time delay if the number of QMF connections is different. This paper presents a new architecture for SBC with flexible band assignment, less arithmetic calculations, and small storage memory. In this SBC architecture, a finite impulse response filter (FIRF) is used instead of the QMF. The center frequency of each band is selected according to the equation $F_s/F_k = D*n/m$, where F_s is the sampling frequency of the input signal, F_k is the center frequency of the k th band, D is the decimation factor, and m, n are integers. According to this relationship, frequency shift and the LPF can be performed at the same time. In this case, data memory for the FIRF is drastically reduced because the input of the FIRF is the same for every band. With the new SBC architecture, the 16- to 12-kbps sub-band coder/decoder was implemented on a single DSP. Band selection and suitable bit allocation are discussed.

U49. Using dynamic time warping to formulate duration rules for speech synthesis. Marion J. Macchi, Murray F. Spiegel, and Karen L. Wallace (Bell Communications Research, Morristown, NJ 07960-1910)

Dynamic time warping (DTW), a speech recognition technique, was used to time align long-duration syllables with shorter-duration renditions of those syllables. A male speaker of English produced stressed syllables with long duration, like "libe," and the same syllables with shorter durations in nonsense words like "libous," "libegous," "libegass," and "libotous." Syllables were formant tracked and a frame-by-frame Euclidean bark-formant distance between each long syllable and each of its shorter counterparts was computed. DTW was applied to the resulting distance matrix. The DTW path, referenced to the long syllable, indicates which portions of the long syllable are shortened (and by how much) to match a shorter version of the syllable. The DTW paths showed regions of maximal and minimal shortening. These reactions were approximately aligned across syllable durations. The DTW paths for syllables of different durations were distinguished from one another primarily in regions of maximal shortening. The regions of maximal shortening were aligned with regions in the syllable with little formant movement. Consequently, a spectral-movement time function derived from the templates in an inventory could serve as the basis for duration rules for demisyllable speech synthesis. Further, this technique provides a method for making duration measurements automatically.

U50. Mixed spectral representation—Formants and LPC coefficients. Joseph P. Olive (AT&T Bell Laboratories, Murray Hill, NJ 07974)

A cascade formant model is well suited to describe certain speech segments, such as vowels and vowel-like sounds. The formant model is also useful because the relationship between formants and the vocal tract configurations are well understood; however, this model is not adequate for other speech sounds, such as stops, fricatives, nasals, etc. On the other hand, LPC analysis, of sufficiently high order, can adequately describe the spectrum of any speech sound, but the relationship between the LPC parameters and the spectrum or vocal tract configuration is not obvious. This paper describes a speech analysis/synthesis scheme that uses both formants and LPC parameters for different sections of a speech signal. Thus, in some regions, the benefit of the formant model can be utilized, while, in other regions, the LPC representation can be used to obtain a good description of the speech spectrum. The analysis algorithm resolves the problem of discontinuities that arise from using the two different spectral representations. The method of speech analysis described in this paper produces resynthesized speech of the quality of multipulse LPC.

U51. Speechreading sentences I: Development of a sequence comparator. L. E. Bernstein (Center for Auditory and Speech Sciences, Gallaudet University, Washington, DC 20002), Marilyn E. Demorest (Department of Psychology, University of Maryland—Baltimore County, Catonsville, MD 21228), and Silvio P. Eberhardt (Jet Propulsion Laboratory, Pasadena, CA 91109)

Previous research suggests that many lexical errors in the speechreading of sentences can be explained in terms of visual phonemic errors. However, description and quantification of perceptual errors at the phonemic level requires specification of stimulus-to-response alignments. Because speechreading produces numerous errors, including phoneme insertions, deletions, and/or substitutions, alignment is a nontrivial problem. This paper describes development of a sequence comparator that can be used to obtain alignments automatically for phonemically transcribed sentences. The comparator employs a weights matrix that reflects presumed visual distances between all possible segmental stimulus–response pairs to find the alignment that minimizes overall stimulus–response distance. Initially, the comparator used weights based on viseme groupings, but these weights resulted in multiple, equal-distance, alternative alignments. More effective weights were obtained empirically via multidimensional scaling of phonemic confusions. Vowel data were obtained from Montgomery and Jackson [J. Acoust. Soc. Am. **73**, 2134–2144 (1983)] and consonant data from a nonsense syllable identification task, which employed 22 consonants spoken by the same talkers who produced the sentence stimuli for this study [Bernstein *et al.*, J. Acoust. Soc. Am. **85**, 397–405 (1989)]. [Work supported by NIH.]

U52. Speechreading sentences II: Application of a sequence comparator to data on CID sentences. Marilyn E. Demorest (Department of Psychology, University of Maryland—Baltimore County, Catonsville, MD 21228) and L. E. Bernstein (Center for Auditory and Speech Sciences, Gallaudet University, Washington, DC 20002)

When subjects speechread sentences, their performance is typically evaluated in terms of the number of words or keywords correct. A comparator that aligns stimulus and response sequences is being used as a heuristic for studying relationships between the lexical and the phonemic level of speechreading. Toward this end, a corpus of response sequences, previously analyzed in terms of words correct, was reanalyzed with the comparator. Stimuli were 50 CID sentences spoken by a male and a female talker and recorded on video laserdisc. Subjects were normal-hearing college students. One result showed that when sentences are short, computed visual stimulus-response similarity is well-correlated with number of words correct. But with few exceptions, when sentences are long, the correlation is reduced, suggesting that the two measures provide complementary information. The sequence comparator appears to be a useful tool for elucidating patterns of speechreading performance. [Work supported by NIH.]

U53. Investigating randomness in foot timing patterns in English. Briony Williams and Steve Hiller (Centre for Speech Technology Research, 80 South Bridge, Edinburgh EH1 1HN, Scotland)

Isochrony has been considered only in terms of stressed syllables. However, it may also be a random property of unstressed syllables, and a control experiment was deemed necessary. A hand-transcribed database of 98 sentences, each produced by three speakers, formed the input to an algorithm calculating durations of feet, number of syllables per foot, and mean syllable duration within each foot. In each output dataset, feet were based on one of the following: stressed, tense, unreduced, random, or arbitrary syllables (the latter based on ordinal numbers of syllables within the utterance). Calculations were made of the correlations between foot duration and number of syllables per foot, and between foot duration and mean syllable duration. The first correlation was significant for all foot types; the second was significant (and negative) for all except the random and arbitrary types. The conclusion was that, although the mechanism of the tendency toward isochrony had by no means been discovered, it had been shown that the tendency was nonrandom and was due to linguistic rather than arbitrary factors.