Intelligibilities of Filtered Chinese Mandarin Sentences

Chen Yafan
Institute of Information and Engineering
Chinese Academy of Sciences
Beijing 100093, China

Xu Yong and Yang Jun
Key Laboratory of Noise and Vibration Research
Institute of Acoustics, Chinese Academy of Sciences
Beijing 100190, China
jyang@mail.ioa.ac.cn

***Abstract*—Determining the intelligibilities of filtered speech with various spectrum regions is of great significance for communication research and applications. The importance of different spectral components upon the speech intelligibility for western languages has been studied extensively, but not for Chinese. The current study investigated the relative spectral contributions of Chinese Mandarin sentences to speech intelligibility. Sentences were low-pass, high-pass or bandpass filtered and restricted to different frequency regions. Listening experiments demonstrated that the frequency regions below 1100 Hz contained more important cues for the intelligibility of Mandarin Chinese sentences than higher frequencies. The low and middle frequencies (<2100 Hz) provided greater benefit for intelligibilities of Mandarin sentences than monosyllables. While compared to the English sentences, it can be seen that intelligibility redundancy in Chinese was much lower than that in English. And the first two formant frequencies made important contributions to the sentence intelligibility for both Chinese and English.**

***Index Terms*—Speech intelligibility, frequency regions, Chinese Mandarin sentences**

I. Introduction

Speech intelligibility, which is related to the numbers of speech items recognized correctly by listeners, has received remarkable attention for several decades owing to its important role in speech communication and the evaluation for speech systems. How information pertaining to speech intelligibility is distributed among the various frequency bands of speech spectrum is always an arresting question in many applications. The study began with Bell Labs in the early years of 20th century, who quantified the bandwidth of telephone line for speech communication [1], [2]. Whats more, frequency importance functions derived from similar researches were used for the development of objective intelligibility such as the Articulation Index (AI) [3] and the Speech Intelligibility Index (SII) [4].

The intelligibility of western languages subjected to various types of spectral filtering has been investigated extensively. French and Steinberg investigated the intelligibility of low-pass and high-pass filtered speech [5] using meaningless monosyllables of the vowel-consonant-vowel type as their test materials. The results were expressed as the percentage of correct-syllables by subjective listening tests. They reported a low intelligibility of 30% for low-passed speech at 750 Hz, and a relatively high intelligibility of 90% for those at 3300 Hz. But in the high-pass filtering condition, speech signals cut off at 750 Hz were 90% intelligible and that cut-off at 2900 Hz decreased to roughly 20% intelligible. Jeff Rodman discussed the relationship between the bandwidth and intelligibility for the telephone network [6]. He reported a low intelligibility score below 20% for single syllables low-pass filtered at 300Hz, a relatively high intelligibility of 75% for the low-pass filtered at 3300 Hz, and the speech at 7000 Hz cut-off frequency yielded a ceiling accuracy of over 95%, which suggested that extending the bandwidth could increase the speech intelligibility. Warren showed that everyday sentences were all above 95% intelligible when sentences had been passed through a 1/3-octave bandpass filter with center frequencies of 1100, 1500 and 2100 Hz respectively [7]. A lower intelligibility of 25% was obtained for sentences filtered at the center frequencies of 370 and 6000 Hz.

Most of the preceding studies investigated the effect of different frequency bands on the speech intelligibility of English databases (e.g., the IEEE corpus) including syllables, words and sentences. However, different languages are characterized by various specific acoustic and phonetic features. For example, F0 information in English conveys the emotion and innovation which contribute little to the speech intelligibility in quiet, but can do help to the segregation from many talkers [8]. While F0 information in Japanese can be used to distinguish different accents and meanings and that are important to the intelligibility. F0 contour in Chinese carries the tone information and contributes a great deal to the overall speech recognition [9].

So it is unclear that whether the primary findings about western languages are applicable to other languages, such as Mandarin Chinese. The contact of Chinese initials, finals and tones articulation with different frequency bands was investigated by Siyu Zhang and Zihou Meng under seven bandpass filtering conditions [10], using KXY monosyllables lists [11] as their testing material. Hui Song and Zihou Meng further designed experiments to examine the word articulation under low-pass and high-pass filtering conditions [12]. They evaluated the subjective intelligibility of filtered monosyllable words at twenty cut-off frequencies ranging from 160 to 12500 Hz. The above mentioned researches focused on the intelligibility of Chinese syllables but not sentences. However, intelligibilities will perform differently when speech materials change, and measures based on sentences are closer to the

TABLE I
CUT-OFF AND CENTER FREQUENCIES FOR FILTERS

Condition	Cut-off /Center Frequency (Hz)
Low-pass	250 370 530 750 1100
High-pass	1500 2100 3000 4200 6000
Band-pass(1/3-octave)	370 530 750 1100 1500 2100 3000 4200 6000
Band-pass(1/2-octave)	370 530 750 1100 1500 2100 3000 4200 6000

common communication scenarios in our daily life.

A Mandarin sentence intelligibility subjective test was implemented in the present research. The contributions of this research are as follows: First, to our knowledge this is the first study to investigate the filtered Chinese sentence intelligibility; second, the present study compared the intelligibility between English sentences and Chinese sentences bandpass filtered; third, a direct comparison for speech intelligibility between Chinese monosyllables and Mandarin sentences in low-pass and high-pass filtering conditions was also provided in this work; more importantly, the contribution of different frequency components to Mandarin sentence intelligibility was analyzed. The work could provide valuable information and understandings for the perception of Chinese and English, especially for the further study of the Chinese speech intelligibility.

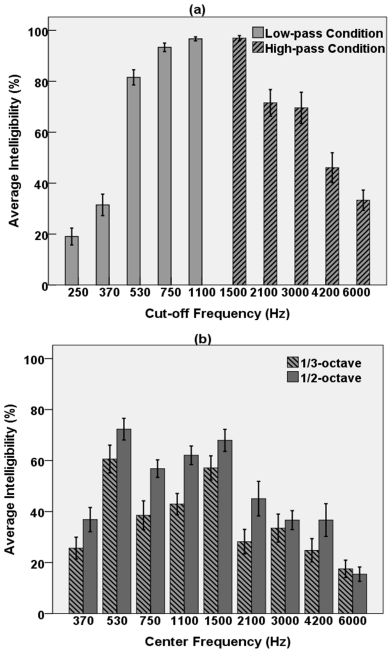

Fig. 1. Average intelligibility for low-pass, high-pass, and bandpass filtered Chinese Mandarin sentences. Results for (a) low-pass and high-pass conditions and (b) bandpass conditions are displayed. Error bars indicate standard errors of the mean.

II. EXPERIMENT

A. Methods

1) Participants: Twelve normal-hearing participants (6 females and 6 males) were all native speakers of mandarin, who were recruited from the University of Chinese Academy and Sciences. Their ages were from 22 to 29 years old. None of the listeners were familiar with corpus used in this experiment. They were paid proper money for their participation.

2) Stimuli: The corpus employed in this study was Mandarin Chinese sentences database from the Chinese PLA General Hospital [13]. The test material consists of 32 phonetically balanced lists, with each list containing 9 five-word to ten-word sentences, totally 50 keywords for scoring the intelligibility per list. The sentences are all simple everyday meaningful speeches, e.g., (translated), The butterfly is so beautiful. Additional ten sentences served as the practice stimuli were also taken from the original corpus. All sentences were produced by a male speaker with no evident regional accent. They had been recorded in digital form, and they were sampled at 44.1 kHz with 16-bits quantization [14]. The stimuli were processed by passing through five low-pass, five high-pass and eighteen bandpass (1/3-octave passbands and 1/2-octave passbands) filters respectively. The low-pass and high-pass filters were designed by the MATLAB function ellip. And the bandpass filters were designed by the MATLAB function buttord. The parameters of frequencies were chosen similar to the experiments designed by Warren [7]. All the cut-off and center frequency values are shown in Table I. And all filtered sentences were normalized to a same overall RMS value.

3) Procedure: The testing was completed in a sound-proof room. The participants listened to stimuli via a RME Fireface UCX soundcard and AKG 550 pro headphone at 65 dB sound pressure level (SPL). Listeners were required to write down each sentence they heard and were encouraged to guess if unsure. All the listeners first completed a practice session to familiarize with the task. During the tests, each subject was presented with all the 28 filtered conditions (totally 252 sentences) in a fully random order and was provided with the opportunity to listen each sentence several times before responding. The total processing time of the experiment session was 120 minutes. The intelligibility score of each filtering condition for every listener was calculated by the percentage of key words written correctly for each list. And we finally got the average intelligibility after all the 12 listeners completed the tests.

B. Results

Fig. 1(a) describes the results for the low-pass and high-pass filtering conditions, which draws the intelligibility changing with cut-off frequencies. The higher intelligibility was obtained for low-passed sentences with higher cut-off frequencies. The mean score of the speech signals cut off at 1100 Hz was 96.7%, which means fully intelligible for listeners. When the low-pass cut-off frequency dropped below 530 Hz, the intelligibility for sentences was severely decreased. The average intelligibility of 31.5% was obtained at the 370 Hz cut-off frequency. The intelligibility was as low as 16.4% when the low-pass cut-off frequency was 250 Hz. As for the high-

TABLE II
INTELLIGIBILITY VALUES IN THE EXPERIMENT

Center /Cut -off Frequency (Hz)	Low-pass	High-pass	Bandpass (1/3-octave)	Bandpass (1/2-octave)
250	16.4	-	-	-
370	31.5	-	25.6	36.9
530	81.5	-	60.6	72.3
750	93.4	-	38.5	56.8
1100	96.7	-	43.5	62.1
1500	-	97.0	57.1	67.9
2100	-	71.5	21.9	45.0
3000	-	69.6	33.5	36.7
4200	-	46.1	24.8	36.6
6000	-	33.3	17.5	15.4

pass filtering, the result in the Fig. 1(a) reflects the decreasing intelligibility changing with the increasing cut-off frequencies. It was high intelligible (97%) for those high-passed sentences with the 1500 Hz cut-off frequency. When the high-pass cut-off frequency was as high as 6000 Hz the average scores decreased to 33.3%, and it was really rough to understand what the sentences meant.

Fig. 1(b) presents the average intelligibility scores and standard errors for the bandpassed sentences with various center frequencies. As shown in the figure, the overall trends of the intelligibility scores were the same for two kinds of passband conditions. The results demonstrated that speeches centered at 6000 Hz with 1/3-octave and 1/2-octave passbands both had a very low intelligibility level (<20%). More importantly, the sentence intelligibility scores were highest with the 530 and 1500 Hz center frequency, which yielded average sores of around 60% 70%. The detailed values of results for low-pass, high-pass and bandpass filtering conditions can be found in Table II.

Two one-way analyses of variance (ANOVAs) were run separately for the low-pass and high-pass conditions. Main effects of the low-pass cut-off frequency [$F(4, 55)=161.081$, $p<0.05$] and high-pass cut-off frequency [$F(4,55)=23.433$, $p<0.05$] were both significant. For the bandpass conditions, results also demonstrated significant main effects of the center frequency for 1/2-octave bands [$F(8,99)=10.704$, $p<0.05$] and 1/3-octave bands [$F(8,99)=7.572$, $p<0.05$]. Subsequent Turkey HSD testing revealed that sentence intelligibility for both 530 Hz and 1500 Hz center frequency conditions performed best, and a nonsignificant effect ($p>0.05$) was obtained between the mean intelligibility of 530 and 1500 Hz center frequency bandpass conditions for 1/3-octave and 1/2-octave bands.

III. General Discussion

Many studies have confirmed that different frequency bands do not contribute to the speech intelligibility equivalently in English and other western languages. The present experiment investigated the effect of various frequency components on the intelligibility of Chinese everyday sentences.

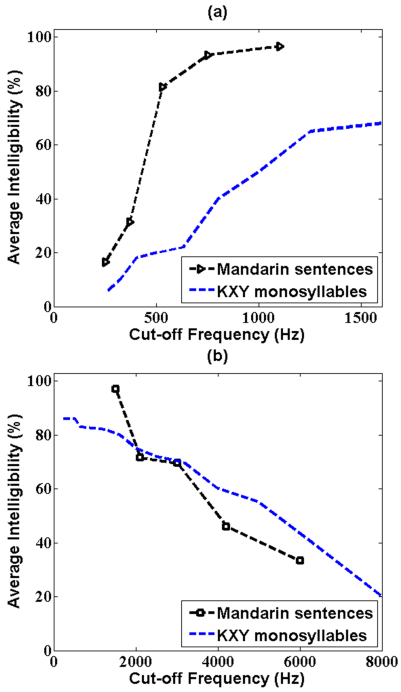

Fig. 2. Average intelligibility scores for Mandarin sentences and KXY monosyllables The left figure (a) shows results about the low-pass conditions, and the right figure (b) shows results about the high-pass conditions.

For low-pass and high-pass filtering conditions, we made the comparison of the results about Chinese Mandarin sentences in this study and the results about Chinese monosyllables measured by Hui song and Zihou Meng [12]. The different average percent intelligibility scores for both monosyllable words and everyday sentences are shown in Fig. 2, under low-pass and high-pass filtering conditions.

As in Fig. 2(a), which displays results about the low-pass filtering condition, the intelligibility of monosyllables was roughly lower than 10% when the cut-off frequency was 250 Hz, and the intelligibility of sentences was 16.4% in the same filtering condition. Monosyllables could not be understood for the low intelligibility of around 20% at 530 Hz cut-off frequency, while sentence intelligibility remained as high as 81.5% under the same filtering condition. When the cut-off frequency rose to 1100 Hz, an intelligibility of around 60% was obtained for monosyllable words, and the sentences were almost fully intelligible with the scores of 96.7%. Obviously, the higher intelligibility was achieved for either Chinese syllables or sentences low-passed at the higher cut-off frequency. Results also demonstrated that Chinese everyday sentences in the present study were much more intelligible than KXY monosyllable words low-passed with the same cut-off frequency. The outcome was consistent with the conventional views that the sentences can be recognized more easily than insolated words [15]. So the primary factor contributing to the higher intelligibility of Chinese sentences is the linguistic context [16]. The everyday sentences are composed of meaningful words with semantic contexts, while the

TABLE III
FORMANT FREQUENCY

Formant	Mean (Hz)	StDev (Hz)	Range (Hz)
F0	149	14	6 ~ 495
F1	534	61	367 ~ 720
F2	1738	126	1443 ~ 2044

independent monosyllables eliminate the syntactic or prosodic cues and can be hardly guessed. Moreover, results for low-passed sentences demonstrated that the frequency regions below 1100 Hz contributed more important speech information for the sentence intelligibility than monosyllabic words.

From Fig. 2(b) it can be seen that the intelligibility decreased with the increasing high-pass cut-off frequency for both monosyllables and sentences. The sentences high-pass filtered at 1500 Hz yielded a high intelligibility of 97%, which were more intelligible than high-passed monosyllables filtered at the same cut-off frequency (about 80%). As the cut-off frequency increased above 2100 Hz, intelligibilities for Mandarin sentences performed a little poorer compared with those for Chinese monosyllables. Mandarin sentences decreased to an intelligibility of around 33.3%, and the monosyllables were about 50% intelligible when they were high-passed at the 6000 Hz cut-off frequency. The phenomena indicated that the low and middle frequency regions (<2100 Hz) carried major semantic information in Mandarin sentences which was also supported by the previous finding about the spectral contributions to Mandarin sentence intelligibility [17]. And compared to the results about monosyllables, we could conclude that advantages of sematic cues on improving the sentence intelligibility were eliminated by the absence of those spectral bands below 2100 Hz.

For the bandpass filtering condition, it could be seen that average intelligibilities of 60%-70% were yielded for the bandpass filtered sentences centered at 530 and 1500 Hz no matter with 1/2-octave or 1/3-octave nominal narrower passbands. The result reflected that Mandarin sentences with narrow passbands could still maintain relatively high intelligibilities though their frequency components were severely absent. More importantly, two peaks in the figure suggest that those bands centered at 530 and 1500 Hz made greater contributions to the sentence intelligibility than other bands. In order to analysis why higher average intelligibility scores were obtained for bandpassed Mandarin sentences centered at around 530 and 1500 Hz, we calculated the fundamental frequency and formant frequency values for all the experimental sentences. Mean values, ranges and standard deviations for the fundamental frequency and the first two formants are provided in Table III. The mean fundamental values for each sentence were calculated using the normalized ACF method, and the mean formant frequency values for each sentence were calculated using the Formant Tracker toolbox [18], which was written by Satrajit Ghosh in a modified LPC method. From Table III, it suggested that the highest intelligibility scores

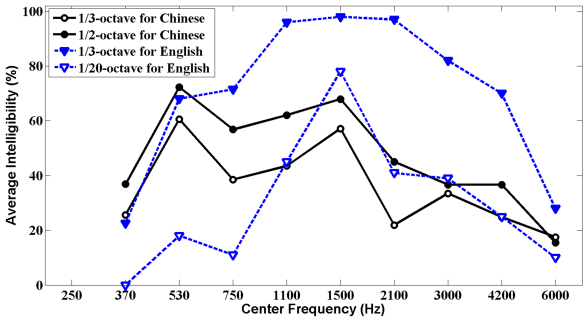

Fig. 3. Average intelligibility for bandpass filtered Chinese and English everyday sentences. The blue dashed lines show the English intelligibility, and the solid lines show the Chinese intelligibility.

were obtained for bands centered around 530 and 1500 Hz, which were located just near the first two formants. Meanwhile according to low-pass and high-pass conditions, it can be seen that low-passed sentences were less intelligible when the cut-off frequencies got lower than 530 Hz, while the intelligibility of high-passed sentences decreased significantly when the cut-off frequencies improved higher than 1500 Hz. These findings combined demonstrated that the first two formants played important roles in the speech intelligibility for Chinese Mandarin sentences.

Fig. 3 presents the Mandarin sentence intelligibility in the present study and English sentence intelligibility from Warrens study [7] in bandpass filtering conditions. The intelligibility performance got better when the nominal passbands increased, for both English (1/20-octave to 1/3-octave) and Chinese (1/3-octave to 1/2-octave). Contrasts demonstrated the significant difference between the English and Chinese everyday sentences of 1/3-octave bands. When the center-frequency ranged from 750 to 6000 Hz, Mandarin sentence intelligibility displayed much lower than English sentence intelligibility. Especially, a high intelligibility of more than 90% was obtained by English sentences of 1/3-octave centered at 1100, 1500 and 2100 Hz. And Chinese Mandarin sentences centered at 1500 Hz yielded a highest score of 57.1% , which was even lower than English sentences of extremely narrow 1/20-octave bands (>70%). It could be concluded that the intelligibility redundancy on the frequency region about Chinese sentences was much lower than the redundancy about English sentences. This result was mainly attributed to the distinctions on syllable structures, the lexical resources and grammars between English and Chinese. Besides, when English sentences were restricted to extremely limited frequency ranges of 1/20-octave bands, higher intelligibility scores were yielded for bands centered at 530 and 1500 Hz, with two peaks located around the male speakers first two formant frequencies. The intelligibility performance for English sentences of 1/20-octave bands showed similar to that observed for Mandarin sentences of 1/3-octave bands, and the consistent phenomenon revealed that the first two formants [19] contributed a lot to the sentence intelligibility for no matter English or Chinese.

IV. Conclusion

This study investigated the effect of various frequency regions on the intelligibility of Mandarin Chinese sentences. The subjective experiment was performed under 28 listening conditions (288 "everyday" sentences), containing low-pass, high-pass and bandpass filtering conditions by varying the cut-off and center frequencies. And the intelligibility was obtained by calculating the average scores for 12 listeners. The conclusions of the present study can be drawn as the followings.

For Chinese Mandarin sentences, the intelligibility increased with the higher low-pass cut-off frequency, and deceased with the increasing high-pass cut-off frequency. The cut-off frequency for low-pass and high-pass filtering conditions yielded significant effects on the intelligibility of Mandarin Chinese sentences, and there also existed a significant effect of the bandpass center frequency on the sentence intelligibility. Furthermore, sentences performance best centered at 530 and 1500 Hz in bandpass conditions.

Comparisons of the Mandarin sentence intelligibility and KXY monosyllable intelligibility revealed that the low frequencies less than 1100 Hz contained more cues for the intelligibility of Mandarin Chinese sentences than higher frequency regions. Whats more, the low and middle frequency regions ($<$2100 Hz) made greater contribution to the intelligibility of Mandarin sentences than the intelligibility of monosyllables, which provided advantages of lexical and semantic information on understanding Mandarin sentences.

Relatively high intelligibility (60%-70%) could be obtained in several conditions (e.g., center frequency 530Hz) for Chinese sentences limited to narrow spectral regions which could reflect the intelligibility redundancy for Chinese. Results about Chinese and English sentences demonstrated that the intelligibility redundancy on the frequency domain for Chinese was much lower than the redundancy for English. In addition, we could conclude that the speakers first two formants were important to the sentence intelligibility for no matter English or Chinese.

Different materials with various phonemic features on frequencies affect the speech intelligibility. From the experiment results and direct comparison, the effect of various bands in the spectrum could be obtained preliminarily. Furthermore, how listeners combine information across frequency regions is important for understanding Chinese perception, which was not taken into consideration in the experiment. It is possible to be further studied by investigating the intelligibility of band-stop filtered Chinese speech for future.

Acknowledgment

This work was supported by National Natural Science Fund of China (Grant No.11474307).

References

[1] H. Fletcher, *Speech and Hearing*. 1929.

[2] H. Fletcher and R. H. Galt, "The perception of speech and its relation to telephony," *Journal of the Acoustical Society of America*, vol. 108, no. 2816, pp. 89-151, 1950.

[3] ANSI S3.5. American National Standard Methods for the Calculation of the Articulation Index. New York: Institute American National Standards. 1969.

[4] ANSI S3.5. American National Standard Methods for Calculation of the Speech Intelligibility Index. New York: Institute American National Standards. 1997.

[5] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 90-119, 1947.

[6] J. Rodman, "The effect of bandwirth on speech intelligibility," *White Paper the Polycom Corporation*, 2003.

[7] R. M. Warren, K. R. Riener, J. A. Bashford, and B. S. Brubaker, "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," *Perception & Psychophysics*, vol. 57, no. 2, pp. 175-182, 1995.

[8] T. Banziger and K. R. Scherer, "The role of intonation in emotional expressions," *Speech Communication*, vol. 46, pp. 252-267, 2005.

[9] J. Li, R. Xia, D. Ying, Y. Yan, and M. Akagi, "Investigation of objective measures for intelligibility prediction of noise-reduced speech for Chinese, Japanese, and English," *Journal of the Acoustical Society of America*, vol. 136, no. 6, pp. 3301-3302, 2014.

[10] S. Zhang, H. Song, and Z. Meng, "Relationship between Chinese Mandarin intelligibility and speech transmission index STIPA under simulated tranmission conditions," in *Proceeding of the IEEE China Summit and International Conference on Signal and Information Processing*, 2015.

[11] D. Ma and H. Shen, *Acoustic Manual,Chinese Science*. 2014.

[12] H. Song, S. Zhang, and Z. Meng, "The effect of frequency filteing on the Chinese articulation," in *Proceedings of China Communication Conference on Audio Engineering*, 2015.

[13] X. Xi, A. Chen, J. Li, and et al., "Standardized Mandarin Sentence Perception in Babble Noise Test Materials for Children," *Journal of Audiology & Speech Pathology*, vol. 17, no. 4, pp. 318-322, 2009.

[14] B. Jiang and J. Yang, "Preferred frame length for the short-time magnitude spectrum on speech intelligibility and speech quality," in *Proceedings of the 8th International Conference on Information, Communications and Signal Processing*, 2011.

[15] J. Local, "Variable domains and variable relevance: interpreting phonetic exponents," *Journal of Phonetics*, vol. 31, pp. 321-339, 2003.

[16] J. Coleman, "Discovering the acoustic correlates of phonological contrasts," *Journal of Phonetics*, vol. 31, pp. 351-372, 2003.

[17] F. Chen, L. N. Wong, and E. Wong, "Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility," *Journal of the Acoustical Society of America*, vol. 134, no. 2, pp, EL178-184, 2013.

[18] http://www.cns.bu.edu/ speech/frack.php.

[19] D. Fogerty and L. E. Humes, "Perceptual weighting of individual and concurrent cues for sentence intelligibility: frequency, envelope, and fine structure," *Journal of the Acoustical Society of America*, vol. 129, no.02, pp. 977-988, 2011.