



**SPEAKER ADAPTIVE PHONEME RECOGNITION IN CONTINUOUS SPEECH BASED
ON VECTOR QUANTIZATION**

Katsuhiko Shirai, Kazunori Mano and Koji Sano

ABSTRACT

To realize a large vocabulary continuous speech recognition, phoneme level recognition and speaker adaptation techniques are fundamental. And also, it is necessary to develop an effective procedure of selecting possible word candidates from the large word dictionary.

This paper proposes a new method applicable for continuous speech recognition which is based on a vector quantization (VQ) technique.

Basic acoustic quantity to characterize speech signal has been rather restricted to the pattern of power spectra or that equivalents. The novel aspect of this research is to explore new acoustic feature space.

The most effective feature newly utilized in this paper is the density function of the VQ code. The definition of the density function is straightforward. If a speech input includes totally N_a frames and N_i frames of them are quantized to i -th centroid of the codebook, the density, $p(i)$, of the i -th code is represented by $p(i) = N_i/N_a$, and the density function is defined as a vector, $(p(1), p(2), \dots, p(N))$. Using this function, we can characterize easily spectral difference among speakers and also words can be effectively classified. In the preliminary experiment, this density function has been shown useful for the speaker identification. In this paper, it will be described that this feature can be successfully applied for the phoneme recognition as well as the word recognition for unspecified speakers.

Department of Electrical Engineering, Waseda University, 3-4-1
Ohkubo, Shinjuku-ku, Tokyo 160, JAPAN