International Journal of Innovative Computing, Information and Control
Volume **4**, Number **7**, July **2008**

 ISSN 1349-4198
pp. 1695–1709

EMOTION RECOGNITION AND EVALUATION FROM MANDARIN SPEECH SIGNALS

Tsanglong Pao, Yute Chen and Junheng Yeh

Department of Computer Science and Engineering
Tatung University
Taipei 104, Taiwan
tlpao@ttu.edu.tw; {d8906005, d9306002}@ms2.ttu.edu.tw

Received May 2007; revised September 2007

Abstract. *The exploration of how human beings react to the world and interact with it and each other remains one of the greatest scientific challenges. The ability to recognize affective states of a person we face is the core of emotional intelligence. In the past, several classifiers were adopted independently and tested on several emotional speech corpora with different language, size, number of emotional states and recording method. This makes it difficult to compare and evaluate the performance of those classifiers. In this paper, we implemented a weighted discrete k-nearest neighbor (weighted D-KNN) classification algorithm and compared it with KNN, M-KNN and SVM classification methods by applying them to a Mandarin speech corpus. This speech corpus contains of five basic emotions: anger, happiness, boredom, sadness and neutral. The results of experiments and McNemar's test revealed that the implemented weighted D-KNN method performed best among these classifiers and achieved an accuracy of 81.4%. Besides, we implemented an emotion radar chart which is based on weighted D-KNN and can present the intensity of each emotion component in the speech in our emotion evaluation system. Such system can be further used in speech training, especially for hearing-impaired to learn how to express emotions in speech more naturally.*

Keywords: Emotional speech recognition and evaluation, Radar chart, Weighted D-KNN classifier, McNemar's test

1. **Introduction.** Communication is a highly complicated inter-person process through conversation, writing, gestures, appearance, behavior, and, at times, even silence. Communication involves much more than just words. What is said and how it is said are equally important. The two basic modes of communication are verbal and nonverbal. Verbal communication is either spoken or written and involves the use of words and means what is said. Nonverbal communication, on the other hand, does not involve the use of words and means how it is said. The forms of nonverbal communication include kinesics (gestures, facial expression, body movement and posture), vocalics (volume, rate, pitch, pausing and silence), physical appearance (hairstyle, clothing, cosmetics and fragrance), haptics (such as frequency, intensity and type of touch), proxemics (spatial cues), chronemics (waiting time and punctuality), and artifacts (manipulative objects in the environment) [1].

The expression and perception of emotion are salient aspects of human vocal communication. In human speech, emotion can be conveyed in two important ways: by content and by prosody. The content refers to the semantic message of what is said. Prosody refers to the way a message is expressed in acoustic terms. Detecting the emotional state of a person has numerous potential applications, such as helping autistic people, consumer feedback, emotion recognition game, emotion recognition software for call center,