International Symposium on Chinese Spoken
Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

ACOUSTICAL STUDY ON SUB-HARMONIC OF GLOTTAL SOURCE IN MANDARIN TONES*

Jiangping Kong

Dept. of Chinese Language and Literature
Center for Chinese Linguistics
Peking University, Beijing

ABSTRACT

This paper is concerned with the acoustical analysis on sub-harmonic of glottal source in Mandarin tones. The methods used in this research are: 1) extracting glottal source of tones by inverse filtering; 2) analyzing sub-harmonic and spectrum tilt by FTT; 3) simulating the double peak pulse by 4 functions and describing the natures of them in both time and frequency domains. There are 3 conclusions: 1) the double peak pulse produce sub-harmonic in glottal source of Mandarin tones; 2) sub-harmonic influences the spectrum tilt; 3) the double peak pulse can be simulated and modeled mathematically.

1. INTRODUCTION

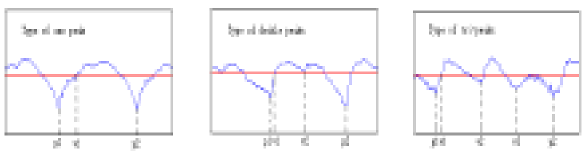
Chinese is a typical tone language, which has 4 basic tones and 20 diatones including the 4 compositions of 4 basic tones and the neutralized tones. The 4 basic tones are: 1) tone 1 or 'high level' tone described as 55 by the 5 letter tone system by Chao Yuanren; 2) tone 2 or 'rising' tone described as 35; 3) tone 3 or 'low' tone described as 214; 4) tone 4 or 'falling' tone described as 51. The neutralized tone is usually called tone 5 whose value depends on that of the basic tone before it. Since the most important parameter of tone is fundamental frequency (F0), almost all researchers pay their attention to F0 of Mandarin tones, and many contributions have been achieved in the phonetic study.

According to the theory of acoustics, speech production can be divided into 3 parts, which are 1) speech source, 2) resonance and 3) lip radiation (Fant, 1960). Tones depend on speech source, since the fundamental frequency is an acoustical parameter of speech source in time domain. As is well known, there are many other parameters of speech source, such as phonation parameters, which can reflect very important nature of speech source and are significantly important in speech perception and synthesis.

The phonation feature of glottal source in Mandarin has been studied in recent years. The patterns of tones and diatones through the signal of electroglottograph (EGG) have been established (Kong Jiangping, 1998). This pattern is modeled by the parameters of F0, open quotient (OQ) and speed quotient (SQ). The physiological phonation natures of the tones and diatones in Mandarin have been studied through the high-speed digital image system, by which the vibration of vocal folds have been observed and studied in quantity through the image signal processing and parameter extracting (Kong Jiangping, 2001). The phonation natures of glottal source have also been studied in time domain acoustically through the parameters extracted from glottal source of tones and diatones in Mandarin (Kong Jiangping, 2003, 2004).

The glottal excitation has been studied and modeled by many researchers. The classical models are the one-mass model and two-mass model. These models have been established theoretically. Acoustical models based on speech acoustics have also been developed. There are 6 models which are considered important. Among these models, the LF-model is the most popularly used one (G. Fant et al, 1985).

The models talked above are all established on the single peak pulse. According to the study on the changes of phonations in the tones and diatones in Mandarin, the glottal pulse can also be double peak pulse and tri-park pulse (Kong Jiangping, 2004). See Fig.1.



a b c

Fig.1: 'a' displays a typical single peak pulse. 'b' displays a typical double-peak pulse. 'c' displays a typical tri-peak pulse.

Persons all use single peak pulse in their speech. For some persons, especially the male voice with low pitch, double peak pulse and tri-peak pulse are used. From the viewpoint of tones, the double peak pulse usually appears at the beginning period of tone 2 and the ending period of tone 4, where the pitch is relatively low. The tri-peak pulse usually appears in the middle period of tone 3, where the pitch is very low and the phonation is creaky voice and vocal fry. This paper is concerned with the study on the natures of double peak pulse in frequency domain. In detail, we discuss the natures of sub-harmonic of the double peak pulse.

2. RESEARCH METHODS

The methods used in this research are: 1) extracting glottal source of tones by inverse filtering; 2) analyzing sub-harmonic and spectrum tilt by FTT; 3) simulating the double peak pulse by 4 functions and describing the natures of them in both time and frequency domains.

2.1 Method of glottal source extracting

The method of inverse filtering has 2 parts: 1) extracting the parameters of F0, formants and bandwidth by LPC; 2) inverse-filtering the speech sound through a group of filters. The LPC model is an 'AR (auto-regressive) model'. The inverse filtering used in this study is the cascade inverse filtering (Gold and Rabiner, 1968). The sampling rate of speech sound is 11025. There are 2 main steps in the inverse filtering: 1) initializing the 5 formants and 5 bandwidths by co-variance LPC; 2) modifying these parameters automatically or manually according to FFT spectrum till the speech glottal source are obtained.

2.2 Method of analyzing harmonic and spectrum tilt

In the study of glottal harmonic, we use FFT to analyze the glottal source, which has been extracted from the speech sound through inverse filtering.

In the study of the spectrum tilt of glottal source, we first use FFT to analyze the source, and then detect the maximum of each harmonic. After that, these parameters have been treated by polynomial curve fitting and then the function with curvature, slope and intercept is got. Then the spectrum tilt can be easily calculated.

2.3 Method of double peak pulse simulation

The double peak pulse is simulated by 4 functions of parabola. To change the parameters of these functions, different double peak pulses can be simulated. After the double peak pulses simulated, we use FFT to analyze the double peak pulses. Since the natures of these double peak pulses can controlled, the natures of sub-harmonic can be discussed through the analysis of the simulated double peak pulses. In this way, the sub-harmonic of double peak pulse of glottal source in Mandarin tones can be studied theoretically.

3. ACOUSTICAL ANALYSIS ON HARMONIC

In the acoustical analysis, two methods have been used. One is FFT and the other is spectrum tilt. The samples are the 4 basic tones in Mandarin. These samples are from the 800 persons' glottal source database. The sampling rate is 11025. The FFT point is 256. The samples shown below are from a male voice.

3.1 Harmonic analysis

Figure 2 shows the spectrum of the middle period of tone 1, where the glottal pulse is single peak pulse. From the spectrum, we can see that harmonic is regular and clear, though there are still effects of resonance, which can be found in the figure. The unit of 'x' axes in the figures of this sub-section is 'Hz'. The unit of 'y' axes in the figures of this sub-section is 'dB'.

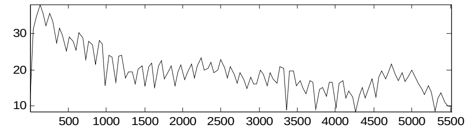


Fig. 2: The spectrum of the middle period of tone 1.

Figure 3 shows the spectrum of the beginning period of tone 4, where the glottal pulse is single peak pulse. From the spectrum, we can see that harmonic is regular and clear, though there are still the effects of resonance, which can be found in the figure.

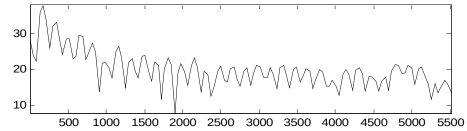


Fig. 3: The spectrum of the beginning period of tone 4.

Figure 4 shows the spectrum of the beginning period of tone 2, where the glottal pulse is double peak pulse. From the spectrum, we can see that harmonic is not regular and clear. There are great effects of sub-harmonic, but it is little difficult to see the sub-harmonic clear.

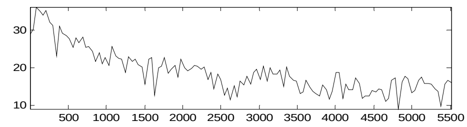


Fig. 4: The spectrum of the beginning period of tone2.

Figure 5 shows the spectrum of the ending period of tone 4, where the glottal pulse is double peak pulse. From the spectrum, we can see that harmonic is not regular and clear. There are great effects of sub-harmonic. Although it is not very clear, the two different harmonics can be seen.

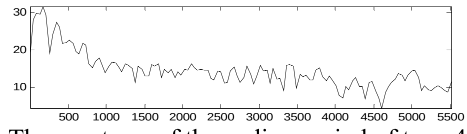


Fig. 5: The spectrum of the ending period of tone 4.

Figure 6 shows the spectrum of the middle period of tone 3, where the glottal pulse is tri-peak pulse. From the spectrum, we can see that harmonic is not regular and clear. There are great effects of sub-harmonic. From this figure the sub-harmonic can be easily seen.

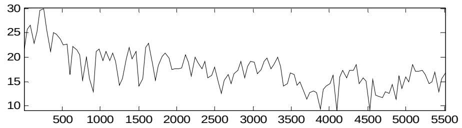


Fig. 6: The spectrum of the middle period of tone 3.

To compare these figures, we can see that the harmonic of high frequency appears in the period of tones whose F0 is low, such as the beginning period of tone 2, the ending period of tone 4, and the middle period of tone 3. After the analysis, we can say that double peak pulse do produce sub-harmonic, but it is difficult to explain them theoretically.

3.2 Spectrum tilts of 4 basic tones

In this section, the spectrum tilts of the 4 basic tones in Mandarin are analyzed and studied. The FFT point used in the spectrum tilt analysis is 255 and the sampling rate is 11025. The parameters of the spectrum tilt are extracted with pitch synchronization.

Figure 7 shows the spectrum tilt of tone 1. From this contour, it is difficult to find the pattern, because the contour is complex. If we take it as a whole, we can found that the contour is 'level'. The unit of 'x' axes in the figures of this sub-section is 'period'. The unit of 'y' axes in the figures of this sub-section is 'dB'.

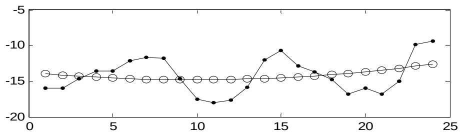


Fig. 7: The spectrum tilt contour of tone 1

Figure 8 shows the contour of spectrum tilt of tone 2. From this contour, we can found that the contour is in a pattern of 'Rising'.

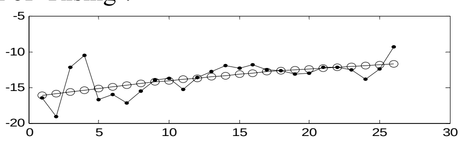


Fig. 8: The spectrum tilt contour of tone 2.

Figure 9 shows the contour of spectrum tilt of tone 3. From this contour, we can found that the contour is in a pattern of 'Rising-falling'.

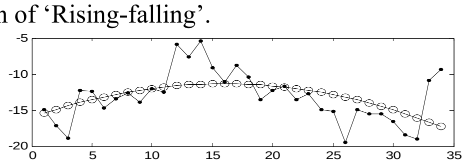


Fig. 9: The spectrum tilt contour of tone 3.

Figure 10 shows the contour of spectrum tilt of tone 4. From this contour, we can found that the contour is in a pattern of 'Rising-falling' or 'Falling'. The spectrum tilt pattern of tone 4 looks a little bit like that of tone 3. The values of them are different.

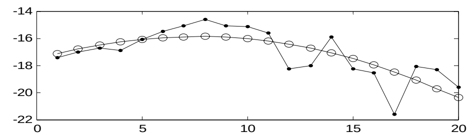


Fig. 10: The spectrum tilt contour of tone 4.

According the analysis above, it can be found that the spectrum tilts in different tones are very complicated. We can not clearly see the relationships of these spectrum tilts and the other phonation parameters.

4. NUMERICAL SIMULATION

In order to dig out the nature and relationship of harmonic and sub-harmonic in one period, we simulate two double peak pulses. The construction of the two double peak pulses is based on the analysis of real samples. The following are the basic parameters. The two double peak pulses are all constructed by 4 half parabolas. The sample rated is 10000. The parabola is normalized between 0 and 1. The period is constructed with 100 point, so the F0 is 100 Hz. The first, second and third half parabolas all have 20 points. The fourth half parabola has 40 point. The first and second parabolas construct the first peak and the third and fourth construct the second peak.

4.1 Sub-harmonic of the first double peak pulse

For the first sample, the magnitude of the first period is 1.0, the magnitude of the second period is 0.2, the magnitude of the third period is 0.1 and the magnitude of the fourth period is 0.8. See Fig. 11. The unit of 'x' axes is 'point' in figure 11 and 14, and is 'Hz' in figure 12, 13, 15 and 16. The unit of 'y' axes is 'magnitude' in Fig.11 and 14, and is 'dB' in Fig. of 12, 13, 15 and 16.

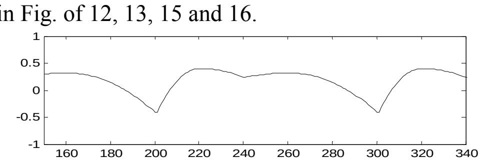


Fig. 11: A typical double peak pulse.

Figure 12 shows the spectrum of the first sample. The FFT point is 256. From this figure, we can see the harmonic and the sub-harmonic, though the sub-harmonic is not very prominent. Theoretically speaking, the fundamental frequency of sub-harmonic is 250 Hz. We can see that there are 4 sub-harmonics within the frequency of 1000 Hz. The spectrum tilt is -22.15dB/oct.

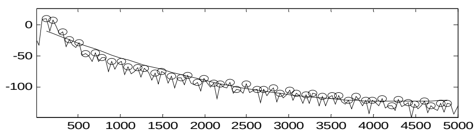


Fig. 12: The spectrum of the first double peak pulse analyzed by 265 point of FFT.

Figure 13 shows the spectrum of the first sample. The FFT point is 512. From this figure, we can see the harmonic, which is 100 Hz, so there are 10 harmonics within the frequency of 1000 Hz. The sub-harmonic, though it is not very prominent, can also be seen from the values of the maximum of each harmonics. The spectrum tilt is -9.03dB/oct.

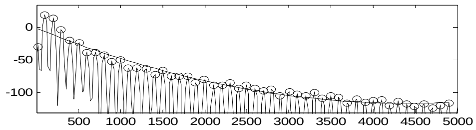


Fig. 13: The spectrum of the first double peak pulse analyzed by 512 point of FFT.

4.2 Sub-harmonic of the second double peak pulse

For the second sample, the magnitude of the first period is 1.00, the magnitude of the second period is 1.0, the magnitude of the third period is 0.9 and the magnitude of the fourth period is 0.9. See Fig. 14.

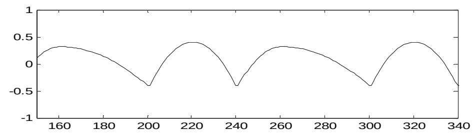


Fig. 14: The second double peak pulse.

Figure 15 shows the spectrum of the second sample. The FFT point is 256. From this figure, we can see the sub-harmonic very clearly. Theoretically there should be 4 harmonics within 1000 Hz, because the first pulse in the period has 40 points. The frequency of sub-harmonic should be 250 Hz. Now we can see that there are 4 sub-harmonics within the frequency of 1000 Hz. The spectrum tilt is -24.35dB/oct.

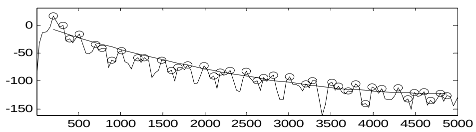


Fig. 15: The spectrum of the second double peak pulse analyzed by 256 point of FFT.

Figure 16 shows the spectrum of the second sample. The FFT point is 512. From this figure, we can see that the harmonic of F0 is very clear. There are 10 harmonics within 1000 Hz. The sub-harmonic can also be seen very clearly from the values of maximum of each harmonic. This result accords with the theory of speech acoustics. The spectrum tilt is -8.50dB/oct.

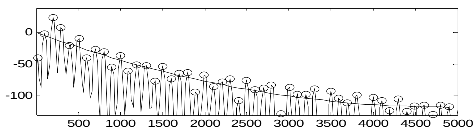


Fig. 16: The spectrum of the second double peak pulse analyzed by 512 point of FFT.

5. CONCLUSIONS AND DISCUSSION

According to the results of this research, there are 3 conclusions: 1) the double peak pulse can produce sub-harmonic in glottal source of Mandarin tones; 2) sub-harmonic influences the spectrum tilt of glottal source; 3) the double peak pulse can be simulated and modeled acoustically and mathematically.

As is well known, the acoustic parameters are the surface form of the speech in the study of speech acoustics and phonetics. The most important thing is how to connect the acoustical parameters to the perception of speech source and even the perception of tones in Mandarin. From this research, we can see that glottal source may have different source qualities of their own, which may arouse difference in the perception of tones. So this research shows us a new aspect in the study of glottal source in Mandarin tones.

6. REFERENCES

[1] B. Gold And L.R. Rabiner, "Analysis of digital and analog formant synthesizers," IEEE Trans. Audio Electroacoust. AU-16, 81-94, 1968.

[2] G. Fant, "Acoustic Theory of Speech Production", Mouton, The Hague, 1960.

[3] G. Fant, J. Liljencrants and Q. Lin, "A four parameter model of glottal flow", STL-QPSR, No. 4, 1985, pp 1-13, 1985.

[4] Kong Jiangping, "EGG model of ditoneme in Mandarin", Proceedings of 5th International Conference on Spoken Language Processing. Tu5P16, Sydney, Australia, 1998.

[5] Kong Jaingping, "Study on Dynamic Glottis through High-speed Digital Image", PhD. Dissertation, City University of Hong Kong, 2001.

[6] Kong Jiangping, "Study of phonation features of tones in Mandarin", Proceedings of 15th International Congress of Phonetic Sciences, 2003.

[7] Kong Jiangping, "Phonation models of tone and diatone in Mandarin", International Symposium on Tonal Aspects of Languages (TAL'2004), Beijing, China, 2004.

The research is funded by National Natural Science Foundation of China (No: 10274105).