



SPEAKER-ADAPTATION OF A CODE BOOK OF VECTOR QUANTIZATION

Yasuhisa Niimi and Yutaka Kobayashi

ABSTRACT

This paper describes a speaker-adaptive word recognition system in which word templates are indirectly atuned to a new speaker through an adaptation of the code book for vector quantization. The adaptation of the code book is made by the following four steps.

- (1) For each word class, word patterns spoken by several speakers are time-warped with the DP-algorithm and averaged to produce a speaker-independent word template. Averaged spectral patterns, denoted by P_j ($j=1-n$), are also prepared for typical stationary phonemes like vowels, nasals and fricatives.
- (2) A speaker-independent code book for vector quantization is calculated using sample points selected from the averaged word templates. Let c_i ($i=1-N$) be the i -th code point.
- (3) The averaged word templates are coded using this code book.
- (4) For each of the phonemes selected in the step (1), a spectral pattern, denoted by p_j' ($j=1-n$), is extracted from a small set of training data spoken by a new speaker. A new code point c_i' for c_i is computed by the following formula;
$$c_i' = c_i + w_{ij}d_j \quad (i=1-N)$$
where d is a difference vector between p_j and p_j' , and w_{ij} is a weighting coefficient inverse to a distance between c_i and p_j . Thus the word templates represented by the sequences of the codes have indirectly been adapted to the speaker.

In the recognition phase, an incoming word is vector-quantized with the new code book and matched against each of the references. We carried out the experiments in which 65 names of Japanese cities were tried to be recognized, and found the proposed method for a speaker-adaptation promising.

Department of Computer Science, Kyoto Institute of Technology,
Matsugasaki, Salyo-ku, Kyoto 606, JAPAN