

NON-VERBAL VOICE EMOTION ANALYSIS SYSTEM

SHUNJI MITSUYOSHI¹, FUJI REN^{1,2}, YASUTO TANAKA³ AND SHINGO KUROIWA¹

¹Faculty of Engineering
Tokushima University
Tokushima 770-8506, Japan
Shunji@agi-web.co.jp; { ren; kuroiwa }@is.tokushima-u.ac.jp

²School of Information Engineering
Beijing University of Posts and Telecommunications
Beijing 100876, P. R. China

³Brain Information Group
National Institute of Information and Communications Technology
588-2, Iwaoka, Nishi-ku, Koube-shi, Hyougo, Japan
ytanaka@po.nict.go.jp

Received November 2005; revised March 2006

ABSTRACT. *A non-verbal voice analysis system that recognizes, separates and ranks concurrent emotions in real time has potential application in various fields, yet such a system that could delineate emotion based solely on the sound of a human voice has not been successfully demonstrated before. Here, we propose a system that recognizes human emotion by means of analyzing the fundamental frequency of a human voice taken from continuous natural speech. The system detects robust fundamental frequencies and intonations by parameterizing them into pitch, power, and deviation of power. Based on these parameters, data was classified via decision-tree logic into the emotional elements of anger, joy, sorrow, and calmness. Degree of excitement was also extracted. The system was evaluated by third parties by matching the system performance to human subjective classification for each element. Results indicate that overall matching rate was 70%, and the matching rate was 86% when compared to the subjects' assessment of their own voices. Our system performance exceeded the baseline with non-verbal information, which was equivalent to human subjective assessment.*

This is the first report of a system to rigorously analyze human emotion in real time and we expect numerous applications.

Keywords : Comparison of subjective evaluation, Spontaneous emotional speech, Conscious emotional speech, Emotional parameter

1. **Introduction.** Since the invention of voice recognition and production, a number of systems have been proposed including the zero-crossing method, the DP-matching method, the Hidden Markov Model, and the n-Gram method [1]. Recent development shows that a real-time, interactive conversation method was invented especially for application for vehicle navigation systems. However, no system could previously recognize human emotions based on the sound of the human voice.

Emotion is not easy to study. Various researchers have dealt with the problem using rats, cats, monkeys and humans, yet the definition of emotion still differs among studies. W. James [2] proposed the first scientific model of emotion in which humans feel emotion