THE SPEECH VISUALIZATION SYSTEM
― KANNON

Ken Nakamuro (Kenji Nakamuro)

This thesis addresses the speech visualization system ``KanNon'' which supports communication of the deaf people. Especially to the spectral analysis applying the minimum cross entropy method, its application to the speech signal, the speech feature extraction and the speech recognition system based on neural network techniques are presented.

In the First, the concepts and structure of the KanNon system are proposed considering the present situation of deaf people in Japan.

Secondly, as the main part of the feature extraction of the speech, continuous spectral estimation method applying minimum cross entropy (MCE) analysis considering an auditory perceptual property are derived. Then we derived the MCE method with uncertain constrains of autocorrelation function. And the numerical experiments comparing both methods are performed.

Furthermore, a novel AR model parameter estimation method extending the Burg method on the basis of the MCE principle is proposed. In order to apply the proposed method to a spectral estimation of a speech data, we introduce an algorithm to determine the usage of a prior information, based on the divergence measure defined by the Kullback information, since effectiveness of a prior information to spectral estimation results depends on the variation of speech signal. The estimation results for real speech data illustrate improved performance in comparison to the Burg method.

On the one hand, pitch frequency estimation method considering continuity of the pitch frequency is addressed and experimental results are presented.

Finally, the speech recognition system consists of speech/silence, voice/unvoiced, vowel recognition based on the neural network techniques is proposed. And the experimental results of recognition test using real speech data are shown.