

A Study on Hakka and Mixed Hakka-Mandarin Speech Recognition

Tsai-Lu TSAI, Chen-Yu CHIANG, 余秀敏, Lieh-Shih LO, Yih-Ru WANG, Sin-Horng CHEN

Humanities and Social Sciences

kuo@chu.edu.tw

Abstract

A first study on Hakka and mixed Hakka-Mandarin speech recognition (SR) is reported in this paper. The main focus of the study is on solving the problem of the lack of a large text corpus for training a reliable language model. In the Hakka SR, several methods to use the information of part of speech and Hakka-Chinese word translation to assist in language modeling are proposed. For mixed language SR, a method to train a mixed Hakka-Mandarin acoustic model is suggested. Experimental results show that the proposed language and acoustic modeling approaches are promising for Hakka and mixed Hakka-Mandarin SR.

Keyword : speech recognition, language model, acoustic model, Hakka ,
Mandarin SR