Applied Acoustics 114 (2016) 240–243

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Technical note

Study of the formant and duration in Chinese whispered vowel speech

Zhao Yue, Lin Wei *

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

ARTICLE INFO

Article history:
Received 29 May 2016
Received in revised form 30 July 2016
Accepted 4 August 2016

Keywords:
Whisper speech
Duration
Formant
The 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.

1. 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.

* Corresponding author.
E-mail address: WLin@Nuaa.edu.cn (W. Lin).

http://dx.doi.org/10.1016/j.apacoust.2016.08.007

2. 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.

3. The LPC algorithm for root-finding

In order to study the formant excursion, we use the LPC root-finding algorithm to get the frequency of formant and the bandwidth calculation. The principle is as follows:

According to the LPC model of the acoustical signal, the signal s(n) can be expressed as:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \quad (1)$$

Obtained by the operation of differential equation, the corresponding sound channel transfer function H(z) is:

$$H(z) = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-1}} \tag{2}$$

In the formula (2), $z^{-1} = \exp(-j\omega T)$ or $z^{-1} = \exp(-j2\pi f/f_s)$, so its power spectrum P (f) is

$$P(f) = |H(f)|^2 = \frac{G^2}{|1 - \sum_{k=1}^{p} a_k \exp(-j2\pi kf/f_s)|^2} \tag{3}$$

According to the required frequency of using this method, the amplitude-frequency response of the power spectrum is found. On this basis, the frequency of formant and bandwidth will be found from the amplitude-frequency response [8]. If the LPC algorithm is used, it is usually chosen the LPC method of complex-root-finding to get the formant information. And in this article, the LPC method for root-finding also be used.

Set the sound channel transfer function as

$$H(z) = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-1}} \tag{2}$$

And set A(z) as the prediction error filter, its value can be obtained by

$$A(z) = 1 - \sum_{i=1}^{p} a_i z^{-1} \tag{4}$$

Decomposing polynomial coefficients of A(z), the frequency and width of formant can be obtained. The solutions of A(z) are mostly pairs of complex conjugate solutions.

Set

$$z_i = r_i e^{j\theta_i} \tag{5}$$

As arbitrary complex roots, then its conjugate solution is

$$z_i^* = r_i e^{-j\theta_i} \tag{6}$$

This is also a solution of the polynomial. F_i is the corresponding the frequency of formant of z_i, which width is B_i. F_i and B_i can be gotten from the following equation:

$$2\pi T F_i = \theta_i \tag{7}$$

$$e^{-B_i \pi T} = r_i \tag{8}$$

In the formula, T is for signal sampling period. So,

$$F_i = \theta_i/(2\pi T) \tag{9}$$

$$B_i = -\ln\ r_i/\pi T \tag{10}$$

Because the largest number of complex conjugate root is p/2, and according to the requirements, p's values usually can be assigned. So it's easy to judge the relationship between the pole and the formant. At the same time, because the widths of special poles at the un-formants are much less than the formants, therefore they can be excluded easily [8].

4. The preparation of experiment

We analyzed 6 Chinese vowels, including /a/, /o/, /e/, /i/, /u/, /ü/. 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 ($6 \times 4 \times 2 \times 10$).

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.

5. The results and the discussion

5.1. The contrast with the frequency of formant in the normal vowels and whispered vowels

Table 1 is the contrast table of the frequency of first three formant in Chinese normal vowels and whispered vowels. Table 2 is the contrast table of the bandwidth. From the table, 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 /ü/, 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 10–15% 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 [9,10]. It defined as follows:

$$RF_i = \frac{whispered\ F_i}{normal\ F_i}; i = 1, 2, 3; \tag{11}$$

$$RBw_i = \frac{whispered\ Bw_i}{normal\ Bw_i}; i = 1, 2, 3; \tag{12}$$

The average ratio of frequency and bandwidth are respectively shown in Tables 3 and 4.

From Table 3, 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

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

Syllable	Normal vowel (Hz)			Whispered vowel (Hz)		
	F1	F2	F3	F1	F2	F3
/ā/	928.257	1391.78	2892.53	987.93	1509.66	2843.77
/á/	968.888	1415.31	2875.56	1046.46	1523.31	2893.69
/ǎ/	983.106	1357.76	2798.72	1035.09	1467.67	2881.56
/à/	945.932	1354.21	2862.24	1022.49	1512.9	2800.26
/ō/	473.658	890.237	2901.39	688.418	1020.29	2883.81
/ó/	462.458	879.508	2856.58	706.9	1080.92	2912.59
/ǒ/	480.976	869.058	2826.94	697.183	1055.26	2888.1
/ò/	493.395	888.623	2889.06	725.411	1198.94	3009.55
/ē/	511.268	1328.56	2909.56	629.223	1388.74	2907.36
/é/	471.959	1361.82	2905.83	652.279	1395.87	2928.54
/ě/	489.51	1311.44	2871.18	718.49	1382.01	2913.09
/è/	508.5	1377.14	2927.37	625.12	1392.39	2928.09
/ī/	279.083	2520.99	3219.56	447.852	2512.47	3192.96
/í/	273.262	2509.55	3227.09	515.372	2507.19	3228.45
/ǐ/	300.518	2407.72	3223.94	664.438	2555.07	3205.6
/ì/	295.554	2595.44	3257.98	581.352	2568.97	3227.96
/ū/	341.202	761.881	2739.45	554.867	1174.94	2692.43
/ú/	347.686	759.288	2674.53	587.445	1120.63	2741.59
/ǔ/	365.52	761.664	2643.36	463.418	933.883	2716.94
/ù/	354.393	781.011	2745.11	467.923	939.113	2720.07
/ǖ/	276.457	2192.41	2649.76	592.842	2184.87	2674.95
/ǘ/	286.209	2187.31	2597.67	639.46	2216.8	2696.93
/ǚ/	301.033	2185.28	2595.28	417.481	2110.97	2616.41
/ǜ/	298.697	2159.54	2551.63	445.014	2155.35	2601.46