ent limitations or accuracy of these procedures. When speech in an office environment is processed using reasonable quality microphones, tape recorders, analog to digital recording, and digital processing methods, the results are frequently disappointing [Holmes, J. Acoust. Soc. Am. 58, 747–749 (1975)]. The purpose of this paper is to describe several of the reasons why this disappointment occurs and to show how a combination of high quality direct recording into the computer and linear phase filtering, in conjunction with the covariance method at automatically derived anchor points, can result in theoretically expected glottal volume velocity waveforms for high intensity vowel sounds where glottal closure is known to occur. Timing relations with respect to the speech waveforms are also compared with other measures of glottal closure [Strube, J. Acoust. Soc. Am. 56, 1625–1629 (1974)].

2:42

RR7. Automatic recognition of semivowels in word context. Hiroya Fujisaki and Yasuo Sato (Department of Electrical Engineering, Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo, 113 Japan)

Based on an approximate formulation of the coarticulatory process at the acoustic level and a method for adaptation to individual differences, a scheme for reliable segmentation and recognition of connected vowels has already been established [H. Fujisaki *et al.*, Speech Communication Seminar, Stockholm 1974]. The present paper describes a study for the extention of the scheme to recognition of vowels and semivowels in word context. Values of formant targets which represent a command for each phoneme, its duration, as well as the rate of transition between successive phonemes are extracted as parameters from time-varying patterns of formant frequencies. Among these parameters, the command duration is shown to be most effective in discriminating semivowels /j/ and /w/ from vowels /i/ and /u/. A scheme for recognition of vowels, semivowels, and their sequences based on formant targets and command duration is then proposed and tested experimentally using meaningful and nonsense words uttered by four speakers.

2:54

RR8. Continuous speech recognition via centisecond acoustic states. R. Bakis (Computer Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598)

Continuous speech was treated as if produced by a finite-state machine making a transition every centisecond. The observable output from state transitions was considered to be a power spectrum—a probabilistic function of the target state of each transition. Using this model, observed sequences of power spectra from real speech were decoded as sequences of acoustic states by means of the Viterbi trellis algorithm. The finite-state machine used as a representation of the speech source was composed of machines representing words, combined according to a "language model." When trained to the voice of a particular speaker, the decoder recognized seven-digit telephone numbers correctly 96% of the time, with a better than 99% per-digit accuracy. Results for other tests of the system, including syllable and phoneme recognition, will also be given.

3:06

RR9. Vocabulary and syntactic complexity in speech understanding systems. Gary Goodman and D. Raj Reddy (Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213)

Comparing the realtive performances of speech-understanding systems has always been difficult and subject to speculation. Different tasks naturally require different vocabularies with varying acoustical similarities. Moreover, constraints imposed by the syntax may make recognition easier, even for vocabularies with high ambiguity. We define "inherent size" as a measure of vocabulary complexity and investigate its relation to recognition rates and the (apparent) vocabulary size. Word recognition is modeled as a probabilistic function of a Markov process using phoneme confusion probabilities derived from an articulatory position model. Multiple pronunciations and allophonic variations are allowed. Analysis of maximal word confusions leads to lower bounds for expected word recognition rates. Inherent vocabulary size is derived from these bounds. To evaluate syntatic constraints for finite state languages, an inherent size is computed for each state based on the subset of words which lead from that state. Then, average path complexity is computed for the language. This language complexity, when compared with the total vocabulary complexity and average path length, yields the degree of syntactic constraint. Analysis of several tasks will be reported.

3:18

RR10. Harpy, a connected speech recognition system. Bruce P. Lowerre and B. Raj Reddy (Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213)

The Harpy connected speech recognition system is the result of our attempt to understand the relative importance of various design choices of two earlier speech recognition systems developed at Carnegie-Mellon University, The Hearsay-I system (D. R. Reddy, *et al.*, IEEE Trans. AU, 229–238 (1973); L. D. Erman, Technical Report, Computer Science Department, Carnegie-Mellon University, (1974) and the Dragon system [J. K. Baker, Ph.D. Thesis (Carnegie-Mellon University)]. Knowledge is represented as procedures in Hearsay-I and as a Markov network with *a-priori* transition probabilities between states in Dragon. Hearsay-I uses a best first search while Dragon searches all the possible (acoustic syntactic) paths through the network to determine the optimal path. Hearsay-I uses segmentation and labeling and Dragon is a segmentation-free system. Systematic performance analysis of various design choices resulted in the Harpy system which represents knowledge as a finite state transition network but without the a-priori transition probabilities, searches only a few "best" paths, and uses segmentation to reduce the number of state probability updates that must be done. The system achieves between 70% and 100% sentence accuracy, depending upon the task, and runs between 1.5 and 10 times real time. Complete details of design, implementation, and experimental results are given in B. P. Lowerre, Ph.D. Thesis (Computer Science Department, Carnegie-Mellon University, 1976).

3:30

RR11 Parameter-independent techniques in speech analysis. Henry C. Goldberg and D. Raj Reddy (Computer Science Department, Carnegie-Mellon University, Pittsburgh, PA 15213)

Most present programs for feature extraction, segmentation, and phonetic labeling of speech are highly dependent upon the specific input-parametric representation used. We have been developing approaches which are relatively independent of the choice of parameters. Vectors of parameter values are dealt with uniformly by statistical pattern-recognition methods. Such an approach permits systematic evaluations of different representations (be they formants, spectra, LPC's, or analog filter measurements). Cost can be balanced against performance. Given the large number of design choices in this area, careful attention must be paid to methods which are invariant under change of parametric representation. For segmentation, regions of acoustic change are detected by functions of parameter vector similarity and by amplitude cues. The basic model of signal detection theory can be applied to quantify the missed/extra-segment error trade-off. Error rates of 3.7% missed for 19% extra have been achieved. To account for allophonic variability when acquiring labeling knowledge (training), a clustering algorithm was developed to empirically discover the acoustic variations inherent in each phonetic class. Labeling then proceeds by nearest-neighbor match. Accuracy for 29 phonetic classes was 35% correct in the first choice and 60% correct in the first three choices. For details, see H. G. Goldberg, Ph.D. Thesis (Carnegie-Mellon University, 1975).

3:42

RR12. Improvement of intelligibility and individuality of helium speech by use of autocorrelation function. J. Suzuki and M. Nakatsui (Radio Research Laboratories, Koganei, Tokyo 184 Japan)

Helium speech is hardly intelligible owing to the expansion of