is difficult to represent hidden linguistic structure and make an N-gram word prediction model using traditional stochastic approaches. In this paper, two neural network models that can learn hidden linguistic structure are proposed. These models can easily be expanded from Bigram to N-gram networks. They were tested by training experiments with an open English text database. The Trigram word category prediction rates show that neural network models are comparable to stochastic models. Trigram neural network models compress information about 150 times, which is the ratio of the Trigram stochastic model free parameters ($89^3 = 704\,969$) to the neural network model link weights (4649). In addition, this paper proposes a new method that dynamically controls the training parameters, updating step size and momentum. These techniques are effective for calculating the efficiency of this system.

U11. Connectionist techniques for speaker-independent recognition of isolated utterances. Michael A. Franzini (Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213)

The performance of previous speech-recognition systems has been limited to a large extent by the incomplete structural theories of speech upon which the systems were based. The work reported here suggests that connectionist learning procedures provide an effective method for generating *speaker-independent* recognition systems without the need for any *a priori* model of human speech production. The backpropagation learning algorithm is applied to the problem of isolated-utterance speech recognition, and the resulting recognition rates approach those of the best conventional systems. Several studies were peformed using computer simulations of various networks trained by backpropagation. With 1 s of digitized speech as input, the networks had to generate as output the appropriate labels, which in these studies were letters of the alphabet. The accuracy for speaker-independent recognition of the whole alphabet reached 89%, which is higher than the accuracies achieved by several more conventional recognizers using the same database [K. F. Lee, Carnegie Mellon Univ. Tech Rep. 85-181 (1985)]. Recognition rates for speaker-dependent recognition of the whole alphabet reached 99% and, for speaker-independent recognition of confusable letter sets such as b–p–e–v–d, the networks achieved 94%. These studies demonstrate that networks with simple task-independent learning procedures can perform as well as systems that explicitly implement procedures such as dynamic time warping and vector quantization of inputs.

U12. Speaker identity feature combination with a neural net. George Velius (Bellcore, 445 South Street, Room 2E-244, Box 1910, Morristown, NJ 07960-1910)

This study compares the performance of a neural net approach with conventional linear methods as applied to the problem of feature combination in the domain of speaker identity verification (SIV). The experiment endeavors to combine features consisting of LPC-cepstral coefficient differences and pitch differences for isolated words in a template-matching scenario. The signal features are analyzed for 30-ms frames every 10 ms. The pitch estimate is based on the cepstrum of the LPC residual. Previous work [G. Velius, ICASSP **88**, 583–586 (1988)] showed that the Fisher linear discriminant (FLD) was better at feature weighting (for cepstral coefficients only) than several other common linear methods. Results show that, when feature combination is done by the neural net, the SIV task is performed significantly better than when the feature combination (i.e., weighting) is done by the FLD. The neural network architecture used in this experiment was in no way "optimized" for the specific task at hand. An additional finding is that the pitch feature used here, in conjuction with the cepstral coefficients, contributes significantly to the SIV task; that is, the error rate is reduced by 13%.

U13. The LPC trace as an HMM development tool. George R. Doddington, Joseph Picone, and John J. Godfrey (Speech Research Branch, Computer Science Center, Texas Instruments Incorporated, P. O. Box 655474, Mail Stop 238, Dallas, TX 75265)

Hidden Markov models have gained wide acceptance in speech recognition due to the ability to construct optimum (maximum likelihood) models automatically from speech data. Understanding how recognition performance is related to a model's structure is not a simple matter, however, especially where the structure is complex. Thus analysis of errors, especially in terms of the properties of the speech data, is not often undertaken. This paper describes a method for relating speech recognition performance to HMM structure. The basic tool is a best-path (Viterbi) trace of the model through the input data, which is represented using well-known LPC parameters. Linear predictor parameters are used as auxiliary model parameters, independent of the HMM output parameters used for recognition, in order to take advantage of established LPC speech synthesis and LPC speech spectrogram utilities. The trace can then be used to gain insight into error mechanisms, often leading to improvements in model structure and system performance. The description will include techniques for creating the LPC auxiliary model, spectrographic and auditory output from supervised and unsupervised model traces, and an evaluation of the impact of this technique on the development of an HMM-based speech recognition system.

U14. Improved training procedures for hidden Markov models. Lawrence R. Rabiner, Chin-Hui Lee, Biing-Hwang Juang, David B. Roe, and Jay G. Wilpon (Speech Research Department, AT&T Bell Laboratories, Murray Hill, NJ 07974)

Techniques for training hidden Markov model (HMM) parameters from a labeled training set of data are well established and include the forward–backward algorithm as well as the segmental K-means algorithm. These algorithms have been shown to be capable of estimating the parameters of an HMM based on mathematically well-founded techniques. In practice, however, difficulties are often encountered when estimating some of the HMM parameters. These difficulties are generally the result of having insufficient training data to give robust and reliable parameter estimates. Typically, the model parameters most affected by having insufficient training data are the spectral parameter variance estimates, and the estimates of parameters related to the modeling of state duration. Although techniques have been proposed for improving estimates of the variances due to the effects of insufficient training data, the results have not proven adequate in some cases. As such, improved training techniques (which give better recognition performance) have been devised for controlling the minimum variance estimate of any spectral parameter, and for thresholding and clipping state duration parameter estimates. These improved training methods have been tested on several databases with good success. In addition, advanced techniques for creating multiple HMMs from the training data (i.e., for speaker independent recognition) have been devised and have proven successful for modeling large databases of training material.

U15. Duration control methods for HMM phoneme recognition. Toshiyuki Hanazawa, Takeshi Kawabata, and Kiyohiro Shikano (ATR Interpreting Telephony Research Laboratories, Twin 21 Building MID Tower, 2-1-61 Shiromi, Higashi-ku, Osaka, 540 Japan)

Two kinds of duration control for HMM (hidden Markov model) phoneme recognition are proposed: phoneme duration control for an HMM phone model and an event duration control for an HMM state. The phoneme duration control is carried out by combining an HMM output probability with a phoneme duration probability. The phoneme duration