INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Intelligibilities of Mandarin Chinese Sentences with Spectral “Holes”

Yafan Chen[1,3], *Yong Xu*[2], *Jun Yang*[2,1,3]

[1]Institute of Information and Engineering, CAS, Beijing, China
[2]Key Laboratory of Noise and Vibration Research, Institute of Acoustic, CAS, Beijing, China
[3]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

chenyafan@iie.ac.cn, jyang@mail.ioa.ac.cn

Abstract

The speech intelligibility of Mandarin Chinese sentences of various spectral regions, regarding band-stop conditions (one or two “holes” in the spectrum), was investigated through subjective listening tests. Results demonstrated significant effects on Mandarin Chinese sentence intelligibilities when a single or a pair of spectral holes was introduced. Meanwhile, it revealed the importance of the first and second formant (F1, F2) frequencies for the comprehension of Mandarin sentences. More importantly, the first formant frequencies played a more primary role rather than those of the second formants. Sentence intelligibilities declined evidently with the lacking of F1 frequencies, but the effect became small when the spectrum holes covered more than 50% of F1 frequencies, and F2 frequencies came into a major play in the intelligibility of Mandarin sentence.

Index Terms: speech intelligibility, spectral holes, Mandarin Chinese sentence

1. Introduction

Speech intelligibility 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]. And these 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 investigated the intelligibility of low-pass and high-pass filtered meaningless monosyllables of the vowel-consonant-vowel type. 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 [5]. Warren [6] 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 Hz, 1500 Hz and 2100 Hz, respectively. Lippman [7] investigated the consonant recognition with the “one-hole” spectrum that a band was removed in the spectrum. And he reported that consonants could be maintained high recognized when the middle frequencies were removed, ranging from 800 Hz to 4000 Hz. Kasturi [8] assessed intelligibilities of vowels and consonants with multiple spectral holes. He reported that consonants were less affected than vowels for one or two holes in the spectrum, and that middle and high frequencies were important for the consonant recognition.

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 intonation which contribute little to the speech intelligibility in quiet [9]. While F0 contour in Chinese carries the tone information and contributes a great deal to the overall speech recognition [10].

Owing to language-specific characteristics, it is unclear that whether the primary findings about western languages are applicable to other languages, especially to Chinese. Song designed experiments to examine the word articulation under low-pass and high-pass filtering conditions [11]. Whats more, the asoication of Chinese initials, finals and tones articulation with different frequency bands was investigated by Zhang [12] under seven bandpass conditions using KXY monosyllables lists [13]. As we know, various speech materials always turn in different performances on the intelligibility [14], and measures based on sentences are closer to common communication scenarios in our daily life. Chen investigated the intelligibility of low-passed, high-passed and band-passed Mandarin Chinese sentences and analyzed the contribution of various frequency regions [15]. However, the above-mentioned researches about Chinese focused on low-pass, high-pass and bandpass filtering conditions, and there are not many studies on the intelligibility of the band-stop filtered speech (“holes” in the spectrum), especially Chinese Mandarin sentences.

In the present research, a subjective listening test on Mandarin sentence intelligibility was implemented. Contributions of this research are as follows: To our knowledge, this is the first study to investigate the Chinese sentence intelligibility with a single or a pair of “holes” in the spectrum. Meanwhile, the contribution of various frequency components to sentence intelligibility is analyzed.

2. Experiment

2.1. Participants

Fifteen normal-hearing participants (23-29 ages) were all native speakers of Chinese, who were recruited from the University of Chinese Academy and Sciences. None of the listeners were familiar with corpus used in this experiment.

2.2. Stimuli

The corpus was Chinese Mandarin sentence database from the Chinese PLA General Hospital [15, 16, 17]. It consists of 32 phonetically balanced lists, with each list containing 9 five-word to ten-word everyday sentences, totally 50 keywords for scoring the intelligibility per list. All sentences were uttered by a male

Copyright © 2017 ISCA

http://dx.doi.org/10.21437/Interspeech.2017-281

Mandarin-Chinese speaker, which were recorded at a sampling rate 44.1 kHz with 16-bits quantization and down sampled to 16 kHz.

Aiming to explore the intelligibility of Mandarin sentences with holes in the spectrum, according to the research by Kasturi [8], we processed the filtered speech into six octave-band channels using six-ordered Butterworth filters after a pre-emphasis filtering with the cut-off frequency of 2000 Hz. Detailed center frequencies (ranging from 200 Hz to 6300 Hz) and frequency boundaries for the six bands are given in Table 1.In each band, a modulated sinusoidal signal was produced with the corresponding envelope information, which was extracted using a rectification and a low-pass filter. Meanwhile, the discrete Fourier transform (DFT) of the same band's speech segments was used to estimate phases of the modulated signal. Finally the generated sinusoids of different channels were summed as the synthesized speech.

Table 1: *Center frequencies and boundaries of the 6 bands.*

Band ID	Frequency Boundaries(Hz)	Center Frequency(Hz)
1	[141 282]	200
2	[282 562]	400
3	[562 1122]	800
4	[1122 2239]	1600
5	[2239 4467]	3150
6	[4467 8000]	6300

We considered all possible combinations for one or two bands missing in the spectrum. So 6 one-hole conditions imply that only the band N (N=1-6) was lost. And 15 two-hole conditions mean that the band N and another band M were simultaneously omitted (N, M=1-6, NM). To introduce the spectral holes, we set the amplitude of corresponding bands (e.g., band 1-6) to zero, and synthesized speech with the remaining other bands. In this experiment, 21 conditions were totally produced followed by normalizing them to a same overall root-mean-squared (RMS) value.

2.3. Procedure

The testing was completed in a sound-proof room. Participants listened to stimuli via a laptop, RME Fireface UCX soundcard and AKG 550 pro headphone at a comfortable sound level (70 dB SPL). Listeners were required to write down each sentence they heard on the GUI. All listeners first completed a practice session to familiarize with the task. During the tests, each subject was presented with 21 conditions (totally 189 sentences) in a fully random order. The listener was provided with a maximum of three times to listen each sentence before responding and was encouraged to guess the meaning. During the 120-minute experiment, a break of 5 minutes was given every 30-minute test to avoid the listening fatigue for participants. The intelligibility score of each condition for every listener was obtained by calculating the correct percentage of key words for each list. And the average intelligibility after all listeners completed the test was computed as the final intelligibility scores.

3. Results

The average intelligibility scores and standard deviations for the sentences with a single hole in the spectrum are shown in Figure 1. Mean scores dropped below 80% for sentences when the spectrum of channel 1, 2, 3 or 4 was lost. A one-way analysis of variance (ANOVA) suggested a significant effect on the intelligibility when a hole existed in the spectrum [F(5,84)=2.745,p<0.05]. A post hoc Tukey HSD test showed that losing the channel of 1, 2, 3 or 4 significantly affected the intelligibility, compared with losing the channel 5 or 6. We could conclude that the damage to sentence intelligibility caused by losing the first four bands (center frequencies are 200 Hz, 400 Hz, 800 Hz and 1600 Hz) is much more than other higher frequency bands.

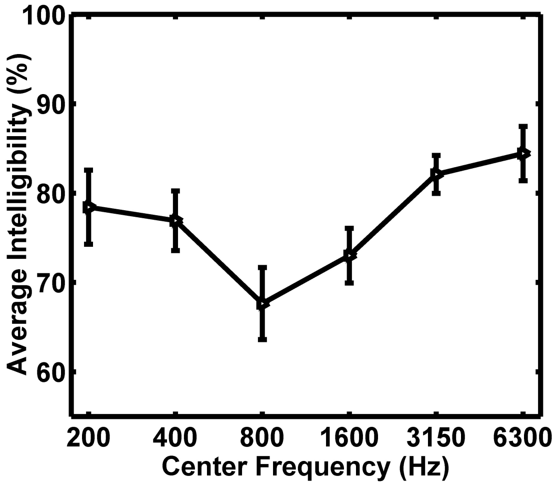

Figure 1: *Average intelligibility for Mandarin sentences with a single hole in the spectrum. Error bars indicate the mean standard deviations which are no more than 4.1.*

Figure 2 displays the results about all two holes conditions. As shown in the figure, relatively low intelligibility scores (under 60%) were obtained in the conditions (12), (13), (23), (34), (35) and (36). A one-way ANOVA showed significant differences in Mandarin-Chinese sentence intelligibilities among the 15 two-hole conditions [F(14,180)=7.884,p<0.05]. A post hoc Tukey HSD test revealed no distinct differences among the six conditions above, but significantly lower intelligibility scores for them than for the remaining nine two-hole conditions: (14), (15), (16), (24), (25), (26), (45), (46) and (56). Overall, it can be found that conditions with channel 3 lost obtained low intelligibility scores, while conditions with channel 4, 5 or 6 lost obtained relatively high scores.

4. Discussion

Results reported that no matter whether one or two holes in the spectrum were introduced, sentence intelligibilities were significantly affected. Statistical analysis demonstrated a significant effect on sentence intelligibility when the band 1, 2, 3 or 4 was removed for one-hole conditions. And it shows agreement with findings [15] that the low and middle frequency regions (below about 2000 Hz) contribute more important information for Mandarin sentences.

In order to analyse the effect of the location of the second spectral hole when it was introduced, we displayed all possible combinations for the band missing in Figure 3. And the subpicture (N) shows results for all conditions in which the band N was lost, and (0N) means conditions that only the band N

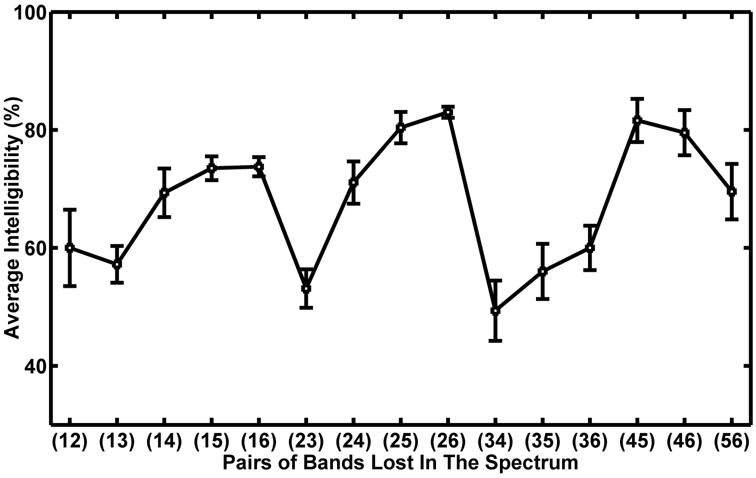

Figure 2: *Average intelligibility for Mandarin sentences with a single hole in the spectrum. Error bars indicate standard deviation of the mean which are no more than 4.1.*

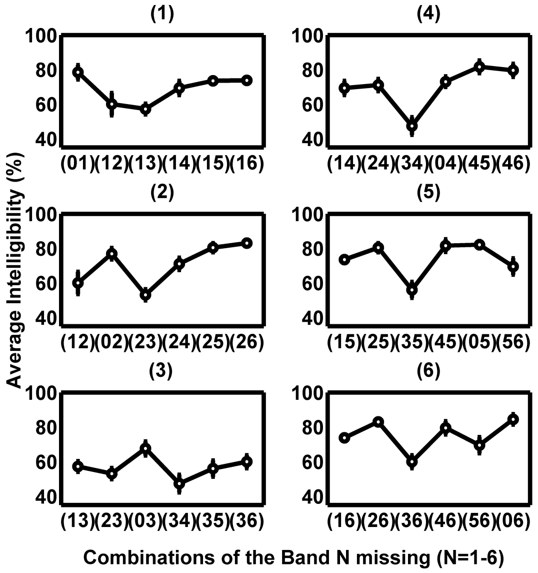

Figure 3: *Average intelligibilities and standard deviations for all combinations of one or two holes in the spectrum.*

missing. Six one-way ANOVAs for each subpicture condition demonstrated that the existence of the second hole showed significant effects on the sentence intelligibility ($p<0.05$). Post hoc Tukey HSD tests for the six subplots showed that lowest intelligibility scores were obtained when the second hole introduced was in band 3, which reinforces the results for all two-hole conditions above. And it also suggested several conditions in which two adjacent bands were removed (e.g., (12), (23), (34) and (56)) had a significant drop in performance except band 4 and 5 (e.g., (45)), which need to be further studied.

For two-hole conditions, results showed that several pair combinations significantly dropped the intelligibility: (12), (13), (23), (34), (35) and (36). To further investigate main factors, we tracked each sentences formant frequencies using the Formant Tracker toolbox, which was written by Satrajit Ghosh in a modified LPC method [18]. Means and average standard deviations of the first and the second formant (F1, F2) frequencies, percentages that F1 and F2 fall into each band respectively are shown in Table 2.

Table 2: *Means and standard deviations of F1 and F2, percentages they falling into the six bands and intelligibility scores for onehole conditions.*

		F1	F2	Intelligibility
	Mean	534Hz	1738Hz	
	StDev	61	126	
Band ID	1	16%	0	78.4%
	2	32%	0	76.9%
	3	48%	13%	67.6%
	4	4%	80%	73%
	5	4%	80%	73%
	6	4%	80%	73%

It revealed that bands 1, 2, 3 and 4 covered the most of F1-F2 frequency ranges, as shown in Table 2. Combined with results about one-hole conditions, the importance of F1 and F2 to the sentence intelligibility was declared. And results in the previous study about the band-passed Mandarin sentences [15] indicate the consistency. The possible reason is that vowels are more sensitive to the loss in the spectrum and carries essential cues for the sentence intelligibility [19].

Percentages of two-hole combinations covering the first two formant frequencies and intelligibility scores were displayed in Table 3. Obviously, these conditions (12), (13), (23), (34), (35) and (36), which decreased intelligibilities the most, covered at least 48% of F1 frequencies. At the same time, the lowest scores were obtained when band pairs (23) or (34) were removed. And it can be seen that the combination (23) covered 80% of F1 information, and the combination (34) covered not only 52% of F1 but also 93% of F2 frequencies.

Furthermore, when the missing spectrum contained most of the F2 frequency ranges, but not too much F1 information, such as conditions (14), (24), (45) and (46), whose lost spectral bands covered no more than 36% of F1, the sentence intelligibility did not perform a significantly drop. Overall, it is safe to conclude that the first formant frequencies play a more dominant role on the understanding of Mandarin Chinese sentence than the second formant frequencies. In addition, results between conditions (12) (losing F1 48%), (13) (losing F1 64%, F2 13%), and (23) (losing F1 80%, F2 13%) showed no big differences. While conditions (34) (losing F1 52%, F2 93%) were less intelligible than (23). And it indicated that when those spectral "holes" covered 48% of F1 frequency ranges, sentence intelligibilities are comparable to that when holes covered 64% or even 80% of F1. So it could be noted that an upper bound for the lost F1 to maintain the primary effect on Mandarin sentence intelligibility. F2 frequency ranges come into play when the percentage of the lost F1 reached to the bound (around 50%), and the more F2 lost, the lower sentence intelligibility obtained.

Table 3: *Intelligibility scores and percentages of F1 and F2 that these two-hole combinations covering.*

Combinations	F1	F2	Intelligibility
(45)	4%	87%	81.3%
(46)	4%	80%	79.6%
(14)	20%	80%	69.3%
(24)	36%	80%	71.1%
(12)	48%	0	60%
(36)	48%	13%	60%
(35)	48%	20%	56%
(13)	64%	13%	57.2%
(23)	80%	13%	53.1%
(34)	52%	93%	47.4%

5. Conclusions

Mandarin Chinese sentence intelligibilities with different spectral regions were investigated in this work. We focused on band-elimination conditions (i.e., speech with holes in spectrum), especially on single-hole and two-hole conditions. Experiments showed that sentence intelligibilities were significantly affected when one or two bands were removed in the spectrum. Conditions in which band 3 was removed from the spectrum resulted in a big drop in the Mandarin sentence intelligibility. Meanwhile,the important contributions to the sentences intelligibility for the first two formant frequencies can be obtained. And we can conclude that F1 frequency ranges play a more dominant role on Mandarin sentence intelligibilities than F2. Furthermore, analysis showed that when F1 coverage within a lost band exceeded an upper bound of about 50%, increasing F1 coverage did not further depress intelligibility and F2 became to play a main part in the Mandarin sentence comprehension.

This study could provide valuable information and understanding for the perception of sentences on the further study of Chinese speech intelligibility. And the future direction of this work would be to explore the combined effect of multiple bands regarding joint and disjoint bands and estimate the frequency importance function based on these experiments using Mandarin sentences.

6. Acknowledgements

We would like to thank all the participants in our experiments and the support by National Natural Science Fund of China (Grant No.11474307, 11674348, 11504405 and 11404367).

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

[7] R. P. Lippmann, "Accurate consonant perception without mid-frequency speech energy," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 1, pp. 66-69, 1996.

[8] K. Kasturi, P. C. Loizou, M. Dorman, and T. Spahr, "The intelligibility of speech with 'holes' in the spectrum," *Journal of the Acoustical Society of America*, vol. 112, no. 1, pp. 1102-1111, 2002.

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

[10] 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.

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

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

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

[14] J. Chen, Q. Huang, and X. Wu, "Frequency importance function of the speech intelligibility index for Mandarin Chinese," *Speech Communication*, 2016.

[15] Y. Chen, Y. Xu, and J. Yang, "Intelligibilities of filtered Chinese Mandarin sentences," *Processings of the IEEE International Conference on Signal and Processing*, 2016.

[16] X. Xi, A. Chen, J. Li, F. Ji, M. Hong, S. Yang, and D. Han, "Standardized Mandarin Sentence Perception in Babble Noise Test Materials for Children," *Journal of Audiology and Speech Pathology*, vol. 17, no. 4, pp. 318-322, 2009.

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

[18] S. Ghosh. Formant Tracker Toolbox. [Online]. Available:http://www.cns.bu.edu/ speech/frack.php.

[19] 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.