Short-Time “Cepstrum” Pitch Detection

A. M. Noll and M. R. Schroeder

Citation: The Journal of the Acoustical Society of America **36**, 1030 (1964); doi: 10.1121/1.2143271
View online: https://doi.org/10.1121/1.2143271
View Table of Contents: https://asa.scitation.org/toc/jas/36/5
Published by the Acoustical Society of America

sions reached and theoretical considerations are presented. A tape demonstration follows. [A portion of this study supported under contract No. AF19(628)-586.]

N2. Digital Equalizer and Deequalizer for Speech. J. L. Flanagan, D. I. S. Meinhart (nonmember), and P. Cummiskey (nonmember), *Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.*—In the analysis and processing of speech, it is sometimes desirable to preemphasize the signal spectrum so that formant maxima are brought to nearly the same amplitude. A simple lumped-constant network is described for accomplishing this. A deemphasis network—the inverse of the former—is also described for recovering the original signal. The transfer functions for both networks exhibit four real-axis poles and four real-axis zeros. Sampled-data equivalents of both networks are derived, and the difference equations are programmed for the digital computer. Because the transfer functions have significant response at high-frequencies, the bilinear z transform is used to derive the difference equations. Errors arising from spectral folding about the sampling frequency are thereby avoided. Sampled and quantized speech signals are digitally processed by the program. The computer results are compared to circuit results obtained in the laboratory.

N3. Speech Energy Rate of Change as a Function of Its Spectral Location. Louis R. Focht, *Philco Corporation, Philadelphia, Pennsylvania.*—An investigation of the first derivative of instantaneous speech power at any point in the spectrum was made by first assuming that the total energy variance is produced by two components: the amplitude modulation of the excitation wave and the movement of the formants. It was found that the effect of the amplitude-modulation component is to produce a uniform first derivative of power over the entire speech spectrum. The maximum rate of change produced by this component is ½ the fundamental pitch of the speaker. The change produced by the formant-movement component, however, does not exhibit a uniform rate of power change over the speech spectrum. Rather, it was found to fall off at the high and low end of the spectrum by more than 3 to 1 and attained a maximum rate of change, at 1000 cps, which was less than ⅓ that of the amplitude-modulation component. These studies were performed by using test sentences obtained from 5 male and 5 female speakers. No significant difference was found between male and female. [Work supported in part by the U. S. Air Force under contract No. AF19(628)-586.]

N4. Pitch Extractor, Using Clippers. Jerome E. Toffler and Fred B. Wade (nonmember), *Hughes Communications Division, Los Angeles, California.*—Several previous pitch extractors have been constructed, based on the principle of nonlinear operations on the input-speech wave. Typical schemes have included envelope detectors, full-wave linear rectifiers, and square-law devices. This paper describes a new experimental pitch extractor in which the nonlinear elements are amplitude limiters, or clippers. Advantages claimed for the system are the following: (1) operation over a wide dynamic range of input amplitude due to the clipping employed, (2) wide frequency range suitable for male and female speakers, and (3) regeneration of the fundamental pitch frequency from its harmonics. Test results are presented showing the performance of the pitch extractor with a variety of speech signals.

N5. Short-Time "Cepstrum" Pitch Detection. A. M. Noll and M. R. Schroeder, *Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.*—A short-time cepstrum analyzer for vocal-pitch detection has been simulated on an IBM-7094 digital computer. (The cepstrum of a signal is defined as the square of the Fourier transform of the logarithm of the amplitude spectrum of the signal.) Since temporal periodicities in a speech signal cause periodic ripples in the amplitude spectrum, Fourier transformation of the spectrum gives the "frequency" of the ripple, which is inversely proportional to the fundamental frequency of the speech. Taking the logarithm of the amplitude spectrum makes the effects of the vocal tract (spectrum envelope) and the vocal source (spectrum fine structure) additive, thereby separating the low "frequencies" of the spectrum envelope from the usually higher frequencies of the spectrum fine structure. Cepstrum pitch detection is insensitive to phase distortion, amplitude distortion, additive noise, and the absence of the fundamental speech frequency. A general description of the technique and some recent results and examples are given.

N6. Pitch-Period Reiterating Speech-Compression System. Joseph J. Megna (nonmember), *Northeastern University, Boston, Massachusetts 02115* and *Communications Communications and Data Processing Operations, Equipment Division, Raytheon Company, Norwood, Massachusetts.*—In this paper, the performance of a speech-compression system, in which certain pitch periods are eliminated and replaced by the latest transmitted pitch period, is studied. The pitch-period representations are digitized and various bit rates and quantizing combinations may be used. Using the various combinations, the experimental results, based on listening tests and intelligibility scores, are presented.

N7. Note on Buzz-Hiss Detection. Bernard Gold, *Lincoln Laboratory* Massachusetts Institute of Technology, Lexington, Massachusetts.*—Speech sounds can be approximately divided into two categories, buzz or periodic sounds and hiss or aperiodic sounds. Vocoder buzz-hiss detectors have discriminated between the two classes by comparing the high-frequency and low-frequency portion of the spectrum. This technique succeeds during most of the input speech because most voiceless sounds contain a preponderance of high-frequency energy whereas the voiced sounds emphasize the lows. However, voiceless sounds such as aspirates and plosives often contain large amounts of low-frequency energy. Also, microphone distortions can produce low-frequency aperiodic sounds. This paper describes a periodicity meter that processes the output from six pitch detectors to produce another measure appropriate for buzz-hiss detection. Preliminary results are presented.

* Operated with support from the U. S. Air Force.

N8. Compact Pitch Computer. N. L. Daggett (nonmember) and B. Gold, *Lincoln Laboratory* Massachusetts Institute of Technology, Lexington, Massachusetts.*—The operational rules of the pitch computer used in the Lincoln Laboratory experimental vocoder have been described at the Fourth International Congress on Acoustics, Copenhagen, 1962 (Bernard Gold, "Description of a Computer Program for Pitch Detection," Paper G34). These rules, and some additional rules for buzz-hiss detection, described in a companion paper, have been implemented by a combination of analog and integrated circuit digital techniques. The complete pitch computer is packaged in a subrack 5×19×22 in. The basic storage element is an 8-bit circulating shift register. An array of 24 such registers provides the input memory. Use of this technique allows serial computations to be executed, simplifying the arithmetic portion of the computer and greatly reducing the density of interconnecting wiring. It is becoming increasingly clear that accurate pitch detection noticeably improves the quality of a channel vocoder. The work reported in this paper is a step toward realizing the improved vocoders as practical devices.

* Operated with support from the U. S. Air Force.