SCIENCE

VOL. 102 FRIDAY, NOVEMBER 9, 1945 No. 2654

SCIENCE: A Weekly Journal, since 1900 the official organ of the American Association for the Advancement of Science. Published by the American Association for the Advancement of Science every Friday at Lancaster, Pennsylvania.

Editors: JOSEPHINE OWEN CATTELL and JAQUES CATTELL.

Policy Committee: MALCOLM H. SOULE, ROGER ADAMS and WALTER R. MILES.

Advertising Manager: THEO. J. CHRISTENSEN.

Communications relative to articles offered for publication should be addressed to Editors of Science, The Science Press, Lancaster, Pa.

Communications relative to advertising should be addressed to THEO. CHRISTENSEN, Advertising Manager, American University, 3801 Nebraska Ave., NW, Washington 16, D. C.

Communications relative to membership in the Association and to all matters of business of the Association should be addressed to the Permanent Secretary, A.A.A.S., Smithsonian Institution Building, Washington 25, D. C.

Annual subscription, $6.00 Single copies, 15 cents

VISIBLE PATTERNS OF SOUND

By RALPH K. POTTER

BELL TELEPHONE LABORATORIES

THE automatic representation of speech sounds by visible traces or symbols has long been a subject of interest to acousticians and phoneticians, and especially to those concerned with the development of electrical communication. Techniques for automatically recording the wave forms of sounds have been very highly developed; but there has remained unsolved, until recently, the problem of recording sounds in a manner permitting their ready visual interpretation and correlation with the auditory sense. An outstanding difficulty with the interpretation of the records of wave forms is the effect of phase relationships between fundamental and harmonics. These effects may produce a marked difference in the appearance of the wave forms of two sounds that are quite indistinguishable to the ear. Consequently, wave traces of even simple vowel sounds do not permit of easy identification by the eye.

The facts are that wave traces contain too much information. To portray sound in a form that the eye can encompass in a glance requires that some means be provided for selecting the essential information and displaying it in an orderly fashion. A form of display that meets these requirements has been developed in the Bell Telephone Laboratories as described below.

The work here described was begun before the war. Because of related war interests it was given official rating as a war project, and has progressed far enough during the war period to justify its being brought now to public attention.

The possible uses of an automatic system for translating sound into patterns which may be readily interpreted by the eye are very numerous. It opens the prospect of some day enabling totally deaf or severely deafened persons to use the telephone and the radio or to carry on direct conversation by visual hearing. [The latter, incidentally, was an objective of the early researches of Alexander Graham Bell.] It suggests the possibility of printing words phonetically and of the automatic retranslation of such printed symbols into understandable sound. It opens the way to the selective operation of automatic devices by voice sounds. It promises to be particularly useful in the specialized fields of phonetics, philology and music. But most immediately, and from humane considerations most importantly, it opens a new avenue of help to the totally and severely deafened—help to learn to speak, and for those who already speak, help to improve their speech. It is to this problem of aid to the deaf that we have first directed our efforts.

It is too early to evaluate the results of these efforts with certainty. That there is a firm basis underlying the legibility of the visible speech patterns which have been obtained can hardly be doubted, but many questions remain concerning the design of practical translating equipment, the time and effort necessary to acquire a reading vocabulary, the effects of transmission and reception conditions on pattern legibility, and the special needs of the equipment for speech teaching and rehabilitation. These questions can not and should not be answered hastily, even though they are naturally urgent to those afflicted with a serious hearing loss. Their answering requires, during the developmental stages of the equipment, the cooperative efforts of the engineer and the groups concerned with the problems of deafness. The purpose of the present paper is to introduce the subject of visible patterns of sound and to describe briefly some of the more general aspects of the work with them.

In Fig. 1 are shown two forms of the new sound patterns. Both represent the same words—"This is visible speech"—and both show the three basic dimensions of sound—frequency, time, and intensity. It should be noted that the words associated with the different sections of the pattern, are inserted at the top of the figure. Time extends horizontally, the total length of each record being roughly 2½ seconds. Frequency is spread out vertically from substantially zero at the bottom of each record to about 3,500 cycles at the top. Intensity is shown by the varying shades of gray. Resolution of the frequency dimension is the significant difference between this and other more familiar displays of sound such as the oscillogram. Such a display provides for the eye the frequency analysis which is natural in aural perception and necessary for an understanding of sounds.

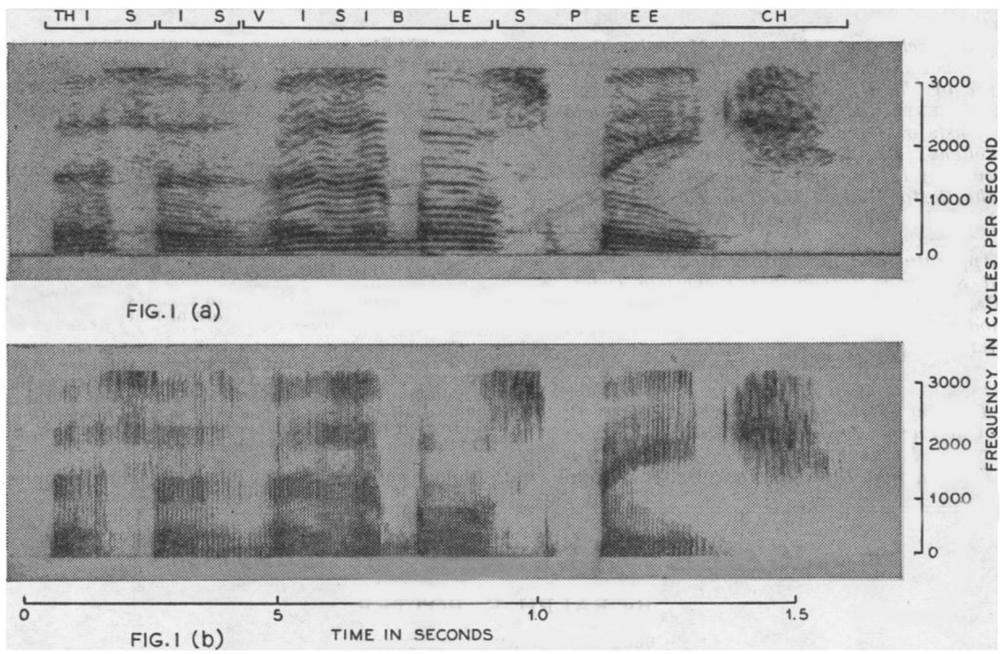

FIG. 1(a). High-detail sound spectrogram of the words "This is visible speech," showing the darker regions of mouth cavity resonance superimposed upon the detail of harmonics in voiced sounds.

FIG. 1(b). Low-detail sound spectrogram of the same words, showing the resonant regions as dark bands and eliminating detail.

Patterns of the type illustrated by Fig. 1(a) are of interest in studies of speech characteristics, while the type shown in Fig. 1(b) are of interest in visual hearing and phonetics. The former pattern shows the frequency composition in great detail, so that the individual harmonics of voiced sounds may be seen, and the manner in which the frequency of the harmonics varies with time. Patterns of the second type show only the broad frequency and time distribution of energy resulting from selective modulation. It is such modulation, produced by variations in the voice cavities accompanying the formation of word sounds, that conveys the information in speech. Incidentally, the vertical striations appearing in Fig. 1(b) are produced by beats between adjacent harmonics so that in the voiced parts their density or frequency of occurrence is a measure of pitch, increasing as the pitch increases.

The patterns of Fig. 1 and others described later were made by an instrument that we have called the sound spectrograph. In this instrument, the sound to be pictured is recorded initially on a loop of magnetic tape and played back repeatedly into a scanning filter, the pass band of which is moved slowly across the frequency spectrum. The scanning filter output is connected to a stylus that makes a trace upon a loop of electrically sensitive paper. Recording paper and magnetic tape loops are moved in a fixed relation so that successive scanning cycles are recorded side by side, thus building up a frequency-time-intensity picture.

If the words pictured in Fig. 1 were repeated at different times by the same speaker, the repetitions would look much alike unless a deliberate attempt were made to change the voice. If the words were spoken by different individuals they would also have a similar appearance, although the pattern shapes would vary with individual characteristics in much the same way that handwriting varies among individuals. To the extent that words sound alike they will also look alike in visible speech form, and to the extent that they sound differently they will look differently. That is, of course, to be expected if the portrayal is accurate.

The similarity in word patterns for various individuals is illustrated by Fig. 2. Here are shown six enunciations of the word "speech" by as many speakers. As indicated, the upper three are female voices and the lower three male. The brief descriptions at the right are only intended to give an impression of the wide range of voice quality represented and should not be interpreted as meaning that all speakers with a "Throaty Voice" or a "Scotch Irish Accent" or an "English Accent" would produce patterns that are just like those so identified. The important point in this illustration is the evidence that characteristic differences in pronunciation do not overshadow the similarities that enable one to recognize particular words. Such similarities illustrate the possibilities of these patterns for visual hearing.

In the studies of sound patterns as applied to the problems of the deaf, the objective thus far has been to determine whether the patterns are sufficiently intelligible for practical use. This has involved a careful study of the patterns of different sounds and sound combinations with respect to both their similarities and their differences. Two methods of investigation were utilized—the first by training individuals to read patterns and following their progress, and the second by what are called "visual discrimination tests" that require no training.

The training group formed for the purpose of learning to read the patterns originally included six girls with normal hearing. This group was assigned to the work on a part-time basis during the summer of 1943. Teachers were selected with experience in the fields of phonetics and education of the severely deafened, but they started the instruction with no previous knowledge of the speech patterns. In addition to this handicap, the first months of training were based entirely upon the sound spectrograph patterns which require several minutes for completion.

A method of instantaneous translation was needed badly, but equipment of this type was not available for class use until the fall of 1944. The first direct translator utilizing a new form of moving screen cathode-ray tube was entirely too large and complicated for any but experimental instruction use. There has since been constructed a much smaller translator, approximately the size of a portable typewriter, with the speech patterns displayed on a moving drum of phosphorescent material. While this more recent unit approaches a practical size, it is still very much in the experimental stage and far from a finished design.

There is also under experimental development a large screen translator of the same general design as the small unit but using a belt of phosphorescent material in place of the drum. In all three of these translators, transient speech patterns are formed by tracing on a moving screen the frequency distribution of speech energy as determined by a bank of fixed band pass filters. The frequency scale used for the speech patterns is linear, although patterns with other scales, including the logarithmic type, have been produced.

One objective in the development of the small translator is to provide a type of unit that might ultimately be associated with telephones in such a way as to permit the very deaf to carry on telephone conversations by seeing, rather than hearing, the speech signals.

Due, at least in large part, to the exploratory nature of the training program and the time required to develop adequate equipment, the ability of the experimental class to read visible speech increased at what would be considered a slow rate for learning lip reading, shorthand, or foreign languages. However, the learning rate improved considerably as the training methods and translators were improved. Recently a congenitally deaf engineer, who depends

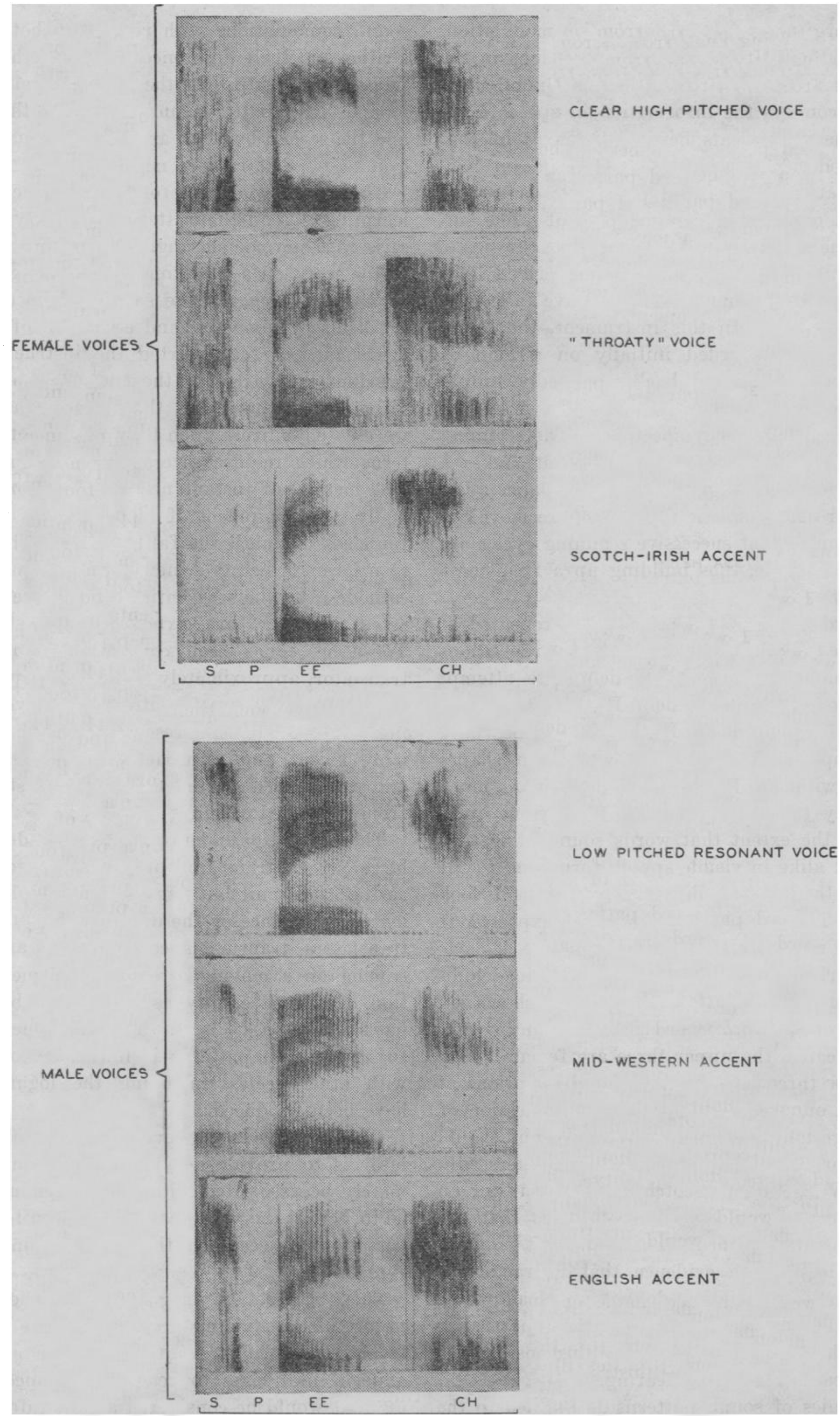

FIG. 2. Patterns of the word "speech" by six different speakers, including male and female voices and a wide range of pronunciation.

entirely upon lip reading, has been added to the class. His learning rate for visible speech has so far compared favorably with that for lip reading. During his training he has been tested regularly on his ability to read word patterns in the visible speech form and by reading lips. Although adept at lip reading, and the lip reading tests were carried on under exceptionally favorable conditions, his score on reading visible speech has stayed well above that for reading lips. Incidentally, his is probably the first case in which a person with substantially no hearing has been enabled to talk over an ordinary telephone circuit without the aid of a human "interpreter."

Thus far the results of the experimental training encourage confidence that these speech patterns may provide a practical form of directly translated visible speech, but much more training experience is needed for complete confirmation.

A second way to determine the legibility of word patterns, by means of so-called "visual discrimination tests," was initially devised to get around a difficulty. It had been assumed at the start of the visual hearing project that the trainees would, after a period of training, be able to say whether one pattern presentation was better than another, and how the many variables that appear in the translator design should be treated. But this assumption proved to be mistaken. After training with one form of pattern there was a tendency to dislike others. It soon became obvious that an unbiased evaluation method was necessary as a guide to development.

The visual discrimination tests produced to meet this situation depend upon an assumption that any language, aural or visible, is made up of many patterns; and that the relative merit of different languages or of different representations of a single language depends upon the ease with which patterns that make up equivalent vocabularies may be identified. In the visible speech case, one test method is to select words that in certain respects look alike, and arrange them in what are called "similarity series." Examples are "man, ran, van, tan," etc. The words are spoken in groups of three, such as "van-tan-tan" or "van-van-tan." An observer of the patterns produced by these words is simply asked to check whether the middle pattern is more nearly like the first or last in a group. It is not required that the observer have any knowledge of the meaning of the patterns. Ratings are in terms of the percentage of correct pattern identifications, taking into account the fact that 50 per cent. accuracy represents pure guesswork. The figure so derived is called the "discrimination index." Poor patterns result in a low DI, or if the translator fails to show certain sounds clearly, the DI for word groups containing these sounds will be relatively low. Figures of this kind permit a quantitative appraisal of the performance of different translators.

Any one who considers carefully the relation between aural and visual perception is likely to question the need for using a three-dimensional form of pattern for visible speech. The ear hears two-dimensionally in frequency and intensity. The time dimension is supplied by the memory. Why not then show only a frequency-intensity speech pattern and make similar use of the visual memory? The question is a perfectly valid one and has received considerable thought during the investigation under discussion. In fact, an experimental study of two-dimensional patterns is being carried on simultaneously with the three-dimensional studies. At the present time it is confined to visual discrimination tests of various two-dimensional displays. Earlier, one girl studied a particular form of pattern for a few months, but when the discrimination test methods were developed this training was discontinued in order to concentrate the available effort upon the more fundamental aspects of two-dimensional display.

Both the limited training and the visual discrimination tests made so far seem to indicate that it will be more difficult to read the two-dimensional patterns than it is to read the three-dimensional type. Three-dimensional speech patterns are analogous to print moving from right to left on a telegraph tape, while two-dimensional patterns are analogous to seeing this moving print through a narrow slit, only as wide as the lines that form the letters. No doubt we could learn to read print moving past such a narrow slit, but it would rather obviously be more difficult than unrestricted reading because we normally perceive whole words rather than bits of letters. It may well be that this manner of reading is a result of the requirement that the eye be focused upon the print in order to obtain a satisfactory memory impression. Although the focusing requires visual effort, large pattern areas may be recorded in a single "exposure" so that the effort is not excessive. But to record patterns a bit at a time, or two-dimensionally, by the visual process would require almost continual concentration and should therefore exact more effort for the same accomplishment.

It is possible then that even though the eye should in theory be able to understand two-dimensional patterns of speech it may be inefficient for this purpose.

In addition to the problem of understanding others, deaf and severely deafened persons are faced with the problem of controlling their own speech. Making sounds without ears to hear them is somewhat like drawing pictures with no eyes to see the results. When the hearing is largely absent in a child, it is necessary to teach speech without the normal control exercised by the hearing. Even in cases

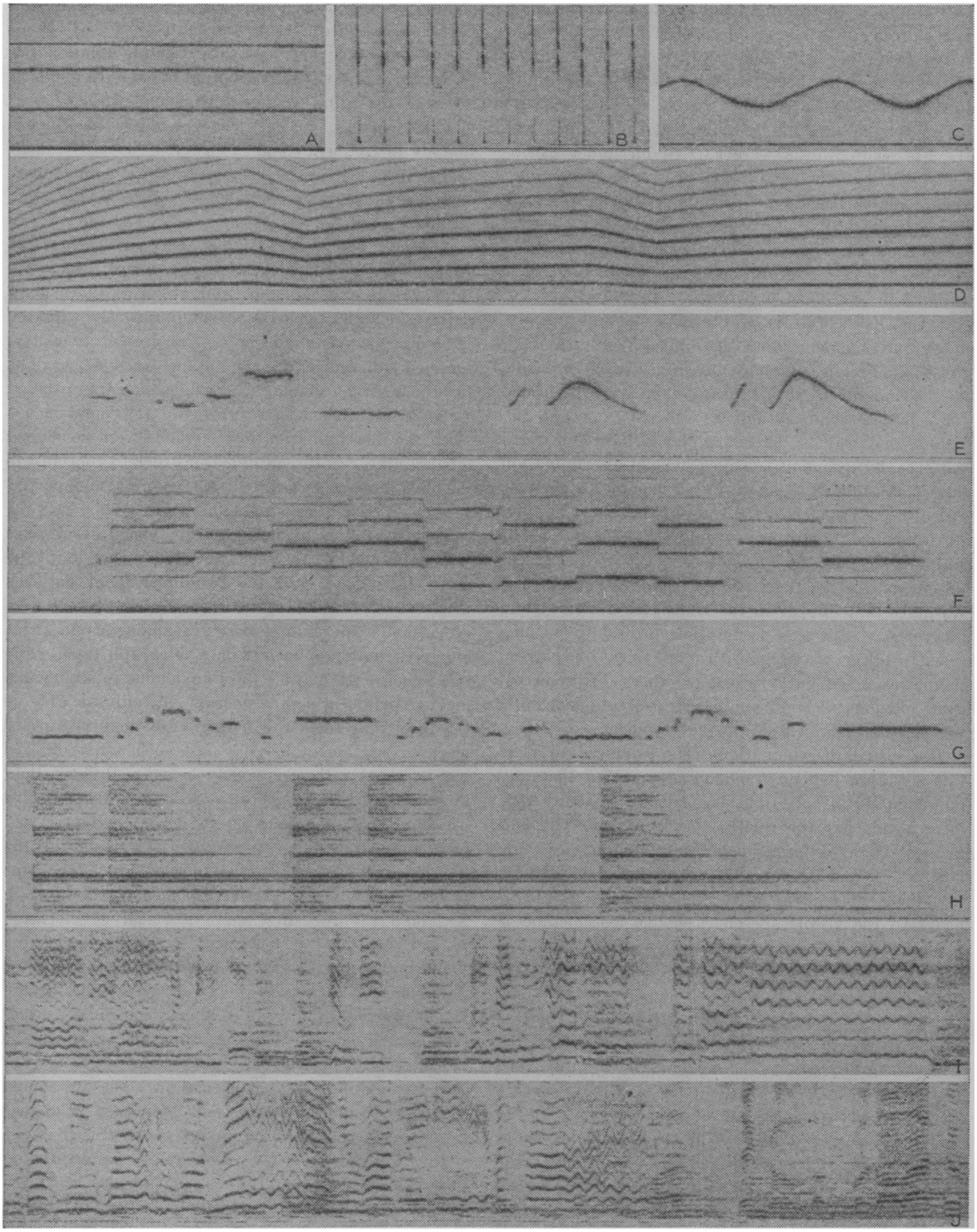

FIG. 3. Sound patterns of:

A—Three steady tones
B—Hammer clicks
C—Wobbled tone
D—Siren
E—Man whistling
F—Oboe solo
G—Ocarina solo
H—Ship's bell ("5 bells")
I—Tenor and orchestra
J—Soprano and baritone

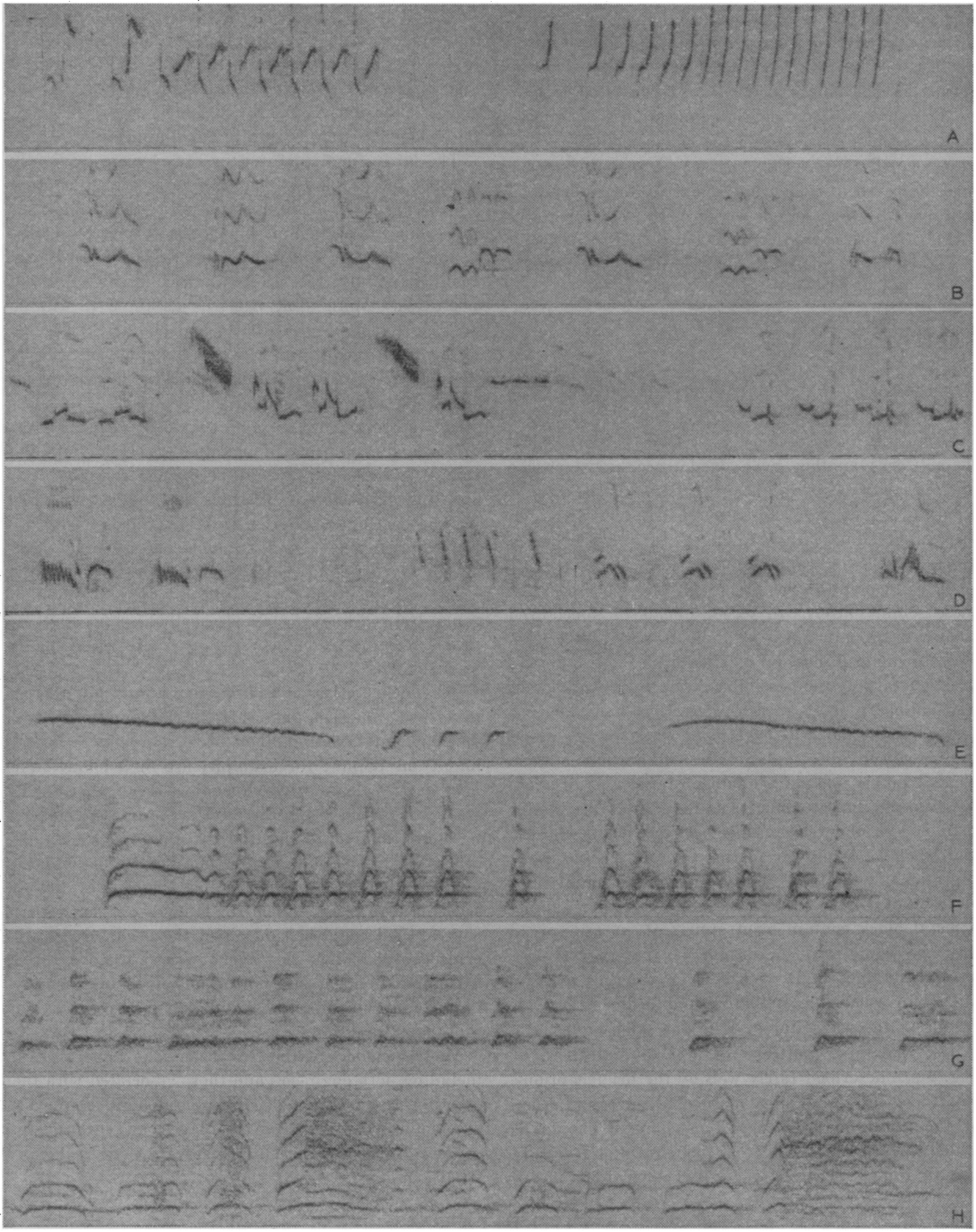

Fig. 4. Sound patterns of:

A—Cardinal
B—Robin
C—Mocking bird
D—Brown thrasher
E—Screech owl
F—Large dog barking
G—Small dog yapping
H—Baby crying

where a vocabulary is developed before severe loss of hearing sets in, there is a gradual deterioration of the speech. In both cases visible speech in the forms described here should be of considerable help in speech training and rehabilitation. This has proven true in the case of the congenitally deaf engineer noted above. During the past year, his speech has improved considerably and can be understood quite well by the average person with whom he comes in contact. His work with the translator patterns, however, has emphasized the desirability of providing an indication of the voice pitch.

Unless the acoustically handicapped are able to control the pitch and the volume of the voice as well as the positions of the articulators, their speech will sound unnatural although it may be intelligible. The best ways to display pitch are still uncertain. One possibility is by means of a wave trace perhaps below the pattern; another by a fixed trace, in which the line intensity is varied in proportion to the fundamental frequency. The latter is simpler from an apparatus standpoint, but in a first experimental use seemed less acceptable as an indication.

There remains to be discussed, potential uses for these sound patterns in various fields of specialized acoustics for purposes of analysis and illustration. The foregoing discussion of speech patterns indicates in a general way the possibilities as applied to phonetics, philology, and speech correction and development, but the patterns have many uses for the visual interpretation of complex waves other than speech. Figs. 3 and 4 include a number of pictures of more or less familiar sounds that may illustrate better than speech the relationship between what we hear and what we would see in this form of visible interpretation. Sustained tones produce horizontal lines as in A of Fig. 3. The clicks of a hammer against a metal block contain brief spurts of energy spread over the whole frequency range, so that they appear as vertical lines in B of Fig. 3. Swinging the frequency of a variable oscillator up and down the scale results in the wavy line of C in the same figure.

The remaining patterns of Fig. 3 and those of Fig. 4 are described fairly well by the brief captions. For those interested in a more detailed examination of the time and frequency dimensions it should be added that in Fig. 3 (A to G inclusive and I and J) and in Fig. 4 (E to H inclusive) the length is approximately 9.4 seconds and the vertical scale includes a frequency range of zero (at the black base line) to 3,200 cycles per second at the top. In H of Fig. 3 and in the first four patterns of Fig. 4 (A to D inclusive) the length is approximately 4 seconds and the frequency scale is zero to 7,500 cycles per second.

The bird songs[1] pictured in Fig. 4 were originally selected for use as test material because they contain a wide variety of tone modulation without the complications appearing in sounds that are rich in harmonics. But these song patterns are obviously revealing and illustrate well the possibilities of sound portrayal. With such patterns as these it will be possible to analyze, compare, and classify the songs of birds, and, of even more importance, it will be possible to write about such studies with meaningful sound pictures that should enable others to understand the results. The same argument applies to an almost endless variety of inaudible as well as audible sounds of both natural and mechanical origin. Even such low frequency oscillations as those accompanying the beat of the heart may be recorded slowly and converted to the sound spectrogram form by high speed reproduction. Also frequencies beyond the upper range of the ear may be shifted to the audible range by well-known methods so that sound spectrograms may be made in a region where recording is less difficult.

In conclusion, it is well to point out that this is necessarily an incomplete story of the sound portrayal development. Nothing has been said about several important points: for example, use of a logarithmic frequency scale, and a frequency selection that corresponds more closely to the aural experience; and better amplitude representation by contours, color, and other means. Some interesting results have been obtained in developments along these lines but it seems best to reserve them for later discussion. Also, a great deal could be said about the need for a modernized alphabet in this age of speed when a rapid exchange of ideas and information is increasingly important, but this too had better be reserved for some later occasion.

Many members of the Bell Telephone Laboratories, including both engineers and those associated with the experimental training, have cooperated in this development work. All have displayed an enthusiastic interest and it is a pleasure to acknowledge their contributions.

OBITUARY

WALTER BRADFORD CANNON
1871–1945

A FEW months before his death Dr. Cannon's last book was published. It is called "The Way of an Investigator" and is an autobiography in the best sense, since it presents those aspects of his life which in long view he felt most useful to others. This book,

[1] These sound spectrograms were made from the Cornell Bird Song Records (Albert R. Brand Bird Song Foundation, Laboratory of Ornithology, Cornell University).