Short-Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection

A. Michael Noll

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received 7 October 1963)

A spectrum analyzer based on a definition of short-time power spectra has been designed and simulated on a digital computer. The analyzer is primarily intended for use in speech analysis. It has been designed to operate in real time, and to produce high-resolution spectra without utilizing either heterodyning methods or bandpass filter banks. The logarithm of each consecutive amplitude spectrum thus obtained can be used as the input to a second similar spectrum analyzer. The output of this analyzer is then the "cepstrum" or power spectrum of the logarithm spectrum. The cepstrum of a speech signal has a peak corresponding to the fundamental period for voiced speech but no peak for unvoiced speech. Thus, a cepstrum analyzer can function both as a pitch and as a voiced–unvoiced detector. Cepstral pitch detection has the important advantages that it is insensitive to phase distortion, and is also resistant to additive noise and amplitude distortion of the speech signal. The method does not require the presence of the fundamental frequency in the speech signal, and will give several separate cepstral peaks if several different pitch periods are present. Cepstral techniques appear to be even more reliable and efficient than visual methods for pitch detection. The short-time spectrum and cepstrum analyzers described in this paper were simulated by a sampled-data system on an IBM-7090 digital computer. The simulation was programmed with the assistance of a special block-diagram compiler.

INTRODUCTION

MOST presently available speech-spectrum analyzers can be divided into two groups: heterodyne analyzers and filter-bank analyzers. The heterodyne analyzer mixes the signal to be analyzed with a sine wave produced by a tunable oscillator. The resulting difference frequencies are then applied to a fixed-frequency bandpass filter to select the frequency band containing the component to be determined. In effect, the input signal is frequency-swept through a fixed-frequency filter. This type of analysis produces spectra with high resolution. However, a relatively long time (more than a minute) is required to analyze only a few seconds of speech. Faster heterodyne analyzers are available, but elaborate schemes are necessary to correct the distortion resulting from the high sweep rate of the input signal through the bandpass filter.[1]

The second type of analyzer consists of a stationary bandpass filter bank. The output of each filter corresponds to that component of the spectrum lying within the passband of the filter. This type of device produces a coarse spectrum lacking in good resolution, unless an impractical number of filters is used. The advantage of this type of spectrum analyzer is that it operates in real time, i.e., the output is instantaneous to within the averaging times of the filters.

Spectrum analyzers combining the advantages of real-time operation and high resolution have been designed.[2–5] A more direct approach to spectral analysis based on short-time spectral techniques at baseband (without heterodyning) might be desirable. The short-time spectrum analyzer described in this paper was designed to operate in real time and to produce high-resolution spectra; it could be implemented by either analog or sampled-data circuitry. A sampled-data version was simulated on an IBM-7090 digital computer.

An extension of the short-time spectral technique, suggested to the author by M. R. Schroeder, gives an interesting method for pitch detection. The quasi-periodic repetitions of the waveforms in a voiced interval of speech cause periodic ripples in the speech spectrum

[1] L. L. Beranek, *Acoustic Measurements* (John Wiley & Sons, Inc., New York, 1949), pp. 537–543.

[2] E. Meyer, *Electroacoustics* (G. Bell & Sons Ltd., London, 1939), pp. 39–45.

[3] J. Capon, "On the Properties of an Active Time-Variable Network: The Coherent Memory Filter," in *Proceedings of the Symposium on Active Networks and Feedback Systems* (Polytechnic Institute of Brooklyn, New York, 1960).

[4] J. S. Gill, "A Versatile Method for Short-Term Spectrum Analysis in 'Real-Time,'" Nature **189**, No. 4759, 117–119 (14 Jan. 1961).

[5] D. E. Wood and T. L. Hewitt, "New Instrumentation for Making Spectrographic Pictures of Speech," J. Acoust. Soc. Am. **35**, 1274–1278 (1963).

or its logarithm. The frequency spacing between the peaks of these ripples equals the fundamental frequency of the speech. If the power spectrum of the logarithm of the speech spectrum is computed, a peak will appear corresponding to the pitch period of the speech; the absence of such a peak would indicate an unvoiced speech interval. The term "cepstrum" is applied to the results obtained by computing the power spectrum of the logarithm of the power (or amplitude) spectrum.[6] In this manner, pitch and voiced–unvoiced detection can be performed by short-time cepstral analysis of the speech signal.

This paper describes the defining equations for short-time spectra and then applies them to the actual design of a spectrum analyzer. Short-time cepstral techniques are introduced and are applied to pitch detection. A cepstrum analyzer is described as a simple modification of the short-time spectrum analyzer.

DEFINITION OF SHORT-TIME SPECTRAL ANALYSIS

The power spectrum $G(\omega)$ of some function of time $f(t)$ is defined as

$$G(\omega)\equiv\left|\int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt\right|^2. \qquad (1)$$

This definition, however, is unsuitable for experimental measurements because it requires both integration over an infinite time interval and a knowledge of the future. In practice, it is possible only to perform this integration up to the present time, which is continuously changing. What is desired is a running power spectrum with time as a second variable. The required modifications of Eq. (1) were made by Fano[7] and later generalized by Schroeder and Atal,[8] who defined a short-time power spectrum $G(\omega,t)$ as

$$G(\omega,t)\equiv\left|\int_{-\infty}^{t} r(t-x)f(x)e^{-i\omega x}dx\right|^2, \qquad (2)$$

or, equivalently,

$$G(\omega,t)=\left|\int_{0}^{\infty} r(\tau)f(t-\tau)e^{-i\omega(t-\tau)}d\tau\right|^2, \qquad (3)$$

where t is real time, and x and τ are variables of integration. The function $r(\tau)$ is the impulse response of a physically realizable but otherwise arbitrary system. Since $r(\tau)$ multiplies the input signal, it can be visualized as a "window" through which a part of the input signal is viewed. For this reason, $r(\tau)$ is sometimes called a lag window. $G(\omega,t)$ can also be considered as $|F(\omega,t)|^2$, where $F(\omega,t)$ is the short-time Fourier transform:

$$F(\omega,t)\equiv\int_{-\infty}^{t} r(t-x)f(x)e^{-i\omega x}dx. \qquad (4)$$

The experimental measurement of $F(\omega,t)$ could be carried out by passing $f(t)$ through a pair of bandpass filters having impulse responses $r(\tau)\cdot\cos(\omega\tau)$ and $r(\tau)\cdot\sin(\omega\tau)$.

The definition of the short-time spectrum can be further modified if the lag window $r(\tau)$ is defined to be zero for $|\tau|\geq\tau_M$. If it is further assumed that $r(\tau)$ is symmetric about $\tau=0$, the signal under analysis is time-limited to an interval $2\tau_M$ in duration. For this type of time-limited signal, the Fourier transforms of the real and imaginary parts of $F(\omega,t)$ are "time-limited" from $-\tau_M$ to $+\tau_M$ sec. By Nyquist's sampling theorem, applied to the frequency domain, $F(\omega,t)$ is completely described by samples spaced every $\Delta f=(2\tau_M)^{-1}$ cps. This means that ω can be represented as $\omega=2\pi n\Delta f$, $n=0, 1, 2, \cdots$. Since it is known that a function cannot be both time-limited and band-limited, the question arises whether $F(\omega,t)$ can be assumed to be approximately band-limited. A consideration of this question follows.

Consider some function defined as the product of the input signal $f(t)$ and the lag window $r(t)$. Taking the Fourier transform of this function results in a spectrum that is the convolution of the spectra of $f(t)$ and $r(t)$:

$$\mathfrak{F}[f(t)\cdot r(t)]=\frac{1}{2\pi}F(\omega)*R(\omega), \qquad (5)$$

where $F(\omega)$ and $R(\omega)$ are the Fourier transforms of $f(t)$ and $r(t)$, respectively; $\mathfrak{F}$ and $*$ indicate Fourier transformation and frequency convolution, respectively. Thus, the Fourier transform of the lag window determines to some extent the shape and resolution of the spectrum. Because of its conceptual similarities with the lag window, the Fourier transform of the lag window is usually referred to as the spectral window.

If the spectral window $R(\omega)$ has a narrow initial lobe and very small side lobes and if $F(\omega)$ is band-limited to frequencies below f_c cps, then the convolution of $R(\omega)$ and $F(\omega)$ will be approximately band-limited to a 0 to f_c cps finite interval. Under these conditions, a finite number of samples of ω spaced every $(2\tau_M)^{-1}$ cps will approximately describe the real and imaginary parts of $F(\omega,t)$ in the range of from 0 to f_c cps. The analysis of the ω variation of $F(\omega,t)$ is now complete. It has been shown that ω can be represented as $2\pi n\Delta f$, where $n=0, 1, 2, \cdots, N$; $N=f_c/\Delta f$, and $\Delta f=(2\tau_M)^{-1}$.

The time variation of $F(\omega,t)$ is now considered. The approach is first to describe a particular form of window–time displacement and then offer a theoretical justifica-

[6] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, in *Proceedings of the Symposium on Time Series Analysis*, edited by M. Rosenblatt (John Wiley & Sons, Inc., New York, 1963), Chap. 15, pp. 209–243.

[7] R. M. Fano, "Short-Time Autocorrelation Functions and Power Spectra," J. Acoust. Soc. Am. 22, 546–550 (1950).

[8] M. R. Schroeder and B. S. Atal, "Generalized Short-Time Power Spectra and Autocorrelation Functions," J. Acoust. Soc. Am. 34, 1679–1683 (1962).

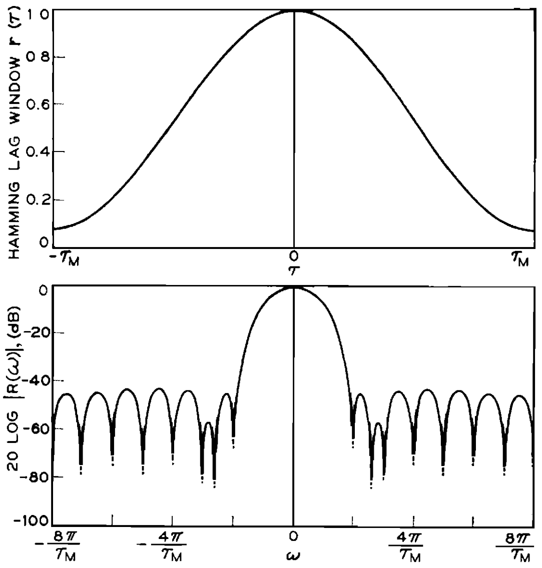

FIG. 1. Hamming lag window and its corresponding spectral window. The spectral window is plotted as $20\log|R(\omega)|$, where $R(\omega)$ is the Fourier transform of the lag window.

tion. The definition of $F(\omega,t)$ as stated in Eq. (4) implies that the lag window is shifted continuously across the input signal and that $F(\omega,t)$ is, therefore, continuous with respect to t. Instead of this continuous window shifting, the lag window of $2\tau_M$ width is shifted τ_M sec for each spectral analysis. In effect, the short-time spectrum is being sampled at intervals of τ_M sec, and this requires some justification. If the time variation of $F(\omega,t)$ is such that appreciable changes in the spectra do not occur for time intervals of less than τ_M sec, then the spectra can be time-sampled. That this is the case here can be verified by performing the Fourier transform with respect to t of $F(\omega,t)$ at a constant ω. The result equals the product of the spectral window $R(\Omega)$ and the spectrum $F(\omega+\Omega)$:

$$\int_{-\infty}^{\infty} F(\omega,t)e^{-j\Omega t}dt=R(\Omega)\cdot F(\Omega+\omega). \qquad (6)$$

Thus, the Fourier transform with respect to t of $F(\omega,t)$ is approximately band-limited by the bandwidth of the spectral window $R(\Omega)$. Consider a spectral window such that its amplitude response is down by a factor of 2 from the peak at frequencies of $\pm(2\tau_M)^{-1}$ cps. This corresponds to a 6-dB bandwidth of $(\tau_M)^{-1}$ cps. The Fourier transform with respect to t of $F(\omega,t)$ is, therefore, approximately band-limited to $\pm(2\tau_M)^{-1}$ cps, and $F(\omega,t)$ can be sampled at intervals of τ_M sec. Hence, the variable t can be represented as $k\tau_M$, where $k=1, 2, \cdots, K$; and $(K+1)\tau_M$ is the maximum time length of the signal $f(t)$.

The material of the preceding paragraphs can now be incorporated into some modifications of the definition of short-time Fourier transformation as stated in Eq. (4). The notation $F_k(\omega)$ is introduced to represent the kth spectrum over a time interval of $2\tau_M$ sec centered at $(k-1)\tau_M$ sec and at radian frequency ω. The first modification involves the concepts of a time-limited lag window $r(\tau)$ and a time-sampled short-time spectrum. The result of this modification is

$$F_k(\omega)\equiv\int_{-\tau_M}^{\tau_M} r(\tau)f[(k-1)\tau_M-\tau]e^{-j\omega\tau}d\tau, \qquad (7)$$

for $k=1, 2, \cdots, K$. The lag window $r(\tau)$ is defined such that $r(\tau)=0$ for $|\tau|\geq\tau_M$, $r(\tau)$ is symmetric about $\tau=0$, and $r(\tau)$ possesses a spectral window with a bandwidth of $(\tau_M)^{-1}$ cps and very small side lobes.

For a time-limited lag window, the ω variation of $F_k(\omega)$ can be sampled at intervals of $(2\tau_M)^{-1}$ cps. This means that Eq. (7) becomes

$$F_k(n\Delta\omega)=\int_{-\tau_M}^{\tau_M} r(\tau)f[(k-1)\tau_M-\tau]e^{-j\tau n\Delta\omega}d\tau, \qquad (8)$$

for $k=1, 2, \cdots, K$; $n=0, 1, 2, \cdots, \omega_c/\Delta\omega$; and where $\Delta\omega=2\pi/(2\tau_M)$. The input signal $f(t)$ is assumed to be band-limited to $\omega_c/(2\pi)$ cps. The spectrum analyzer described later in the paper was based upon this equation.

The preceding lag-window conditions are satisfied by the "hamming" lag window that was actually used as $r(\tau)$ in the spectrum analyzer.[9] The hamming lag

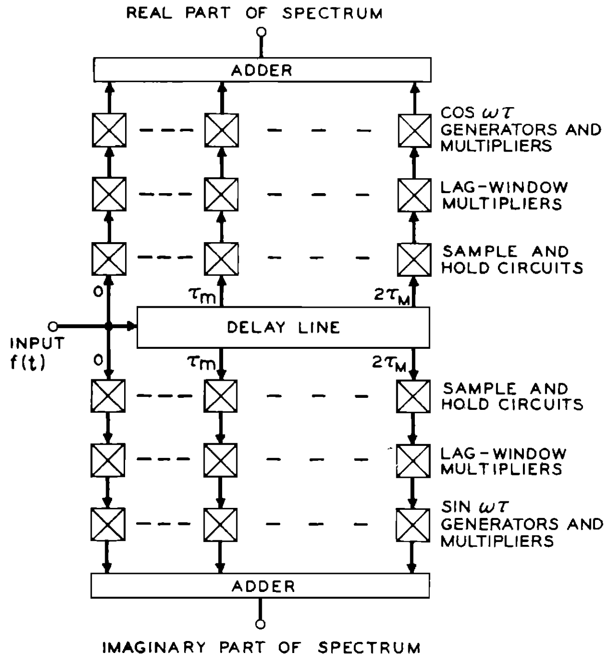

FIG. 2. Block diagram of short-time spectrum analyzer. The radian frequency ω is sampled so that $\omega=2\pi n/\Delta f$ for $n=0$, 1, 2, $\cdots$, $f_c/\Delta f$, where $\Delta f=(2\tau_M)^{-1}$ and the signal is restricted to frequencies below f_c.

[9] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra* (Dover Publications, Inc., New York, 1958).

window is a raised cosine function with a small pedestal, or

$$\begin{aligned} r(\tau) &= 0.54+0.46\cos(\pi\tau/\tau_M) && \text{for} \quad |\tau|<\tau_M, \\ &= 0 && \text{for} \quad |\tau|>\tau_M. \end{aligned} \qquad (9)$$

This empirically derived lag window has a spectral window with a maximum side lobe 44 dB below its peak response. The hamming lag window and its corresponding spectral window are shown in Fig. 1.

SAMPLED-DATA SHORT-TIME SPECTRUM ANALYZER

The short-time spectrum analyzer described in this paper is for use with an input signal initially band-limited to frequencies below f_c and sampled at the Nyquist interval, $\Delta T=(2f_c)^{-1}$. This sampled-data requirement means that the integration indicated in Eq. (8) can be approximated by a finite summation over τ from $-\tau_M$ to $+\tau_M$ sec in steps of ΔT sec. In this manner, the experimental measurement of $F_k(2\pi n\Delta f)$ as indicated by Eq. (8) is realized by a sampled-data short-time spectrum analyzer. The delayed input signal $f(t-\tau_m)$ is obtained at the mth tap of a delay line and is held for τ_M sec by a sample-and-hold circuit (see Fig. 2). In this manner, the function $f[(k-1)\tau_M-\tau_m]$ is obtained at the output of the mth sample-and-hold circuit. Multiplication by the lag-window coefficient $r(\tau_m)$ is then performed. The exponential in Eq. (8) is separated into its real and imaginary parts, and $r(\tau_m) \cdot f[(k-1)\tau_M-\tau_m]$ is multiplied by either $\sin[(2\pi n\Delta f)\tau_m]$ or $\cos[(2\pi n\Delta f)\tau_m]$. The final output after summation over τ_m is either the real or imaginary part of the spectrum at radian frequency $2\pi n\Delta f$. The parameter n then changes value to $n+1$; the sine and cosine multiplication is performed; and summation over τ_m produces either the real or imaginary part of the spectrum at radian frequency $2\pi(n+1)\Delta f$. The parameter n can be stepped from 0 to $f_c/\Delta f$ in any manner or time interval. The only requirement is that the time for the total sweep must be less than or equal to τ_M sec. If not, the outputs of the sample-and-hold circuits would change value before completion of a spectral analysis.

The cosine (sine) functions are symmetric (antisymmetric) about the center tap of the delay line. The lag window $r(\tau)$ is also symmetric, and, therefore, the delay line can be "folded" about its center and symmetrically located tap outputs either added or subtracted, as shown in Fig. 3. This procedure results in a 2:1 saving in the number of multiplier and sample-and-hold circuits.

CEPSTRAL TECHNIQUES FOR PITCH DETECTION

The spectrum of speech equals the product of the spectra of the vocal source and of the vocal tract. This relationship can be expressed in magnitude form as

$$|F(\omega)|=|S(\omega)|\cdot|V(\omega)|, \qquad (10)$$

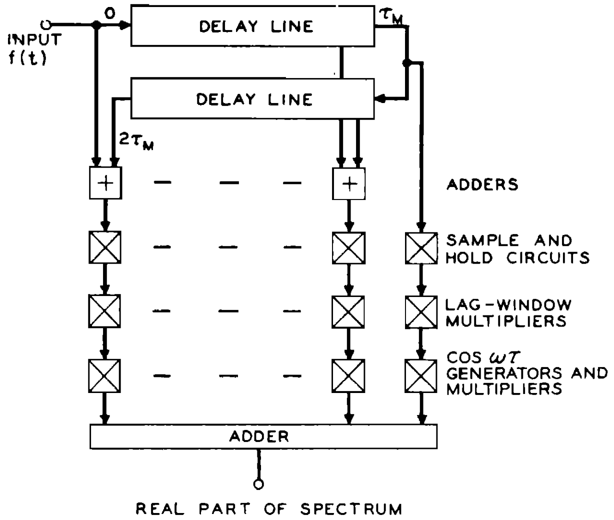

FIG. 3. Block diagram of short-time spectrum analyzer for the real part of the spectrum. The dealy line is folded about its center for a symmetric lag window.

where $|F(\omega)|$ is the speech spectrum, $|S(\omega)|$ is the vocal-source spectrum, and $|V(\omega)|$ is the magnitude of the Fourier transform of the impulse response of the vocal tract. For pitch and voiced–unvoiced detection, information about the vocal source is desired. Therefore, the effects of the vocal tract should be separated from the source. This can be accomplished by taking the logarithm of the speech spectrum,

$$\log|F(\omega)|=\log|S(\omega)|+\log|V(\omega)|. \qquad (11)$$

Since the vocal tract forms a resonator, its log spectrum consists of peaks located at the resonant frequencies or formants of the vocal tract. These formant frequencies are usually widely separated in the spectrum. The effect of the vocal tract, then, is to produce "long-wavelength" ripples in the logarithm spectrum.

The vocal source can be represented as a periodic buzz. Consider only two pitch periods of this source signal. The magnitude of the spectrum of the source signal over these two pitch periods is

$$|S(\omega)|=|S_T(\omega)||1+e^{-i\omega T}|, \qquad (12)$$

where T is the pitch period and $S_T(\omega)$ is the spectrum of a single pitch period of the source signal. Taking the logarithm of Eq. (12) and simplifying gives

$$\log|S(\omega)|=\log|S_T(\omega)|+\tfrac{1}{2}\log[1+\cos(\omega T)], \qquad (13)$$

within an additive constant. The $\log[1+\cos(\omega T)]$ term has spectral peaks every T^{-1} cps. Thus, the effect of the vocal source is to produce periodic "short-wavelength" ripples in the logarithm spectrum. Using more than two pitch periods of the vocal source only makes the spectral peaks more pronounced; the peaks still occur every T^{-1} cps.

The preceding paragraphs have demonstrated that

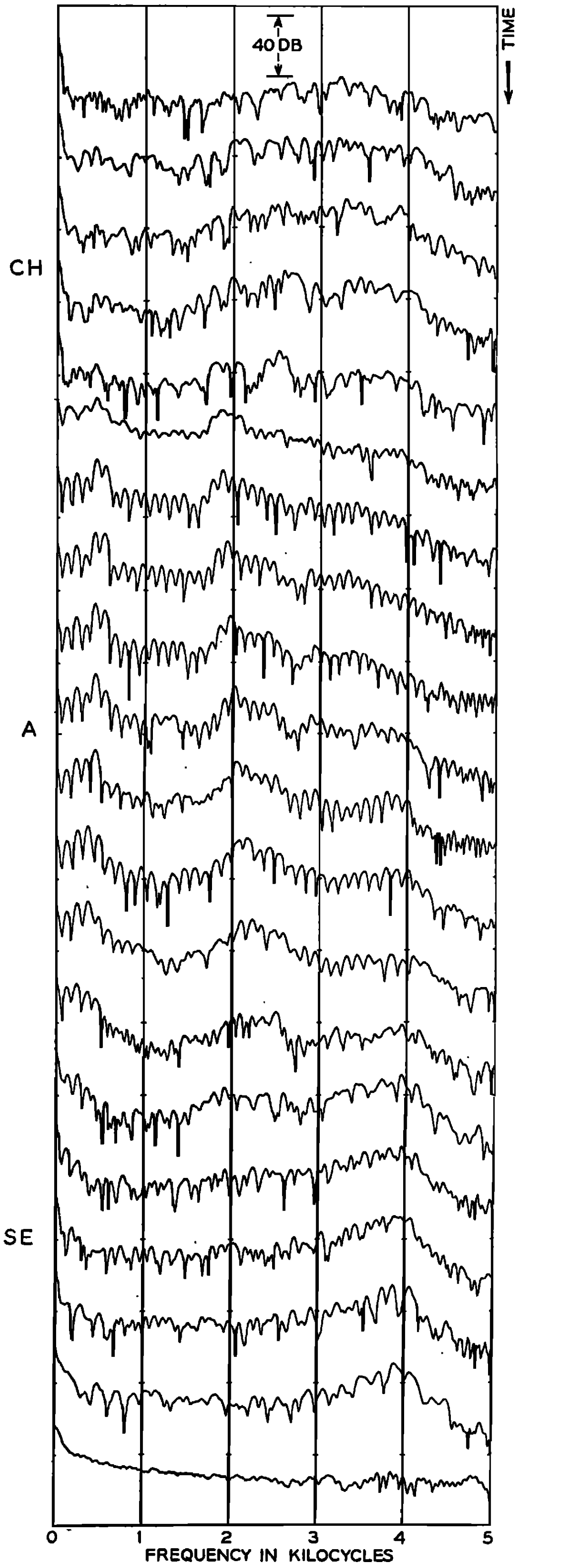

Fig. 4. Logarithm of the short-time amplitude spectra of the word "chase" spoken by a male. The spectra were computed on an IBM-7090 digital computer and were automatically plotted on microfilm.

the logarithm spectrum of voiced speech consists of "short-wavelength" ripples along the ω axis superimposed on long-"wavelength" ripples. *Wavelength* is enclosed in quotations because the ripples occur in the spectrum and are a function of frequency, not time. To prevent this ambiguity, the spectral ripples have been called "quefrency" components by J. W. Tukey.[6] The quefrency of a spectral ripple is measured as the reciprocal of the "period" (frequency spacing in cps) of the ripple and has the dimension of time. The *shorter*-"wavelength" spectral ripples correspond to *high*-quefrency components. The logarithm spectrum of speech consists then of high-quefrency components caused by the periodic movement of the vocal cords superimposed on the low-quefrency components of the formant structure.

As mentioned previously, the pitch of speech results in spectral peaks spaced every T^{-1} cps, and the quefrency of this spectral ripple is equal to T, the pitch period. If the logarithm spectrum is visualized as a new signal upon which a spectral analysis can be performed, a quefrency-analysis scheme emerges. Such an analysis produces the cepstrum or the power spectrum of the logarithm spectrum, once again using Tukey's terminology. While frequency was the independent variable in the spectrum, quefrency becomes the independent variable in the cepstrum. A cepstral analysis of a spectrum gives the amplitudes of the various quefrency components present in the spectrum. For voiced speech, the cepstrum contains a large peak at the quefrency corresponding to the pitch period of the vocal source T. The absence of such a cepstral peak indicates an unvoiced speech interval. If two or more different pitch periods occur during the analyzing interval, separate cepstral peaks will appear corresponding to each individual pitch period. In this manner, a cepstral analysis functions as an effective pitch and voiced–unvoiced detector, with the important advantage that it is independent of the presence of the fundamental frequency component of the speech.

There is a certain similarity between cepstrum analysis and autocorrelation analysis. The autocorrelation function of a signal is defined as the Fourier transform of the power spectrum, while the cepstrum is defined as the square of the Fourier transform of the logarithm of the power (or amplitude) spectrum. The essential difference between the two techniques is the logarithmic operation. The logarithmic operation accounts for the superiority of cepstral techniques for pitch and voiced–unvoiced detection. The reason for this is that in the autocorrelation function the effects of the vocal source and the vocal tract are convolved with each other since their respective spectra are multiplicative in the speech spectrum [see Eq. (10)]. However, in the cepstrum the effects of the vocal source and the vocal tract are separated, since their respective spectra are additive in the logarithm of the speech spectrum [see Eq. (11)].

SAMPLED-DATA CEPSTRUM ANALYZER FOR USE IN PITCH DETECTION

The short-time cepstrum of some short-time spectrum $F_k(\omega)$ is defined as

$$C_k(\tau)\equiv\left|\int_0^{\omega_c} s(\omega)\log|F_k(\omega)|\cos(\omega\tau)d\omega\right|^2. \quad (14)$$

The function $s(\omega)$ is an arbitrary even "weighting" function that could be of the same form as $r(\tau)$ in Eq. (7). Since the amplitude spectrum and $s(\omega)$ are both even functions, it is necessary to multiply by only $\cos(\omega\tau)$ and not by $e^{-i\omega\tau}$ as in Eq. (7).

The similarities between the definitions of the short-time spectrum [Eq. (7)] and the short-time cepstrum [Eq. (14)] immediately suggest the use of a modified short-time spectrum analyzer to obtain cepstra. Thus, it is necessary only to take the logarithm of the power spectrum and to use this signal as the input to a real-part spectrum analyzer.

In order to use the cepstrum analyzer for pitch analysis, it is important that the maximum quefrency be larger than any expected pitch period. However, as mentioned previously, the Fourier transform of the short-time spectrum is time-limited to $\pm\tau_M$ sec; i.e., the maximum useful quefrency in the cepstrum is τ_M sec. As an example, detection of pitches down to 100 cps would require a window at least 20 msec wide.

COMPUTER SIMULATION AND RESULTS

The spectrum analyzer and cepstrum analyzer were simulated on an IBM-7090 digital computer, using the BLODI compiler.[10] The input to the spectrum analyzer consisted of speech band-limited to 4 kc/sec and sampled at a 10-kc/sec rate ($\Delta T=10^{-4}$ sec).

The number of delay-line taps and the corresponding maximum quefrency were determined from considerations of the speech pitch period and the lowest pitch frequency to be analyzed. It is desirable to produce spectra at intervals of one or two pitch periods, since speech spectra can change appreciably within these periods. However, decreasing these time intervals necessitates decreasing the number of dealy units and also decreases the cutoff quefrency, thereby making pitch analysis for long pitch periods impossible. A compromise is indicated, and a maximum delay of 30 msec was chosen. This allows pitches down to 67 cps to be detected. A complete spectrum is produced every 15 msec.

The output of the simulated analyzer is displayed on a General Dynamics S-C 4020 microfilm printer. Frequency is plotted along the abscissa, amplitude along the ordinate, and each successive plot is the spectrum averaged over a 30-msec real-time interval. As an

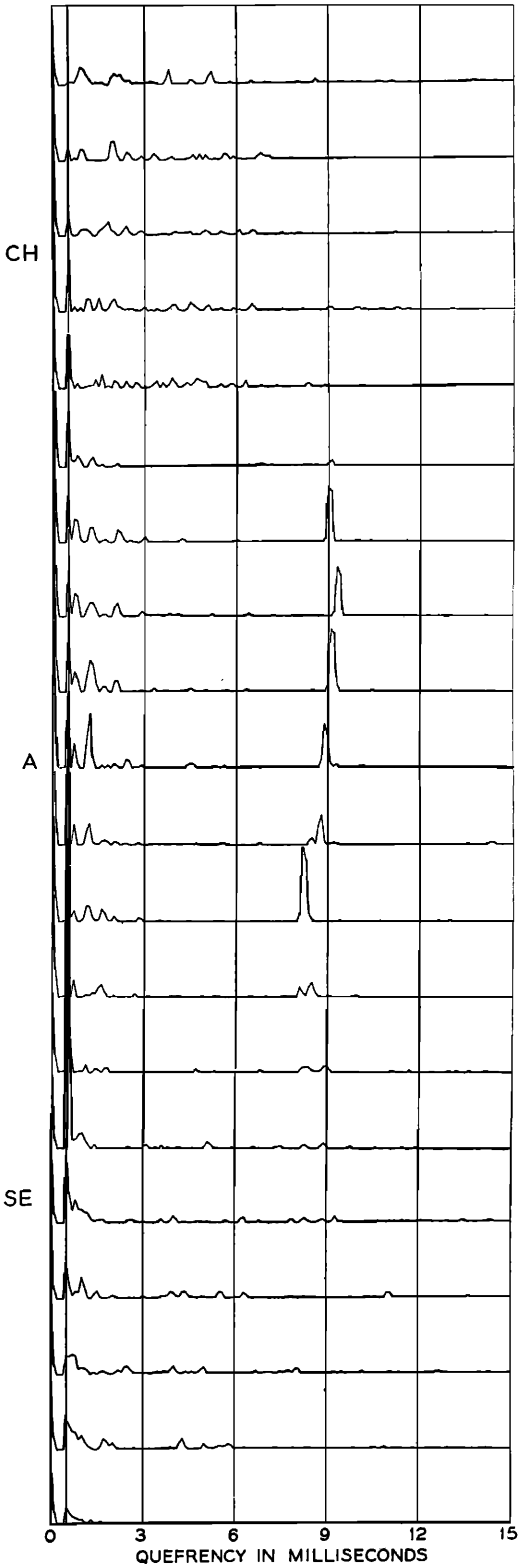

FIG. 5. Short-time cepstra of *chase*. The cepstral peaks clustered around 9.0 msec correspond to the pitch periods of the speech. The absence of these peaks indicates unvoiced speech intervals or silence.

[10] John L. Kelly, Jr., Carol Lochbaum, and V. A. Vyssotsky, "A Block Diagram Compiler," Bell System Tech. J. 40, 669–677 (1961).

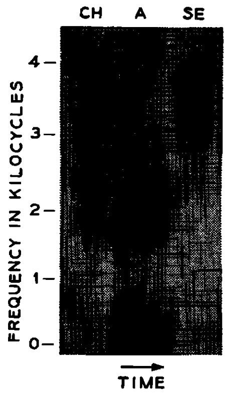

FIG. 6. Sound spectrogram of *chase*.

illustration, Fig. 4 is a sample output of the analyzer for the logarithm of the power spectrum. The utterance is *chase* spoken by a male. The corresponding "visible-speech" spectrogram is shown in Fig. 6.

Figure 5 depicts the cepstrum of *chase*. The peaks around 8.2 to 9.3 msec correspond to pitch frequencies of 108 to 122 cps. Since the logarithm of the spectrum is usually always positive, a large zero-quefrency component is expected. For this reason, quefrencies below 0.5 msec are reduced in amplitude; this scale change is indicated by the vertical line at 0.5 msec in Fig. 5. The cepstra with no pronounced peaks between 3 and 15 msec (333 and 67 cps) correspond to intervals of unvoiced speech or silence.

CONCLUSIONS

The successful computer simulation of the sampled-data spectrum and cepstrum analyzers demonstrates the feasibility of real-time pitch detection by cepstral techniques. Voiced–unvoiced decisions are possible based on the presence or absence of a cepstral peak.

The cepstral method of pitch and voiced–unvoiced detection performs its analysis on the log-amplitude spectrum and is, therefore, insensitive to phase distortion. The method does not depend upon the presence in the spectrum of the fundamental pitch frequency and will, therefore, pitch-detect such signals as telephone band-limited speech. Amplitude distortion of the speech signal usually preserves any periodicities present in the signal, and a cepstral peak is still obtained. Additive wide-band noise fills in only the gaps between the spectral peaks, and the cepstral method is, therefore, relatively noise-resistant. Narrow-band noise, in effect, produces only a single spectral peak, and, therefore, there is no pronounced cepstral peak. A cepstrum analysis produces nonaveraged pitch information in the sense that, if more than one pitch period is present in the signal interval under analysis, separate cepstral peaks will be obtained for each pitch period.

The final judge of a pitch detector is its ability to perform satisfactorily in an actual vocoder. Work is presently in progress to incorporate a cepstral pitch detector into a spectrum-channel vocoder.

ACKNOWLEDGMENTS

The author is especially indebted to M. R. Schroeder for suggesting this study and for his impetus and guidance. Many helpful discussions and suggestions were given by C. H. Coker and M. M. Sondhi. Computer-programming consultation was offered by Carol M. Bird. The author gratefully acknowledges their contributions.