INTERACTIVE SPEECH CONVERSION AND MULTI-LEVEL SPEECH QUALITY EVALUATION TOOL

Kwang Myung Jeon[1], *Woo Kyung Seong*[1], *Hong Kook Kim*[1], *and Sung Dong Jo*[2]

[1] School of Information and Communications, Gwangju Institute of Science and Technology (GIST)
{kmjeon, wkseong, hongkook}@gist.ac.kr
[2] Info-Communications Development Team, Hyundai Motor Company
sdjo@hyundai.com

ABSTRACT

In this demonstration, an interactive speech conversion and multi-level speech quality evaluation tool is proposed. As shown in Fig. 1, the proposed tool first converts recorded speeches by changing a level of tone and a speaking rate, which can be selected by a user. Next, the perceived quality and naturalness of converted speeches are evaluated on the basis of signal-level, frame-level, and sentence-level measurement.

In particular, the proposed speech conversion tool uses a vector quantization (VQ) based conversion parameter control technique [1] to obtain converted speeches with various speaking characteristics and high naturalness. In other words, the VQ-based parameter control technique extracts feature parameters from both input and reference speeches per each frame, and then concatenates them into a feature parameter vector. Next, all the feature vectors are clustered into a certain number of representative vectors by using VQ. Finally, the vectors are applied as conversion parameters to speech conversion method based on pitch synchronous harmonic and non-harmonic model (PS-HNH) [2]. Note here that the control parameters can also be applied to a segment of a speech utterance while maintaining the rest of the utterance without any change.

After the speech conversion is done, the proposed speech quality evaluation tool measures signal-level and feature-level distances from converted speeches, such as the degree of voicing, pitch/formant trajectories, spectral distortion [3], and delta cepstrum distance [4]. In addition, automatic speech recognition (ASR) and single-sided speech quality measures (3SQM) [5] are performed to measure sentence level quality of converted speeches.

The proposed tool can be used not only for recorded speeches of any duration, but also for preloaded input speech DB with length of hours. Thus, the proposed tool successfully can help to develop a commercial speech recognizer and synthesizer by providing large-scale conversion DB with various speaking characteristics and high naturalness to them.

Figs. 2 and 3 illustrate graphic user interface (GUI) of a speech conversion and a speech evaluation part of the proposed tool, respectively.

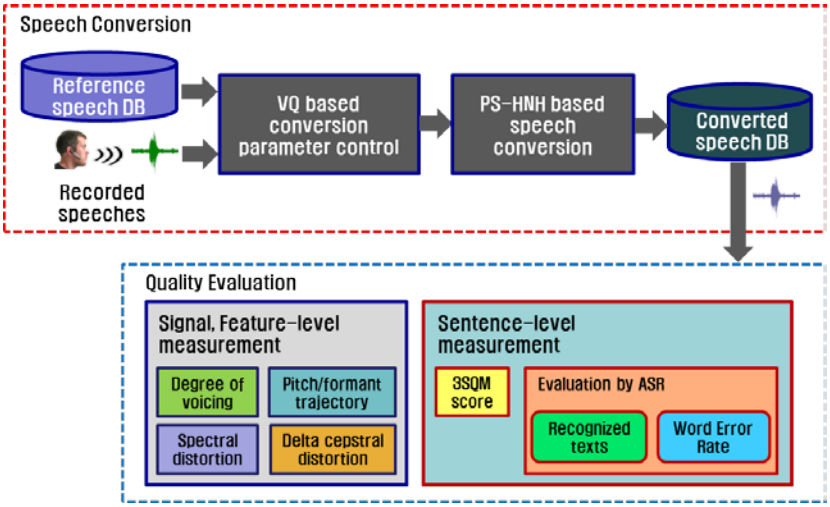

Fig. 1. Procedure of the proposed speech conversion and multi-level speech quality evaluation tool.

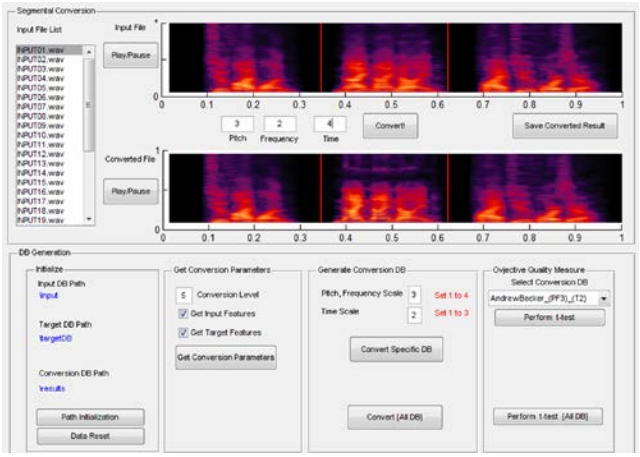

Fig. 2. GUI of a speech conversion part of the proposed tool.

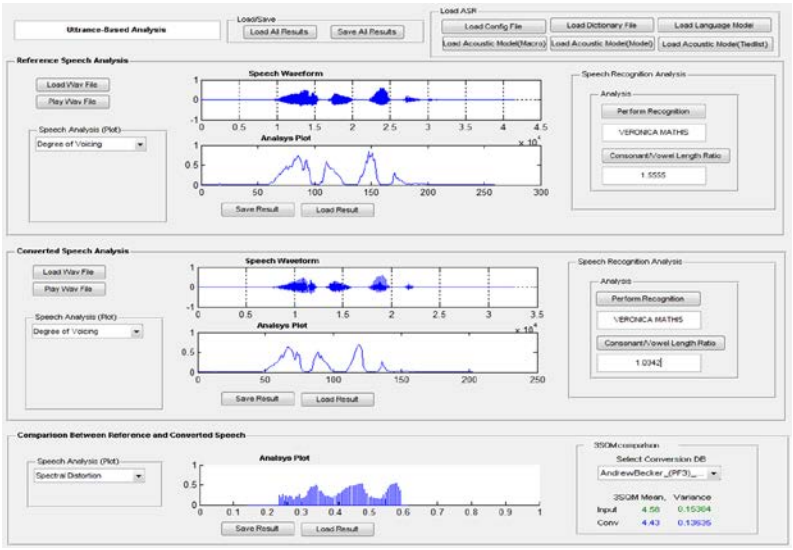

Fig. 3. GUI of a speech quality evaluation part of the proposed tool.

ACKNOWLEDGEMENT

This work was supported in part by Hyundai Motor Company, the National Research Foundation of Korea (NRF) grant funded by the government of Korea (MSIP) (No. 2012-010636), and the MSIP, Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-2005) supervised by the NIPA (National IT Industry Promotion Agency).

REFERENCES

[1] K. M. Jeon, W. K. Seong, H. K. Kim, and S. D. Jo, "Vector quantization-based parameter control for speech conversion with improved naturalness," accepted in *International Conference on Convergence and its Application (ICCA)*, Jeju Island, Korea, Nov. 2013.
[2] K. M. Jeon and H. K. Kim, "High-quality speech modification based on pitch-synchronous harmonic and non-harmonic modeling of speech," *Advanced Science and Technology Letters*, vol. 14, no. 1, pp. 176-179, Aug. 2012.
[3] A. H. Gray and J. D. Markel, "Distance measures for speech processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 5, pp. 380-391, Oct. 1976.
[4] E. Vincent, R. Gribonval, and F. Cédric, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462-1469, July 2006.
[5] A. W. Rix, J. G. Beeerends, D. S. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality – technology and applications," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1890-1901, Nov. 2006.