Adaptive Control of a Vocal Chord and Vocal Tract for Computerized Musical Singing Instruments

Hideoaki Sawada* and Shuji Hashimoto
Department of Applied Physics, School of Science and Engineering, WASEDA University
* Research Fellow of the Japan Society for the Promotion of Science

Abstract

We are constructing a phonetic machine having a vocal chord and a vocal tract based on mechatronics technology, and have so far developed a pitch generation part as a subsystem for melody synthesis to sing in humming. In the pitch generation, the analysis and mechanical simulation of the behavior of the vocal chords are required. The fluid mechanical system, however, is too complicated to make the control difficult. This paper presents a singing instrument system together with an adaptive tuning algorithm of the physical mechanism. The mechanical method is considered to be promising to generate more natural voice than algorithmic sound synthesis methods.

1. Introduction

Although an algorithmic sound synthesis is a popular technique for the computerized speech production [Ilseos, 92(Rodet et al., 89)](Depal et al. et al., 94), a mechanical approach using a phonic or vocal model would be an interesting research [Flanagan, 72] not only as an acoustic speech generation but also as a study of human acquisition of speaking and singing skills.

In human vocalization, a vibration of vocal chords generates a source sound, and then the sound wave is lead to a vocal tract which works as a filter to determine a spectrum envelope. The fundamental frequency and volume of the sound source can be varied by the change of the physical parameters such as the stiffness of the vocal chords and the amount of airflow from a lung, and these parameters are uniquely controlled when we utter a song. On the other hand, the spectrum envelope is necessary for the pronunciation of lyrics consisting of vowels and consonants is formed according to the inner shape of a vocal tract and mouth which is governed by the complex movements of a tongue and muscles.

This paper reports some experimental results on the adaptive control of a vocal chord and a vocal tract for computerized musical singing instruments.

2. Human Vocalization

Human voice is generated by the vocal apparatus [Hayashi, 79]. The voice sound is perceived in verbal communication as words which consist of vowels and consonants, and is produced by the relevant operation of the physical apparatus such as the lung, trachea, vocal chords, vocal tract, tongue and muscles. The lung has the function of air tank, and the airflow through the trachea causes the vocal chord vibration as the source sound of voice. The glottal wave is lead to the vocal tract which works as a sound filter to form spectrum envelope of the voice. The characteristics of the sound filter can be varied by the inner shape of the vocal tract with the help of the movements of the tongue and the jaw. The vowel sounds are radiated by the relatively stable configuration of the vocal tract, while the consonants are produced generally by the short time dynamic motions of the vocal apparatus. Furthermore, the dampness and viscosity of the organs have a great influence on the timbre of the sounds being produced so that we may experience when we have a sore throat. Appropriate configurations of the vocal tract for the production of syllables are acquired as infants grow by repeating trials and errors of heaving and uttering voice sounds.

3. Mechanical Model for Vocalization

3.1. Artificial Vocal Chord

The characteristics of the glottal wave which determines the pitch and volume of the human voice is governed by the complex behavior of the vocal chords. It is due to the oscillatory mechanism of human organs consisting of the mucous membrane and muscles excited by the airflow from the lung. Although several research about the computer simulation of the movements are found [Hashizaka et al., 72], we are trying to realize the wave by a mechanical model.

---

*Sawada & Hashimoto 444 ICMC Proceedings 1996
We employed an artificial vocal chord used by the people who had to remove their vocal chords because of glottal disease. Structural view of the artificial vocal chord is shown in figure 1. The vibration of the rubber attached over the plastic body makes vocal sound source. Its waveform shown in figure 2 is similar to the actual one.

![Sound Wave of Artificial Vocal Chord](image)

By considering the simplified dynamics of the vibration of a strip of rubber of the length $l$, the fundamental frequency $f$ of the rubber vibration is given as,

$$f = \frac{1}{2L} \sqrt{\frac{D}{S}}$$  (1)

This equation shows that the fundamental frequency varies according to the tension $S$ and the density $D$ of the material.

The tension of the rubber can be manipulated by applying tensile force. We measured the relationship between the tensile force and fundamental frequency produced by the artificial vocal chord. Figure 3 shows typical experimental results. Although the relation does not fit the equation (1) due to the change of the density in addition to the tension, the fundamental frequency varies from 130Hz to 520Hz by the manipulation of force applying to the rubber. Furthermore, the relation between the produced frequency and the applied force is not linear but tends to change with the repetition of experiments. The artificial vocal chord is, however, considered to be suitable for our system not only because of its simple structure, but also its frequency characteristics to be easily controlled by the tension of the rubber.

![Relation between Tensile force and Fundamental Frequency](image)

3.2. System Configuration

As shown in figure 4, the mechanical vocal system mainly consists of an air compressor, an artificial vocal chord, a harmonizing tube and a microphone connected to a FFT analyzer, which correspond to a lung, a vocal chord, a vocal tract and an ear, respectively.

The air in the compressor is compressed to 5 hpa, on the other hand the pressure of air from our lungs is about +0.2 hpa larger than the atmospheric pressure. We apply a pressure reduction valve at the outlet of the air compressor so that the pressure will

![System Configuration](image)
be nearly equal to the pressure of the air through the trachea. The valve is also effective to reduce the fluctuation of the pressure in the compressor. The decompressed air is lead to the artificial vocal chord via an airflow control valve which works to control the voice volume. The harmonizing tube is attached to the artificial vocal chord for the modulation of sound envelope to sing in harmony. The FFT analyzer plays a role of an auditory system. It realizes the fast Fourier transformation of produced sound in real-time and extracts the fundamental frequency and spectrum envelope which are necessary for the auditory feedback. The system controller manages the whole system with the motor controller and FFT analyzer through GP-IB interface.

For the fundamental frequency and volume adjustments, two stepping motors are employed: one is to manipulate the screw of the airflow control valve, and the other is to apply tensile force to the rubber of the vocal chord for the tension adjustment.

4. Voice Generation and Adaptive Tuning Algorithm

As mentioned in the section 3-1, not only adjusting but also maintaining the output frequencies is not an easy task due to the dynamical mechanics of vibration which is easily affected by fluctuations of tension and airflow. Stable output has to be obtained no matter what kind of disturbance applies to the system. Constructing adaptive control mechanism would be a good solution for robust system.

We introduce an adaptive tuning algorithm for the production of singing voice using the mechanical vocal system. The algorithm consists of two phases. First in the learning phase, the system acquires a position-pitch-map which associates motor positions with fundamental frequencies by comparing the pitches of output sounds with the desired pitches included in melody lines. Then in the performance phase, the system sings according to the obtained map while voice pitches are adaptively maintained by bearing its own outputs.

4.1. Adaptive Tuning in Learning Phase

Figure 5 shows a schematic diagram of the adaptive tuning which arises in the pitch learning phase. The algorithm simulates the process of our pitch learning in practicing singing. The algorithmic process of pitch acquisition in the system controller is shown in dotted lines. The tuning manager manages the behaviors of all the other units presented in boxes. The mechanical vocal system starts its action by receiving a present-position-vector \( p = (p_1, p_2) \) as a command for the movements of the motor1 and motor2. Two elements \( p_1 \) and \( p_2 \) of the vector are the desired positions of the motor1 and motor2, respectively, and can be obtained by the calculation of the frequency-to-position translator, which is trained in advance to output desired motor positions from frequencies according to the relations between tensile force and fundamental frequency shown in figure 3.

At first, the system controller starts with setting arbitrary values as a present-position-vector to send to the vocal system. The fundamental frequency of the generated sound is calculated by the FFT analyzer of the auditory system. To compare with the desired pitch, the difference between the two pitches is obtained according to the go-signal triggered generated by the tuning manager. The go-signal, in

![Figure 5 Diagram of Adaptive Tuning](image-url)
the same instant, drives the frequency-to-position-translators so that the two motors work. The frequency difference decreases as the feedback process repeats. When the frequency difference becomes smaller than a predetermined threshold value which is currently set at 2 Hz, the judgment-signal unit arises, so that the present position vector is associated with the target pitch and stored as the position-pitch map.

The result of the feedback process is shown in figure 6, in which the system acquired the sound pitches from C to G. In several repetitions of auditory feedback, it can learn each pitch as designed in figure 5.

4.2. Adaptive Control in Performance Phase

The schematic diagram of the singing performance is shown in figure 7. The performance control manager takes charge of two tasks: one is for the performance execution presented by solid lines in figure 7, and the other is for the adaptive control of pitches during the performance. The score-information unit sends melody lines as sequential data of pitches and duration.

Singing performance is executed according to the go signals generated by the performance control manager. The manager has the internal clock and makes a temporal planning of note outputs with the help of the duration information in the score-information unit. The note information is translated into a present-position-vector by referring to the position-pitch map.

During the performance, unexpected changes of air pressure and voice force cause the fluctuations of sound output. The adaptive control with the auditory feedback is introduced in this mechanical system by hearing the output sound. The auditory units observe the pitch errors so that the system can start fine tuning of the pitch by receiving the tuning-signal trigger under the control of the performance manager. The position-pitch-map can be also renewed by the map-repair signal. The system can sing in humming by the adaptive mechanical control using the auditory feedback.

5. Conclusions

In this paper we introduced a phonic machine having a vocal chord and a vocal tract. With the adaptive tuning of the physical mechanism, it acquires the pitch generation to sing in humming.

By improving the vocal tract to be adaptively controlled, we can realize an acoustic musical instrument controlled by a computer that will shout and sing like human.

References

[Flanagan, 72]

[Ichikawa et al., 72]

[Hayashi, 79]

[Rodet et al., 89]

[Hirose, 92]

[Depalle et al., 94]