

Title	A Study on Correlates of Acoustic Features to Emotional Singing Voice Synthesis
Author(s)	Nguyen, Thi Hao
Citation	
Issue Date	2018-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15462
Rights	
Description	Supervisor: 赤木 正人, 先端科学技術研究科, 修士 (情報科学)

A Study on Correlates of Acoustic Features to Emotional Singing Voice Synthesis

Nguyen Thi Hao (1610058)

School of Information Science, JAIST, s1610058@jaist.ac.jp

Extended Abstract

Singing voice analysis and synthesis become an interesting topic recently. In this context, expressiveness plays an important role in obtaining a high quality of singing voices. Expression control conducts a set of acoustic features that are related to emotions, styles or singer individualities. Listeners, depending on their moods and situations, would like to hear a song with different emotional expressions. It therefore needs to have a computer-based application in singing voice performances, in order to satisfy such purpose.

M. Alonso tries to generate emotional singing voices using a rule-based approach. The advantage in this method is that the obtained rules are relatively simple and deterministic. However, it takes a long time for the analysis and synthesis phases to get the rules, and the system is yet difficult to control. Another singing voice synthesizer is proposed by M. Umberto, which applies a unit-selection methodology. It constructs implicit rules based on a number of units of a singing voice. Hence, it requires a huge unit database to process. Besides, HMM-based approaches (Hidden Markov Model) are also taken into account for synthesizing expressive singing voice, by using statistical model to model important features from database. This method, nevertheless, has a parameters over-fitting problem. A novel speech-to-singing system, proposed by Saitou et al., has applied the performance-driven approaches, which can produce a singing voice from simple resources: (i) a speaking voice reading a song's lyric, and (ii) its musical score. It succeeded in synthesizing a neutral singing voice. Nonetheless, the expressiveness was not taken into consideration.

Our work aims to investigate the correlations of acoustic features to emotional singing voices. Two sub-goals are consisted in this study: (i) analyzing a set of acoustic features that are strongly related to emotions, and (ii) conducting experimental examinations to evaluate the importance of each acoustic feature in the emotional singing voice. By achieving the first sub-goal, we determine which features are most significant to emotional expressions in singing voices. Regarding the second sub-goal, we propose a method to modulate the amplitude envelope based on the entire F0 contour, to have a higher naturalness in a singing voice. Our experimental results show that the spectral feature is the most affected acoustic feature to the emotions of a singing voice. However, in order to obtain high naturalness and singing-ness in synthesized voices, it is necessary to manipulate all three features, including F0 contour, amplitude envelope and spectral sequences.