Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 22, No. 6, 761–762, November, 1950

Spectrum Analysis

FRANKLIN S. COOPER
Haskins Laboratories, New York, New York
(Received August 9, 1950)

THE study of speech sounds from the point of view of spectrum analysis has at least three things to recommend it. They are, in order of increasing importance, the fact that familiar techniques and concepts are utilized. Second, that spectrum analysis is a "natural" way in which to think of auditory phenomena, since the ear itself appears to perform frequency analyses. But perhaps the most important reason depends upon the fact that the end result of spectrum analysis is a spectrogram. In 1934, Steinberg published a hand-drawn spectrogram of that classic sentence recounting the fate of father's shoe bench. Since then, the development of the sound spectrograph at the Bell Telephone Laboratories has resulted in spectrum analysis becoming almost a standard procedure for speech studies.

It is true that the sound spectrogram contains a great deal of detailed information about speech waves. Its real value, however, lies in the organization of that information into a "picture." In this way, complex sound patterns, which we readily perceive as entities by ear, have been so processed that they are now readily perceived also by eye. This is to say, one can now compare the same phenomena when they are presented via two separate sensory modalities.

It is this cross comparison, carried to its logical completion, which is the underlying rationale of the special techniques which I should like to describe and demonstrate for you. The experimental procedure involves two steps, first the representation of speech in terms of a sound spectrogram, and second, the reconversion of these visual patterns—after appropriate simplifications and modifications—back into sound for evaluation by ear. In this way, one can check the conclusions based on the visual interpretation of spectrograms in much the same sense that a chemist turns to synthesis as the final proof of analysis. Thus, statistical description and synthesis are complementary procedures.

The spectrograph which we have built for this work does not differ in principle from the recording spectrograph developed by Dr. Potter and his co-workers. There are very marked differences in engineering design, since our spectrograph is intended primarily to produce photographic transparencies for use in the synthesis of speech sounds. I shall not spend much time on a description of the equipment, since a number of you have already heard it described; moreover, an account of the constructional details will soon be published. The principal design objectives were to secure a linear representation of intensities over a dynamic range of about 40 db, and to have the spectrogram appear as a transparency.

The spectrograms obtained with this equipment are large in size to permit easy changes and hand painting of modified copies. The time scale is 7.2 in./sec.; the frequency scale is 1200 cycles/in. with a maximum frequency of about 7200 cycles; a dynamic range of 40 db is represented by a density range of 2. A single spectrogram represents a recording time of 12 sec.

The operating principle of the pattern playback employs about the simplest possible way of reinserting the frequencies specified by the spectrogram. This principle has been described in the patent literature by Dr. Potter. The tone wheel which we are now using supplies the first 50 harmonics of a fundamental frequency of 120 cycles, thus giving a frequency range of 0 to 6000 cycles. Obviously, such a machine will "speak" only in a monotone; thus far, this has not been a serious restriction. The spectrogram from which the sound is reconstituted can be a photographic copy of an original spectrogram, in which case the appropriate frequencies are transmitted by the film to the light collector and photo-cell. For synthetic spectrograms, on the other hand, it is most convenient to paint only what one wishes to hear. Hence, flat white paint on clear cellulose acetate is used, and the useful light is that which is scattered back into the light collector by diffuse reflection from the painted areas. Convenience of operation has proved to be a most important consideration in the design of equipment of this sort.

What are some of the uses to which such equipment can be put? It was designed primarily to follow the rapid changes of connected speech—to permit a study of the interactions between the individual sounds of speech—although it can also be used to synthesize single phonemes. But it is in the capacity to deal with the dynamics of speech that the major virtue of the equipment lies. At the present time, it is being used for both steady-state and dynamic studies. Professor Delattre has underway an analysis of French nasal vowels. Also, some work on connected speech is being done by Dr. Alvin Liberman, Mr. John Borst, and the author. This has, quite naturally, started with tests of the intelligibility of sentences and of nonsense syllables—a necessary step in checking the performance of the equipment. Moreover, it provides textual material which can then be degraded in various ways, with an indication of the consequent loss of intelligibility. Interactions between vowels and consonants can readily be studied, and there are some very interesting problems in this area. The simplification of speech in various ways is an aspect to which we have already given some attention. Simplification can take the form of the selective omission of various components, or omission combined with schematization, or in the extreme case, a reduction to standard symbols. You will soon have a chance to hear some of the results of our preliminary experiments in schematization. What is perhaps the central problem of speech analysis at present lies in finding the invariants of speech, that is to say, in finding generalized characterizations of the various phonemes. A related problem is to determine the range of acoustic variation which does not trespass across phonemic boundaries. Obviously, the ability to modify, or even to synthesize sound patterns is very useful in this connection.

To recapitulate, the principal virtue of spectrum analysis lies in the cross comparisons which it permits between audible and visible patterns. The synthesis of speech sounds provides a most useful, if not indeed a necessary, counterpart to analysis and statistical description.

I should like to say that we have been greatly aided in this work by Professor Delattre of the University of Pennsylvania, and that the work itself was made possible by funds granted by the Carnegie Corporation of New York.

(The above paper was followed by slides and sound recordings to demonstrate the operation of the pattern playback and the extent to which the spectrographic patterns could be simplified without serious loss of intelligibility.)