Speech Analysis and Synthesis by Linear Prediction of the Speech Wave

B. S. Atal

Citation: **47**, 65 (1970); doi: 10.1121/1.1974658
View online: http://dx.doi.org/10.1121/1.1974658
View Table of Contents: http://asa.scitation.org/toc/jas/47/1A
Published by the Acoustical Society of America

theory of acoustics and also of the Brillouin scattering of light has been based on the Stokes–Kirchoff method of solving the linearized differential equations of fluid dynamics; modes of propagation or decay of sound and components of the Brillouin spectrum are identified by roots of a secular equation. It has then been assumed that the general solution is the sum of the particular solutions, making the modes normal and the components separable. This assumption is shown to be not correct, in general, for a given real value of the propagation constant, and errors that arise through its use in the analysis of Brillouin spectra of relaxing fluids are examined. Different sets of approximate normal coordinates are required for the adequate description of spectra, depending on the ratio of the frequencies of relaxation and acoustic resonance.

4:45

4J8. Brillouin Scattering in Liquids of High Viscosity. A. B. Bhatia and E. Tong, *Physics Department, University of Alberta, Edmonton, Canada.*—The spectral distribution of scattered light is calculated for the model of a liquid in which the shear and bulk relaxation processes are each characterized by a continuous set of relaxation times. The treatment is a generalization of that given by the authors previously [Phys. Rev. **173**, 231 (1968)] in which bulk relaxation was ascribed to the relaxation of a thermodynamic order parameter relaxing with a single relaxation time. It is found that, in general, one has to introduce two independent distribution functions associated with the bulk relaxation times. Only when all the bulk relaxation processes are of the same types (pressure induced or temperature induced, etc.) does the expression for the spectral function contain just one distribution function, say $g_d(\tau)$, for bulk relaxations. In contrast, the bulk relaxation in ultrasonic work can always be characterized by just $g_d(\tau)$. The significance of these reuslts in relation to experimental data on Brillouin scattering and ultrasonic absorption in glycerine are discussed.

5:00

4J9. Absorption of Light by the Sound Waves of Disordered Crystals. E. Whalley, D. D. Klug, and P. T. T. Wong, *Division of Chemistry, National Research Council, Ottawa 7, Canada.*—Photons and phonons can interact in disordered crystals without conservation of wave vector. The absorption of light by the acoustic waves of orientationally disordered crystals, such as ice, and of vitreous silica are discussed in detail. The effect of short-range correlation of orientations has been examined theoretically. When the wavelengths of the excitations are very long compared to a lattice spacing, the absorption of light can be described in terms of a kind of piezoelectric coefficient, the root-mean-square value of which is not zero for orientationally disordered crystals.

Tuesday, 4 November 1969 — Cotillion Room 1, 2:00 p.m.

Session 4K. Acoustical Analysis of Speech

James L. Flanagan, *Chairman*

Contributed Papers (12 minutes)

2:00

4K1. Determination of the Vocal-Tract Shape Directly from the Speech Wave. B. S. Atal, *Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey 07974.*—Sound propagation through the human vocal tract is studied by quantizing the vocal tract into a finite but arbitrarily large number of contiguous uniform cylindrical sections. The transfer function between volume velocities at the glottis and lips and the corresponding impulse response are obtained by matching the pressure and the volume velocity at the junctions between adjacent sections. It is shown that the frequencies and bandwidths of the poles of the transfer function determine uniquely the areas of the cylindrical sections. The vocal-tract area function is determined directly from the autocorrelation function of the impulse response and the calculation does not involve any iterative or search procedure. The application of this procedure to determine the vocal-tract shape for connected speech is discussed.

2:15

4K2. Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. B. S. Atal, *Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey 07974.*—A method of representing the speech signal by time-varying parameters relating to the shape of the vocal tract and the glottal-excitation function is described. The speech signal is first analyzed and then synthesized by representing it as the output of a discrete linear time-varying filter, which is excited by a suitable combination of a quasiperiodic pulse train and white noise. The output of the linear filter at any sampling instant is a linear combination of the past output samples and the input. The optimum linear combination is obtained by minimizing the mean-squared error between the actual values of the speech samples and their predicted values based on a fixed number of preceding samples. A 10th-order linear predictor was found to represent the speech signal band-limited to 5kHz with sufficient accuracy. The 10 coefficients of the predictor are shown to determine both the frequencies and bandwidths of the formants. Two parameters relating to the glottal-excitation function and the pitch period are determined from the prediction error signal. Speech samples synthesized by this method will be demonstrated.

2:30

4K3. Speech Segmentation. W. B. Newcomb, W. D. Larkin, and R. A. Houde, *Research Department, Electronics Division of General Dynamics, Rochester, New York 14609.*—This paper reports a study of speech segmentation for bandwidth compression by segment coding. A segmentation procedure was desired that would segment continuous speech into units that were suitable for both classification and subsequent synthesis. Several measures obtainable for short-time spectrum information were considered: over-all amplitude, the difference in over-all amplitude between successive short-time spectra (normalized and nonnormalized) and the spectrum change between successive short-time spectra (normalized and nonnormalized). Representative speech samples were selected to include stop, fricative, nasal, and semivowel boundaries and digitized to 12 bits for computer entry. The short-time spectral data were obtained every 10 msec using a 40-msec Hanning weighting function and an FFT algorithm on the computer. The components obtained were summed to effect the desired spectrum resolution, and the desired time averaging was accomplished by summing successive short-time spectra. The above-listed measures were studied as the parameters of spectrum weighting, frequency resolution, and time averaging were varied. The analysis of the resulting data is presented.