ARCHIVES OF ACOUSTICS
34, 2, 127–135 (2009)

Formant Frequency Estimations of Whispered Speech in Chinese

Gang LV, Heming ZHAO

Soochow University
School of Electronic Information
Suzhou, 215021, P.R.China
e-mail: lvgang@suda.edu.cn

(*received September 13, 2008; accepted February 16, 2009*)

Formant frequencies are important cues for characterizing whispered speech. However, it is difficult to exactly estimate its formant by the conventional linear prediction coding algorithm. The main reason is that the formant bandwidth of a whisper is wider than that of voiced speech. This brings up the pole interaction problem that then leads to the result that one or more real roots are regarded as spurious and deleted from the original LP polynomial. To reduce the degradation of pole interactions, an improved root-finding formant estimation algorithm has been proposed. In this algorithm, the whisper formant bandwidth is modified to make the spectral energy of the remained formant polynomial equal to that of the original LP polynomial. Experimental results with six Chinese whispered monophthong phonemes show that the formant frequencies obtained by the proposed algorithm produce a more reliable formant spectrum than the one that does not consider the pole interaction effect.

Keywords: whispered speech, formant, linear prediction, pole interaction.

1. Introduction

Whispering is a common mode of speaking to communicate quietly and privately. Since F0 is absent in whispers [1], formant frequency estimation becomes prominent in its analysis and recognition. However, previous research has shown that exact abstracting of the formant information is extremely difficult [2, 3].

Linguistic studies reveal that whispered speech sources are distributed through the lower portion of the vocal tract, resulting in speech that is completely noise-excited. So, compared with voiced speech, the first and second formant frequencies of whispered speech shifts to the higher frequencies and the first formant is closed to the second formant [4, 5]. Thus, the pole interaction problem arises [6], Fig. 1

shows that the spectrum of whispers is flatter than the voiced spectrum; this leads to the result that one or more real roots, regarded as spurious roots, are deleted from the original LP polynomial by the conventional root-finding formant estimation algorithm. Therefore, an improved algorithm for reliably computing whisper formants is proposed. Based on the rule that the spectral energy of the remained whisper formant polynomial is equal to that of the original LP polynomial, this algorithm decreases the effect of the pole interaction problem by modifying the bandwidths of remaining whisper formant poles before their selection.

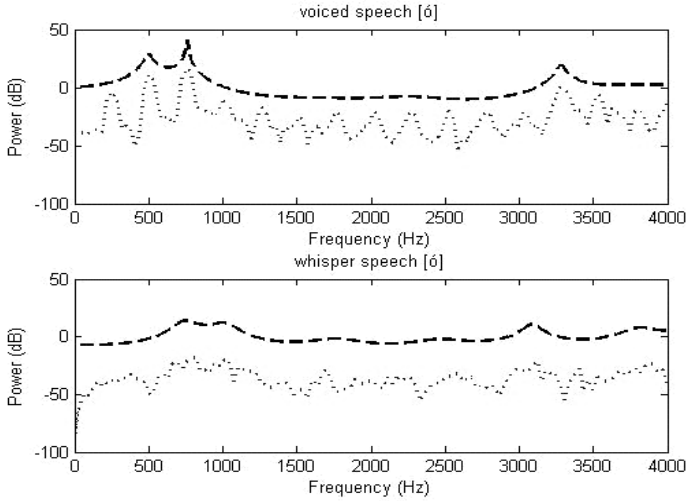

Fig. 1. A comparison between the spectrum of whisper and voiced speech.

The pole interaction problem and the relationship between a pole and its resonant bandwidth are discussed in Sec. 2. Section 3 dwells on the improved formant estimation algorithm, which is based on modifying bandwidths using the pole interaction factor. Section 4 shows the experimental results. Finally, the conclusions are included in Sec. 5.

2. Pole interaction problem

In the frequency domain, if the sampling frequency is F_s, the formant F_i and the 3 dB 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:

The radiation angle of the pole:

$$\phi_i = 2\pi \frac{F_i}{F_s}. \tag{1}$$

The radius of the pole:

$$r_i = e^{-B_i * \pi / F_s}. \tag{2}$$

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

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

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

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

For the convenience of this discussion, we first suppose two poles z_1 and z_2, so the power $\left| H(e^{j\phi_1}) \right|^2$ at the radiation angle ϕ_1 is

$$\frac{1}{(1 - r_1)^2} \cdot \frac{1}{1 - 2r_2 \cos(\phi_1 - \phi_2) + r_2^2} = \frac{1}{(1 - r_1)^2} \cdot \Delta |H|, \tag{5}$$

where $\Delta |H|$ is defined as the pole interaction factor (PIF) of pole z_2 with pole z_1 and can be used to measure the interaction effect.

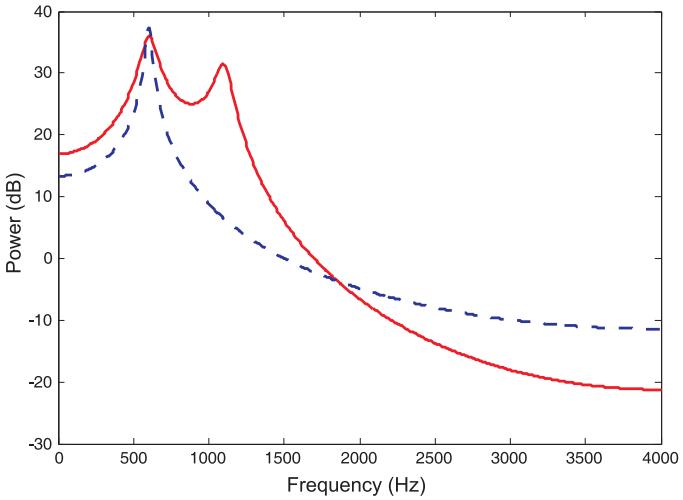

Fig. 2. An example of pole interaction problem.

In the z domain, when poles z_1 and z_2 gradually converge, the differences in their radiation angles would be reduced, according to formula (5), the angle difference between these two poles is diminished, the PIF is increased, and the corresponding spectral peak at the angle ϕ_1 is raised. Otherwise, the PIF is decreased and the corresponding spectral peak is lowered.

Here we have conducted an experiment. First, we supposed that there is one pole z_1, whose power spectrum is shown by a broken line in Fig. 2. After that, a pole z_2 is added. Being a pole interaction factor, the added pole affected the

bandwidth of pole z_1. As shown by a solid line in Fig. 2, the bandwidth of pole z_1 has been broadened; it will lead to pole z_1 and may be deleted as a spurious root by the conventional root-finding formant estimation algorithm.

Since the conventional root-finding formant estimation algorithm derives the appropriate formant pole from the LP coefficients according to its bandwidth, the pole interaction problem will obviously affect the selected results for whispering, because the first two formants' positions shift together and their bandwidths increase [4, 5].

3. Formant estimation algorithm

As mentioned above, the conventional root-finding algorithm suffers the problem that one or more real roots are deleted from the original whispered LP polynomial. To reduce the degradation of pole interaction, we proposed an improved algorithm. In the new design, the radii of the remaining poles are modified to make the spectral energy of the formant polynomial equal to that of the original whispered LP polynomial at the formant frequencies.

Suppose 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

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

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 on other reserved poles when changing the radius of a pole. So, the formula (6) could be extended as

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

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

The algorithm is realized in the following seven ways:

1. Create a whisper spectrum through pre-filtering;
2. Determine LP roots from the LP polynomial;
3. Calculate the bandwidth for each root by using formula (2);
4. Descend the bandwidths and classify the root with the largest bandwidth as the deleted pole;

5. Start with the root whose bandwidth is the smallest and obtain the new root by using formula (6);
6. Use formula (7) to modify the rest of the formant roots;
7. Continue this process from Step 3 to Step 6 until only the first three formants remain.

An example comparing the ability of abstracting formants from whispers [ó] between the proposed algorithm and the conventional LPC is given in Fig. 3. The detailed calculation process for the improved algorithm is shown in Table 1.

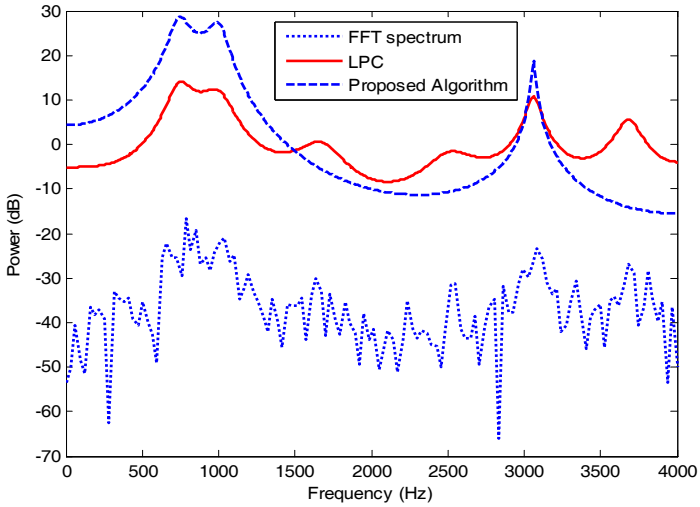

Fig. 3. A comparison of the ability of abstracting formant from whisper [ó] between the proposed algorithm and the conventional LPC.

Table 1. The calculation process of the proposed algorithm.

Repeat times		roots							
		1	2	3	4	5	6	7	8
1	Formant	3055	736	3673	1020	1671	2506	0	0
	Bandwidth	84	143	146	183	262	308	733	17942
2	Formant	3055	736	3673	1020	1671	2506	0	
	Bandwidth	83.7	142.7	145.7	182.7	262.3	307.6	732.6	
3	Formant	736	1020	3055	3673	1671	2506		
	Bandwidth	78.1	128.5	138.2	258.6	286.2	464.1		
4	Formant	3055	736	1020	1671	3673			
	Bandwidth	57.8	92.4	132.4	171.1	211.8			
5	Formant	3055	736	1020	1671				
	Bandwidth	26.9	164.6	224.4	235.8				
6	Formant	3055	1020	730					
	Bandwidth	26.7	110.4	112.6					

According to voiced speech, the real second formant of a whisper [ó] at 1020 Hz is close to the first formant, so the bandwidth of the second formant is bordered because of the pole interaction. We can see from the first row in Table 1 that the second formant at 1020 Hz lays at fourth if it is sorted by bandwidth and the corresponding root is deleted by conventional LPC.

To reduce the degradation of pole interaction, the improved algorithm deletes the root with the largest bandwidth and changes the bandwidth of remaining roots by using formula (7). The calculation is made according to the order from left to right of the first row. After all the remaining formant roots are modified, we sort roots by bandwidth again and get the data for the second row. This process is repeated until only the first three formants remain. As Table 1 shows, after three repetitions of the calculation, the spurious formant at 3673 Hz is identified and the correct formants are at last identified.

4. Experimental tests

Chinese vowels /a/, /e/, /i/, /o/, /u/ and /ü/, which possess four pronunciations including the level tone, rising tone, falling-rising tone and falling tone, were pronounced by a male speaker, in order to evaluate the effectiveness of the proposed method.

The 24 vowels were recorded in a studio, digitized at 16 kHz, and downsampled to 8 kHz. We performed a signal preemphasis by calculating the first-order difference of the sampled speech signal. Every 8 ms, a 32 ms Hamming window is applied to overlapping speech segments, and a 14th order LP polynomial is used to estimate the power spectrum. The frequency range from 0–4 kHz was used for formant estimation.

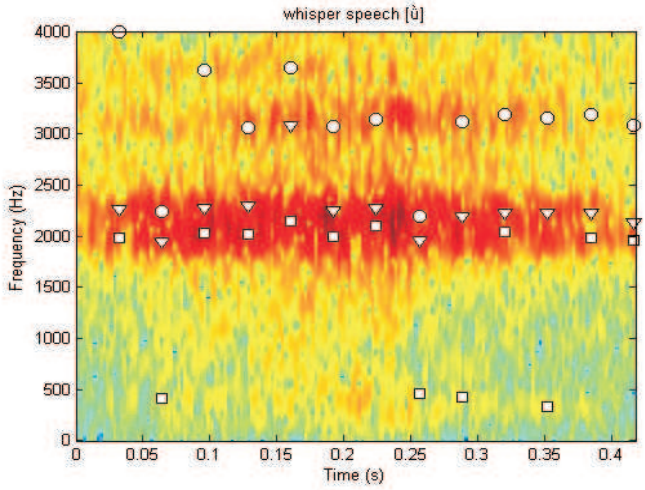

Fig. 4. Formant track of whisper [ù] using LPC algorithm.

Figure 4 shows the first three formants' tracks of whispers [ù], as determined by a conventional root-finding formant estimation algorithm. From the spectrogram, we can see the first formant frequency should lie at about 2000 Hz, while the second is about 2200 Hz, and the third is about 3100 Hz. The first formant is close to the second formant. It is easy to observe that the algorithm abstracts wrong formants in Frames 1, 2, 3, 5, 8, 9, and 11.

Figure 5 shows the first three formants' tracks of whispers [ù], as determined by a modified algorithm. Compared with Fig. 4, we can observe that the improved algorithm captured the correct formants in every frame.

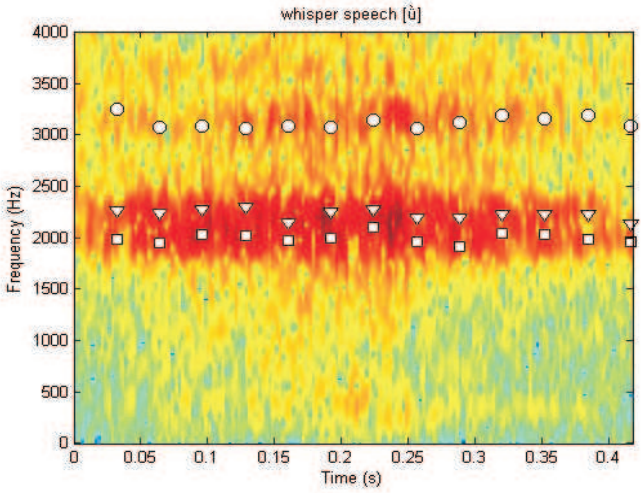

Fig. 5. Formant track of whisper [ù] using the proposed algorithm.

To evaluate the performance of the proposed algorithm, we used the statistical software SPSS for the data analysis. For example, we captured the data of the first three formants of whispers [ù] in every frame. It is shown in Table 2.

Table 2. The first three formants of whisper [ù] got by the proposed algorithm and the conventional LPC.

	Formant	1	2	3	4	5	6	7	8	9	10	11	12	13
Conventional LPC	F1	1986	416	2024	2015	2144	1991	2103	461	421	2043	335	1988	1956
	F2	2259	1952	2271	2294	3081	2249	2274	1954	2189	2234	2231	2230	2140
	F3	4000	2239	3627	3060	3643	3068	3142	2198	3118	3184	3150	3187	3081
Proposed algorithm	F1	1986	1952	2024	2015	1976	1991	2103	1954	1918	2043	2025	1988	1956
	F2	2259	2239	2271	2294	2144	2249	2274	2198	2189	2234	2231	2230	2140
	F3	3251	3077	3079	3060	3081	3068	3142	3060	3118	3184	3150	3187	3081

After that, we calculated the average F_i and standard deviation σ_i, and defined the significance level at $\mu = 0.05$. The results given by the software were

that six frames are rejected by the conventional LPC and zero when using the proposed algorithm. If we define

$$Error\_rate = \frac{rejected\_frame}{full\_frame}. \tag{8}$$

The error with the conventional LPC is $\frac{6}{13}$, and the error with the proposed algorithm is $\frac{0}{13}$.

A performance comparison of the conventional LPC-based algorithm, spectral segmentation based algorithm mentioned in reference [7] and the proposed algorithm is given in Table 3. It shows that the proposed method brings about a considerable improvement in general over other method in the formant frequency results, especially for the vowels /a/, /e/, and /u/.

Table 3. Comparative performances of the proposed algorithm and other method ($\mu = 0.05$).

	Tone	a	e	i	o	u	ü
Errors with the conventional LPC (%)	Level	27.27	20.00	41.18	33.33	12.50	43.75
	Rising	11.11	43.75	50.00	26.08	25.00	27.78
	Falling-rising	34.78	30.44	50.00	30.77	42.11	16.00
	Falling	36.36	27.27	33.33	63.64	30.77	46.15
Errors with the spectral segmentation algorithm (%)	Level	9.09	16.00	17.65	33.33	12.50	18.75
	Rising	16.67	6.25	22.22	17.39	12.50	27.78
	Falling-rising	26.09	26.09	33.33	15.39	36.84	20.00
	Falling	45.46	27.27	41.67	27.27	15.39	7.69
Errors with the proposed algorithm (%)	Level	0	4.00	0	19.05	6.25	12.50
	Rising	5.56	0	5.56	4.35	0	5.56
	Falling-rising	0	0	8.33	3.85	0	16.00
	Falling	9.09	9.09	8.33	9.09	0	0

5. Conclusions

This paper proposed an improved algorithm for formant extraction for Chinese whispered speech. This algorithm changes the bandwidths of the formant poles to solve the interaction problem while maintaining the formant structure.

In order to evaluate the features of the proposed algorithm, we have constructed a Chinese whispered database and compared the error rates. The experiment results show that the formant frequencies obtained by the proposed algorithm produce a lower error rate than one that does not consider the pole interaction effect.

Acknowledgments

The authors would like to express thanks to J.X. Liu for his assistance in providing the earlier research data. We also wish to acknowledge that the comments from anonymous reviewers on the paper's early version were very helpful in making this paper more readable. This work is supported by the National Natural Science Foundation of China, under Grant No. 60572076 and the University Natural Science Research Project of Jiangsu Province of China, under Grant No. 05KJB510113.

References

[1] Tartter V.C., *What's in a Whisper?*, Journal of Acoustical Society of America, **86**, 1678–1683 (1989).

[2] Itoh T., Takeda K., Itakura F., *Analysis and recognition of whispered speech*, Speech Communication, **45**, 139–152 (2005).

[3] Morris R.W., Clements M.A., *Modification of formants in the line spectrum domain*, IEEE Signal Processing Letters, **9**, 19–21 (2002).

[4] Ding H., Li X.L., Xu B.L., *Initial/Final Segmentation of Chinese Whispered Speech based on the Auditory Model*, Acoustic Application, **23**, 20–25 (2004).

[5] Li X.L., Ding H., Xu B.L., *Entropy-based Initial/Final Segmentation for Chinese Whispered Speech*, Acta Acustica, **30**, 69–75 (2005).

[6] Kuwabara H., Ohgushi K., *Contributions of vocal tract resonant frequencies and bandwidths to the personal perception of speech*, Acoustica, **63**, 120–128 (1987).

[7] Gong C.H., Zhao H.M., Lv G., Liu J.X., *Formant estimation of whispered speech based on spectral segmentation*, IEEE International Symposium on Signal Processing and Information Technology, 562–566 (2006).