Singing Synthesis Based on Resonance Characteristics of Vocal Tract

Kazuya Nagata(1310049)

Japan Advanced Institude of Science and Technology School of Information Science

2015 年 2 月 12 日

Keywords: singing synthesis, source-filter model, resonance characteristics of vocal tract, formant analysis, formant control.

Number of expression styles of singing voices is more than that of speaking voices. Singers can perform those expression styles by his/her production system of the voices. However, structure of the system is not known completely. Studies on singing voices have been conducted for a long time. Particularly, the study on opera song is conducted flourishingly in Europe. The system for synthesizing singing voicessy called "Farinelli" was introduced by Institut de Recherche et Coordination Acoustique/Musique (IRCAM). This system can reproduce singing voices of the deknackered opera singer. The proposed synthesis system in this study is based on resonance characteristic of vocal tract. It makes not only synthesized singing voices natural but also resonance characteristics of vocal tract unravel.

There have been two major ways to synthesyze singing voices. The first method is measuring physical parameters of singers. This method uses laboratory equipment like MRI. The method controls a physical model according to the measurement results. However, the measurement cause big burden on subject and physical model is complicated. Thus, this study uses another method. The alternative method analyzes recorded singing voices and controls parameters of acoustic characteristics based on the results.

There are some techniques for the singing voice synthesis. The techniques are such as a wave pattern concatenation method like VOCALOID and Hidden Markov Model (HMM) synthesis method. However, those techniques do not consider about production system of the voices. Therefore, the study based on those techniques cannot obtain knowledge of the mechanism of production system for the voices. This study uses a source-filter model that can simulate vocal fold vibrations and vocal tract shapes of humans.

There are some reports using a source filter model. One of the report is based on foundamental frequency (F0) control model. This study controls F0 dynamic character-

istics. Those characteristics are vibrato, overshoot/undershoot, and preparation. The vibrato is a phenomenon that F0 and amplitude envelop vibrate in modulation frequency between 4Hz to 6Hz. Another report is based on Liljencrants-Fant (LF) model. The second study can express three vocal registers when controlling LF model. However, the resonance characteristics of vocal tract have not been examined enough in these studies.

Therefore, this study propose a singing voice synthesis system by considering resonance characteristics of vocal tract. This characteristic is not considered in previous studies enough. In addition, this study makes not only synthesized singing voices natural but also knowledge about resonance characteristics of vocal tract unravel. In this study, Klatt Formant Synthesizer is used as a synthesizer. This synthesizer can control parameter values for each formant independently. This synthesizer is suitable to express resonance characteristics of the vocal tract as controlling resonators independently. In addition, this study uses the F0 control model and LF model for expressing dynamic characteristics and vocal register. These controls make note into sound source.

Recorded singing voices are analyzed to constract a formant control model. Number of the recorded sounds is 60 in total. Those sounds are singing voices and speaking voices uttered by one male who has vocal music training experience. The singer wears EGG. It can record the maximum opening phases and the maximum closing phases of the glottis.

Then, a formant analysis is carried out. Sounds for analysis are 108 sounds that are divided singing voices into every note. Sampling frequency of the sounds is 8 kHz. This analysis uses humming window and LPC. Order of LPC is 10. After the analysis, medians are calculated by tenth median filter to show changes of formants. As a result, F1 and F2 move characteristic in different four are when F0 move. The first section of F0 is the range with less than 300Hz. This is the range sung in modal. In this section F1 decreases 0.32Hz per 1Hz of F0 and F2 increases 0.41Hz per 1Hz of F0. The second section of F0 is the range from 300Hz to 336.64Hz. This is the range sung in modal and falsetto. In this section F1 decreases 3.32Hz per 1Hz of F0 and F2 decreases 4.23Hz per 1Hz of F0. The third section of F0 is the range from 336.64Hz to 454.64Hz. This is the range sung in falsetto and low pitch. In this section F1 increases 0.15Hz per 1Hz of F0 and F2 decreases 0.82Hz per 1Hz of F0. The fourth section of F0 is the range with over than 454.64Hz. This is the range sung in falsetto and high pitch. In this section F1 increases to F0 and median of between F1 and F0 is 15Hz. F2 increases to second harmonics and median of between F2 and second harmonics is 35Hz. The formant control model is built for four sections based on these analysis.

Next, an evaluation experiment is carried out with six listeners to evaluate effectiveness of formant control model. In this experiment, four kinds of synthesized singing voices are used. Three of them are synthesized following the previous studies, and one of them is

synthesized by the proposed technique. In this experiment, paired stimuli were presented to the listeners. The linsteners judged which one is more natural singing voice that shift modal to falsetto. As a result, one subject judge synthesized sound by propose method is best. However, propose method is less natural than synthesized precedent study in total.

Result of analysis shows extreme change is happened in amplitude. Then amplitude of propose method is controled fitting to judged most natural sound. As a result, amplitude controlled sound is not only more natural than sound D but also same natural as sound C. The result shows effectiveness of formant control model.