Singing Synthesis Controlled Fundamental Frequency and Spectrum Envelop

Ichiro Shimizu (110057)

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

February 14, 2003

Keywords: spectrum envelop,fundamental frequency,singing synthesis,speaking voice,spctrum control.

This thesis purposes to achieve a high-quality singing voice synthesis from speaking voices and the melody to the singing voice. It proposes the technique by which the singing voice is synthesized from voices while controlling a fundamental frequency and a spectrum sequence by using STRAIGHT of the source filter model.

It is one of important feelings expressions by which feelings and the desire are told that the person sings and it is important to research what the song is. Human rapidly changes to put together to the melody keeping his tone when singing. In a word, human is keeping tone while greatly changing a fundamental frequency when singing.

The spectrum sequence relates with a singing voice tone is described. In the speaking voices, it is said that the first formant and the second formant decides the phoneme and the formant more than the third formant does not change so much and are related to a personal feature. On the other hand,in the singing voice it is said that the first two formants are necessarily to decide whether the vowel is. For instnce, the first formant frequency and second formant frequency of vowel [i:] when a male opera singer sang corresponds to [y:] in the vowel chart for the speaking voices. Moreover, since female soprano singer's high voice is sung with a fundamental frequency of 1000Hz or 1400Hz, the phenomenon that the fundamental frequency is

higher than the first formant frequecy happens. Then, the operation by which the first formant is moved to the neighborhood of a fundamental frequency is done. The difference of the tone by the voice timbre type pays attention to the difference off formant of the vowel about the singing voice of each voice kind and is researched and it is generally described higher voice timbre type are ,higher formant frequencies move to bias. It is guessed the spectrum has to be changed to keep toning changing in a fundamental frequency. Then, the spectrum sequence has to be controlled corresponding to the change in a fundamental frequency.

To synthesize a high-quality singing voice from a speaking voice, speaking voice and the singing voice were analyzed and compared. In the speaking voices and the singing voices , formant frequencies versus F0 were analyzed. In the speaking voice,it was disposed the first formant and the third formant rose and the second formant descended when the fundamental frequency rose. In the singing voice,it was disposed the first the second and the third formants rose when the fundamental frequency rose.

And, equation $I = 0.13\bar{F}_0 P + 0.84P$ by which singing voice spectrum I was calculated from singing voice F0 ratio versus speaking F0, $\bar{F}_0$ and speaking spectrum P, was derived by the minimum mean square method. Moreover, it has been understood that there is a correlation on F0 of singing voice vibrat and a minute change of F1 without time lag. A spectrum in the freqency domain is controlled from the result of obtaining by this analysis. The control of a spectrum in the frequency domain is that singing voice spectrum is made by mapping speaking voices spectrum of each 1ms by using the equation of $I = 0.13\bar{F}_0 P + 0.84P$. And, it is added to F1 as a minute change of F1 synchronizing with the shake corresponding to vibrat element of F0. Next, because the length of the syllable in the singing voice is decided by the note duration, the duration of the vowel to be decided if the consonant duration of a singing voice is understood. And the ratio of the duration of the consonant of singing voices to the duration of consonant of speaking voice is analyzed. It has been understood to only have to make the duration of a singing voice consonant 1.28 times at voices friction sound, 1 time at the explosion sound, 2.37 times at the semivowel,1.43 times at the nasal, and /y/ 1.22 times by the analysis in each modulation method of the consonant. The method of controlling

the spectrum sequence in the time domain was designed by using these analysis results. Segmentation is done to the the stable consonant portion and modulation uniting part (40ms) and the regularity the vowel portion for the transformation in time domain from the speaking voice spectrum to the singing voice spectrum. The control of a singing voice spectrum uses the modulation uniting part and the speaking voices one is used as it is and extends the speaking voice spectrum of the regularity part of voices vowel and the consonant. The length of a singing voice consonant is decided from the analysis result depending on duration and the modulation method of speaking voices consonant. The length of a singing voice vowel only has to subtract the length of the consonant from note duration. It proposes three as a technique by which the time of the regularity part of the vowel and the consonant is extended. As the first method,we make a singing stable voice part by getting one point of voices regularity part, and arranging only the length of time to want to extend it. As the second method ,to make it to the duration to want to extend the voices regularity part analyzed in the FFT 8192 point, we linear interpolate each channel each frequency in the time direction. As the third method,to make it to the length of time to want to extend the voices regularity part analyzed in the FFT 8192 point, we interpolate the spline in each channel of each frequency in the time domain. We controlled the spectrum by three methods and made the made singing voice and synthetic sound. We listened to three synthesized singing voice and synthetic sounds and evalated that the one that the spline was interpolated has the most high-quality and adopted this as a control of spectrum sequence in the time domain.

We made a singing voice and synthetic sound from speaking voices and the melody controlling singing voice F0, spectrum envelop in the time domain and the frequency domain. To examine useful to be high-quality a singing voice and synthetic sound, and to synthesize the singing voice with a high-quality spectrum envelop control, the listening experiment was done. The singing voice synthetic sound made by the technique of these researches was evaluated by using Scheffe's method of paired comparison.

Effectiveness of the time direction control and thefrequency direction control was shown from the listening experiment. Moreover, the effectiveness of the technique by which the singing voice was synthesized from the

speaking voice and the melody was able to be shown.