Correspondence

Comments on "Spectral Analysis of the Calls of the Male Killer Whale"

In the above paper,[1] Singleton and Poulter point to an abundance of harmonics present in spectral analyses of many animal sounds. They comment that workers in bioacoustics have believed that some of these harmonics may be artifacts of the equipment, and cite Busnel and Watkins. Busnel[2] does not refer to artifacts in analysis and does not even mention harmonics; he referred to frequency shifts, that he thought might have been caused by Doppler, as artifacts. The reference to my work is to a recent paper.[3] I did not suggest either in my paper or in the oral presentation "that harmonics induced by altering the signal are in any way spurious or nonexistent," as Singleton and Poulter infer. I did state that a knowledge of the analyzing filter bandwidth (which directly affects response time) is critical for the interpretation of spectrographic analysis. For example, a train of pulses may be portrayed in the analysis either by discrete pulses or by its equivalent harmonic structure, depending entirely on the analyzing filter bandwidth employed.

WILLIAM A. WATKINS
Woods Hole Oceanographic Institution
Woods Hole, Mass. 02543

Manuscript received November 13, 1967.

[1] R. C. Singleton and T. C. Poulter, *IEEE Trans. Audio and Electroacoustics*, vol. AU-15, pp. 104–113, June 1967.

[2] R.-G. Busnel, "Information in the human whistled language and sea mammal whistling," in *Whales, Dolphins, and Porpoises*, K. S. Norris, Ed. Berkeley, Calif.: University of California Press, 1966, pp. 544–568.

[3] W. A. Watkins, "The harmonic interval: fact or artifact in spectral analysis of pulse trains," in *Marine Bioacoustics II*, W. N. Tavolga, Ed. New York: Pergamon, 1967, pp. 15–42. (Paper presented at the 1966 Symp. on Marine Bioacoustics.)

Authors' Reply[4]

Watkins' paper[3] was unavailable at the time of writing our paper, and apparently our memory of his oral presentation was faulty. Perhaps we were misled by his title, but in any event we apologize for misinterpreting his remarks.

We would agree with Watkins that the effective filter bandwidth, or equivalent information, should be included in reporting spectral analysis results. In the case of a digital analysis, this information can be conveyed by giving the time duration of the data set, the data window function used, and the smoothing applied to the estimated power spectral density function. With a sample of length T, the spectral estimates are spaced $1/T$ apart in the frequency domain. In the analysis of a nonstationary signal, as for example speech data, the choice of T will represent a compromise between resolution in time and in frequency.

R. C. SINGLETON
T. C. POULTER
Stanford Research Institute
Menlo Park, Calif. 94025

[4] Manuscript received December 18, 1967.

Adaptive Spectral Analysis for Speech-Sound Recognition

Abstract

A pattern recognition algorithm has been used to compare the usefulness of two types of spectrum analyzers for speech recognition.

A number of speech researchers have noted that it might be useful to allow the parameters of speech spectrum analyzers to vary according to the characteristics of the speech sounds being processed. In particular, it seems clear that the use of narrow-band filters should enhance the formant pattern of vowels, and that wide-bandwidth filters should bring out the noise-like structure of consonants.

The possible uses of a variable parameter spectrum analyzer for the purpose of recognizing simple speech sounds have been investigated. A pattern recognition algorithm has been used to compare an analyzer having 32 active Gaussian filters of 100-Hz bandwidth and one having 16 filters of 200-Hz bandwidth but twice the time resolution. Experimental results with the four stop consonants /p/, /b/, /t/, /d/ taken two by two in initial position, and with the pairs of vowels /a/, /ɔ/ and /e/, /i/ have been obtained. Narrow-band filters have been found to be preferable for recognizing vowels, and wide-band filters appear to be better for classifying st consonants when only the data or main body of the consonant are u

Manuscript received May 29, 1968.

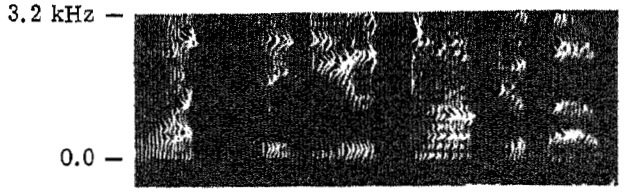

Fig. 1. Spectrogram of "What did you talk about?" with 100-Hz filters. Horizontal scale: 10 ms/div.

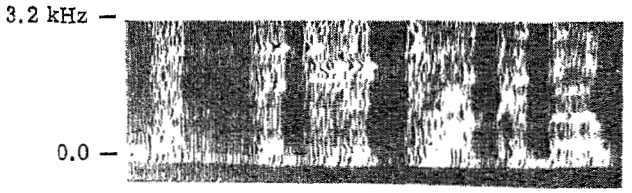

Fig. 2. Spectrogram of "What did you talk about?" with 200-Hz filters. Horizontal scale: 5 ms/div.

TABLE

Number of Sounds Correctly Classified on a Given Total (t_0 Starting Point of the Consonant)

Sounds	Time Interval		Recognition Performance	
	From (ms)	To (ms)	100 Hz	200 Hz
p–b	t_0	t_0+40	24/52	40/55
p–t	t_0	t_0+20	47/90	62/90
t–d	t_0	t_0+40	27/46	39/49
	t_0+30	t_0+50	34/57	47/61
b–t	t_0	t_0+20	79/96	71/96
	t_0+20	t_0+40	47/64	39/64
d–b	t_0+20	t_0+60	68/90	52/90
Vowels	t_0+40	t_0+80	125/154	110/154

An adaptive spectral analyzer is an instrument which can match its parameters to those of an incoming signal. A spectral analyzer with variable parameters has been designed by Thomas [1]. It is an analyzer of the bank-of-filters type with active filters which are read and quenched simultaneously at fixed intervals.

Increasing the rate of quenching (the time resolution) increases the frequency bandwidth of the filters (decreases the frequency resolution). The active filters have a sin $(x)/x$ frequency response but, when the input signal to the analyzer is multiplied by a raised cosine synchronized to the quenching pulse and having a period equal to the duration of the interval between two successive readings, the filters have effectively nearly Gaussian time and frequency characteristics. The output of the analyzer can be seen on a display oscilloscope as shown in Figs. 1 and 2; the frequency axis is vertical, the time axis goes horizontally from left to right. The energy present in the different regions of the spectrum is indicated by intensity modulation as well as some horizontal deflection. Figs. 1 and 2 give an idea of the differences between the two analyzing modes selected for the same speech sounds; it can be seen that the formant structure is much better defined in the 100-Hz mode (Fig. 1), while the noise structure is enhanced in the 200-Hz display.

Arguments in favor of adaptive spectrum analysis can be drawn from the study of speech characteristics and of the human hearing system. Gabor [2], having applied to signal analysis Heisenberg's uncertainty principle, showed that the ear can trade time and frequency resolution between the limits of an uncertainty relation of the form

$$\Delta f \cdot \Delta t = \text{constant}.$$

Different authors [3]–[5] have supported his conclusions by taking measurements of phenomena related to the uncertainty relation. Others [6]–[8] came to similar conclusions through studies of auditory signal detectability in noise; Creelman [7], for example, interprets his results "as suggesting that observers are able, with high amplitude complex signals as well as with sinusoids, to match their receptive systems to the signals to be detected." It appears then that the existence of an auditory uncertainty relation applicable to the whole audio-frequency range and at least in the time interval from 3 to 300 milliseconds is virtually certain, although an exact measure of this relation has yet to be found. It should be mentioned also that phoneticians and speech researchers use large-bandwidth filters in spectrum analysis when they want to study the fine details and the temporal effects in the speech sounds, and narrow-bandwidth filters when studying the formant structure.

In the first phase of the experiments, data were recorded on the same speech sounds for different analyzing modes; a few speech sounds were then selected for

learning, and the discriminant function finally arrived at was tested.

A linear hyperplane method was chosen for this work. It has been proposed and used by Braverman [9] for the recognition of printed digits. Turski [10] has published a modification of the algorithm and large parts of it were used for the present work. The hyperplanes were drawn perpendicular to and at the middle of the "hyperline" joining two "hyperpoints" of different affiliation. Once all the speech samples reserved for learning had been used, the regions without affiliation were incorporated in a neighboring region affiliated to some category.

Four speakers were used; most of the experiments were done with the stop consonants because, while it is questionable whether it would really be profitable to use large-bandwidth filters to recognize consonants from one another, it seems certain that the recognition of vowels will become more difficult under these conditions. Because of the limitations in computer time, it was resolved to take the consonants by pairs. About 35 samples of each consonant were recorded, 5 of which were reserved for learning.

The starting point of the stop consonants was taken to be the earliest point at which two adjacent samples at frequencies higher than 600 Hz would equal or exceed a threshold just high enough not to be exceeded by occasional noise. The data on the precursive voicing were disregarded.

A portion of the speech sound had to be selected for study, and the segmentation had to be carried out not only between the phonemes but also inside the phoneme. For example, one part of the stop consonants consists of a burst of noise; another contains the frequency transitions from the burst to the vowel. It is normal to expect that the optimum mode of the analyzer would not always be the same for all parts of the consonant. One had, at first, to proceed by trial and error; after some time, it became convenient to study the first 20 ms of the consonant, the first 40 ms, and the second part of the consonant going from 20 or 30 ms to 50 or 60 ms after the start of the vowel. Every pair of stop consonants was tested in these three time intervals. The experiments were repeated with new learning samples. Table I summarizes the results in the cases where significant differences were obtained.

A conclusion which can be drawn from these results is that the 100-Hz mode has shown itself superior to the 200-Hz mode for recognizing vowel sounds and for classifying consonants with the help of the frequency transitions between consonants and vowels. In other words, the 100-Hz representation has shown itself more capable of classifying sounds correctly in these cases where, according to our knowledge of speech characteristics, sounds are characterized and should be differentiated by their resonance pattern.

The 200-Hz mode has, on the whole, shown itself superior to the 100-Hz mode when considering the start of stop consonants and, in particular, when trying to distinguish between two stop consonants whose loci were in the same frequency region. For the initial part of the six pairs of stop consonants studied, the 200-Hz mode proved to be significantly superior to the 100-Hz mode for three pairs (p–b, p–t, d–t); both displays seemed to be nearly equivalent in two other cases (p–d, b–t); and the 200-Hz display appeared somewhat better for the other pair (d–b). In general, it can be said that the 200-Hz mode was at an advantage when considering the initial part of the stop consonants, although this might not be true in every case.

This study is incomplete in this sense, that it represents only a first step in the study of an important problem. It might have been more interesting to compare 100-Hz and 400-Hz filters. However, it can already be seen that small differences in the design of a spectrum analyzer can produce substantial differences of recognition performance. The fact that one cannot conclude that one mode always performs better in given conditions makes it difficult to implement an automatic adaptive spectrum analyzer. It might be a better idea to use in parallel a few spectrum analyzers with different filtering characteristics.

MICHEL LECOURS
Laval University
Quebec, Canada
J. J. SPARKES
University of Essex
Essex, England

REFERENCES

[1] R. S. Thomas, "A real-time audio spectral analyser using active filters with adjustable parameters," Ph.D. dissertation, University of London, London, England, 1964.
[2] D. Gabor, "Theory of communication," *J. IEE (London)*, pt. 3, vol. 93, p. 429, 1946.
[3] R. Oetinger, "Die Grenzen der Hörbarkeit von Frequenz- und Tonzahländerungen," *Akustica*, vol. 9, p. 431, 1959.
[4] Chic-an Liang and L. A. Chistovich, "Frequency difference limens as a function of tonal duration," *Soviet Phys.—Acoust.*, p. 75, July 1960.
[5] A. R. Sekey, "A study of auditory perception in the time frequency domain," Ph.D. dissertation, University of London, London, England, March 1962.
[6] D. M. Green, T. G. Birdsall, and W. P. Tanner, "Signal detection as a function of signal intensity and duration," *J. Acoust. Soc. Am.*, vol. 29, p. 523, 1957.
[7] C. D. Creelman, "Detection of complex signals as a function of signal bandwidth and duration," *J. Acoust. Soc. Am.*, vol. 33, p. 89, 1961.
[8] D. M. Green, "Auditory perception of a noise signal," *J. Acoust. Soc. Am.*, vol. 32, p. 121, 1960.
[9] F. M. Braverman, "Experiments on machine learning to recognize visual patterns," *Automation and Remote Control*, vol. 25, p. 315, 1962.
[10] W. Turski, "A learning automaton for solving stability problems of differential equations," *Computatio*, vol. 1, p. 57, 1964.
[11] M. Lecours, "Adaptive spectral analysis for speech-sound recognition," Ph.D. dissertation, University of London, London, England, May 1967.

The ac Resistance of Carbon Microphones

IEEE Standard No. 258[1] gives in paragraph 5.3 a method for determining the impedance of a carbon microphone. The measurement taken, namely the ratio of the dc voltage drop across the microphone and the dc exciting current, is actually the effective dc resistance of the microphone, acoustically excited. This has been called the "speaking resistance." Since a carbon microphone is virtually nonreactive, it is often wrongly assumed that its speaking resistance is equal to its effective ac source resistance or its impedance. Similarly, under "quiet" conditions, i.e., with very low

Manuscript received June 5, 1968.

[1] "IEEE Standard on Test Procedure for Close-Talking Pressure-Type Microphones," *IEEE Trans. Audio and Electroacoustics*, vol. AU-14, pp. 156–162, December 1966.