

Study of the formant and duration in Chinese whispered vowel speech

Zhao Yue^{1,a}, Lin Wei^{1,a}

¹ Department of Applied Physics, College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 211106,

^a zhaoyue93214@163.com

Keywords: Whisper speech; Duration; Formant; Acoustical characteristics

Abstract. Study on the acoustical characteristic is important to speech and speaker recognition in Chinese whispered speech. In this paper, the characteristics of whispered speech are introduced and the acoustical characteristics in Chinese whispered speech are discussed. There is no fundamental frequency in the whispered speech, so other characteristics such as the duration and frequency of formant are extracted and analyzed. From experiments with six simple Chinese whispered vowels, it is proved that the duration and the frequency of formant can be used as the main acoustical characteristics in the Chinese whispered recognition.

Introduction

The whisper is a kind of natural voice that is not used very often. Because the speakers' vocal cords don't vibrate, fundamental frequency lacks and the low acoustic energy when using whisper to communicate. But people communicate using the whisper without any obstacles, because the enough information can be gotten from whisper carries. [1] This phenomenon is aroused the interest of linguists. Moreover, the secret requirements of communication also make the whisper to be use more and more frequently, because of the wide application of mobile communication in the public. [2] It is sometimes needed to convert whispered voice to normal voice, it also requires the study of the acoustical characteristics of whispered speech, especially the frequency characteristics such as the position and the bandwidth of formant. However, there are some problems in the study of the whisper. Such as, which acoustical characteristics can be used for the recognition and reconstruction of whisper, these characteristics are how important in the speech recognition, and if the whisper compared with normal tones has more changes on these characters.[3] When refactoring the whisper into normal tones, these problems have to be faced. So it is very important to the study of acoustical characteristics in the whispered speech. And this article is based on the experimental research and analysis about the characters of the duration and the frequency of formant in whispered speech.

The characteristics of the whisper

When people use the whisper to communicate, the glottis is ajar, exhaled air through the narrow glottis produces a turbulence caused noise. [4][5] Compared the speech spectrogram of in the whispered consonant with the normal consonant, there is a small difference. But there is a huge difference between the normal and whisper vowel. Because there is no vocal-cord vibration, the whispered vowels have no quasi-periodic and fundamental frequency (F0). But they also have the formant. [6][7] It is the weak-coupling property of transfer function that caused the vocal cords to be narrow and the glottis system changed. So the first and second frequency of formant (F1 and F2) will have a phenomenon that the frequency shifted to higher and bandwidth expanded to wider. On the other hand, due to noise excitation, whispered speech has lower acoustical power and longer duration than the normal tones.

The preparation of experiment

We analyzed 6 Chinese vowels, including [a]、 [ɔ]、 [ɿ]、 [i]、 [u]、 [y]. The experiment selected

10 Chinese adults (each half of men and women) to collect speech samples, that including normal vowels and whispered vowels. In an ordinary room, speakers spoke 24 vowels in turn (6 vowels, each vowel in four tones), the normal and whispered vowels each were recorded in twice. The number of the normal and whispered vowels is 480 separately (6×4×2×10).

Firstly, the LPC algorithm for root-finding is used to find trajectories of the first three formant in the Chinese whisper vowels, all the frequency and bandwidth of formants in the normal and whispered vowels are obtained the average to get the frequency characteristic in the whisper different from the normal tones. In addition, we also calculated the characteristics of various durations as a reference.

The results and the discussion

From the experiment, it can be found that the frequency of the first formant (F1) in whispered vowels is higher than normal vowels, and most of the frequency of the first and the second formant in whispered vowels is higher than normal vowels, but the difference is in reducing. And the frequency of the third formant is generally lower than normal vowels, but for these two syllables, [u] and [y], this trend is not so obvious. At the same time, we also compared the difference between speakers in the different genders about the frequency of three formants. The calculation shows that the frequency of formant in female speech is averagely 10% to 15% higher than in male speech, in both normal and whispered vowels.

In order to compare the situation that the frequency of formant in whispered vowels shifted to higher than the normal vowels, we calculated the average ratio of the frequency and bandwidth about the same kind of frequency of formant in whispered vowels and normal. [8][9]It defined as follows:

$$RF_i = \frac{\text{whispered } F_i}{\text{normal } F_i}; i = 1, 2, 3;$$

$$RBw_i = \frac{\text{whispered } Bw_i}{\text{normal } Bw_i}; i = 1, 2, 3;$$

The average ratio of frequency and bandwidth are respectively shown in table 1 and 2.

Table 1. The table of the average ratio of frequency of first three formant in Chinese normal vowels and whispered

	RF1	RF2	RF3
female	1.37	1.09	0.98
male	1.56	1.06	0.98
average	1.46	1.08	0.98

Table 2. The table of the average ratio of bandwidth of first three formant in Chinese normal and whispered vowels

	RB1	RB2	RB3
female	4.33	1.38	1.04
male	3.63	1.19	0.87
average	3.98	1.28	0.95

From Table 1, it can be found for all speakers that the average ratios of frequency are RF1 = 1.46, RF2 = 1.08, RF3 = 0.98. The difference about the ratio between male and female is not obvious. From this, we can find that the first and second frequency of formant in Chinese whispered vowels

are higher than the normal vowels; the third frequencies of formant in whispered and normal vowels are approximately at the same frequency.

From Table 2, it can be found for all speakers that the average ratios of bandwidth are $RB1 = 3.98$, $RB2 = 1.28$, $RB3 = 0.95$. The bandwidth expansion of the female is bigger than the male. Like the frequency shift, the first and second bandwidth of formant in Chinese whispered vowels are also bigger than the normal vowels; the third bandwidth of formant in whispered and normal vowels are approximately same.

In general, the frequency shift of formant in Chinese whisper is similar to other languages. The difference between male and female is in a different way about the style of pronunciation for showing the tone.

In this paper, we measured the duration of syllables in the database. The results are shown in Table 3. The durations of whispered vowels are slightly longer than the normal. After observation, the durations of the normal vowels in female speech are generally 4% to 6% longer than the male, but the durations of the whispered' in male speech are generally 10% to 19% longer than the female. Because the pronunciation of whisper is with airflow, it needs a lot of airflow, So the speed of whispered is slow, and the duration is long.

Table 3. The average of the duration in the different genders of the Chinese normal and whispered vowels unit: second (s)

syllable	female		male	
	Normal	Whispered	Normal	Whispered
[a]	0.35613	0.36761	0.34069	0.47217
[ɔ]	0.37788	0.38205	0.34980	0.44250
[ɿ]	0.36923	0.40783	0.35499	0.45564
[i]	0.37085	0.41503	0.38988	0.49632
[u]	0.38675	0.40625	0.37098	0.43572
[y]	0.39665	0.42752	0.37588	0.44461

In order to quantify the increase of the duration in the whisper, the duration of whispered vowels in each tones divided by the duration of normal vowels. We finally got the growth ratio, shown in Table 4. By observation, the change about the rising tone in female speech is the most, and the falling tone is next. And the change about the falling tone in male speech is the most, and the rising tone is next.

Table 4. The average ratio of the duration in the Chinese normal and the whispered vowels

tone	female	male
level tone	1.002	1.243
rising tone	1.117	1.284
falling-rising tone	1.070	1.195
falling tone	1.085	1.335

Conclusion

In this paper, the characteristics about the frequency of formant and the duration in Chinese whispered vowels is studied. Based on it, these characteristics are compared with the Chinese normal vowels using statistical method. Study shows that first, second and third frequency of

formant in Chinese whispered vowels are respectively upward shift 1.46 times, 1.08 times and 0.98 times that compare with the normal. Accordingly, the bandwidths respectively expand 3.98, 1.28 and 0.95 times. The durations of the whispered vowels are 7% to 33% longer than the normal vowels. Thus, compared with the Chinese normal vowels, the frequency of formant in Chinese whispered vowels become higher. Moreover the bandwidth of the Chinese whispered vowels become wider and duration of the Chinese whispered vowels become longer. The change rule is basically similar with the other languages. There is no fundamental frequency in the whispered speech, so other characteristics such as the duration and frequency of formant are extracted and analyzed. From experiments with six simple Chinese whispered vowels, it is proved that the duration and the frequency of formant can be used as the main acoustical characteristics in the Chinese whispered recognition. The results can help to improve the accuracy about recognition of Chinese whisper and transformation into the Chinese normal tones.

Acknowledgements

This work was financially supported by the Nanjing University of Aeronautics and Astronautics Basic Scientific Research Expenses (XAA150791).

References

- [1] L.L. YANG, W. LIN, B.L. XU: Isolated word recognition Chinese whispered speech. *Applied Acoustics*, 2006, 25(3): 187-192.
- [2] G.L. DING: *Research on Whispering Speaker Recognition*. Suzhou University, 2009.
- [3] X.Y. PAN : *Characteristic Analysis and its Application of Chinese Whispered speech*. Suzhou University, 2007.
- [4] W. LIN, L.L. YANG, B.L. XU: Speaker Identification in Chinese Whispered Speech Based on Modified-MFCC. *Journal of Nanjing University Natural Sciences Edition*, 2006, 42(1): 54.
- [5] W. LIN, L.L. YANG, B.L. XU: A new frequency scale of Chinese whispered speech in the application of speaker identification. *Progress in Natural Science*, 2006, 16(10): 1072-1078.
- [6] W. LIN: *Speaker Recognition in Chinese Whispered Speech*. Nanjing University, 2006.
- [7] D.Q. SHA, X.L. LI, B.L. XU: Research on the characteristics of whispered tone. *Elementary Electroacoustics*, 2003 (11): 4-7.
- [8] Y. ZHAO, L. ZHAO, C.R. ZOU: A survey of whisper analysis and its processing. *Technical Acoustics*, 2008, 27(4): 562-569.
- [9] L.L. YANG, Y. LI, B.L. XU: The Establishment of a Chinese Whisper Database and Perceptual Experiment. *Journal of Nanjing University (Natural Sciences)*, 2005, 41(3): 311-317.