A Study on Estimation of Glottal Source Waveform and Vocal Tract Shape for Singing Voice with Wide Frequency Range

Kyoko Takahashi (1610111)

Graduate School of Advanced Science and Technology,
Japan Advanced Institute of Science and Technology

February 9, 2018

Keywords: singing-voice analysis, vocal register, glottal source waveform, ARX-LF model.

Singers perform good singing using some different vocal qualities and vocal registers. Singing voices are characterized by temporal fluctuations of glottal vibration and vocal-tract shapes. Some studies reported that each vocal qualities and vocal registers are different in temporal fluctuations of glottal vibration. Detailed investigation of the features is required for singing voice synthesis with vocal quality and vocal register. However, direct observation of vocal folds is difficult. Singing voice analysis and synthesis have been studied for vocal quality and vocal register using speech waveform.

Based on source-filter theory, speech and singing voice are defined as output of vocal tract filter with input signal as glottal source. Temporal fluctuations of glottal vibration and vocal-tract shapes can be obtained by estimating glottal source waveform and vocal tract filter. Several analysis methods have been proposed based on the source-filter theory. Ohtsuka and Kasuya proposed a speech analysis-synthesis method based on auto-regressive with exogenous input (ARX) model, which was a model of vocal tract filter, and Rosenberg-Klatt (RK) model, which was a model of glottal source waveform. They reported that the method can analyze female- and children-speaking voices as well as male ones. Vincent et al. proposed another method for speech analysis and synthesis based on ARX model and Liljencrants-Fant (LF) model. They analyzed simulated speech data and female speaking voices. They reported that the method can lead good results on speech data with low fundamental frequency (fo).

Several methods based on the ARX-LF model have been reported for singing voice analysis and synthesis. Lu and Smith Ⅲ focused on glottal aspiration noise in singing voice. They proposed the method for extraction and synthesis of the glottal aspiration noise in singing voices. Motoda and Akagi investigated features of glottal source waveform of each vocal register by analyzing singing voices using ARX-LF model. As their results, differences of glottal source waveforms were found among vocal registers.

The previous methods, however, suffered from accurate analysis of singing voices with high fo over 400 Hz. Singers can sing songs fluently with changing fo freely in not only one vocal register but also inter-registers. Therefore, estimation of glottal source waveform and vocal-tract shape accurately in wide fo range is required to obtain the characteristics of such singing.

The objective of this paper is to propose a method of accurate estimation of the glottal source waveform and the vocal tract shape for singing voices which has widely-ranged fo. In this paper, effects of forwarded periods are considered as a cause of inaccurate estimation of singing voice with high fo. In singing voices with high fo, the settling time of the vocal tract filter exceeds

a length of each periods and the responses of the vocal tract filter of the forwarded periods leak into the target period. Thus, singing voice is re-synthesized using the estimated ARX-LF parameter values to estimate the leaked components.

In the proposed method, parameters of the ARX-LF model were estimated accurately with exhaustive search in determined range and a simulated annealing method. Additionally, singing voice was re-synthesized using the estimated results of the vocal tract filter and periodic glottal source waveform with a length of settling time for considering the effects from forwarded periods.

Two experiments for evaluation were carried out using the simulated singing voice which has low, middle and high fo, and the real singing voice. Reduction of the power of the minimized error was found in the results of the simulated singing voice. As a result, it is indicated that the proposed method could accurately estimate the glottal source waveform and the vocal tract filter of singing voice with wide range of fo. The results of the real singing voice indicated that the proposed method could be applied to estimate the parameter values in real singing voices.

The features of glottal source waveform were found in vocal registers as modal and falsetto with analyzing using the proposed method. The experiment for analysis was carried out using the real singing voice with one register. The trends of features were found as with the reports of previous studies.

The temporal features of glottal source waveform were found in the singing voice which widely-ranged fo including two registers with analyzing using the proposed method. The fluently temporal fluctuations of ARX-LF parameter values were found in the results of the real singing voice with changing fo from modal to falsetto.