

E14. A single-boarded isolated word recognizer using LPC cepstrum. M. Hamada, Y. Bessho, T. Norimatsu, and A. Yamada (Acoustic Research Lab., Matsushita Electric Ind. Co., 1006 Kadoma, Osaka 571, Japan)

Minicomputer simulation was carried out to design an optimum system based on currently available signal processing LSI's. First, finite-word-length effects of Levinson-Durbin (LD) algorithm and Le-Roux (LR) algorithm [J. Le-Roux *et al.*, IEEE Trans. Acoust. Speech Signal Process. ASSP-25, 257-259 (1977)] for extracting PARCOR coefficients were investigated regarding (1) the PARCOR/AR/cepstrum coefficient error, (2) the difference in LPC cepstrum distance between the top two candidates, and (3) recognition rate. LR was found to be almost always better than LD by each of above measures. Second, the effects order of analysis, number of template bits, and the template normalization method were examined to minimize memory size. It was shown that the number of template bits of each cepstrum coefficient can be reduced to four with little decrease of recognition rate as compared to the system with floating point number templates. A single-boarded recognizer using TMS320 for LPC analysis and MN1263 for DP matching was implemented. The overall recognition rate of on-line test in speaker-dependent mode was 99.4% for a ten-word vocabulary (total of 1000 tokens of ten speakers). Multiple template speaker-independent mode achieved a recognition rate of 97.0% for an eight-word vocabulary.

11:23

E15. Network representation of templates in word recognition. Kai-Fu Lee (Computer Science Department, Carnegie-Mellon University, Pittsburgh, PA 15213)

Reference set generation is a crucial step in template-based isolated word recognition. Some common techniques include casual training, selection, clustering, and averaging. These single-template systems do not capture the variations in speech. Alternatively, multiple-template techniques lead to additional storage and recognition time. Network representation combines the performance of multiple-template techniques and the efficiency of single-template techniques. Using the network approach, words are divided into segments, and different examples of the same or different words can share segments. This not only reduces storage required, but also enables the system to focus on acoustically dissimilar segments. Network representation has not been popular because of the difficulties in (1) segmenting correctly and recovering from segmentation errors, and (2) creating and modifying the network automatically. A word recognition system has been designed and implemented to facilitate network training by providing (1) relatively reliable segmentation, (2) segment-based warping algorithm that tolerates inexact segmentation, and (3) incremental network generation. Preliminary results show that network training is superior to all of the above-mentioned methods for speaker-dependent and independent recognition. [Work supported by NSF.]

E16. The use of allophonic variations of /a/ in automatic continuous speech recognition of French. Jacqueline Vaissière^{a)} (Room 36-541, Research Laboratory of Electronics, MIT, Cambridge, MA 02139)

The acoustic characteristics of the vowels in continuous speech are affected by their duration, the stress condition, and the surrounding phonemes. The purpose of this study is to investigate the extent of the allophonic variations of the most frequent and most variable vowel in French, /a/, and the integration of such variations in an automatic speech recognition system. The preliminary corpus consists of 200 vowels /a/ extracted from 128 sentences uttered by two speakers. Using the SPIRE facilities, the sentences were digitized and segmented [Zue and Leung, J. Acoust. Soc. Am. Suppl. 1 75, S59 (1984)]. The segments labeled as SONORANT corresponding to the occurrence of /a/ were extracted. The values of F_1 and F_2 at the onset, offset, and middle of the segments were estimated using automatic peak tracking from the LPC spectra. Our results suggest that: (1) the MANNER of articulation of the preceding and following consonant plays a negligible role on F_1 and F_2 onset and offset (in contrast with other data published for English and Swedish); (2) if the PLACE of articulation of the consonants is divided into FRONT, MID, and BACK, then the difference ($F_2 - F_1$) at /a/ onset allows the unique specification of the place of articulation of the preceding consonant in 86% and 57% of the cases, for SP1 and SP2, respectively, and there is no overlap between FRONT and BACK consonants; (3) ($F_2 - F_1$) at /a/ offset indicates that F_2 of the following vowel has to be taken into account before interpretation; (4) determination of the thresholds (for MID and BACK consonants) can be done on a few well selected words, where the effect of the context is known to be maximal. Detailed descriptions of our findings extended to the study of more speakers, and of their consequences for the coding of the permissible variations of the vowel /a/ for each word of the vocabulary will be presented. [Work supported in part by the Office of Naval Research under contract N00014-82-K-0727.]^{a)} On leave from CNET, Lannion, France.

11:47

E17. Phonetic string alignment. Kathleen M. Goudie-Marshall, Joseph Picone, and William M. Fisher (Texas Instruments Inc., P. O. Box 226015, M/S 238, Dallas, TX 75266)

This paper will describe a newly developed expert system which uses linguistic knowledge to align the phonetic content of two different words or phrases and score their similarity. The development of the system arose out of a need to have an automatic scoring algorithm for intelligibility testing for text-to-speech systems and other synthetic-speech coding and recognition systems. The system uses an automatic text-to-phone algorithm to translate the input and reference ASCII text strings to phonemic units, aligns them using linguistic knowledge-based decision criteria and a dynamic programming optimization, and outputs the aligned strings as well as tabulating the phoneme confusions.