phoneme string, the synthesizer selects the corresponding sequence of nuclei and transitions and interpolates between each of the 14 cross-sectional areas, producing a model of the shape of the vocal tract changing in time. The filtering process of this vocal tract model is identical to the optimum inverse filter of linear prediction analysis, allowing direct conversion to linear predictive coding (LPC) synthesis. Experiments were performed which tested the effect of varying certain attributes of the synthesizer, viz., linear versus spline interpolation, one point versus three point transitions, nucleus versus no nucleus, and different gain rules. Results will be discussed and a tape played. [Work supported by Rome Air Development Center under Contract No. F30602-77-C-0056.]

8:33

GGG12. Subjective quality testing of a new source model of LPO vocoders. A. W. F. Huggins, R. M. Schwartz, R. Viswanathan, and J. Makhoul (Bolt Beranek and Newman Inc., Cambridge, MA 02138)

Our source model, reported at an earlier meeting, excites the LPO speech spectrum with a low-frequency band of pulses mixed with a high-frequency band of noise. Pulses are low-pass filtered and noise is high-pass filtered at the same frequency, to yield a flat source spectrum. The cutoff frequency of the filters, a continuous variable, replaces the usual binary voiced/voiceless decision. Thirty-six phoneme-specific test sentences were processed through a single high-quality vocoder (5 kHz bandwidth, 11 poles, no quantization, 100 frames/s), which was excited in turn by both the usual pulse/noise source and by the new source. Subjects rated the resulting speech separately on eight-point buzziness and breathiness scales. The results show that the new source model greatly reduces perceived buzziness, occasionally at a cost of slightly increased breathiness. Any remaining inadequacies can probably be ascribed to the algorithm that extracts the cutoff frequency during analysis, rather than to the model itself. [Work supported by ARPA-IPTO.]

8:36

GGG13. Listener preferences and spectral similarities of eigenparameter quantization schemes for speech. William J. Strong and Kem E. Robinson (Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602)

Several quantization schemes for eigenparameters of speech have been studied as part of an effort to determine a set of spectral patterns that will be adequate to synthesize speech having the characteristics of a particular talker. This paper describes an experiment in which sentence material from each of two male and two female talkers was analyzed using an autocorrelation method to obtain 12 log area ratios at 10 millisecond intervals. An eigenparameter analysis was then performed on the log area ratios. The eigenparameters were quantized in each of six ways based on their variance and standard deviation. Ten versions of speech were created for each talker: one was natural, three were synthesized from nonquantized parameters, and six were synthesized from quantized parameters. Each version of speech was paired with all other versions and listeners were asked which of the two versions in a pair they preferred. The same quantization schemes did not produce the same listener preferences for all talkers. Average spectral similarities of each quantized version (relative to a 12 parameter version) were also calculated using the minimum prediction residual and an ad hoc measure that maintains distance symmetry between "sample" and "template." The results will be considered in terms of how well measures of spectral similarity correlate with listener preferences.

8:39

GGG14. Measuring the effects of echoes on speech intelligibility. Paul G. Lacroix (Naval Submarine Medical Research Laboratory, Groton, CT 06340) and James E. Atkinson (Naval Underwater Systems Center, New London, CT 06320)

Speech intelligibility was measured as a function of the time interval between the onset of an original signal and the onset of its equal amplitude echo. Scores were obtained on three different speech tests for delay intervals ranging from 50 to 700 ms. Tests included the monosyllabic Modified Rhyme Test (MRT), the multi-word CID sentences and the NSMRL Tri-word Test of Intelligibility (TTI). Delay conditions were produced by a specially constructed digital delay line. Intelligibility was most degraded when echoes occurred 200 to 250 msec after the onset of the original signal. The multi-word tests (CID and TTI) proved more sensitive than the single-word (MRT) test under all conditions. TTI test words were analyzed according to their position within an utterance. Results suggest that the TTI provides information about echo delay conditions as they affect both single and multiple-word tests. The correlation between the means for final word TTI items and the means obtained for the MRT across delay conditions was $r = 0.80$, while the correlation between means for middle word TTI items and the means obtained for the CID sentences was $r = 0.95$. These results are discussed in terms of temporal masking effects.

8:42

GGG15. The intelligibility of simulated underwater voice communication as measured by the NSMRL Tri-Word Test of Intelligibility (TTI). James E. Atkinson (Naval Underwater Systems Center, New London, CT 06320) and Paul G. Lacroix (Naval Submarine Medical Research Laboratory, Groton, CT 06340)

Our previous paper suggested the TTI for use in assessing the effect of temporal distortions on speech intelligibility. Here, the TTI is applied in an investigation of the phenomenon of multipath, a temporal distortion which can severely limit the intelligibility of underwater voice communication. Evidence suggests that intelligibility may be improved if speech is time-reversed prior to underwater transmission. Results are presented for simulated forward and time-reversed speech conditions across an echo delay range of 50–500 ms as a function of the amplitude of an echo relative to that of the original signal. Asymmetric intelligibility functions are demonstrated which are dependent on the length of echo delay interval. These results are discussed in terms of temporal masking and perceptual fusion effects.

8:45

GGG16. Recursive estimation of autoregressive moving-average parameters in a pole-zero representation of the speech production process. Hiroyoshi Morikawa and Hiroya Fujisaki (Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo, 113 Japan)

A method is presented for estimating the orders as well as the parameter values of an autoregressive moving-average process excited by Gaussian white noise or a train of randomly spaced pulses. This process is adopted as a model for the process of speech production. The estimation is accomplished directly from a signal waveform using recursive formulas and is optimum in the sense that the variance of errors in the estimated waveform is minimized. In comparison with methods based on the short-time autocorrelation function, this method is less vulnerable to fluctuations in the fundamental period of the source. The validity of the proposed method is demonstrated by comparing its results with those obtained on the basis of a fixed-order all-pole formulation, both in terms of the mismatch between the input and the estimated speech spectra and in terms of subjective evaluation of the reconstructed speech. [Work supported by Ministry of Education Grant-in-Aid for Scientific Research No. 239005.]