TIME SERIES ANALYSIS OF JITTER IN SUSTAINED VOWELS

Li Dong

Department of Chinese Language and Literature, Peking University, Beijing, China
dongli_pku@yahoo.com.cn

ABSTRACT

This paper proposes a new equation to show the time domain features of jitter: Jitter = $(T_{i+1}-T_i)/T_i$. The subjects are tonal language speakers. Jitter is extracted from EGG signals. The results show that: (1) jitter does not obey Gaussian distribution but multimodal distribution; (2) the first peak position near zero of the jitter distribution curve increases with the pitch; (3) each peak position is an integral multiple of the first peak position; (4) jitter values are around 0, and in most cases, adjacent values have opposite algebraic sign; (5) the zero value percentage and zero-crossing rate of jitter increase with the pitch; (6) the variance of pitch keeps in the allowance of just noticeable difference, which can be attributed to the effect of audible feedback, and it affects the variance of jitter further.

Keywords: jitter, time domain features, distribution

1. INTRODUCTION

Jitter involves small fluctuations of the glottal cycle lengths. P Lieberman was the first to compare the small cycle-to-cycle fluctuations in the fundamental periods of healthy and diseased larynxes [4]. From then on, most studies of jitter focus on using jitter value to distinguish between healthy voice and pathological voice. Another function of jitter is to make the synthesized voice much livelier. Previous studies of jitter can be divided into two categories, mean analysis and time series analysis. Studies using time series analysis method have a small number. Because in time domain, each pitch value is around the mean value, most studies use Gaussian distribution to simulate the distribution of jitter [1]. In [4, 5], jitter value was broken down into predictable and random components.

This article focuses on the time domain features of jitter, and a new equation is proposed to show them. The subjects are native Chinese speakers, for previous studies chose non-tone languages speakers whose jitter may be different from tone languages speakers. In this article, the distribution of jitter is discussed, and jitter values are broken down into predictable and random components.

Previous studies used different equations to calculate jitter. Some couldn't reflect time domain features; some were affected by the pitch; some used the mean value on the local scale. Therefore, a new equation was proposed to show the time domain features of jitter.

$$(1)\ Jitter = \frac{T_{i+1} - T_i}{T_i}$$

In Eq. (1), T_i is the length of the ith pitch period. This equation is free from the pitch effect and can show the local fluctuations exactly.

2. METHODS

Two male subjects and two female subjects participated the recording. They are all college students and the average age is 25. The recordings were made in a sound-proof room. The sample rate is 44.1 kHz and the resolution is 16 bit. The left channel is speech signal and the right channel is EGG signal. Jitter is very small, so that the sample rate is very important. In [1], the average jitter extracted in 40 kHz is close to that extracted in 80 kHz. So the jitter value in this article is reliable.

Some recorded sounds were played back to the subjects and they pronounced with the same pitch. The target frequencies are 95Hz, 113Hz, 131Hz, 149Hz and 167Hz for male subjects and 180Hz, 210Hz, 240Hz, 270Hz and 300Hz for female subjects. Before the test, the subjects' lowest and highest sounds were recorded, and then each step's value was calculated. Each subject pronounced /a/, /i/, /u/ in each frequency twice.

Jitter was extracted from EGG signals. EGG signals are more pure than speech signals, for speech signals contain more information such as formant information. The differentials of the EGG signals were calculated and each peak of the differentials was marked as the initial point of each cycle. Then jitter sequences were calculated.

3. RESULTS

3.1. The distribution of jitter

Jitter does not obey Gaussian distribution but multimodal distribution. The peaks of the multimodal distribution are the foundation of producing jitter. Figure 1 is a typical distribution of jitter. It is /a/ in 270Hz pronounced by a female subject, and it is multimodal distribution. The distribution is symmetrical and the occurrence number peaks when jitter value is equal to zero. To show the distribution clearer, Figure 2 uses the same data as Figure 1 and transforms the scaling to nonmetric order.

Figure 1: The distribution of jitter.

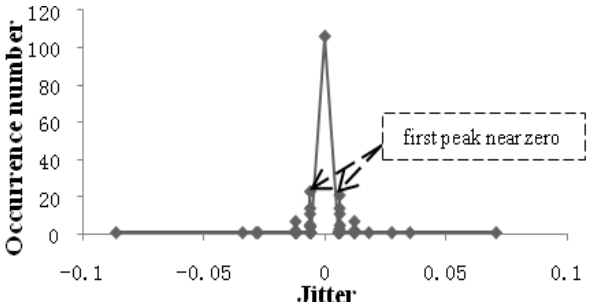
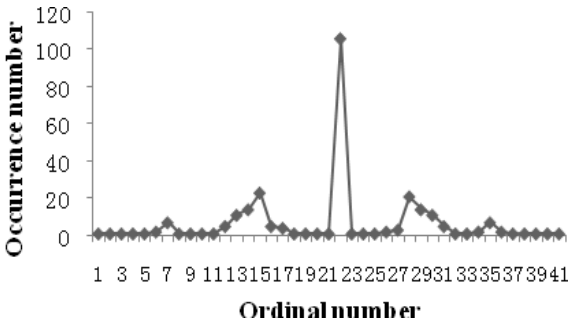
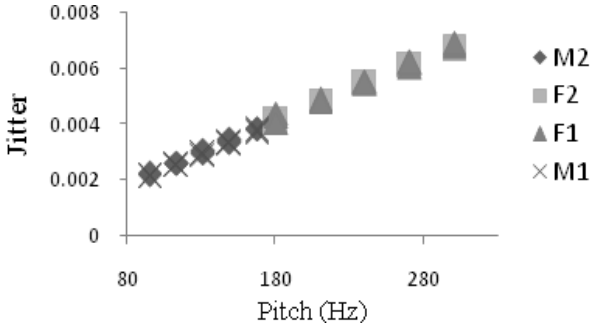


Figure 2: The transformation of the distribution of jitter.



Jitter is related to pitch. The number of the peaks is irrelevant to pitch, but the first peak position near zero (the absolute value of the abscissa value of the first peak position near zero) is positively correlated with pitch. That is to say, the first peak position near zero increases with the pitch. So jitter calculated through Eq. (1) is relative that the positive correlation is not simply caused by the increase of the minuend and subtrahend. Figure 3 shows the first peak positions near zero of /a/, /i/, /u/ in five frequencies of four subjects. Because the distributions are symmetrical, the average absolute value of the left and right first peak positions near zero was used to plot Figure 3. Figure 3 shows that the first peak position near zero is positively correlated with the pitch, and it is not affected by vowels and subjects. Not only that, the slope of the fitted strait line equals to the inverse of the sample rate. Look back to Eq. (1). Then a conclusion was drawn that T_{i+1}-T_i equals to the inverse of the sample rate in most cases no matter how high the pitch is.

Figure 3: The first peak position near zero of the distribution of jitter.



Another result is that each peak position is an integral multiple of the first peak position. Table 1 shows each peak's position and the occurrence number in the jitter distribution of /a/ in 270Hz pronounced by a female subject. The second peak position is twice the first peak position and the third peak position is three times it. Though the left and right fourth peak positions are not equal, they are all at integer multiples of the first peak position.

Table 1: The jitter distribution of /a/ in 270Hz.

Jitter value	The occurrence number
-0.509	1
-0.307	1
-0.093	1
-0.077	1
-0.018	2
-0.012	14
-0.011	1
-0.006	74
0	88
0.006	73
0.012	10
0.018	1
0.019	1
0.036	1
0.103	1
0.635	1

3.2. The time domain features of jitter

Figure 4 shows the time domain features of jitter of /a/ in 300Hz pronounced by a female subject and Figure 5 is the features of /a/ in 95Hz pronounced by a male subject. The ordinate represents jitter value and the abscissa represents the ordinal of the

cycles. In these two Figures, all jitter values are around 0, and in most cases, adjacent values have opposite algebraic sign. In Figure 4, most jitter values are near 0.0068; in some cycles, the jitter values jumped to about 0.0137 in random. In Figure 5, the jitter values are irregular compare to Figure 4. But, adjacent values still have opposite algebraic sign in most cases. By observing all the data, the pitch instead of gender is the key factor that made the curves change.

Figure 4: The time domain features of jitter of /a/ in 300Hz.

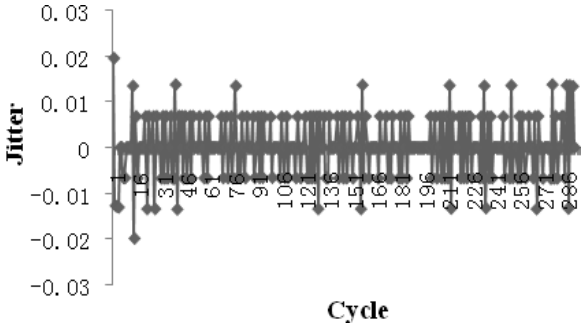
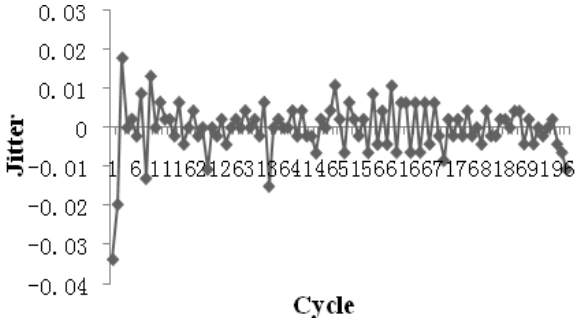


Figure 5: The time domain features of jitter of /a/ in 95Hz.



Two parameters are introduced to show the trends of jitter varying with the pitch. The zero value percentage is the percentage of zero of all the jitter values. The zero-crossing rate of jitter is the proportion of the number that adjacent values have opposite algebraic sign of all the cycles. Figure 6 and 7 show that the zero value percentage and zero-crossing rate of jitter increase with the pitch (A few exceptions on the local scale). The range of zero value percentage is from 10% to 50% and the range of zero-crossing rate of jitter is from 0.6 to 1.

Figure 6: The zero value percentage.

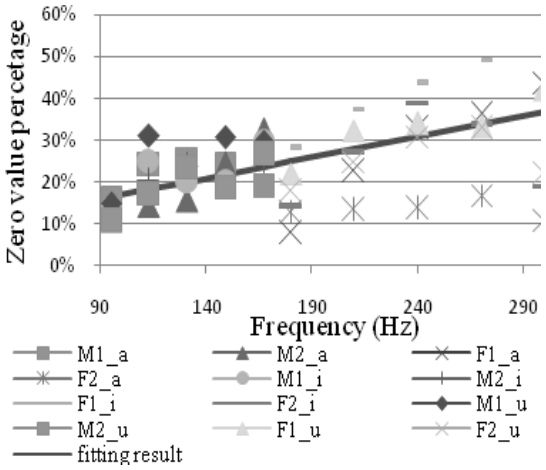
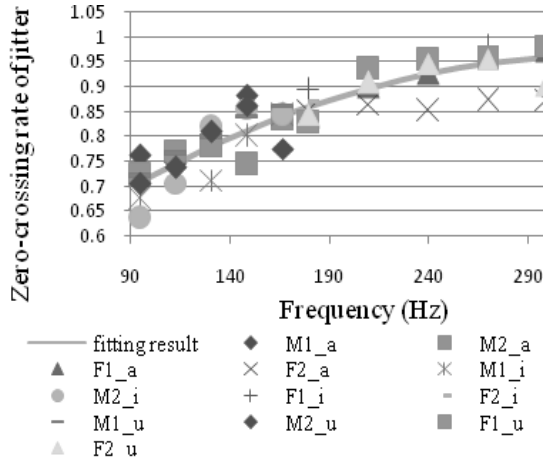


Figure 7: The zero-crossing rate of jitter.



3.3. Jitter and audible feedback

The pitch changes with the jitter, and the jitter's change is limited by the pitch. Figure 8 is the F0 of /i/ pronounced by male subject and Figure 9 is the F0 of /i/ pronounced by female subject. These two F0 curves seem to be limited by two boundaries. When the F0 increases to the upper bound, the F0 starts to decrease. When it decreases near the lower bound, it starts to increase. It can be attributed to the effect of audible feedback. A simple example is that if one can't hear his own voice when he sings, he will be off key. Similarly, when one speaks, there are target tones in one's brain. So when the pitch drifts too far from the target value, the speaker will adjust the jitter value. It can explain why the synthetic voice is unnatural, even if synthetic jitter is small, but the voice is natural, even if the natural jitter is large in some cycle. Just noticeable difference (JND) is a complex variation. Klatt [3] indicated that the subjects could detect a

change of 0.3 Hz in a constant F0 contour when F0 = 120 Hz, but the JND is an order of magnitude larger (2.0 Hz) when the F0 contour is a linear descending ramp (32 Hz/sec). So JND varies with pitches. JND is also different between tone language speakers and non-tone language speakers; because as Wang [6] showed that, for the previous, the perception of tone is category perception that their JND may be larger.

Figure 8: F0 of /i/ pronounced by male subject.

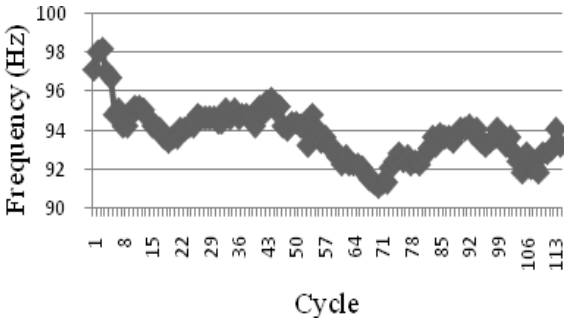
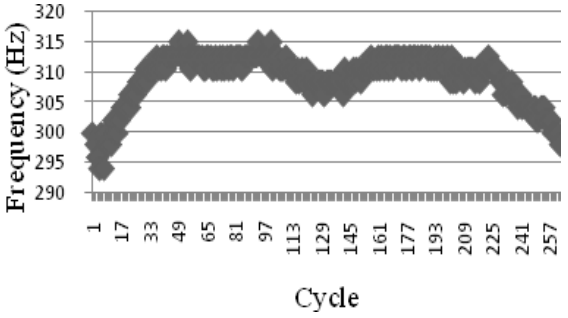


Figure 9: F0 of /i/ pronounced by female subject.



4. CONCLUSION

Jitter value can be broke down into predictable and random components. The predictable components are that: all jitter values are around 0 and in most cases, adjacent values have opposite algebraic sign; the zero value percentage and zero-crossing rate of jitter increase with the pitch; the first peak position of the distribution of jitter increases with the pitch; each peak position is an integral multiple of the first peak position; the variance of pitch keeps in the allowance of just noticeable difference and it affects the variance of jitter further. The random components are that: the zero value position; the number of the peaks of the multimodal distribution; the position that jitter value jumped suddenly.

The analyses above are not conflict with previous study. In [2], jitter in percent (Jitt) was used to explore the relationship between jitter and pitch. The jitter in percent defines the relative variance of periods in a voiced speech sound as:

$$(2)\ Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N}T_0^{(i)}}$$

Where i=1, 2, 3, ... N of $T_0^{(i)}$ is the parameter for extracting period and N is the number.

In [2], Jitt of /a/ decreases along with the increase of pitch. The Jitts of /i/, /u/ and /m/ decrease first and then increase a little along with the increase of pitches. On one hand, this is contrary to the trend of the first peak position of the distribution of jitter. On the other hand, the zero value percentage increasing with the pitch makes the sum of $|T_0^{(i)} - T_0^{(i+1)}|$ decrease with the pitch when the other condition is changeless. The trend of Jitt is caused by many factors. Time series analysis of jitter can help explore more detailed information.

5. ACKNOWLEDGEMENTS

This research was funded by China Social Sciences Funds of the Ministry of Education (Grant 10JJD740007) and China National Natural Science Funds (Grant 61073085). I am grateful to two anonymous reviewers for the comments on the earlier version of the paper.

6. REFERENCES

[1] Heiberger, V.L., Horii, Y. 1982. Jitter and shimmer in sustained phonation. *Speech and Language: Advances in Basic Research and Practice* 7, 299-332.

[2] Jiangping, K. 2008. A study on jitter, shimmer and F0 of Mandarin infant voice by developing an applied method of voice signal processing. *2008 Congress on Image and Signal Processing* 5, 314-318.

[3] Klatt, D.H. 1973. Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *Journal of the Acoustical Society of America* 53(1), 8-16.

[4] Schoentgen, J., Guchteneere, R.D. 1994. Time series analysis of jitter. *Journal of Phonetics* 23(1-2), 189-201.

[5] Schoentgen, J., Guchteneere, R.D. 1997. Predictable and random components of jitter. *Speech Communication* 21, 255-272.

[6] Wang, W.S.Y. 1976. Language change. In Harnad, S.R., Steklis, H.D., Lancaster, J. (eds.), *Origins and Evolution of Language and Speech*. New York: New York Academy of Sciences, 280.