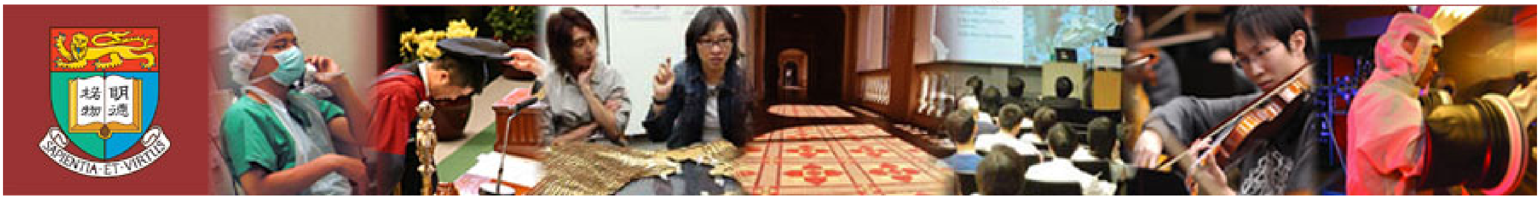

Relative contributions of consonants and vowels to Mandarin sentence intelligibility

Wong Yee Wa

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, 2013.

Abstract

This study investigated the relative contributions of consonants and vowels to Mandarin sentence intelligibility using a noise replacement paradigm. In Experiment 1, 20 young normal-hearing native Mandarin listeners recognized Mandarin sentences with various amounts of segmental information preserved. Results showed that the vowel-only sentences (consonants replaced by noise) yielded a remarkable 3:1 intelligibility advantage over the consonant-only sentences (vowels replaced by noise). This ratio was larger than that found in English, suggesting that vowels contribute more to sentence intelligibility in Mandarin than in English. Intelligibility increased significantly when a little portion at vowel onsets was added to the consonant-only sentences. However, intelligibility of the vowel-only sentences was still maintained high when the same portion of vowel onsets and an equal amount of vowel offsets were replaced by noise, suggesting that vowel onsets contain redundant information to vowel centers for Mandarin sentence recognition. In Experiment 2, the same listeners discriminated tones of vowels with various durations preserved at either onsets or centers. Results were compared with the findings in Experiment 1, and suggested that lexical tones are relatively redundant for Mandarin sentence intelligibility in quiet.

Relative contributions of consonants and vowels to Mandarin sentences intelligibility

Consonants and vowels are two categories of speech sounds which exist in all languages (Ladefoged, 2001). Consonants are characterized by complete or partial vocal tract constriction, high frequency, and short duration; while vowels are characterized by a relatively open vocal tract with sustained voicing, low frequency and long duration.

Theoretically, knowing which segment contributes more to speech intelligibility can tell us how listeners make use of acoustic cues in consonants and vowels to understand speech. This knowledge is also useful in clinical application such as for the design and programming of amplification devices. The specific acoustic feature could be enhanced for better speech perception.

Previous findings about the segmental contributions to English sentence intelligibility

Many researches had been conducted to investigate the relative perceptual contributions of consonants and vowels to English intelligibility and it was consistently found that vowels provide a 2:1 intelligibility advantage over consonants in English sentences.

In a preliminary study investigating segmental contributions to sentence intelligibility, Cole, Yan, Mak, Fanty, and Bailey (1996) replaced consonant and vowel segments in the English sentences extracted from the TIMIT database (Garofolo *et al.*, 1990) by either speech-shaped noise or harmonic complexes. This process is known as the noise replacement paradigm. It was found that the vowel-only sentences (consonant-replaced) yielded a 2:1 intelligibility advantage over the consonant-only sentences (vowel-replaced) under both types of segmental replacement. The intelligibility advantage of the vowel-only sentences persisted even when 20 ms more vowel portion (10 ms from the onset and 10 ms from the offset) was replaced by noise in each vowel.

In a latter study, Kewley-Port, Burkle, and Lee (2007) used similar methodology and stimuli as the study by Cole *et al.* (1996). However, they used speech-shaped noise only for the segmental replacement and controlled the presentation level of the stimuli. Kewley-Port *et al.* (2007) also confirmed the 2:1 intelligibility advantage of vowels for young normal-hearing (NH) listeners at 70 dB SPL and for elderly hearing-impaired (HI) listeners at 95 dB SPL. Although elderly HI listeners had overall poorer performance than young NH listeners, listeners from both groups always showed significantly better performance in the vowel-only condition compared to the consonant-only condition.

Latter, a further study was conducted by Fogerty and Kewley-Port (2009) to investigate how the relative perceptual contributions of consonants and vowels are affected by the co-articulation information at the consonant-vowel (C-V) boundaries using the same English sentences database and noise replacement paradigm as Kewley-Port *et al.* (2007). It is known that acoustic cues for segments can be found across C-V boundaries, and co-articulation means a single segment gives acoustic information about identities of more than one phoneme (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In the study, the C-V boundary was specified by the TIMIT database. Glimpse windows were defined as the preserved speech signal intervals and contained proportional amounts of vowel information at the C-V boundary (i.e. VP) that was either added to consonants (i.e. C+VP) or deleted from vowels (i.e. V-VP). The study once again confirmed the 2:1 intelligibility advantage of vowels over consonants in English sentences. Besides, it was found that sentence intelligibility increased linearly under the C+VP conditions with the increase in VP, and the intelligibility under V-VP conditions was unaffected until 30% VP was replaced by noise. It suggested that information at the C-V boundary may be replicated with that carried by vowel center, and the intelligibility advantage of vowels is robust against deletion.

Recently, Fogerty, Kewley-Port, and Humes (2012) examined the relative contributions of consonants and vowels to English sentence intelligibility in young NH listeners, old NH listeners, and elderly HI listeners. In their study, they used stimuli from the same sentences database as Fogerty and Kewley-Port (2009) and Kewley-Port *et al.* (2007). Segments containing predominantly consonants or vowels were prepared by noise replacement using low-level speech-shaped noise. Again, vowels were found to contribute more than consonants to sentence intelligibility in all groups, suggesting that the vowel intelligibility advantage in sentence recognition is independent of age and hearing sensitivity of listeners.

Difference between English and Mandarin

To date no studies have examined the contributions of vowels and consonants to

Mandarin sentence intelligibility. Although both Mandarin and English have consonants and vowels, they are different in many aspects. First of all, Mandarin is a tonal language in which tone is important for distinguishing lexical meanings. The four lexical tones in Mandarin are characterized by the fundamental frequency (F0) contours of the voiced segment: Tone 1 has flat F0, Tone 2 has rising F0, Tone 3 has falling-rising F0, and Tone 4 has falling F0 (Howie, 1976). Identical syllable with different tones convey different meanings. For example, /ma/ can mean *mother* (Tone 1), *hemp* (Tone 2), *horse* (Tone 3), or *scold* (Tone 4). In contrast, F0 contour in English conveys no lexical meaning.

Secondly, the syllable structure in English is more complex such that consonant clusters can appear in both onset and coda of a syllable. In contrast, Mandarin has no consonant cluster and is made up of primarily consonant-vowel syllables. Li, Tan, McLoughlin, and Teo (2000) proposed that there are just about 415 permutations of Mandarin syllable in common use. Even when lexical tones are considered, there are only about 1200 syllables in Mandarin (Howie, 1976). However, there are far more permutations of syllables in English because of the existence of consonant clusters. There are about 10,000 non-homophonous monosyllables in English according to the CMU Pronouncing Dictionary (Carnegie Mellon University, 2008). Hence, the probability that a listener could accurately identify a syllable based on the vowel information alone would be lower in English because of the existence of the large number of syllable permutations.

Thirdly, Mandarin has a monosyllabic nature while English can be either monosyllabic or multisyllabic. Every Mandarin syllable can be transcribed to at least one corresponding Chinese character; however, syllables within a specific English multisyllabic word must be processed as a whole in order to be recognized. This increases the difficulty for listeners to correctly identify English word than Mandarin word in an intelligibility testing.

Given the difference discussed above, there may be a difference in the relative contributions of vowels and consonants to sentence intelligibility in English and Mandarin. It is also possible that difference in the distribution of vocalic and consonantal information across the C-V boundary exists. Hence, it is necessary to investigate the relative segmental contributions to sentence intelligibility in Mandarin.

Relationship between tone identification and acoustic cues in vowels

There is a close relationship between tone perception and the acoustic cues in vowels. It is known that F0 contour of vowel is the dominant cue for tone perception (e.g. Lin, 1988; Whalen & Xu, 1992) while other cues such as amplitude contour and vowel duration also have some contributions (Fu & Zeng, 2000; Fu, Zeng, Shannon, & Soli, 1998; Liang, 1963; Whalen & Xu, 1992). For example, the amplitude contour is correlated with the F0 contour of vowel and is found to be important to the discrimination of Tone 3 and 4 (Fu & Zeng, 2000; Whalen & Xu, 1992). In addition, the vowel duration makes contribution to Tone 3 identification as the duration of Tone 3 is significantly longer than that of the other three tones (Fu & Zeng, 2000).

Given that tone carries lexical meaning in Mandarin (e.g. Lin, 1988; Wang, 1989) and the close relationship between tone perception and the acoustic cues carried by a vowel, tone may have influence in the contribution of vowels to Mandarin sentence intelligibility. This possible sub-segmental contribution may be unique to tonal language and should be investigated to learn more about the vowel contribution to Mandarin sentence intelligibility.

Purpose of study

There were two experiments in the present study. The first experiment aimed to investigate the relative contributions of consonants and vowels as well as the contribution of acoustic information at the C-V boundary and vowel center to Mandarin sentence intelligibility. The second experiment studied the contribution of tone to Mandarin sentence intelligibility.

Participants

Twenty (9 male and 11 female) young NH native Mandarin listeners were paid to participate. Young NH participants were recruited to ensure their speech recognition would not be affected by hearing loss. The participants' ages ranged from 19 to 32 years old (M = 25.6), with majority being students at The University of Hong Kong. All participants attended pure-tone audiometry to ensure their bilateral pure tone thresholds are not greater than 20 dB HL at octave intervals from 250 to 8000Hz (American National Standards

Institute, 1996). Tympanometry was conducted to ensure all participants had no middle ear dysfunction according to the tympanometric norms for Chinese young adults (Wan & Wong, 2002). The participants participated in both experiments.

Experiment 1: Contributions of segments and C-V boundary

The first experiment investigated the relative perceptual contributions of vowels and consonants to Mandarin sentence intelligibility. As the acoustic characteristics of vowels and consonants are universal across all languages, it was expected that vowels would make greater contribution to sentence intelligibility than consonants in Mandarin, as in English. However, it was expected that difference would exist between English and Mandarin in the magnitude of contribution to sentence intelligibility by vowels and consonants respectively because differences exist between the syllable structures of the two languages.

On the other hand, in order to study the perceptual contribution of co-articulation information at the traditional C-V boundary and that of vowel center, this experiment also attempted to examine the effect of adding various amounts of initial vowel portions to the consonant-only sentences or deleting various amounts of vowel portion from both onsets and offsets of vowels in the vowel-only sentences on Mandarin sentence intelligibility.

Method

Materials. The sentence materials were extracted from the Mandarin speech perception test database (MSP; Fu, Zhu, & Wang, 2011). One hundred sentences from the

10 lists of the database were used, and each sentence contains seven monosyllabic words. The distribution of vowels, consonants, and tones were phonetically balanced based on 3500 commonly spoken Mandarin words, and there was no significant difference in sentence identification across the MSP lists (Fu *et al.*, 2011). In this study, phonemes were specified by an experienced phonetician based on the traditional segmental boundaries and confirmed by Praat, which is a computer software for acoustic analysis of speech (Boersma & Weenink, 2011). Syllables containing both vowel and final nasal consonant were considered as compound nasal vowels because MSP was validated based on this classification. Thus, 35 vowels and 21 consonants were used (see Appendix).

Two strategies were used to prepare stimuli. The first strategy preserved the entire consonants (C) and some proportion of vowel duration (VP) at vowel onsets, with the remaining vowel portion replaced by noise. This stimulus type is denoted by C+VP. The second strategy preserved various portions of vowel (V) at vowel centers but replaced the entire C and some VP from the vowel onset and offset by noise. This stimulus type is denoted by VC, where VC represents vowel center. For C+VP type, five values of VP were chosen to prepare five conditions. They are 0%, 10%, 20%, 30%, and 40%. For VC type, same values of VP were chosen to be deleted from both vowel edges (i.e. onset and offset). Hence, there are also five conditions, preserving 100%, 80%, 60%, 40%, and 20% of vowel segment at vowel center respectively. Figure 1 shows an outline of the 10 conditions.

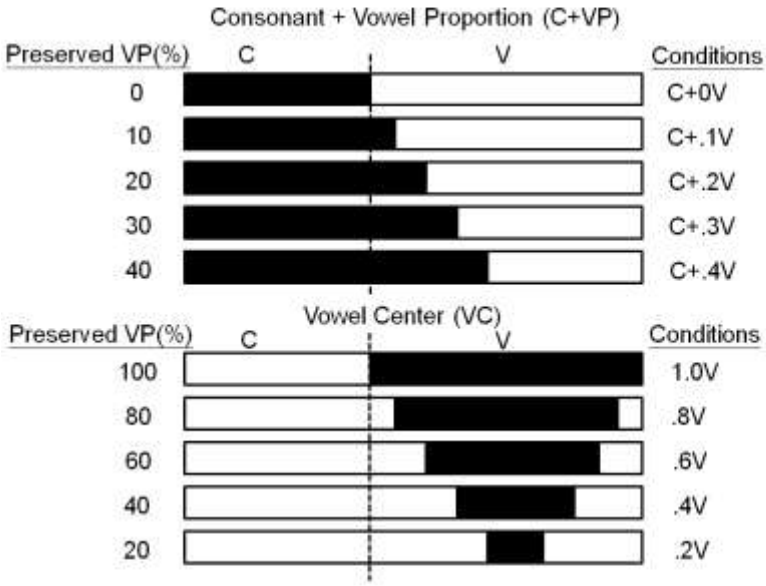

Figure 1. Outline of noise replacement conditions for a syllable within a sentence (not drawn in scale). Black and grey bars represent the preserved and noise-replaced portions respectively. Dashed lines show the boundary of consonant and vowel portions.

All intervals within a sentence were replaced by a speech-shaped noise scaled to -16 dB relative to the average speech level of the intact sentence. This noise level had also been used in previous studies (e.g. Fogerty & Kewley-Port, 2009), and it was chosen to maintain speech continuity and prevent phonemic restoration. It also avoided giving cues to the listeners about the segment amplitude.

Procedure. Participants listened to all sentence stimuli at a comfortable presentation level through a circumaural headphone in a sound-proof booth. They listened to 40 practice sentences before the experiment. Feedback of sentence meaning was given to familiarize the participants with the procedure and sentences processed by the noise replacement paradigm. The experiment started after the participants listened to all practice sentences and told the investigator that they understood the procedure. The order of the 10 experimental conditions was randomized across participants. There were 10 sentences per condition, and no sentence was repeated across conditions. Participants could listen to each stimulus for at most three times. They were asked to repeat all the words they could recognize. Their responses were scored in the session by the investigator. A 5 min break was given in every 30 min.

Results

Sentence intelligibility was calculated based on the percentage correct score obtained by dividing the total number of words correct by the total number of word stimuli. The results are displayed on Figure 2.

To investigate the relative perceptual contributions of vowels and consonants to Mandarin sentence intelligibility, a Wilcoxon signed-rank test was conducted because ceiling effect was noted in the vowel-only condition. Sentence intelligibility was significantly higher in the vowel-only condition (*Mdn* = 100.0%) than in the consonant-only condition (*Mdn* = 37.9%), $z = -3.92$, $p < .001$, $r = -.62$.

As the distribution of score measured in the conditions C+.2V and C+.3V was found to be significantly different from the normal distribution using Kolmogorov-Smirnov test ($p < .05$), intelligibility of the five C+VP conditions was compared using Friedman's ANOVA and suggested significant difference, $X^2(4) = 58.86$, $p < .001$. Wilcoxon tests were used for paired comparisons across the conditions with Bonferroni correction. A p value of .005 level (2-tailed) was used to evaluate significance of findings. It was found that intelligibility measured in the C+0V condition (*Mdn* = 37.9%) was significantly lower than those in the C+.1V (*Mdn* = 85.0%), C+.2V (*Mdn* = 92.9%), C+.3V (*Mdn* = 97.1%), and C+.4V (*Mdn* = 96.4%) conditions, $z = -3.92$, $p < .001$, $r = -.62$. Intelligibility measured in the C+.1V condition was significantly lower than those in the C+.3V and C+.4V conditions, $-3.77 \leq z \leq -3.57$, $p < .001$, $-.60 \leq r \leq -.57$. Intelligibility measured in the C+.2V condition was significantly lower than that in the C+.3V condition, $z = -3.62$, $p < .001$, $r = -.57$.

As ceiling effect was noted in the 1.0V and .8V conditions, Friedman's ANOVA was used and revealed significant difference in performance among the five VC conditions, $X^2(4) = 73.44$, $p < .001$. Wilcoxon tests were applied with Bonferroni correction. A p value of .005 level (2-tailed) was used to evaluate significance of findings. The tests revealed that intelligibility measured in the 1.0V condition (*Mdn* = 100.0%) was significantly higher than those in the .6V (*Mdn* = 89.3%), .4V (*Mdn* = 62.9%), and .2V (*Mdn* = 20.0%) conditions, $-3.92 \leq z \leq -3.55$, $p < .001$, $-.62 \leq r \leq -.56$. Intelligibility measured in the .8V condition (*Mdn* = 98.6%) was significantly higher than those in the .6V, .4V, and .2V conditions, $-3.92 \leq z \leq -3.73$, $p < .001$, $-.62 \leq r \leq -.59$. Intelligibility measured in the .6V condition was significantly higher than those in the .4V and .2V conditions, $-3.88 \leq z \leq -3.92$, $p < .001$, $-.62 \leq r \leq -.61$. Intelligibility measured in the .4V condition was significantly higher than that in the .2V condition, $z = -3.92$, $p < .001$, $r = -.62$.

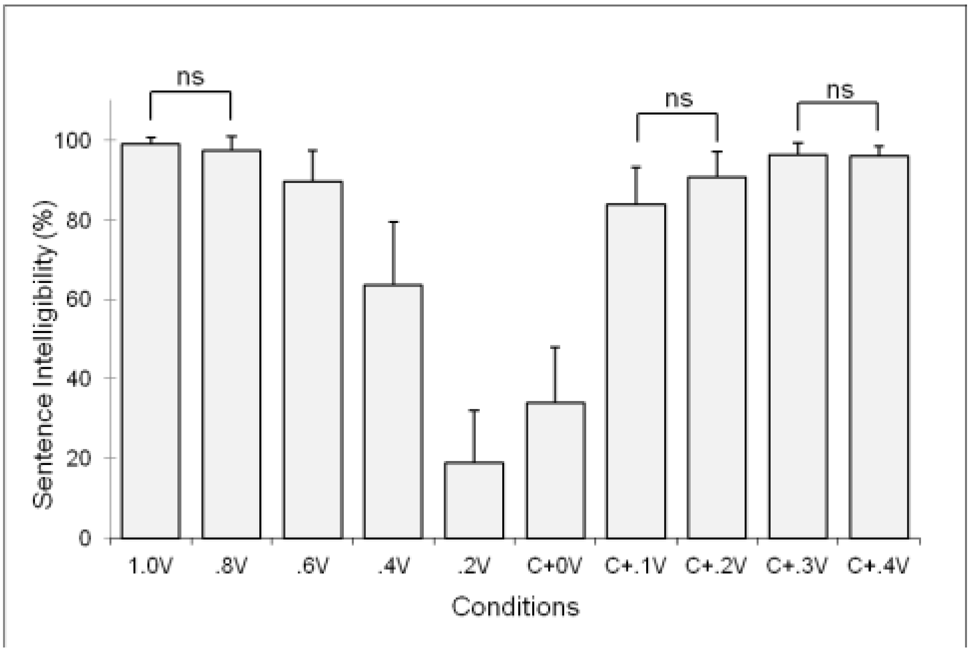

Figure 2. Means of sentence intelligibility of the 10 conditions in Experiment 1. Error bars show standard deviation. "ns" indicates that the difference of the paired intelligibility scores is not significant (p > .005).

Discussion

Relative segmental contribution to Mandarin sentence intelligibility. The result showed that vowels contribute more than consonants to Mandarin sentence intelligibility, which is consistent with previous findings obtained in studies with English (e.g., Cole *et al*., 1996; Kewley-Port *et al*., 2007). It suggested that intelligibility advantage of vowels was not restricted to English but also found in Mandarin, as a tonal language. Possible reason may be that the basic acoustic characteristics of consonants and vowels are universal across languages so they carry same cues in different languages. Some studies of artificial languages have found that vowels carry cues for syntax but consonants carry cues for word identity, supporting that there are functional differences between vowels and consonants (Bonatti, Pena, Nespor, & Mehler, 2005; Toro, Nespor, Mehler & Bonatti, 2008). Hence, vowels may facilitate top-down processing of speech and contribute more to intelligibility in sentence context. However, whether the vowel intelligibility advantage in sentence context can be found in other languages still need further studies with different languages to confirm.

Despite the similarity, there is still difference in the ratio of segmental contributions to sentence intelligibility between English and Mandarin. In English, vowels were found to have a 2:1 intelligibility advantage over consonants (Cole *et al*, 1996; Fogerty & Kewley-Port, 2009; Fogerty *et al.*, 2012; Kewley-Port *et al.*, 2007). However, Mandarin vowels were found to have a 3:1 intelligibility advantage over consonants in this study (Vowel-only condition: M = 99.0%; Consonant-only condition: M = 34.1%).

The acoustic properties of English and Mandarin syllable may contribute to this difference. First, tones are important for lexical contrasts in Mandarin. Cues for tone identification such as F0 contour, amplitude contour, and vowel duration are mainly located in vowel segments (Fu & Zeng, 2000; Fu *et al.*, 1998; Liang, 1963; Lin, 1988; Whalen & Xu, 1992). Hence, Mandarin vowels may have a more important role in speech intelligibility because of their additional role of carrying lexical tones. Second, Mandarin has no consonant clusters and is primarily made up of C-V syllables. Thus, the relative importance of vowels for phonemic contrast in Mandarin should be higher than that in English. Third, there are 21 consonants and 35 vowels in Mandarin according to the classification used by MSP (Fu et al., 2011), but 32 consonants and 20 vowels in English according to the classification used by Fogerty & Kewley-Port (2009), Fogerty *et al*. (2012), and Kewley-Port *et al*. (2007). With more vowels but fewer consonants in Mandarin, Mandarin vowels would naturally contribute more to intelligibility than English vowels for phonemic contrasts.

Furthermore, it was found that the total vocalic interval was shorter than the total consonantal interval in English sentences, but the opposite was found in Mandarin. Ramus, Nespor, and Mehler (1999) found that vocalic interval constituted at least 10% less of the total sentence duration than the consonantal interval in English. However, the average duration of a vowel and a consonant in Mandarin was found to be 187 ms and 76 ms respectively based on the MSP database. Hence, the total vocalic interval would be longer than the total consonantal interval in Mandarin sentences as Mandarin is made up of primary consonant-vowel syllables. Longer vocalic interval in Mandarin sentences may increase the vowel intelligibility advantage in sentence context by providing more acoustic information (e.g. lexical tones, stress) for sentence recognition.

Kewley-Port *et al*. (2007) found that presentation level of stimuli may alter the ratio of segmental contributions to sentence intelligibility. In their study, they found that the ratio between intelligibility of vowel-only and consonant-only conditions was about 2:1 at 70 dB

SPL, and the ratio reduced to 1.26:1 at 95 dB SPL for young NH listeners. Stimuli in the current study were presented at most comfortable listening level as in Cole *et al.* (1996). Although the presentation level was not strictly controlled, Cole *et al.* (1996) also found that English vowels had a 2:1 intelligibility advantage over consonants. Also, Kewley-Port *et al.* (2007) believed that 70 dB SPL was an intensity level close to the comfortable listening level. Thus, presentation level should not have affected the results in this study.

Contribution of C-V boundary to Mandarin sentence intelligibility. For the C+VP conditions, there was a general pattern that sentence intelligibility increased with increased vowel proportion added to consonants. Thus, providing more transitional information at the C-V boundary could increase the sentence intelligibility under consonant-dominant conditions (i.e. C+VP).

However, unlike the pattern found in the English study by Fogerty & Kewley-Port (2009) that the increase was linear, the trend observed in this study was non-linear. The sentence intelligibility increased significantly when 10% VP was added to the consonant-only condition. Then, the increase in intelligibility was not significant until 20% more VP was added to the C+.1V condition. It seems to suggest that the transitional information at C-V boundary is rich with acoustic information of vowels that is redundant with information at vowel center. As a result, there is a large increase when 10% VP was added to the consonant-only condition but the increase became smaller when more VP was added. This hypothesis was supported by the finding that the sentence intelligibility did not decrease significantly when a total of 20% vowel portion was deleted from the vowel onset and offset.

Contribution of vowel center to Mandarin sentence intelligibility. The vowel-only condition yielded a mean of 99.0% accuracy, and the intelligibility did not decrease significantly even when 20% of vowel portion was removed. This finding was different from English studies because vowel-only condition never had more than 90% intelligibility in those studies (Cole *et al*, 1996; Fogerty & Kewley-Port, 2009; Fogerty *et al*., 2012; Kewley-Port *et al*., 2007). Therefore, it seems that vowel-only condition is sufficient for listeners to identify Mandarin sentences in quiet environment with consonant replaced by noise. Even when 10% of vowel portion was deleted from each edge of a vowel so that possible consonant information present at the transitional boundary had been reduced much, the intelligibility did not significantly decreased. Furthermore, the intelligibility of Mandarin sentences can still be maintained at a mean of 89.6% even when only 60% of vowel portion was preserved at the vowel center. Hence, vowel center seems to have an important role in contribution to Mandarin sentence intelligibility.

Experiment 2: Role of tone in vowel contribution

In order to find out the effect of tonal information to vowel contribution to Mandarin sentence intelligibility, the second experiment was conducted. The result obtained was compared with the results of the first experiment so as to investigate the relationship between tone perception and sentence intelligibility.

Method

Materials. There were 48 vowel stimuli generated from six single vowels (/a/, /o/, /e/, /i/, /u/, / ü/) spoken with the four Mandarin tones by a male and a female native Mandarin speakers (i.e. 6 vowels × 4 tones × 2 speakers). The duration of each stimulus was normalized to 400 ms to remove duration cue for tone recognition (e.g., Fu & Zeng, 2000).

The original (FULL) vowels were modified to produce 2 more stimulus types. The first type was Left-only (LO). Variable lengths of the final vowel portions were replaced by speech-shaped noise, producing stimuli that have 40, 80, 120, 160, and 200 ms of the initial vowel portions preserved. The second type was Center-only (CO). Both the initial and final portions of vowels were replaced by noise, yielding stimuli that contain 240, 180, 120, 90, or 60 ms of the vowel centers. Figure 3 shows an outline of the 11 conditions.

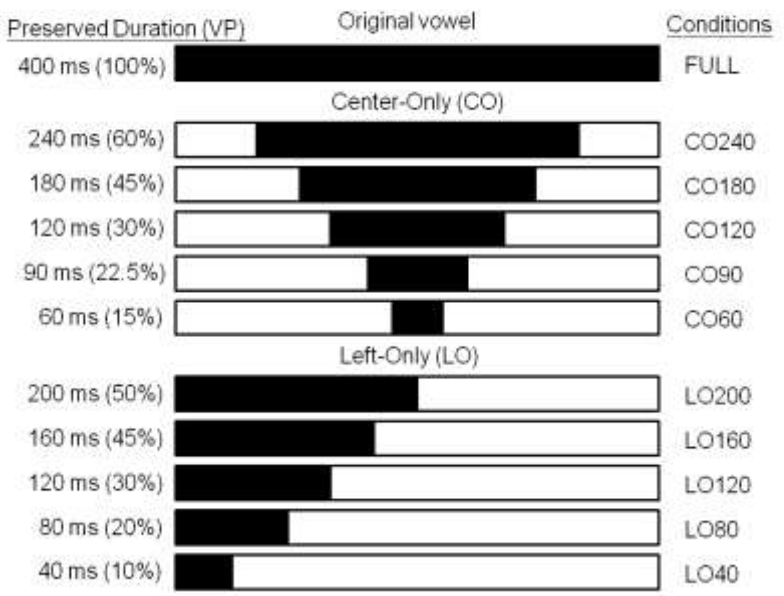

Figure 3. Outline of noise replacement conditions for a vowel (not drawn in scale). Black

and grey bars represent the preserved portions and the noise-replaced portions respectively.

A speech-shaped noise was used for all replacements within a single vowel. The intensity of the noise was scaled to -16 dB relative to the original vowel intensity. The noise replacement was used to maintain speech continuity. This noise level was chosen to avoid phonemic restoration and remove cues of amplitude contours (Whalen & Xu, 1992; Fu & Zeng, 2000).

Procedures. Participants listened to all vowel stimuli at a comfortable presentation level through a circumaural headphone in a sound-proof booth. There were 11 conditions with 48 stimuli (2 speakers × 6 vowels × 4 tones) per condition. Before the experiment, participants listened to 48 FULL stimuli as practice. Feedback of tone number was provided to familiarize the participants with the procedure and vowels which normalized to 400 ms. The order of the conditions was randomized across participants with the exception that FULL condition was always presented first. Participants needed to respond by choosing the correct tone number from four choices displayed on a computer screen using a mouse. Participants were required to obtain at least 90% accuracy in FULL condition after practice in order to proceed to the other 10 conditions. A 5 min break was given in every 30 min.

Results

The participants' identification scores were calculated by dividing the correct response by the total number of stimuli in that condition. The results are displayed on Figure 4.

The effect of perseveration of various amounts of initial vowel portions was investigated by comparing the participants' scores measured under the FULL and the five LO conditions. As the LO40 and LO80 conditions were found to be significantly different from normal distribution using Kolmogorov-Smirnov test ($p < .05$), Friedman's ANOVA was applied and revealed significant difference, $X^2(5) = 94.13$, $p < .001$. Wilcoxon tests were used for paired comparisons across the conditions with Bonferroni correction. A *p* value of .003 level (2-tailed) was used to evaluate significance of findings. It was found that the score measured in the FULL condition (*Mdn* = 95.8%) was significantly higher than those in the LO200 (*Mdn* = 81.30%), LO160 (*Mdn* = 64.6%), LO120 (*Mdn* = 47.9%), LO80 (*Mdn* = 35.4%), and LO40 (*Mdn* = 35.4%) conditions, $-3.93 \leq z \leq -3.92$, $p < .001$, $r = -.62$. In addition, scores measured in the LO200, LO160, and LO120 conditions were significantly higher than those in all of their shorter LO conditions, $-3.93 \leq z \leq -3.57$, $p < .001$, $-.62 \leq r \leq -.56$.

The effect of deletion of various amounts of vowel edges on tone identification was investigated. As the distribution of scores measured in CO180 was found to be significantly different from normal distribution using Kolmogorov-Smirnov test ($p < .05$), Friedman's ANOVA was used and suggested significant difference in scores measured among the 6 conditions, $X^2(5) = 89.59$, $p < .001$. Wilcoxon tests with Bonferroni correction were applied. A *p* value of .003 level (2-tailed) was used to evaluate significance of findings. It was found that performance in the FULL condition (*Mdn* = 95.8%) was not significantly different with that in the CO240 condition (*Mdn* = 93.8%), $z = -2.08$, $p = .037$. However, it was significantly higher than those in the CO180 (*Mdn* = 88.6%), CO120 (*Mdn* = 84.4%), CO90 (*Mdn* = 70.8%), and CO60 (*Mdn* = 55.3%) conditions, $-3.92 \leq z \leq -3.73$, $p < .001$, $-.62 \leq r \leq -.59$. Except the finding that performances in the CO180 and CO120 conditions were not significantly different ($z = -2.40$, $p = .017$), performances in the CO240, CO180, CO120, and CO90 conditions were significantly higher than those in all of their shorter conditions, $-3.92 \leq z \leq -3.31$, $p \leq .001$, $-.62 \leq r \leq -.52$.

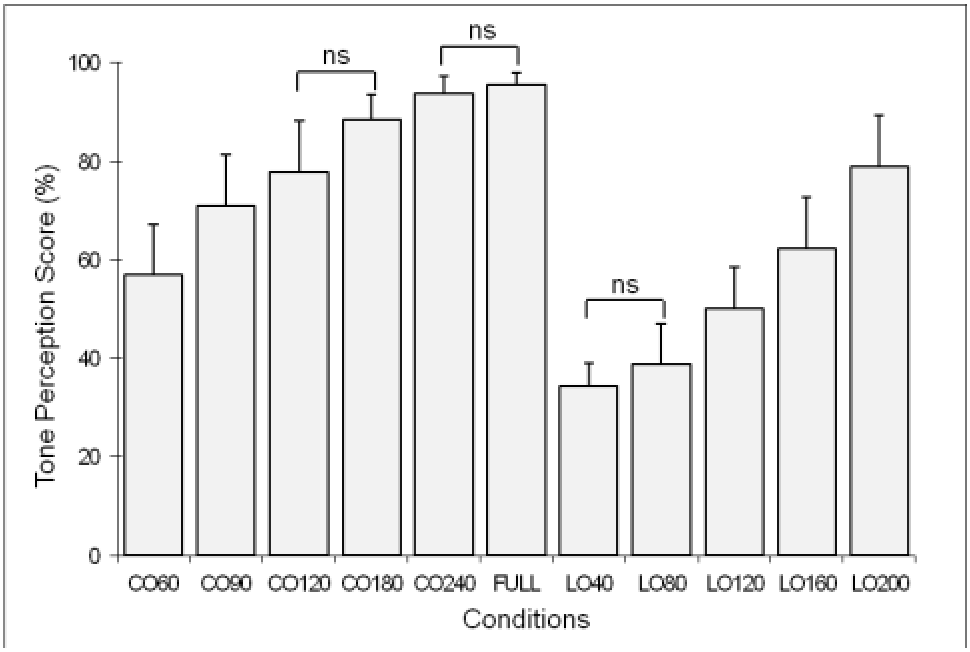

Figure 4. Means of tone percpetion scores of the 11 conditions in Experiment 2. Error bars indicates the standard deviation. “ns” indicates that the difference of the paired tone perception scores is not significant ($p > .003$).

Discussion

It was found that the tone perception at LO40 was not high (*M* = 34.2%) and just slightly higher than the chance level (i.e. 25%). Thus, suggesting that the acoustic information preserved at the initial 40 ms of vowel portion (i.e. initial 10% VP) was not sufficient for reliable tone identification. However, it was noted that the mean sentence intelligibility increased significantly from 34.1% accuracy to 83.8% accuracy when 10% of the initial vowel portion was added to the consonant-only condition in Experiment 1. Therefore, tonal information was not likely to be the contributing factor for the large increase in intelligibility at C+.1V condition. On the other hand, while the tone identification under CO240 condition (equivalent to 60% VP preserved at VC) was not significantly different with that under the FULL condition, sentence intelligibility under the .6V condition was significantly lower than that under 1.0V in Experiment 1. These results suggest that the difference in intelligibility between .6V and 1.0V conditions was independent of tone information, and vowel segment has its own contribution to sentence intelligibility that is not very related to tone.

In addition, it was noted that high sentence intelligibility could be maintained even when the tone identification was poor. For all C+VP conditions except C+0V, all mean values of sentence intelligibility exceeded 80%. However, the mean values of their corresponding tone identification scores in LO40, LO80, LO120, and LO160 were all lower than 65% and were significantly poorer than the scores in FULL condition. This suggested that the contribution of tone to Mandarin sentence intelligibility was small. This result was consistent with the finding by Patel, Xu, and Wang (2010) that Mandarin sentences with flat F0 was as intelligible as normal speech in quiet condition. In their study, Patel *et al.* (2010) found that F0 contour was important only when noise with 0 dB SNR was introduced. Since the contribution of tone was believed to be small in this study, the greater intelligibility advantage of vowels found in Mandarin sentences is likely to be related to other differences in syllable structures between Mandarin and English.

Further study

The present study only investigated the relative perceptual contribution of consonants and vowels to Mandarin speech intelligibility in sentence context. Toro *et al.* (2008) showed that listeners could extract syntactic cues from vowels in connected speech but failed to extract those cues from consonants. Therefore, the relative importance of vowels and consonants to speech intelligibility may be different in word context where listeners cannot benefit from syntactic cues for top-down processing of speech. Some studies found that vowels and consonants contribute almost equally to isolated English word recognition if open-set response and noise-replacement paradigm were used (Fogerty & Humes, 2010; Fogerty *et al.*, 2012). Hence, further study need to examine the segmental contributions to Mandarin speech intelligibility in isolated word context.

Audibility is a main factor for the deterioration of speech understanding in elderly but cognitive factors are also important (Humes, 2007). The elderly listeners may process acoustic cues in a different way from young listeners because of cognitive decline. Therefore, the relative segmental contribution to Mandarin speech intelligibility may be different in elderly listeners. In order to apply our findings to guide the design of amplification devices, elderly NH, and elderly HI listeners should be recruited to investigate the effects of age and hearing loss on the relative contributions of vowels and consonants in Mandarin sentence recognition in further study.

Also, as Patel *et al.* (2010) found that F0 contour was important to speech intelligibility when noise was considered, effect of background noise should be studied to see whether the segmental contributions to Mandarin sentence intelligibility would change at the presence of background noise. Presentation level of stimuli should also be controlled so we can be sure that this factor will not have a significant impact on the relative perceptual contributions of vowels and consonants to Mandarin sentence intelligibility.

Conclusion

To summarize, the relative perceptual contribution of consonants and vowels to Mandarin sentence intelligibility for young normal-hearing listeners was investigated. As in English, vowels make greater perceptual contribution than consonants to sentence intelligibility in Mandarin. However, a 3:1 instead of 2:1 intelligibility advantage of vowels over consonants was found. Perceptual contributions of acoustic information at C-V boundary and vowel center were also studied. Although vowel center and initial vowel portion may contain redundant information for sentence recognition, a little initial vowel portion at the C-V boundary could significantly increase the sentence intelligibility when it was added to the consonant-only sentences. The role of tone in the contribution of vowels to Mandarin sentence intelligibility was also investigated but no obvious influence was found. Further study should be done to investigate the effect of noise on the importance of tone. In order to generalize the findings to design of amplification devices, further study is recommended to include elderly normal-hearing and elderly hearing-impaired listeners. Investigation on the relative segmental contributions to Mandarin speech intelligibility in word context is also recommended.

Acknowledgement

This study was funded by Faculty Research Fund, Faculty of Education, The University of Hong Kong. The author gratefully acknowledges all the subjects participated in this study. In addition, deepest thanks are given to the author’s dissertation supervisors, Dr. Lena Wong and Dr. Felix Chen, for their kind and supportive guidance.

References

American National Standards Institute. (1996). *American national standard specifications for audiometers*. New York: Acoustical Society of America.

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version 5.3) [Computer software]. Retrieved from http://www.praat.org

Bonatti, L. L., Pena, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*(6), 451-459.

Carnegie Mellon University. (2008). *The CMU pronouncing dictionary (Version 0.7a)* [data file]. Retrieved from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

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

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

Fogerty, D., & Kewley-Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *Journal of the Acoustical Society of America, 126,* 847-857.

Fogerty, D., Kewley-Port, D., & Humes, L. E. (2012). The relative importance of consonant and vowel segments to the recognition of words and sentences: Effect of age and hearing loss. *Journal of the Acoustical Society of America*, *132*, 1667-1678.

Fu, Q. J., & Zeng, F. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing, 5*(1), 45-57.

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America, 104,* 505-510.

Fu, Q. J., Zhu, M., & Wang, X. (2011). Development and validation of the Mandarin speech perception test. *Journal of the Acoustical Society of America, 129*(6), EL267-EL273.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., & Dahlgren, N. L. (1990). *DARPA TIMIT acoustic-phonetic continuous speech corpus* [CD-ROM]. Gaithersburgh, MD: National Institute of Standards and Technology.

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

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology, 18*(7), 590-603.

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

Ladefoged, P. (2001). *Vowels and consonants: an introduction to the sounds of languages*. Blackwell: Oxford.

Li, Z., Tan, E. C., McLoughlin, I., & Teo, T. T. (2000). Proposal of standards for the intelligibility tests of Chinese speech. *IEE Proceedings-Vision, Image & Signal Processing, 147* (3), 254-260.

Liang, Z. A. (1963). The auditory perception of Mandarin tones. *Acta Physica Sinica, 26*, 85-91.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of speech code. *Psychology Review, 74*(6), 431-461.

Lin, M. C. (1988). The acoustic characteristics and perceptual cues of tones in Standard Chinese. *Chinese Yuwen, 204*, 182-193.

Patel, A. D., Xu, Y., & Wang, B. (2010). The role of F0 variation in the intelligibility of Mandarin sentences, *Processing of Speech Prosody 2010*. Retrieved from http://www.phon.ucl.ac.uk/home/yi/yispapers/Patel_Xu_Wang_SP2010.pdf

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal, *Cognition*, *75*(1), 265-292.

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*(2), 137-144.

Wan, K. K., & Wong, L. N. (2002). Tympanometric norms for Chinese young adults. *Ear & Hearing*, *23*(5), 416-421.

Wang, R. H. (1989). Chinese phonetics. In Y. B, Chen & R, H, Wang (Eds.), *Speech signal processing* (pp. 37-64). Hefei, China: University of Science and Technology of China Press.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetic*a, *49*, 25-17.

Appendix: A list of Mandarin consonants and vowels

Twenty one initial consonants and 35 final vowels were used in Experiment 1. The pinyin letters and their corresponding IPA symbols are shown in the following:

1. 21 consonants

   b[p], p[p^{h}], m[m], f[f], d[t], t[t^{h}], n[n], l[l], g[k], k[k^{h}], h[x], j[tɕ], q[tɕ^{h}], x[ɕ], zh[tṣ], ch[tʂ^{h}], sh[ʂ], r[ʐ], z[ts], c[ts^{h}], s[s]

2. 35 vowels

   a. 6 simple vowels

   a[a], o[o], e[ɤ], i[i], u[u], ü[y]

   b. 13 complex vowels

   ai[aɪ], ei[eɪ], ao[ɑʊ], ou[oʊ], ia[ia], ie[iɛ], iao[iaʊ], iou[iəʊ], ua[ua], uo[uo], uai[uaɪ], uei[ueɪ], üe[yɛ]

   c. 16 compound nasal vowels

   an[an], en[ən], ang[ɑŋ], eng[əŋ], ong[oŋ], ian[iɛn], in[in], iang[iɑŋ], ing[iŋ], ion[iʊŋ], uan[uan], uen[uən], uang[uɑŋ], ueng[ʊəŋ], üan[yɛn], ün[yn]