ACOUSTICS2008/1749
Developing a bilingual communication aid for a Japanese ALS patient using voice conversion technique

Akemi Iida[a], Shimpei Kajima[b], Keiichi Yasu[b], John Kominek[c], Yasuhiro Aikawa[d] and Takayuki Arai[e]

[a]School of Media Science, Tokyo University of Technology, 1404-1, Katakura-cho, Hachiouji, 192-0982 Tokyo, Japan

[b]Dept. of Electrical and Electronics Engineering, Sophia University, 7-1 Kiyoi-cho, Chiyoda-ku, 102-8554 Tokyo, Japan

[c]Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Ave. West, Pittsburgh, PA 15213, USA

[d]1-36-2, Shinjuku, Shinjuku-ku, Shinjuku dai-nana-biru 7F, 160-0022 Tokyo, Japan

[e]Dept. of Information and Communication Sciences, Sophia University, 7-1 Kiyoi-cho, Chiyoda-ku, 102-8554 Tokyo, Japan

A bilingual communication aid for a Japanese amyotrophic lateral sclerosis (ALS) patient has been developed. From our previous research, a corpus-based speech synthesis method was ideal for synthesizing speech with voice quality identifiable as the patient's own. However, a recording of a large amount of speech, which is a burden for the patient, is required for such system. In this study, a voice conversion technique was applied so that a smaller amount of recording is needed for synthesis. An English speech synthesis system with the patient's voice was developed using Festival, a corpus-based speech synthesizer with voice conversion technique. Two methods for Japanese speech synthesis were attempted using HTS toolkit. The first used an acoustic model built from all 503 recordings of the patient. The second used an acoustic model built from 503 wavefiles of which voice was converted to the patient's from a native speaker's. The latter method requires fewer recordings of the patient's. The result of the perceptual experiment showed that the voice synthesized with the latter was perceived to have a closer voice quality to the patient's natural speech. Lastly, GUI on windows was developed for the patient to synthesize speech by typing in the text.