



A Modified Adaptive Algorithm for Formant Bandwidth in Whisper Conversion

Gang Lv & Heming Zhao

School of Electronic Information

Soochow University

Suzhou 215021, China

Tel: 86-512-6219-0898 E-mail: lvgang@suda.edu.cn

The research is supported by the National Natural Science Foundations of China under Grant No.60572076 and the University Natural Science Research Project of Jiangsu Province of China under Grant No.05KJB510113. (Sponsoring information)

Abstract

The whisper conversion technology is to transform undistinguished whispers with lower SNR into clear normal speech, and it has important application prospect in mobile communication. Because the whisper speech is stirred by the yawp source, its formant position shifts and its bandwidth increases, which induces the problem of formant combination occurs in the whisper conversion. By analyzing the power spectrum, in this article, we proposed a modified adaptive algorithm for formant bandwidth. Based on the rule that the pole power does not change, the algorithm has resolved the problem of formant combination by modifying the formant bandwidth of whisper before implementing formant conversion. The experimental results with six Chinese mandarin monophthong phoneme conversions proved the validity of the algorithm.

Keywords: Whisper speech, Voice conversion, Formant combination, Bandwidth

1. Introduction

The research to the whisper which is the special speech first rooted in the need for the phonation to the laryngeal excision sufferer (Liang, 1997, P.151-152). With the extensive use of mobile communication tool, whisper has been a sort of effective communication mode which could increase the secret of talk and doesn't influence the hearing environment. How to transform undistinguished whispers with lower SNR into clear normal speech and realize the whisper communication has been more and more noticed by foreign and domestic scholars (Li, Xueli, 2004 & Morris R W, 2002).

Unlike the pronunciation mode of normal speech, the whisper source is yawp and the vocal cords doesn't oscillate, so comparing with normal speech, the whisper has no pitch frequency, the sound level is lower about 20dB (Gao M, 2002). The change of track transfer function makes the formant position of the whisper change, and the corresponding formants that the first and second formants are higher than the normal speech and the phrase spectrums such as the adding bandwidth of formant occur (Jovicic S T, 1998, P.739-743).

According to above acoustics characters, the conversion from whisper to normal speech is mainly realized by adding the pitch frequency and modifying the formant spectrum (Morris R W, 2002, P.515-520). But in the experiment of Chinese mandarin monophthong phoneme conversion, we found that the traditional formant conversion algorithm would meet the problem of formant combination and influence later normal speech combination. In this article, through analyzing the pole power spectrum of formant, we pointed out that the adding whisper bandwidth is the direct reason to produce the problem of formant combination for the whisper conversion, and put forward a sort of modified adaptive formant bandwidth algorithm which could effectively reduce the bandwidth of whisper formant and solve the problem of formant combination of whisper after conversion.

2. Formant combination in whisper conversion

Figure 1 is the former three formants extracting whisper signal [o4] by the linear predictive spectrum algorithm

(McCandless S, 1974, P.135-141), and Figure 2 is the formant of the normal speech [o4] after conversion by the Gauss mixture model (Lv, 2004). From Figure 2, the first formant and the second formant produce the formant combination. And we adopt the method of pole power spectrum to analyze the reasons for above phenomena.

First, we realize the conversion from the frequency domain to the z domain. If the sampling frequency is F_s , the formant F_i and the 3dB bandwidth B_i from the LPC algorithm can be converted to the pole with the angle of ϕ_i and the radius of r_i in the z domain according to the following formulas.

$$\text{The radiation angle of the pole: } \phi_i = 2\pi \frac{F_i}{F_s} \quad (1)$$

$$\text{The radius of the pole: } r_i = e^{-\frac{B_i * \pi}{F_s}} \quad (2)$$

From the radiation angle and the radius, we can obtain the transfer function

$$H(z_i) = \frac{1}{1 - r_i e^{j\phi_i} z_i^{-1}} \quad (3)$$

The power spectrum of the pole z_i in the z domain is

$$|H(e^{j\theta})|^2 = \prod_{i=1}^n \frac{1}{1 - 2r_i \cos(\theta - \phi_i) + r_i^2} \quad (4)$$

Convenient for discussion, we first suppose two poles z_1 and z_2 , so the power $|H(e^{j\phi})|^2$ at the radiation angle ϕ_1 is

$$\frac{1}{(1 - r_1)^2} \cdot \frac{1}{1 - 2r_2 \cos(\phi_1 - \phi_2) + r_2^2} \quad (5)$$

In the z domain, when poles z_1 and z_2 gradually close up, the difference of their radiation angle would reduce, and from formula (5), we can see that the power peak value at the radiation angles ϕ_1 and ϕ_2 will also reduce until two power peaks combines.

The traditional formant conversion algorithm only directly shifts the formant of whisper signal [o4] from $F1=1078\text{Hz}$, $F2=1721\text{Hz}$ and $F3=2718\text{Hz}$ to $F1=670\text{Hz}$, $F2=1173\text{Hz}$ and $F3=3630\text{Hz}$ and doesn't modify the bandwidth, and because of the poles close up, so the problem of formant combination occurs.

3. Modified bandwidth adaptive algorithm

According to above analysis, if we can reduce corresponding formant bandwidth when the formant is converted, so the combination of formant could be eliminated in theory. In the traditional LPC algorithm, when we abstract the formant, we directly delete the formant poles which don't accord with the requirements. Based on the rule that the pole power does not change, in this article, we propose the algorithm which automatically add the energy of the deleted formant to the reserved formant and realize the adaptive formant bandwidth change. The implementing principle of the algorithm includes the reserved formant pole with the angle of ϕ_i and the radius of r_i is z_i , and the deleted formant pole with the angle of ϕ_j and the radius of r_j is z_j . According to formula (5), the power at the angle of ϕ_i is

$$\frac{1}{(1 - r_i)^2} \prod_{j=1}^M \frac{1}{1 - 2r_j \cos(\phi_i - \phi_j) + r_j^2} = \frac{1}{(1 - r_i')^2} \quad (6)$$

Here, r_i' denotes the corresponding new pole radius when deleting pole z_j and reserving the unchanged pole energy, and M denotes the deleted pole amount.

In addition, we must consider the influence to other reserved poles when changing the radius of a pole. So the formula (6) could be extended as

$$\frac{1}{(1 - r_i')^2} \prod_{k=1, k \neq i}^N \frac{1}{1 - 2r_k \cos(\phi_i - \phi_k) + r_k^2} \times \prod_{j=1}^M \frac{1}{1 - 2r_j \cos(\phi_i - \phi_j) + r_j^2} = \frac{1}{(1 - r_i')^2} \prod_{k=1, k \neq i}^N \frac{1}{1 - 2r_k' \cos(\phi_i - \phi_k) + r_k'^2} \quad (7)$$

Here, r_k is radius of other reserved poles, r'_k is the corresponding pole radius after modification, and N is the amount of reserved linear predictive multinomial pole.

4. Experimental results

In the experiment, we select six monophthong whisper speeches including /a/, /o/, /e/, /i/, /u/ and /ü/, and the normal speech as the samples, and every speech possesses four pronunciations including level tone, rising tone, falling-rising tone and falling tone, and the sample amount is 24.

We take the stable speech area sample to implement pre-aggravating and window-adding processing, and the pre-aggravating coefficient μ is 0.975, and we adopt the window of Hamming. The experimental sampling rate is 8kHz, and every frame has 256 sampling points. The experimental results showed that the problem of formant combination also occurs in the whispers [o2], [i2] and [u4] except for whisper [o4].

The experiment result is seen in Figure 3. The point lineation is three formants obtained by traditional LPC algorithm, and the thin real line is the normal speech frequency spectrum curve after conversion obtained by Gauss mixture model, and from the figure, we can intuitively see the combination of formants F1 and F2. The broken line is the whisper frequency spectrum curve obtained by the modified adaptive bandwidth algorithm, and through the comparison, we can see that the new algorithm the 3dB bandwidth of three formants is smaller the bandwidth obtained by the traditional LPC algorithm. The wide real line is the frequency spectrum curve of normal speech through the whisper formant conversion obtained by the new algorithm, and comparing with the conversion result through the traditional method, it eliminates the combination of formants F1 and F2.

5. Conclusions

In the conversion of Chinese whisper, because the formant bandwidth of whisper is wider than the normal speech, it will induce the problem of formant combination when we adopt the conversion method directly shifting the formant. In this article, we put forward the method which first add the spectrum power of the deleted formant pole to the formant of the reserved pole, realize the adaptive adjustment of the formant bandwidth, and utilize the Gauss mixture model to implement the formant conversion. The experimental conversion of Chinese mandarin monophthong phoneme proved the method could better solve the problem of formant combination occurring in the speech conversion.

References

- Gao M. (2002). *Tones in Whispered Chinese: Articulator and Perceptual Cues*. Canada: University of Victoria.
- Itoh T, Takeda K & Itakura F. (2005). Analysis and Recognition of Whispered Speech. *Speech Communication*. No.45(2). P.139-152.
- Jovicic S T. (1998). Formant Feature Differences between Whispered and Voiced Sustained Vowels. *Acustica*. No.84(4). P.739-743.
- Liang, Cifang, Wu, Xiaozhong, Li, Chengtian & Tan, Xiaohui. (1997). The Application of Electric Artificial Larynx in the Phonation for the Suffer without Larynx. *Chinese Journal of Otorhinolaryngology Surgery*. No.32(3). P.151-152.
- Li, Xueli. (2004). *Study on the Transformation from Chinese Whisper Speech to Normal Speech*. Nanjing: Doctoral Dissertation of Nanjing University.
- Lv, Sheng. (2004). *Study on the Method of Human Voice Transformation*. Guangzhou: Doctoral Dissertation of South China University of Technology.
- McCandless S. (1974). An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra. *IEEE Trans. on ASSP*. No.22(2). P.135-141.
- Morris R W, Clements M A. (2002). Reconstruction of Speech from Whispers. *Medical Engineering & Physics*. No.24(8). P.515-520.
- Morris R W. (2002). *Enhancement and Recognition of Whispered Speech*. USA: Doctoral Dissertation of Georgia Institute of Technology.

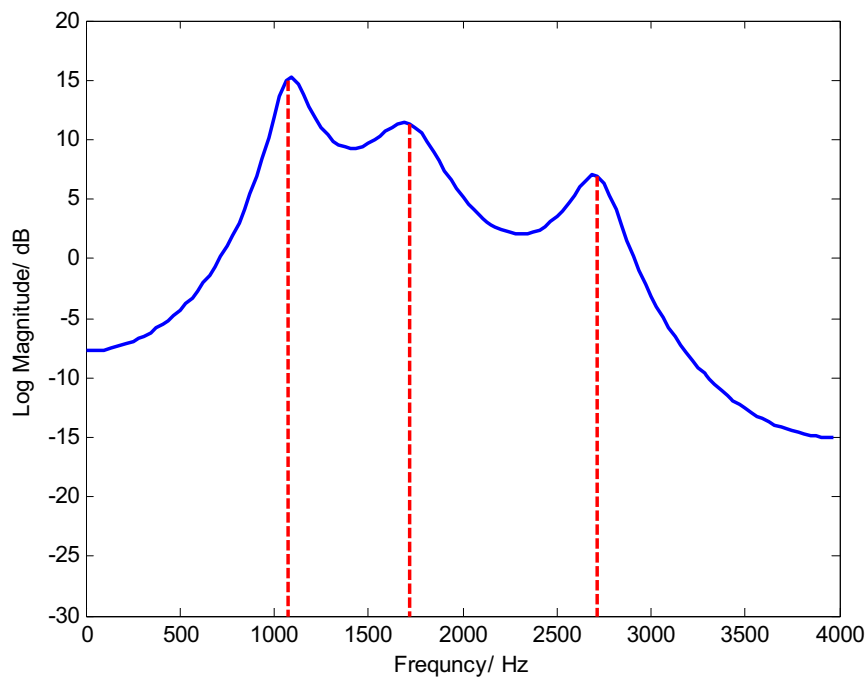


Figure 1. LPC Frequency Spectrum Envelope of Whisper Speech [o4]

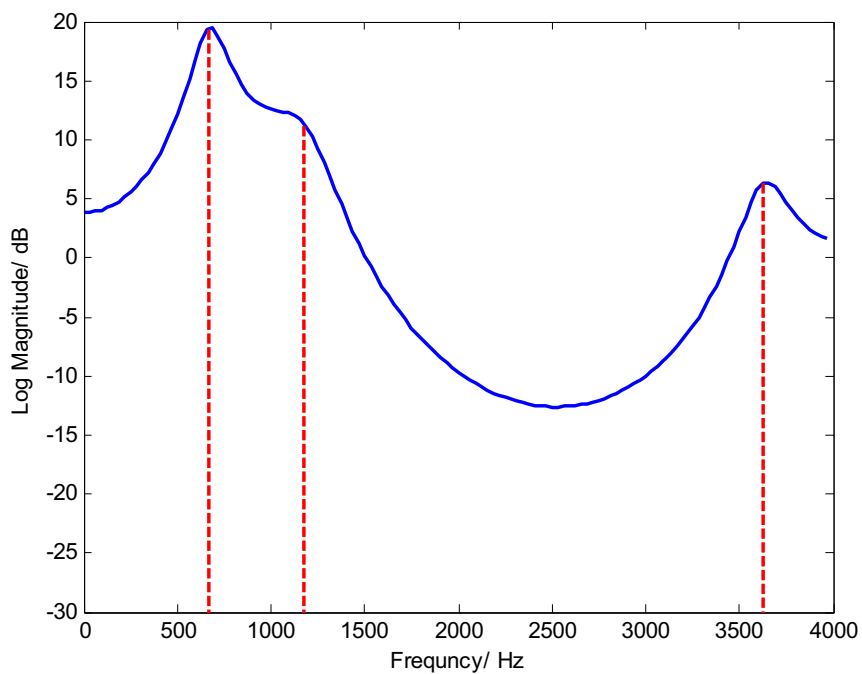
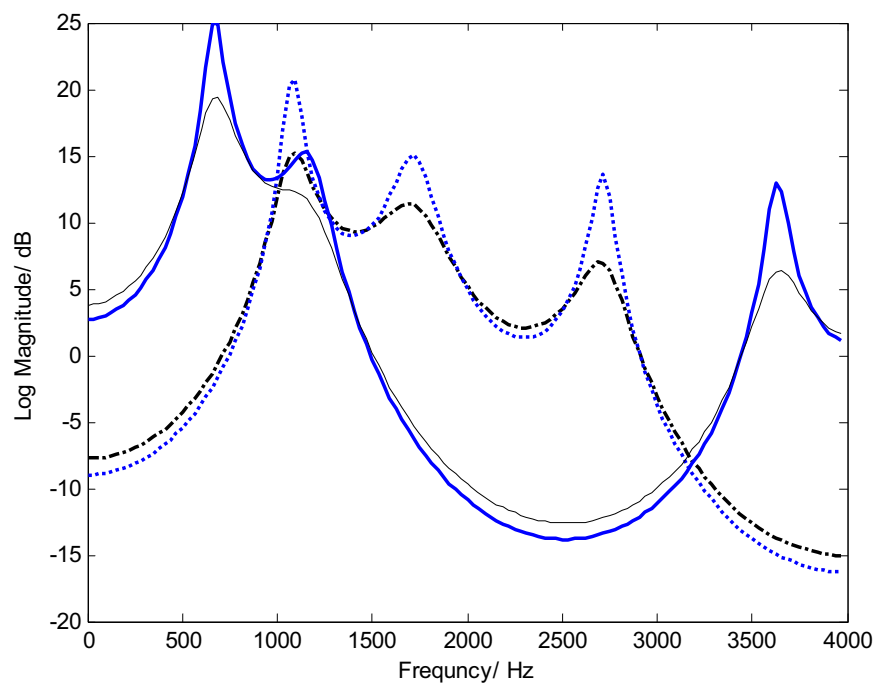
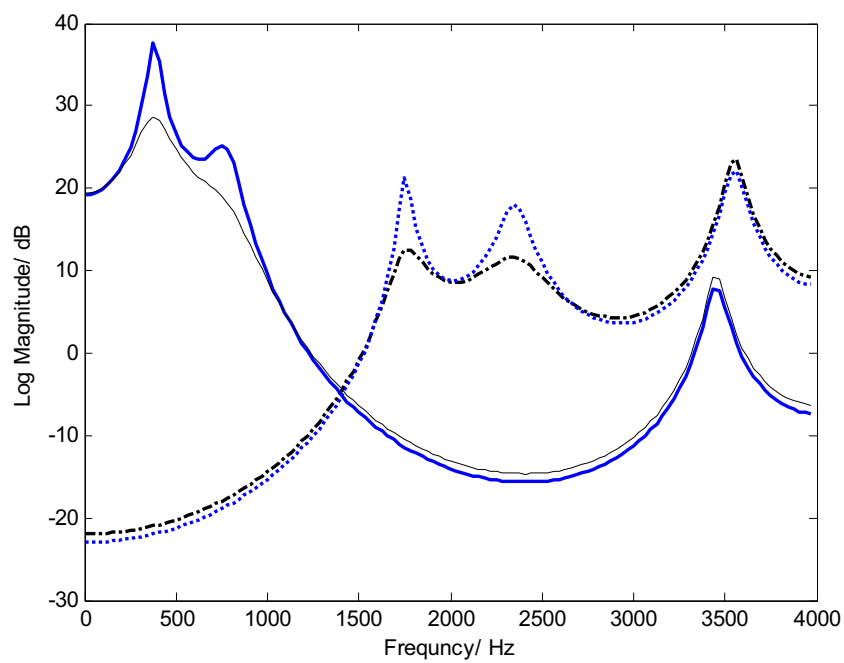


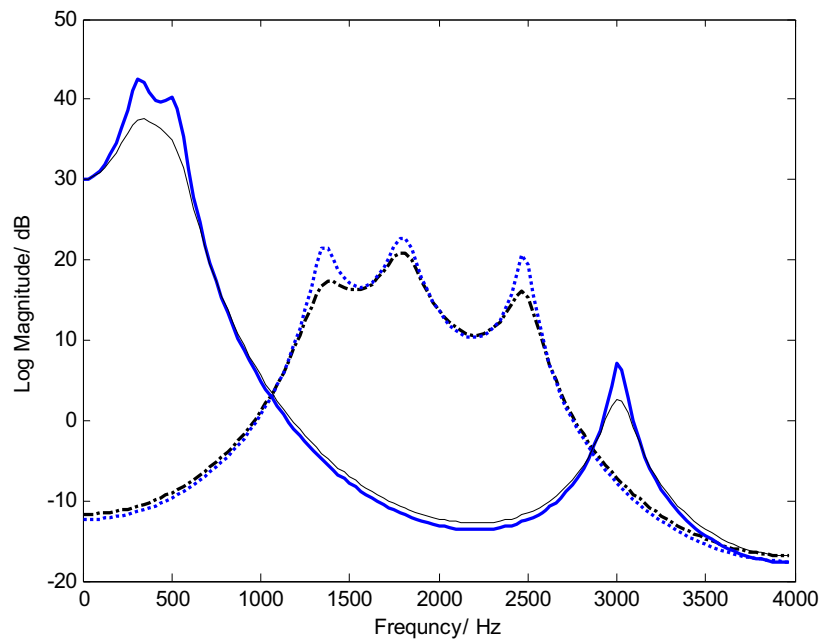
Figure 2. Frequency Spectrum Envelope of Normal Whisper Speech [o4] after Conversion



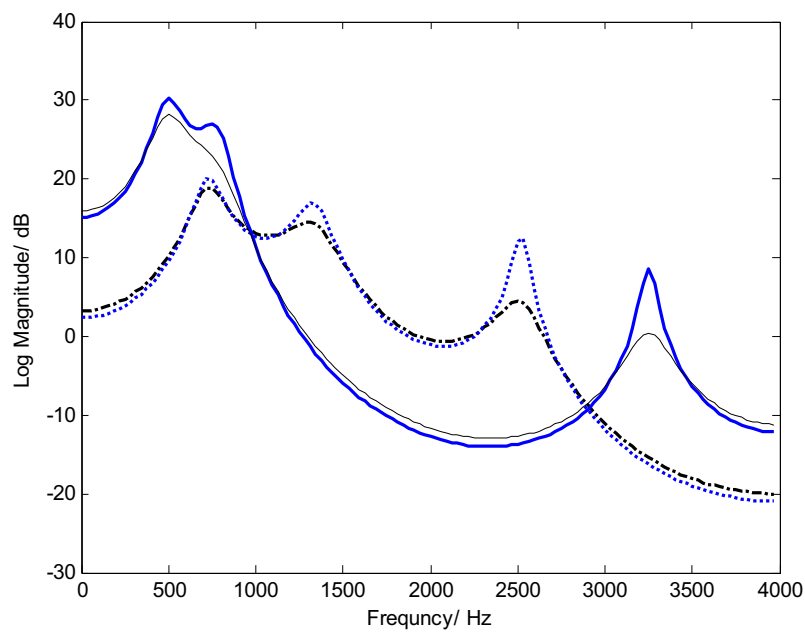
[o4]



[u4]



[i2]



[o2]

Figure 3. Comparison between Proposed Algorithm and LPC Algorithm on the Effect of Frequency Spectrum Envelopes of Normal Whisper Speeches after Conversions