Journal of Network and Computer Applications 34 (2011) 1459–1467

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Robust several-speaker speech recognition with highly dependable online speaker adaptation and identification

Po-Yi Shih [a,*], Po-Chuan Lin [b], Jhing-Fa Wang [a], Yuan-Ning Lin [a]

[a] Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan
[b] Department of Electronics Engineering and Computer Science, Tung Fang Institute of Technology, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 1 March 2010
Received in revised form
6 July 2010
Accepted 15 August 2010
Available online 20 August 2010

Keywords:
Speech recognition
Speaker adaptation
Speaker identification
Dependable adaptation
Confidence score

ABSTRACT

The currently adaptive mechanisms adapt a single acoustic model for a speaker in speaker-independent speech recognition system. However, as more users use the same speech recognizer, single acoustic model adaptation leads to negative adaptation upon switching between users. Such a situation is problematic (undependable adaptation). This paper, considering the situation of a smart home or an office with staff members, presents the speaker-specific acoustic model adaptation based on a multi-model mechanism, to solve the problem of undependable adaptation. First, the identification of the current speaker is confirmed using the SVM classifier, then the corresponding acoustic parameters are extracted and integrated with the speaker-independent acoustic model to yield the speaker-dependent acoustic model and speech recognition accuracy then be promoted for the current speaker. To provide dependable adaptation data to achieve online positive speaker adaptation, a mechanism that measures confidence score is designed to verify each recognition result and determined whether it can be an adaptation datum. The experimental results indicate that the proposed system can effectively increase the average speech recognition accuracy from 62% to 85%. Thus, the proposed system can achieve robust several-speaker speech recognition with highly dependable online speaker adaptation and identification.

1. Introduction

Speaker adaptation is crucial to robust speech recognition systems, which modify an original acoustic model into a specific speaker model based on the speaker's acoustic characteristics. The traditionally popular methods of speaker adaptation have been shown to be effective in recent speaker-independent speech recognition systems (IBM Via Voice; Microsoft Speech Recognition Engine). They include maximum a posteriori (MAP) (Woodland, 2000; Gauvain and Lee, 1994), maximum likelihood linear regression (MLLR) (Woodland, 2000; Leggetter and Woodland, 1995), and eigenspace-based maximum likelihood linear regression (EMLLR) (Chen et al., 2000; Mak et al., 2004).

Although a speaker-independent speech recognition system can be applied in a situation with fixed members, such as a smart home or office, the system is not suited to an environment with several speakers as it has only one acoustic model for adaptation. When several speakers use a single speech recognizer, recognition accuracy cannot be guaranteed for all speakers, and recognition performance is reduced by negative speaker adaptation, which is associated with the adaptation of a single acoustic model. The acoustic model for a current speaker must be retrained and adapted again when the user changes, and no particular adaptation can be sustained to increase recognition accuracy. The speech recognition system is therefore undependable, in that the adaptation mechanism reduces and also yields unstable recognition performance. Accordingly, a robust speech recognition system with highly dependable adaptation that can adapt to each speaker and positively is required. No investigation has yet considered the problem that arises when several speakers use a single acoustic model in speech recognition.

The optimal system should have two properties. First, it will independently recognize every speaker's speech based on user's speaker-specific acoustic model. Second, the speech model will be adapted to the specific speaker's acoustic parameters to improve recognition performance and retrain the identity of the model for the next time it is used by the same speaker. The development of such a system requires that two questions are answered: (1) how is the speaker identified? (2) how are the speaker's acoustic parameters adapted positively for unsupervised demand?

To answer these two questions, "robust several-speaker speech recognition with highly dependable online speaker

* Corresponding author. Tel.: +8866 2084431.
E-mail addresses: poyi.shih@gmail.com (P.-Y. Shih), tony178.lin@gmail.com (P.-C. Lin), wangjf@mail.ncku.edu.tw (J.-F. Wang), yukinaco@hotmail.com (Y.-N. Lin).

1084-8045/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2010.08.007