3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

Synthesis and Analysis of Vocal Source with Vibration of Larynx

Nobuhiro Miki, Naohisa Kamiyama, and Nobuo Nagai
Research Institute for Electronic Science
Hokkaido University
N-12 W-6, Kita-ku, Sapporo 060 Japan
Email : miki@elsip.hokudai.ac.jp

10.21437/Eurospeech.1993-118

Abstract

In this paper we show a speech production model of the vocal tract analogue, which can simulate the propagating mechanical vibrational wave on the tract wall, synthesize vocal sounds with the vibrational wave, and analyze the spectral properties for the synthesized waves. We propose a new lattice filter model for the effect of wall vibration in the larynx, and show that the sound propagation is influenced with the vibration in the result of simulation.

1 Introduction

The vocal tract analog have been used for the simulation of speech synthesis in long time, in which the lung, trachea, vocal cords, vocal and nasal tract, and radiation at lips also are modeled as equivalent circuits. In the recent speech synthesizer, LPC analysis approach is widely used, but the source model is separated from the model of vocal tract or the other parts, and the source is determined by the experimental data from LPC analysis. In this method it is difficult to realize the speech synthesis reflecting the personal feature in the dynamics of articulation.

On the other hand the vocal tract analog have the ability for modeling of the source - tract interaction, the dynamics of articulation, and with the physical parameters of human body. Only the problem is computational cost of simulation in the vocal tract analog model, but from a technological point of view, it should be noted that recent progress of IC technologies must bring the realization of a synthesizer based on these models in the near future.

In this paper we show a speech production model of the vocal tract analogue, which can simulate the propagating mechanical vibrational wave on the tract wall, synthesize vocal sounds with the vibrational wave, and analyze the spectral properties for the synthesized waves. Since the vibrational wave is caused by the mechanical vibration of the vocal cords, this is computed with the two-mass model of vocal cords, and in this simulation some propagation parameters of the wall are based on physical constants. In the wall tissue, the propagational speed and the attenuation constant are influenced with the physical constants.

We propose a new lattice filter model for the effect of wall vibration in the larynx, and show that the sound propagation is influenced with the vibration in the result of simulation.

2 A Model of Wall Vibration (MWV)

Since speech signal processing has been treated as a linear processing of the system, it would seem that the linear circuit modeling is enough for the acoustic propagation in the vocal tract. However, it is difficult to assume that the wall of near the larynx is rigid, since the wall should be assumed as a soft tissue, we should consider a nonlinear influence of the mechanical vibration of the wall caused by vibrating vocal cords. In Fig. 1 we show a schematic representation for a model of mechanical vibration of larynx.

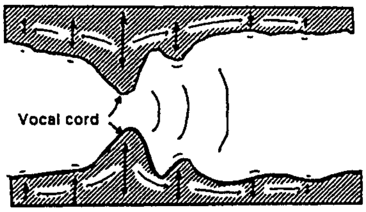


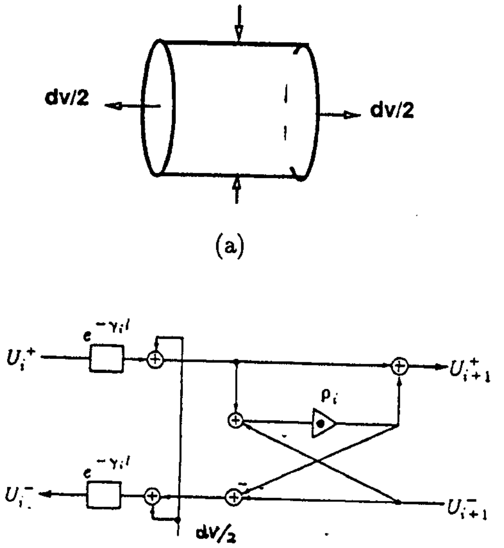
Fig. 1 Model of mechanical vibration of larynx

By the movement of the vocal cord the vibration is propagated in the larynx wall as the elastic wave, and this wave decays gradually in the distance from the vocal cord. Consequently, the cross section of the larynx and the vocal tract is varied with the mechanical vibration, and this vibration influences the air volume in the tract and the acoustic propagation also. In this interaction of the mechanical vibration and the acoustic propagation, since the effect has a nonlinear property, it is difficult to built in the linear-separable transfer model of speech production.

Now we propose a model of driving wall with the mechanical vibration in Fig. 2(a). As it is shown in this figure, the small variation of the air volume dv is yield with the driving force from the vocal tract wall, and is represented as

$$dv = \pi(\delta r^2 + 2r\delta r)l \qquad (1)$$

where, δr is the displacement of the radius in the mechanical vibration. Since the radius is varied, we see that the characteristic impedance of the small section is also varied in the acoustic model, and this should be reflected in the reflection coefficients of our lattice model. Our model is shown in Fig. 2(b) for the model of wall vibration (MWV).



(b)

Fig. 2 Model of driving wall (a),
and lattice model of driving source (b)

3 Simulation of MWV

In our simulation, we employ our digital simulator for speech production [2], and combine this simulator and the additional model mentioned above such as eq. (1) and Fig. 2(b). In the trachea and lung model, the lossy transmission model of 20 sections is used.

The vocal cord vibration is simulated with the 2-mass model [4], and in the propagation of the mechanical vibration the modeling is based on the some physical constants of human skin. The propagation speed and the decay constant have been measured for the arm skin [5] ; the propagation speed $v = 296[cm/s]$, and the decay constant $\alpha = 0.576$. Supposing the pitch $T = 7[msec]$ for the vocal cord vibration, we get the wave length propagating in the wall, $\lambda = vT \approx 2.07[cm]$. If the total length of the vocal tract $L[cm]$, the number of section n, and the length of the one section $l[cm]$ are given, the sampling interval T_o is determined for our simulation as follows,

$T_o = l/c$ where c is the propagation speed of sound.

In the one section of the vocal tract, the traveling time of the elastic wave is given as,

$T_w = l/v$

The amplitude of the wall vibration in the i-th section is represented as y_i , which is approximately given by the glottal side of the section, and the time delay between i-th and i+1 th section is obtained as follows,

$$\frac{T_w}{T_o} = \frac{c}{v} \approx \frac{35600}{256} \approx 120.27 \qquad (2)$$

From the above relation, we can obtain the y_i from the y_{i-1} of the 120 step before,

$$y_i(n) = e^{-\alpha l} y_{i-1}(n-120), \quad \alpha = 0.576 \qquad (3)$$

where n is the time index, and $y_0(n)$ is the amplitude of displacement at the glottis caused by the vocal cord vibration. As shown in Fig. 3, we see that the amplitude of wall vibration is decayed enough at 5 [cm] distance from glottis, thus we can ignore the reflection of the elastic wave from the lips or the lung.

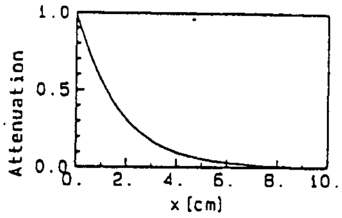


Fig. 3 Attenuation of propagation wave in wall

We show the cross-sectional function of the vocal tract in Japanese vowels in Fig. 4, which is computed form the cross- sectional figure measured using MRI-CT for an adult man [3]

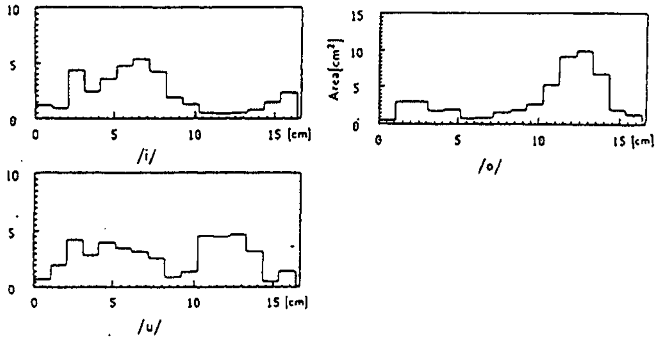


Fig. 4 Cross-sectional functions of vowels

4 Experimental Results

The displacement of the vocal cord is simulated as in Fig. 5, which is computed with the 2-mass model. For the experimental results of Japanese 5 vowels, we show in here the synthesized vowels for /i/, /u/ and /o/, and the spectral analysis results using LPC analysis in the figures, Fig. 6 - 9.

In these results, it is compared with the two kind of models; (a) / (b) shows the result without / with the mechanical wall vibration. In these figure, we see the repetition of the spectra along the time axis , where the spectra with broad band formants are the influence of the subglottic system in the glottal open phase. The difference in the two models is appeared in the transitional time of the subglottal pressure. We see the spectral differences in the figures with the mark.

The results mean that if the subglottal pressure is varied continuously such as the continuous speech of real utterance, then the influence on the spectra is greater than the synthesizer without the MVW.

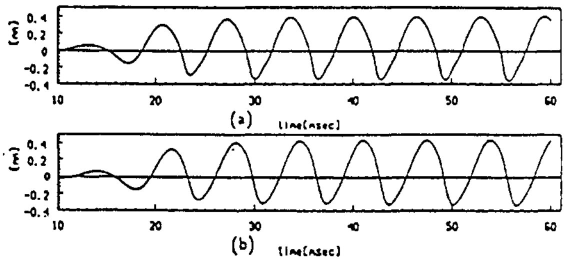
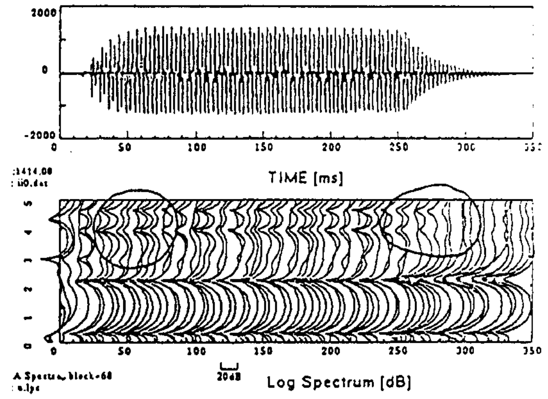
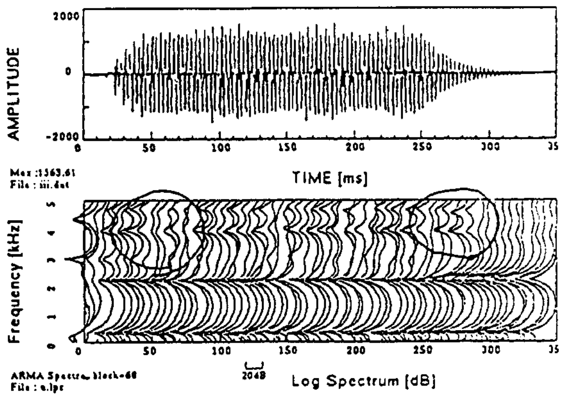


Fig. 5 Displacement of vocal cord

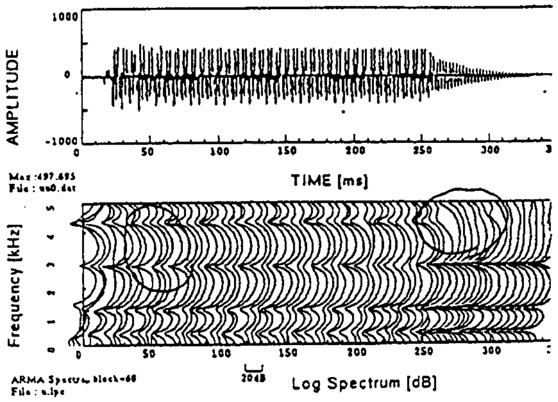


(a) /i/ without MWV

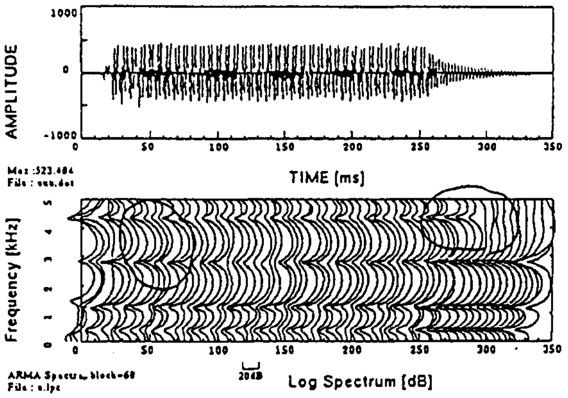


(b) /i/ with MVW

Fig. 6 Synthesis and analysis for /i/

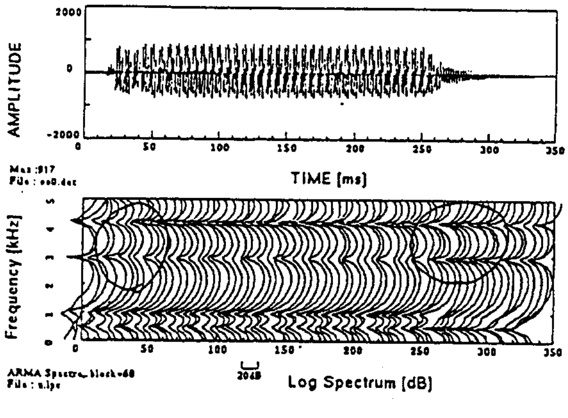


(a) /u/ without MWV

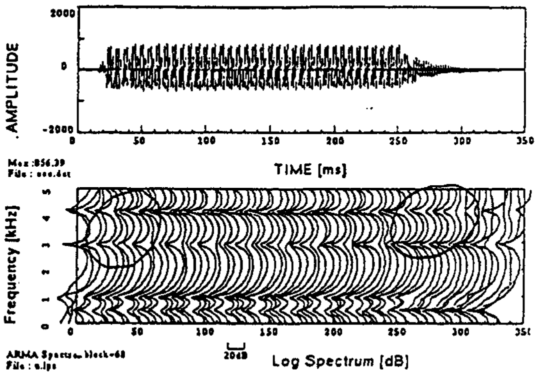


(b) /u/ with MVW

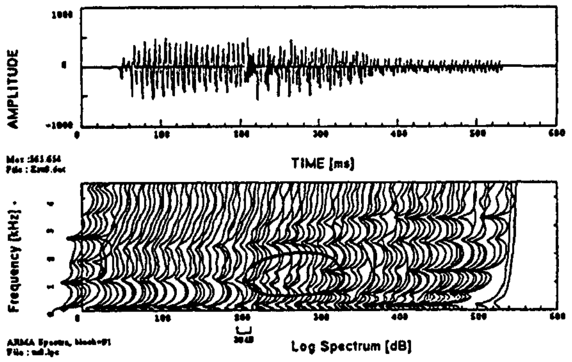
Fig. 7 Synthesis and analysis for /u/



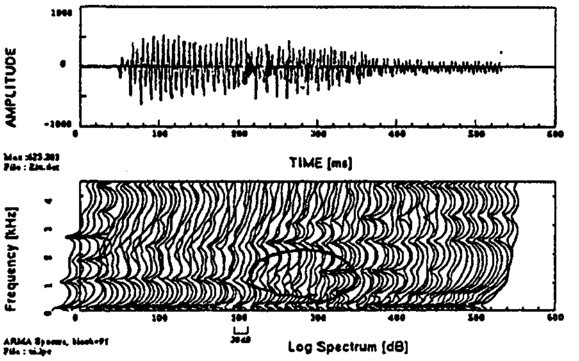
(a) /o/ without MWV



(b) /o/ with MVW
Fig. 8 Synthesis and analysis for /o/



(a) /iu/ without MWV



(b) /iu/ with MVW
Fig. 9 Synthesis and analysis for /iu/

5 Conclusion

We showed a speech production model of the vocal tract analogue, which can simulate the propagating mechanical vibrational wave on the tract wall, synthesize vocal sounds with the vibrational wave, and analyzed the spectral properties for the synthesized waves. From the LPC analysis of the synthesized vowels, we see that the spectral differences are appeared in the transitional time of the subglottal pressure. We can see importance to simulate the larynx vibration in the speech synthesis.

References

[1] N. Miki, K. Motoki, and N. Nagai, "A lattice filter model with accurate lip impedance for dynamic articulatory movement," Proc. IEEE ICASSP87, Dallas, 22A4 (1987)

[2] N. Miki,"Application of ARMA digital lattice filter to speech analysis and reconstruction," in Ed. N. Nagai, Linear circuits, systems and signal processing: Advanced theory and applications, Marcel Dekker, New York (1990)

[3] N. Kamiyama, N. Miki and N. Nagai, "Study of the Vocal Tract Wall Impedance Using Viscoelastic Model of the Wall," (in Japanese) IEICE Trans. A, vol. J75-A no. 11 (1992)

[4] K. Ishizaka and M. Matsudaira, "Fluid mechanical considerations of vocal cord vibration," SCRL Monograph No.8, Speech Communications Research Laboratory, INC. (1972)

[5] T. Irie, H. Oka and T. Yamamoto, "Measurement of Biomechanical Properties on Skin by Impact Response," (in Japanese) IEICE Trans. D-II, vol. J75-D-II, no.4 (1992)

[6] H. Iijima, N. Miki, and N. Nagai, "Glottal impedance based on a finite element analysis of two-dimensional unsteady viscous flow in a static glottis," IEEE Trans. SP. Sep. (1992)

[7] A. Sakakura and H. Takahashi, "Body Wall Vibration in Trained and Untrained Voices," pp. 391 - 401, Vocal Physiology: Voice Production, Mechanisms and Functions, Ed. O. Fujimura, Raven Press (1988)