ISCA Archive
http://www.isca-speech.org/archive

Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 1999)
Firenze, Italy, September 1-3, 1999

SHARP TUNING IN OVERTONE SINGING BY EFFECTIVELY EMPLOYING ANTI-RESONANCES

Yuki Kakita, Kanazawa Institute of Technology, Kanazawa-Minami, 921 Japan.
kakita@his.kanazawa-it.ac.jp, kakita@mail.shs.ohio-state.edu

Osamu Fujimura, Dept. of Speech and Hearing Science, Ohio State University, 1070 Carmack Rd., Columbus, OH43210, U.S.A..
fujimura.1@osu.edu

Abstract

In overtone singing, a melody is produced by selecting an appropriate voice harmonic, one after another, while keeping the fundamental frequency constant. A specific harmonic is selected by changing the shape of the oral cavity which functions as a resonator corresponding to the formant used in vowel production. In addition, attenuation of particular frequency components by antiresonance of the vocal tract seems to occur. Frequently, such suppressed spectral components are observed in the spectrogram, usually in the frequency region immediately above the selected frequency for the tune and sometimes, in skilled performers, in a lower region also. This suppression, presumably associated by additional poles coupled with the zeroes introduced, may perceptually enhance the tuned component. We suggest two hypotheses about the mechanism to produce the pole-zero pairs: one employs an acoustic branch cavity, and the other assumes nonlinear generation of additional sound source at the vocal tract constriction due to positive feedback involving turbulence and possibly wall vibration.

Introduction

The overtone singing is performed in Mongolia, Tuva, and a number of areas of the Eurasian continent. A melody is produced by selecting an appropriate voice harmonic, one after another in temporal sequence, while keeping the fundamental frequency constant. A specific harmonic is selected by changing the shape of the oral cavity which functions as a resonator corresponding to the formant used in vowel production [1]. Frequently, also, obviously suppressed spectral components are observed in the spectrogram, usually in the frequency region immediately above the selected frequency for the tune and sometimes, in skilled performers, in a lower region also. This suppression, presumably caused by zeros associated by additional poles of the vocal tract transfer function, may perceptually enhance the tuned component by increasing saliency due to the contrast of neighboring frequency components.

Signal Analysis

We identified the spectral zeros (anti-resonances of the vocal tract acoustic filter) by a simple method of zero-detection of the LPC residual signal.

Assuming that a speech signal is generated by an all-pole filter with a voice source signal as input, LPC analysis with an appropriate set of parameters can subtract the effect of poles from the speech output signal. The residual signal is supposed not to contain any effective poles. If there are no zeros either, the residual should have a flat spectral envelope. The sound spectrogram of the residual signal then should be a monotonic function of frequency because the radiation characteristics are monotonically increasing frequency function, roughly 6 dB per octave. If there are zeros in particular frequency region of the residual signals, the spectrum should show characteristic valleys [2,3]. The spectrographic analysis also has the advantage that we can visually trace the spectral change of the residual signal in time, in comparison with the spectrogram of the corresponding original speech signal.

Sound signals were digitized (Sampling frequency: 44.1 kHz, Digitization: 16 bit) in a personal computer from a music CD [4] and also from a microphone recording of some performance by one of the authors was used for this study. SoundScope/16 (GW Instruments) was used for the spectrographic analysis and the computation of the residual signal.

Observation

Before applying the zero detection by spectrographic examination of the LPC residual signal, we carefully adjusted the LP parameters, i.e. (a) number of conjugate poles, and (b) the LP analysis time window.

We also applied the same method to human oral vowels, and synthetic vowel signals using a software synthesizer and an acoustic model [5]. We verified that the residual signals of human and synthetic vowels were flat showing no spectral valleys. Figure 1 shows an example of the synthetic vowel obtained by the synthesis software SenSyn (Sensimetrics Corp.). The sound example whose spectrogram is shown in Fig. 1 is a neutral vowel generated by formants (poles) only. The flat spectrum of the residual signal indicates no spectral valleys.

Figure 2 shows an example of Sygyt (a tune with sharp tones in overtone singing, a portion of the 13th tune of Ref. 4). Note that the magnitude of the tune components as marked in the spectrogram exhibits changes in tune (Indicated by arrows with a label $\underline{T}$). In the residual spectrogram, we can observe the valleys (suspected zeros indicated by $\underline{Z_H}$). This occurred in the region above the fundamental frequency of the tune. In some portions, valleys are observed in frequencies below the tune component (indicated by $\underline{Z_L}$).

Discussion

Generation of zeros: theoretical interpretation

The physical mechanism of the generation of zeros are of particular interest, given that the corresponding speech signals, presumably with a similar articulation, do not exhibit such zeros. Possible explanations are: (1) branch cavity, (2) additional sound source near the vocal tract constriction, or both.

(1) Branch cavity

There are published tips for producing overtone singing which suggests the first mechanism:

> <<Divide the mouth into two similar-sized compartments by raising your tongue so that it meets the roof of

your mouth, a bit like you're saying "L".

1. Spread your tongue a bit so that it makes a seal all the way round. At this point, you won't be able to pass air through your mouth.

2. Then, break the seal on the left (or right) side of the mouth, simply to provide a route for the air to get through.>> [6]

These tips suggest that a branch cavity may be the cause of a created pole-zero pair just as in nasalization due to opened nasal coupling at the uvula [7, 8]. This mechanism accounts for the antiformants (zeros of the vocal tract transfer function) of lateral and nasal consonants as well as nasalized vowels [7, 8, 9, 10, 11].

(2) Additional source

Based on the acoustic theory of speech production the vocal tract transfer characteristics can be effectively approximated by an all-pole model if the tract has no branch and when the sound source of voice excitation is located at the innermost end of the tube, namely at the glottis. For the same tube, however, when the sound source is located between the glottis and the mouth opening, the transfer function contains zeros as well as poles. [10, 11, 12]

If a cavity (as part of the vocal tract) exists behind the signal source, then this acts as a back branch for the signal transmitted from this secondary source to the mouth opening exhibiting antiresonance due to the resonance of the back cavity. This output signal is superimposed onto the usual speech signal transmitted from the glottal source to the mouth opening given the same vocal tract acoustic tube.

This is a somewhat novel situation for a totally quasi-periodic signal, if our speculation is correct, , while a similar situation for turbulence noise generation at the vocal tract constriction is well known for voiced fricatives. How would the secondary source occur in overtone singing? The supra-laryngeal source, if it is periodically synchronized with the vocal fold vibration, must be created by some nonlinear interaction between the alternating air flow at the constriction due to the glottal source and some turbulence generated at the constriction. Note that this interaction is not amplitude modulation of the turbulent signal as in voiced fricatives, since the overtone signals are approximately periodic. It calls for a novel nonlinear mechanism that creates a periodic pressure source that is synchronized with the volume velocity fluctuation due to vocal fold vibration.

In the linear acoustic theory, there is no possibility of generating a second sound source, however sharp the resonance may be. When a turbulence or chaos is created by the modulated airflow, and if there is a sharp resonance with a velocity loop at the constriction causing high velocity air-particle movement, there can be a positive feedback process resulting in a nonlinear amplification of the signal. The energy source is the pulmonary DC airflow just as in the case of vocal fold vibration. At the moment, this is solely a speculation based on the observation of the spectral zeros that seem to correspond to the back cavity configuration. The compliant tongue surface serving as a movable wall of the tube may also play a role in this sound generation or enhancement mechanism. In any case, the additional sound signal created at the constriction, just like the turbulent noise, must suffer from frequency selective signal absorption due to back/side branch resonance modes with their velocity loop at the constriction.

Summary

In overtone singing, a specific harmonic is selected by a resonator corresponding to the vowel formant [1]. We observed spectral zeros which may perceptually enhance the tuned component. We proposed two hypotheses about generating zeros: one employs an acoustic branch cavity, and the other assumes nonlinear generation of additional sound source.

References

[1] Y. Kakita, Theoretical consideration on the musical scale of Xoomij singing based on the speech production process, J. Acoust. Soc. Am., 104, p.1768 (1998)(A). ASA Web site: http://www.acoustics.org/136th/kakita3.htm

[2] M. V. Mathews, J. E. Miller, and E. E. David, Strategies for automatic pole-zero analysis of speech, Proc. Speech Communication Seminar, Stockholm 1962, Quarterly Progress and Status Report, Royal Inst. Tech., Speech Transmission Labs, 1963, Vol. 1, 1-7 (B10)

[3] B. S. Atal and M. R. Schroeder, Linear prediction analysis of speech based on a pole-zero representation, JASA 64, 1310-8.

[4] Huun-Huur-Tu: The Orphan's Lament, SHANACHIE 64058. (Music CD)

[5] Y. Kakita: Acoustic Analysis of Speech and Sound, Laboratory Textbook 19-1, Unpublished laboratory textbook, (Kanazawa Institute of Technology, Kanazawa, Japan, 1997

[6] Web site on khoomei: http://www.obsolete.com/pipe/oldpipe/khoomei.html

[7] R. D. Kent and C. Read: The Acoustic Analysis of Speech, (Singular, San Diego, CA, 1992) pp.130-136.

[8] O. Fujimura & J. Lindqvist: Sweep-tone measurements of vocal tract characteristics, J. Acoust. Soc. Am. 49, pp.541-558 (1971).

[9] M. Halle and K. N. Stevens: The postalveolar fricatives of Polish, In S. Kiritani et al. (Eds.), Speech Production and Language: In Honor of Osamu Fujimura, (Mouton de Gruyter, Berlin, 1997) pp. 177-193.

[10] G. Fant: Acoustic Theory of Speech Production, (Mouton, The Hague, 1960).

[11] J. Flanagan: Speech Analysis, Synthesis, and Perception, Second Ed. (Springer-Verlag, New York, NY, 1972).

[12] E. A. Guillemein, Synthesis of passive networks: theory and methods appropriate to the realization and approximation problems. (Wiley, New York, 1957).

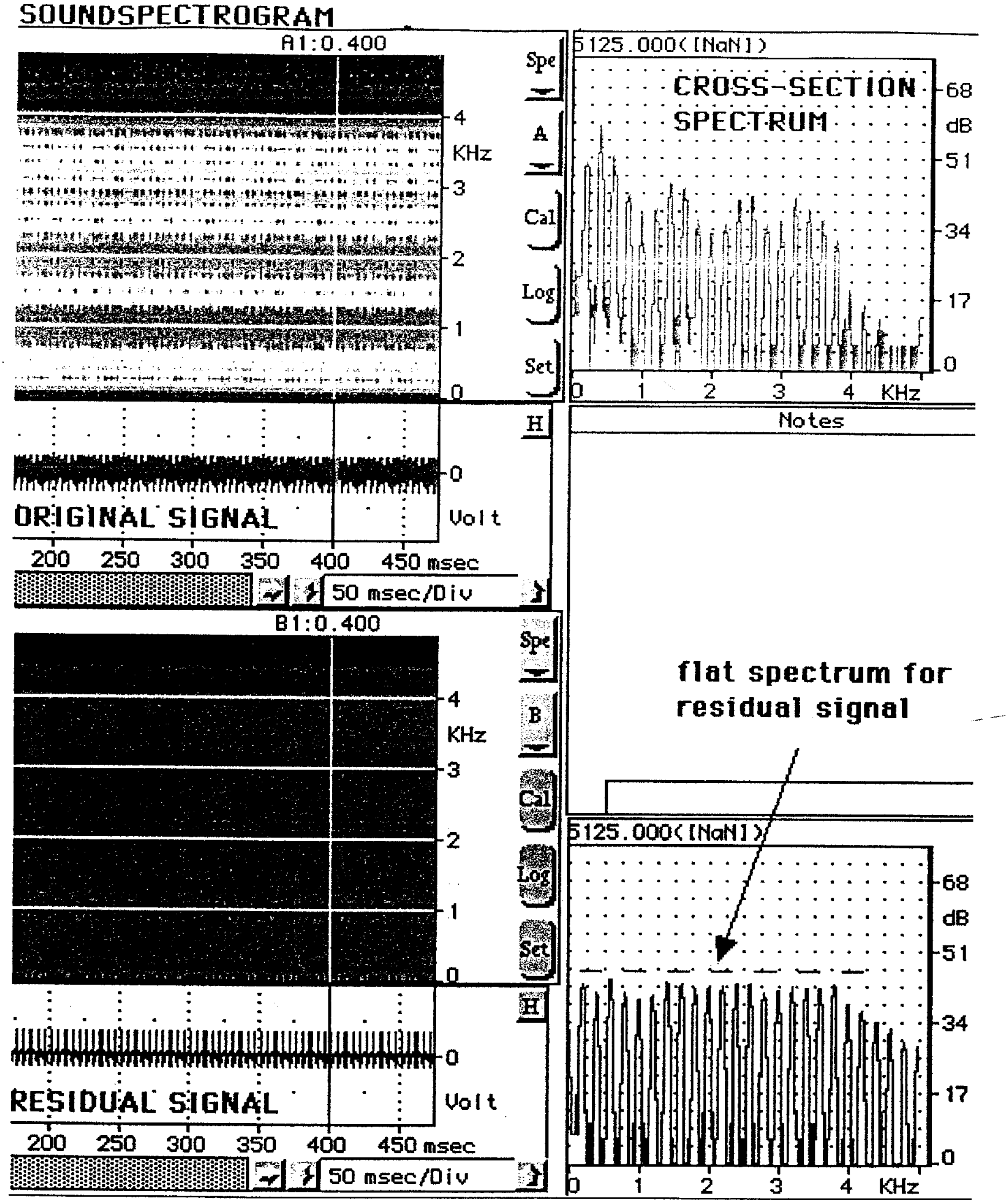

Fig. 1 An example of spectrogram and cross-section spectrum of a vowel like sound generated by the synthesizer. Lower displays are for the LPC residual signal. Notice that the residual signal shows flat with no valleys.

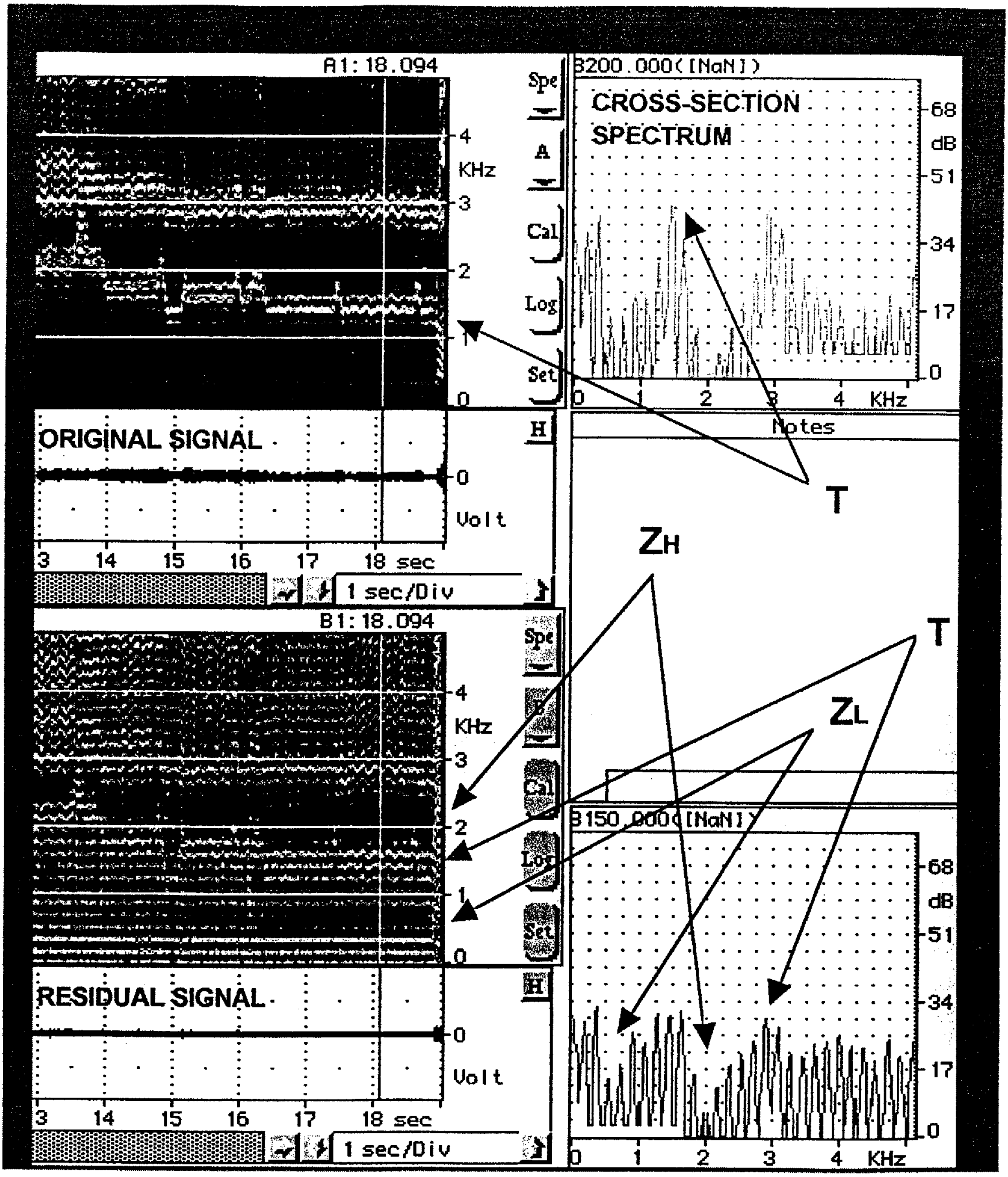

Fig. 2 An example of spectrogram and cross-section spectrum of Sygyt. Lower displays are for the LPC residual signal.