International Journal of Innovative Computing, Information and Control
Volume **1**, Number **3**, September **2005**

ICIC International ©2005 ISSN 1349-4198
pp. 547–561

THE KANNON SYSTEM – REAL TIME SPEECH VISUALIZATION

Ken Nakamuro, Katsuhiro Haruki and Sueo Sugimoto

Department of Electrical and Electronic Engineering
Ritsumeikan University
Noji-Higashi, Kusatsu City, Shiga 525-8577, Japan
sugimoto@se.ritsumei.ac.jp

Received November 2004; revised April 2005

Abstract. *In this paper, we have developed a new real time speech-displaying system called "KanNon" which helps deaf people understand speech contents by spectrogram-reading. The term "spectrogram-reading" means understanding speech contents by seeing formant transition which appears in the sound spectrogram. This idea is based on that speech is characterized by formant transition. The KanNon system displays not only sound spectrogram, but also pitch frequency, loudness of speech and characters by speech-recognition system as real time scrolling image. In order to clearly display formant patterns with high accuracy and estimate the autoregressive model parameters, we applied Burg method combining with the minimum cross-entropy (Burg-MCE) method. We also adopted the mel-scale to the frequency axis in the sound spectrogram instead of the hertz-scale. Finally, we showed that the displaying of the spectrogram-reading in the KanNon system is more effective.*

Keywords: Minimum cross-entropy method, Burg method, Kullback information distance, Pitch estimation, Spectral estimation, Speech visualization

1. Introduction.

When the deaf people communicate with others, they use sign language, transcript or lip reading. However there are difficulties to communicate with normal people (people without hearing disabilities) using these methods in terms of convenience of communication. Against this background, we have developed the KanNon system [1, 2] as a new real time visualization system, which helps deaf people communicate with others using the spectrogram-reading. The KanNon system displays not only sound spectrogram, but also pitch frequency, loudness of the speech and characters by a speech recognition system.

Sound spectrogram is a two-dimensional display of the power spectra with 256 gray-scale image. The elapsed time and the frequencies are described there on the horizontal and vertical axes, respectively. Speech is characterized by the resonance frequency of the vocal tract, which appears as spectral peaks in speech spectrum; called formant. Formants are displayed as black belts in the sound spectrogram. Each phoneme in language system has unique formant pattern, and this fact enables us to understand contents of the speech by spectrogram-reading. So it is important to estimate spectral peak of the vocal tract with high accuracy and display the formant pattern clearly in sound spectrogram to distinguish phonemes by spectrogram-reading.