3837393 SCCS/M : MAJOR : COMPUTER SCIENCE ; M.Sc. (COMPUTER SCIENCE)
KEY WORD : SPEAKER RECOGNITION / SPEAKER IDENTIFICATION
VARAPORN PHOMVI-IN : VOICE RECOGNIZER FOR PERSONAL IDENTIFICATION. THESIS ADVISOR : SUPACHAI TANGWONGSAN Ph.D., DAMRAS WONGSAWANG Ph.D. 65 p. ISBN 974-663-997-8

This research presents a model of a voice recognizer for personal identification. It is the process of automatically determining who the speaker is by matching his/her speech pattern to reference speakers in a database, which is obtained from a group of known speakers. The outcome of speaker identification is the one whose speech pattern is the best match to the known speech samples.

The voice recognizer system in this study consists of three major parts, namely: feature extraction, clustering, and identification. In the feature extraction, we initially detect voiced segments of speech samples by using ZCR (Short-Time Average Zero-Crossing Rate), then given as input to perform DFT (Discrete Fourier Transform) and compute the PSD (Power Spectrum Density). Next, the PSD vectors are used as training samples in a pattern-based training model and also used to calculate the distribution of the acoustic features by employing ML (Maximum Likelihood Procedure) to represent the statistical-based training model. In addition, gender classification was implemented in order to improve the performance during testing. This is based on the pitch frequency value that can be computed by using Auto-correlation function.

From the clustering method, we pre-detected the possible set of speakers who are acoustically similar to the test speaker by combining the scoring procedure and the Mahalanobis distance measure to optimize the distribution of each phonetic data. Finally, in the identification process, we employed the square-error pattern matching to determine the cohort set which is the last and smallest set of possible speakers. Then the decision algorithm is performed to obtain the most likely speaker from the cohort set.

The system is implemented in a stand-alone personal computer, in which Thai language is chosen as the word utterance for the evaluation. For noise-free speech recording environment, the voice recognizer is able to achieve the identification within accuracy up to 95%, while the identification error down to 5%. Concerning the computational time, the system is able to determine the speaker within 30 seconds, which is quite short in the sense of real environment. Moreover, the system requires minimal disk space because of using a small number of speech feature representatives, which significantly reduces the system pre-processing time.