International Journal of Innovative Computing, Information and Control
Volume **7**, Number **7(B)**, July **2011**

ICIC International ©2011 ISSN 1349-4198
pp. **4207–4222**

ENHANCEMENTS OF MAXIMUM LIKELIHOOD EIGEN-DECOMPOSITION USING FUZZY LOGIC CONTROL FOR EIGENVOICE-BASED SPEAKER ADAPTATION

Ing-Jr Ding

Department of Electrical Engineering
National Formosa University
No. 64, Wunhua Rd., Huwei Township, Yunlin County 632, Taiwan
ingjr@nfu.edu.tw

Received May 2010; revised November 2010

Abstract. *This paper presents a fuzzy logic control (FLC) mechanism for the popular eigenvoice-based speaker adaptation scheme. The proposed mechanism regulates the influence of maximum likelihood eigen-decomposition (MLED) when the training data from a new speaker is inadequate. The FLC-MLED method functions by accounting for the adaptation data when estimating the linear combination coefficients for eigenvector decomposition. This approach ensures the robustness of speaker adaptation against data scarcity. The proposed mechanism is conceptually simple, effective and computationally inexpensive. Experimental results indicate that FLC-MLED outperforms conventional MLED, especially when encountering data insufficiency. The proposed approach performs better than maximum a posteriori eigen-decomposition (MAPED) at a much lower computing cost.*

Keywords: Speech recognition, Speaker adaptation Takagi-Sugeno fuzzy logic controller, Maximum likelihood eigen-decomposition, Maximum a *posteriori* eigen-decomposition

1. Introduction.

Computing techniques for automatic speech recognition (ASR) have existed for years [1-4]. As they have matured, these techniques have found more and more applications in everyday life [5]. Nevertheless, the recognition performance of any speech recognition system ever built is undeniably inferior to a human listener [6]. During speech recognition, variations in speech are strange to the system or known to the system only in poor “vocal shape”. These variations often cause a mismatch between the pre-established reference templates and the testing template, compromising recognition performance. Speaker adaptation (SA) sometimes referred to as model-based adaptation, can reduce this mismatch phenomenon. Speaker adaptation is the process of transforming a speaker independent (SI) speech recognition system into a speaker dependent (SD) system. This process achieves SD-like performance by adjusting the acoustic parameters of the SI speech model, typically in the form of hidden Markov models (HMM), with speech samples acquired from a target speaker.

There are three major types of speaker-adaptive techniques: Bayesian-based adaptation, transformation-based adaptation and speaker-clustering-based adaptation. Bayesian-based model adaptation directly re-estimates the acoustic model parameters using maximum *a posteriori* (MAP) adaptation [7,8]. The Bayesian reasoning framework is an example of this approach. Transformation-based model adaptation, such as maximum likelihood linear regression (MLLR) [9] and maximum *a posteriori* linear regression (MAPLR) [10,11], must derive certain appropriate transformations from a set of adaptation utterances from a new speaker and then apply them to clusters of HMM parameters.