

Title	声区表現を可能とする歌声合成を目的としたARX-LFモデルの制御法に関する研究
Author(s)	元田, 紘樹
Citation	
Issue Date	2013-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/11326">http://hdl.handle.net/10119/11326</a>
Rights	
Description	Supervisor: 赤木正人, 情報科学研究科, 修士

# A method for controlling ARX-LF model for synthesizing singing voices characterizing vocal registers

Hiroki Motoda (1110061)

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

February 6, 2013

**Keywords:** Vocal registers, Singing voices synthesis, ARX-LF model, Voice production mechanisms.

Singing voices synthesis is one of the important topics in speech science. Construction of singing voices synthesis system having naturalness and variety contributes to music information processing and elucidation of the relationships between voice production mechanisms, acoustic features, and perceptions. However, singing voices synthesis cannot represent ‘Vocal registers’ appropriately.

Vocal register is a particular series of tones in the human voice that are produced by one particular vibratory pattern of the vocal folds and therefore possess a common quality. Human can sing songs naturally in wide range of frequency by training how to use vocal fold vibrations to represent vocal registers. Many reserchers investigated relationships between vocal registers and characteristics of glottal sources. However, even state-of-the-art singing voices synthesis systems cannot produce vocal registers appropriately. Naturalness of the synthesized singing voices using these systems is degraded in low and high tone ranges. One of the methods for improving naturalness is adding characteristics of glottal sources for each vocal register. This paper considers a model that can represent glottal sources with characteristics of vocal registers for an advanced singing voice synthesis system.

This paper proposes a method for controlling characteristics of glottal sources to synthesize singing voices having characteristics of vocal registers. This paper constructs a singing voice synthesis system with the ARX-LF model that can describe glottal sources for each vocal register by simulating human voice production mechanisms . A model for controlling ARX-LF parameters corresponding to characteristics of glottal sources was constructed in accordance with tone ranges. A singing voices synthesis system for synthesizing singing voices characterizing vocal registers was proposed, and the model for

controlling ARX-LF parameters is constructed. Singing voices were synthesized by the proposed system, and the control model was evaluated by performing both of objective and subjective evaluation. This paper reports these results.

A singing voices synthesis system based on ARX-LF model was constructed to add characteristics of glottal sources for each vocal register, that has analysis - control - synthesis procedures. Singing voice data of each vocal register were analyzed by the ARX-LF model, and ARX-LF parameter values corresponding to glottal source of each vocal register were obtained. The control model was constructed using the results of the analysis. ARX-LF parameters for each vocal register were interpolated in a linear manner.

Singing voices were synthesized by the proposed system, and objective evaluation was performed. Spectral tilt in low frequency ranges was analyzed. As the results, almost the same tendency of spectral tilt were obtained from the synthesized singing voices as those from actual singing voices in each vocal register. Quality of the synthesized voices was evaluated for subjective evaluation. As the results, almost the same impressions were obtained from the synthesized singing voices as those from actual singing voices in each vocal register. Additionally, it was suggested that controlling characteristics of glottal sources of each vocal register is important in high and low tone ranges. Results revealed effectiveness of the control model for synthesizing singing voices to characterize vocal registers.

Methods and findings on this study are thought of as leading to the realization of singing voices synthesis having naturalness and variety, and the elucidation of the relationships between voice production mechanisms, acoustic features, and perceptions.