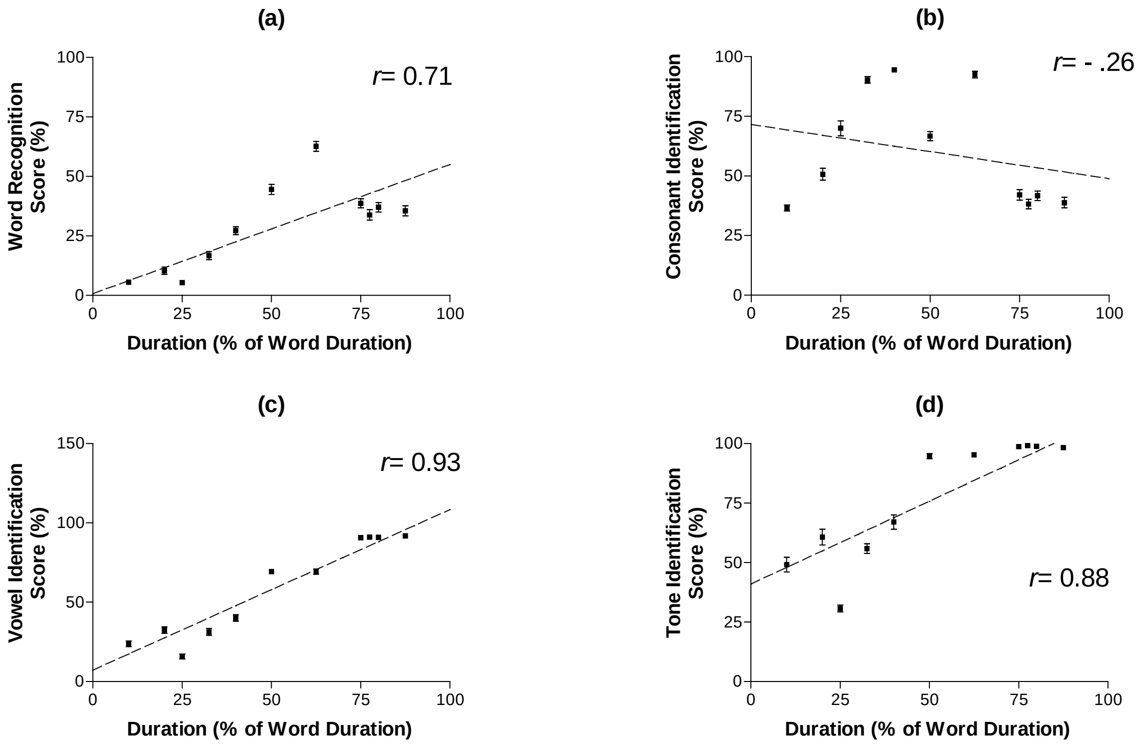

Figure 4. Scatter plots of (a) word recognition scores, (b) consonant identification scores, (c) vowel identification scores, and (d) tone identification scores against the duration of the speech segments presented. The error bars show the standard deviation of the mean scores.

E. Correlation between Consonant/Vowel/Tone Scores and Word Scores

The correlation of word recognition scores with consonant identification scores, vowel identification scores, and tone identification scores were analyzed. Figure 5(a)-(c) are three scatter plots of word recognition scores against consonant identification scores, vowel identification scores, and tone identification scores respectively. Correlational analysis demonstrated that word recognition score was significantly correlated with consonant identification score, *r* = .22, *p* = .001. It was also significantly correlated with vowel identification score, *r* = .73, and tone identification score, *r* = .78 (*p*s < .001). Though all are significantly correlated with word recognition score, it is noted that the correlation coefficients of vowel and tone identification scores are notably higher than that of consonant identification score (.73 and .78 vs. .22).

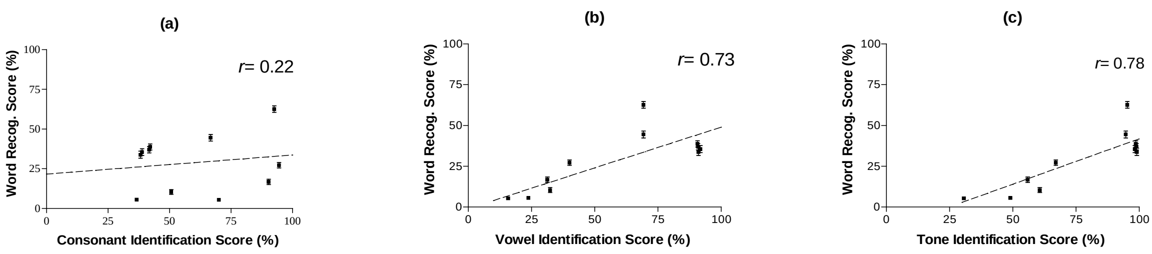

Figure 5. Scatter plot of word recognition score against (a) consonant identification score, (b) vowel identification score, and (c) tone identification score. The error bars show the standard deviation of the mean scores.

Discussion

A. Relative Contribution of Consonants and Vowels in Isolated Word Recognition in Mandarin

The present study reveals that there is a greater contribution of vowels than consonants in recognizing isolated Mandarin word. First, the word recognition score in V-only condition was significantly higher than that in C-only condition (i.e., 38.7% vs. 5.4%). Second, result of one-way repeated measure ANOVA showed a significant main effect of altering proportion factor *p* in word recognition for C + *p*V conditions but no significant main effect was shown for *p*C + V conditions. It demonstrates that adding a proportion of vowel onset improves the performance in word recognition relative to that in C-only condition, but adding a proportion of consonant does not. Third, the correlation between vowel identification score and word recognition score was much greater than that between consonant identification score and word recognition score (i.e., .73 vs. .22). These three findings together indicate that vowels contribute more to isolated word recognition in Mandarin than consonants, which is similar to the result found in Chen, Wong and Wong (2013) studying segmental contribution in Mandarin sentence intelligibility.

B. Comparison of Relative Contribution of Vowels and Consonants to Isolated Word Recognition between Mandarin and English

The result in present study is different from the findings in previous studies in English. The study of Owren and Cardillo (2006) showed that listeners discriminated word meaning better when they were presented with C-only isolated words than when they were presented with V-only isolated words in English using a forced-choice judgment task. The study of Forgerty and Humes (2010) found no significant difference between C-only and V-only words in word recognition performance, but found that the performance with C-only words was more influenced by lexical difficulty than the performance with V-only words, suggesting that consonant has a more important role in lexical access. These two studies in English implied a more important role of consonants than vowels in isolated word recognition in English. The difference between Mandarin and English suggests that the relative contributions of vowels and consonants in isolated word recognition are language-specific, that is, vowels might carry more perceptual contribution than consonants for recognizing isolated Mandarin words.

The difference in the relative contributions of vowels and consonants in isolated word recognition between Mandarin and English may be explained by their different phonological system. Mandarin words are monosyllabic whereas English words can be multisyllabic. Moreover, there are no consonant clusters in Mandarin whereas consonant clusters exist in English (McMahon, 2002; Kiparsky, 1981). The existence of consonant clusters is one possible reason why there is a more important role of consonants in word recognition in English. However, the syllabic structure is probably not the only reason to explain the difference between Mandarin and English. It is because Forgerty and Humes (2010) only used monosyllabic English words with consonant-vowel-consonant (CVC) structure in their study. Although their stimuli were not multisyllabic and did not contain consonant clusters, their finding was still different from that in Mandarin. This suggests the existence of other possible reasons accounting for the difference in the relative contributions of vowels and consonants between Mandarin and English in addition to syllabic structure.

There is a difference in the numbers of consonants and vowels between Mandarin and English. There are 35 vowels and 21 consonants in Mandarin, and the number of vowels is greater than the number of consonants in Mandarin (Chen, Wong & Wong, 2013). In contrast, there are 20 vowels and 24 consonants in English according to the classification in Cambridge

English Pronouncing Dictionary (Jones, 2006), and the number of vowels is fewer than the number of consonants in English. The difference in the ratio of the number of vowels to the number of consonants may partially interpret the finding that vowels had a greater contribution to word recognition in Mandarin whereas consonants had a greater contribution to word recognition in English.

Besides, the average durations of consonants and vowels in present study are different from that in previous study in English. The average durations of vowels and consonants are 350 ms and 118 ms respectively in the present study, and the ratio of average duration of vowel to that of consonant is around 3:1. The average durations of vowels and consonants are 200 ms and 237 ms respectively in the study of Forgerty and Humes (2010). The ratio of average duration of vowel to that of consonant is around 0.8:1. The average duration of vowels is longer than that of consonants in the present study, whereas the average duration of vowels is shorter than that of consonants in the study of Forgerty and Humes (2010). The difference in average duration of vowels and consonants may explain why vowels contribute more than consonants in isolated word recognition in the present study whereas consonants play a more important role in lexical access in the study of Forgerty and Humes (2010). However, it should be noted that Forgerty and Humes (2010) used only monosyllabic English words with CVC structure in their study. Since English words can be multisyllabic, the ratio of average duration of vowels to that of consonants in English may be different from the ratio used in Forgerty and Humes (2010)'s study. Although Owren and Cardillo (2006) used English words with two to five syllables, they did not report the average durations of vowels and consonants in their study. Therefore, whether this difference in average duration of vowels and consonants in isolated words is also true for multisyllabic English words cannot be concluded.

C. Contribution of Consonant-Vowel Transition to Word Recognition

Contribution of CV transition to isolated word recognition was investigated in the present study. Significant main effect was found for *p*C + *p*V conditions, indicating that increasing CV transitional information leads to improvement in word recognition performance. The finding is consistent with the English study of Fogerty and Humes (2010) and the study of Chen, Wong and Wong (2013) which investigated Mandarin sentence intelligibility. It suggests that CV boundary transitional information contributes to not only Mandarin speech (word and sentence) recognition, but also English word recognition. A possible explanation for this contribution is that CV transitional boundary consists of co-articulatory information which helps the identification of adjacent phonemes (Kent & Minifie, 1977; Recasens, 1999).

D. Contribution of Vowels and Consonants to Tone Identification Performance

As Mandarin is a tonal language, tone identification also plays a role in the recognition of isolated words in Mandarin. Correlational analysis demonstrates a great correlation between tone identification score and word recognition score ($r = .78$). Since the perception of lexical tone mainly depends on fundamental frequency, which is carried by the vowel segment (Howie, 1976), it is hypothesized that vowels contribute more to tone identification than consonants. The present study shows that the tone identification score in V-only condition was significantly higher than that in C-only condition (i.e., 98.7% vs. 30.7%), supporting the above-mentioned hypothesis. The greater contribution of vowels in tone identification than consonants in Mandarin may also explain why vowels contribute more than consonants in isolated word recognition in Mandarin, which is a tonal language.

In addition to the findings that vowels contribute more than consonants in tone perception, whether consonants contribute to tone identification in Mandarin was also investigated. The difference between C + *p*V and *p*C + *p*V conditions is the presence of whole consonant versus the presence of a proportion of consonant. Except the comparison of tone identification score between C + 0.2V and 0.2C + 0.2V conditions (i.e., 67.0% vs. 60.7%), there is only marginal or no significant difference in tone identification score between C + *p*V and *p*C + *p*V conditions (i.e., 55.9% vs. 49.1%, and 95.3% vs. 94.7%). It implies that the role of full consonants in tone identification may be limited.

E. Contribution of Segmental Duration to the Identification of Consonants, Vowels, and Tones

Correlational analysis revealed that the correlations of segmental duration with vowel and tone identification scores were greater than that with consonant identification score (i.e., .93 and .88 vs. -.26). It implies that segmental duration provides more acoustic cue for vowel and tone identification than consonant identification.

Limitation

Only 19 participants were recruited in the present study. The sample size is not large, which limits the ability to generalize the findings of the study to the Mandarin speaking population. Furthermore, there are a variety of dialects spoken in mainland China and people from different districts may speak or listen to different dialects in addition to Mandarin in their daily life. Different dialects may have different influences on the perception of Mandarin. Due to time and resource constraint, the participants recruited in this study did not cover every district or dialectal region in Mainland China. Therefore, the finding may not be generalized to the whole Mandarin speaking population in China. A larger scaled study using probabilistic sampling may yield findings that can be more representative of the population.

Future Direction

The present study only investigated the relative contributions of vowels and consonants in Mandarin isolated word recognition in quiet environment. The relative contributions of vowels and consonants in Mandarin isolated word recognition in noise can be further investigated in order to find out the acoustic cues that are more resistant to the influence of noise.

Since only normal hearing listeners were recruited in the present study, further study can investigate the relative contributions of vowels and consonants in Mandarin isolated word recognition in hearing impaired population to find out what kinds of acoustic cue are affected by hearing impairment.

Finally, future work may be proposing a model of spoken word recognition in Mandarin to predict word recognition performance by involving different identification tasks of consonants, vowels, and tones.

Conclusion

In summary, the present study shows that there is a greater contribution of vowels than consonants in isolated word recognition in Mandarin, which is different from the findings in English studies. Possible reasons for this language difference include differences in syllabic structure, numbers of consonants and vowels, and average durations of consonants and vowels between Mandarin and English. In addition, it is also found that there is a great correlation between vowel/tone identification score and word recognition score. The finding that vowels carry more tonal information than consonants may further explain the greater contribution of vowels than consonants in Mandarin isolated word recognition. Increasing CV boundary information improves word recognition performance in Mandarin, which is consistent with the findings in English.

Acknowledgements

The study was supported by Faculty Research Fund, Faculty of Education, The University of Hong Kong and Seed Funding for Basic Research, The University of Hong Kong.

I would like to express my sincere gratitude to my supervisor, Dr. Felix Chen, for his continuous support during the dissertation process. His guidance helped me to go through every stage of the dissertation. I would like to express my gratefulness to all the participants in the study, who have kindly devoted their precious time and effort to the study. Finally, I would like to take this opportunity to thank my family, friends, and classmates for their encouragement and unconditioned support over the past four years. In particular, I am grateful to my classmate Sharon Wong for her stimulating opinions and mental support.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

ANSI (1996). *American national standards specification for audiometers*. [ANSI S3.6-1996 (R1973)]. New York: American National Standards Institute.

Chen, F., Wong, L. L., & Wong, E. Y. (2013). Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility. *The Journal of the Acoustical Society of America*, *134*, EL178-EL184. doi: 10.1121/1.4812820

Cole, R. A., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing,* 853-856. doi: 10.1109/ICASSP.1996.543255

Denes, P., & Pinson, E. (1993). *The speech chain: The physics and biology of spoken language.* New York, NY: W. H. Freeman.

Fogerty, D., & Humes, L. E. (2010). Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors. *The Journal of the Acoustical Society of America*, *128*, 3114-3125. doi: 10.1121/1.3493439

Fogerty, D., & Humes, L. E. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *The Journal of the Acoustical Society of America*, *131*, 1490-1501. doi:

10.1121/1.3676696

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

Janse, E., & Ernestus, M. (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: evidence from listeners with normal and impaired hearing. *Journal of Phonetics, 39*, 330-343. doi: 10.1016/j.wocn.2011.03.005

Jones, D. (2006). *English pronouncing dictionary.* Cambridge: Cambridge University Press.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5*, 115-133.

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech* (2nd ed.). Albany, NY: Singular/Thomson Learning.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 122,* 2365-2375. doi: 10.1121/1.2773986

Kiparsky, P. (1981). Remarks on the metrical structure of the syllable. In W. Dressler (Ed.) *Phonologica 1980*. Proceedings of the Fourth International Phonology Meeting.

Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants*. West Sussex: John Wiley & Sons.

McMahon, A. M. (2002). *An introduction to English phonology*. Edinburgh: Edinburgh University Press.

Nespor, M., Pena, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio*, *2*, 203-230. doi: 10.1418/10879

Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *The Journal of the Acoustical Society of America*, *119*, 1727-1739. doi: 10.1121/1.2161431

Pickett, J. M. (1999). *The acoustics of speech communication: Fundamentals, speech perception theory, and technology*. Boston: Allyn and Bacon.

Recasens, D. (1999). Acoustic analysis. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data, and techniques* (pp. 322-336) . Cambridge: Cambridge University Press.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, *25*, 193-247. doi: 10.1007/BF01708572

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream functional differences between vowels and consonants. *Psychological Science*, *19*, 137-144. doi: 10.1111/j.1467-9280.2008.02059.x

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. M. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34-57). Cambridge: Cambridge University Press.