12✓

Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 28, No. 3, 509–511, May, 1956

Current Publications on Acoustics

FLOYD A. FIRESTONE
57 East 55th Street, New York 22, New York

Book Reviews

The opinions expressed are those of the individual reviewers and are not necessarily endorsed by the Editorial Board of this Journal.

Ultrasonic Engineering

ALAN E. CRAWFORD. Pp. x+344. Academic Press, Inc., New York, 1955. Price $8.00.

High-intensity ultrasonics provides a good illustration of the rapid way in which contemporary physics passes to the technological stage. From a scientific curiosity in the twenties it has, through the agency of electronics, become a significant branch of engineering with an impact on many industrial processes. The present book is an attempt to summarize in moderate compass and relatively simple language the many interesting and valuable practical applications of high-frequency sound.

After a very brief presentation of the theory of ultrasonic radiation and a chapter on cavitation in liquids which might perhaps have found a more appropriate place further on in the volume, the next 100 pages are devoted to the treatment of the generation of high-intensity and high-frequency sound. The remaining 180-odd pages discuss the specific applications including precipitation, emulsification, chemical action, metallurgical processing, coating of metals, as well as biological and medical applications. There is a final chapter on ultrasonic instruments and control gear.

The treatment is descriptive throughout with a minimum of emphasis on analytic detail. Those formulas presented are introduced for the most part from the technical literature without deduction, much of the analysis being replaced indeed by numerous clear and well-drawn graphs. The bibliography, though not extensive, is adequate for the engineering reader who wishes to follow up some special line. The style is clear and readable. The illustrations of ultrasonic equipment are excellent, and the same is true of the typography. All in all, the volume should find wide use among acoustical processing engineers.

R. B. LINDSAY
Brown University

Manual of Phonology

CHARLES F. HOCKETT. Pp. 246, Figs. 31. Indiana University Publications in Anthropology and Linguistics, No. 11, Bloomington, Indiana, 1955. Price $3.50

In recent years many linguists have begun to feel that the results obtained and obtainable in the acoustical laboratory have direct bearing on their own problems. This represents a significant departure from the previous position prevalent among linguists who, in practice, if not always in theory, agreed with Twaddell's opinion that the results of acoustics are of little interest, since "The presence of 'phoneme-features' as positive, additive entities in the sound waves is not demonstrable, and there is no reason to believe that it will be." [W. F. Twaddell, "On defining the phoneme," Language Monograph No. 16 (1935), p. 24.]

The study under review is devoted to the phonic means used for differentiating utterances and to the development of a typological framework for the classification and comparison of phonologic systems of different languages. In its treatment of acoustical problems it represents a full-fledged return to the earlier, negative position, for in Hockett's opinion, "acoustic phonetics can hardly ever be regarded as supplying any evidence for phonologic analysis." (p. 212) In the following I shall deal exclusively with the acoustical problems that arise out of the author's interpretation and review of the available evidence. I shall pass over entirely the many fascinating questions raised by Hockett in other areas of the subject, since these are primarily of a technical, linguistic nature and hence somewhat peripheral to the interests of all but a small minority of the readers of this journal.

Hockett's picture of the acoustical nature of speech is vividly portrayed in the following passage:

> Imagine a row of Easter eggs carried along a moving belt; the eggs are of various sizes, and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between the two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source; the mess that emerges from the wringer represents the output of the speech transmitter. At a subsequent point, we have an inspector whose task it is to examine the passing mess and decide, on the basis of the broken and unbroken yolks, the variously spread out albumen, and the variously colored bits of shell, the nature of the flow of eggs which previously arrived at the wringer. (p. 210)

It is understandable that anyone for whom the acoustical signal is "a mess" of this kind will have no use for it. It is, however, a fact well known to both engineers and linguists that for lack of a code book a prefectly good signal is sometimes mistaken for noise. A different evaluation of the evidence, including data omitted by Hockett, shows that while the acoustical analysis of speech is no simple matter, it is not quite as hopeless as his analogy would suggest. But before attempting a reinterpretation we must examine the basis for Hockett's conclusions, which is found in his picture of the present state and possibilities of acoustic phonetics.

Hockett's discussion of acoustics begins with a presentation of basic concepts. Though entitled "Physical Dimensions of Sound," the section is in reality concerned with both physical and psychophysical aspects of sound. In the light of recent research the wisdom of this move may be questioned. Indeed it would seem that this mixing of universes of discourse is at least partially responsible for not altogether felicitous statements like "the physical dimensions of sound, in addition to *duration* are *frequency* and *amplitude* or *intensity*; any other property of some specific sound can be expressed as a function of these three. There is one additional independent factor, *phase-relationship*, with which it will be convenient to deal a bit later; but there is much evidence to show that human ears do not perceive differences of phase relationships, so that it can hardly play any relevant role in the speech signal." (pp. 183–184)

It is unfortunate that in his commendable desire not to talk over the heads of his audience, which is primarily made up of nonspecialists, Hockett felt constrained to give an over-simplified account of the devices used in the study of speech sounds. "The two chief devices . . . are," according to him, "the oscillograph . . . and the spectrograph." (p. 185) Today the study of the acoustical properties of speech is to a large extent a problem in electrical measurements. No measurements laboratory limited to the two instruments mentioned can hope to contribute much to the solution of the problems.

Among the "methods of attack" Hockett mentions only the following two: (a) the study of "acoustic records (particularly spectrograms)" of carefully enunciated utterances and (b) the study of responses of listeners to "hand-drawn pseudo-spectrograms . . . varying in some systematic way one or more of the features which have been shown . . . to be of probable relevance." (p. 192) This is, in fact, the method of attack employed by the investigators at the Haskins Laboratories. No one acquainted with the field would want to minimize the significance of the work of these investigators, but in a well-balanced picture of the field their results would hardly be emphasized to the exclusion of almost all other work.

In view of this high regard for the work of the Haskins investigators it is difficult to understand Hockett's cavalier treatment of their results. In a discussion of the well-known experiment of the Haskins group [Word **8**, 195–210 (1952)] in which listeners were asked to match synthetic two-formant vowels with the 16 IPA cardinal vowels, Hockett reproduces one of the figures of the original article where the results are represented in the customary F_2 *vs* F_1 plot. The sixteen points on this graph are connected with four horizontal and four more or less vertical lines. It appears, according to Hockett, that the "more nearly vertical lines . . . look straight, but . . . are not, as a larger scale graph would show. However, by making slight adjustments in the second formants . . . we can make the four vertical lines straight." (pp. 195–196) I have tabulated below the results of this "slight adjustment" [the first figure after each vowel symbol is the original (Haskins) F_2 frequency; the second, Hockett's adjustment]:

[e] 2400 2411; [ø] 1650 1675; [ɤ] 1100 1098; [o] 800 813.2
[ɛ] 2000 2011; [œ] 1450 1485; [ʌ] 1150 1145; [ɔ] 950 938.4

One wonders why the recomputation was carried to four significant figures, or what purpose was served by straightening lines that were not crooked to begin with. One's puzzlement at this exercise is hardly lessened by the ingenuous comment that "the cardinal vowels are arbitrary to begin with, and, what is more, the very slight changes in second formant frequencies would scarcely be detectable by ear." (p. 196)

The rest of the section on vowels consists of fairly elementary measurements of vowel formants of French and English, a discussion of the Haskins experiment with one formant vowels and an attempt at correlating formant positions with tongue configurations.

The discussion of consonants which concludes the review of "work done so far" is made up of three parts: "*r*-color and nasalization" are treated according to Joos' "Acoustic phonetics," Language Monograph No. 23 (1948) supplemented by some Haskins data that have been superseded by recent experiments. Fricatives are disposed of in three short paragraphs. Stops are discussed in the light of the well-known experiments of the Haskins group [J. Acoust. Soc. Am **24**, 597–606 (1952) and Am. J. Psychol. **65**, 497–516 (1952)], which are summarized in fair detail.

From what was said above it is evident that Hockett's discussion of acoustical and psychoacoustical problems leaves something to be desired and that except for the work of the Haskins group the rich literature on the acoustics of speech is ignored by the author. Since Hockett's conclusions are based on the evidence reviewed, doubts as to their validity may justifiably arise. Nevertheless one cannot reject these conclusions unless it can be shown that a different interpretation of the facts is not only possible but also preferable. This I shall now attempt to do.

The basic problem of interest to the linguist might be formulated as follows: What are the rules that would make it possible to go from the continuous acoustic signal that impinges on the ear to the symbolization of the utterance in terms of discrete units, e.g., phonemes or the letters of our alphabet? There can be no doubt that speech is a sequence of discrete entities, since in writing we perform the kind of symbolization just mentioned, while in reading aloud we execute the inverse of this operation; that is, we go from a discrete symbolization to a continuous acoustic signal.

A commonly held view pictures the discrete units as norms, as bull's eyes which the speaker is trying to hit and which he misses slightly. (See also Hockett, p. 199 and Fig. 27.) If we admit the validity of this model, we are forced to accept Hockett's conclusion that there is little hope for an acoustic description of the phoneme. It must, however, be added that things are no better on the articulatory level and it is only ignorance of the details of the articulatory process that allows one to affirm that on the articulatory level there is greater uniformity. The excellent x-ray movies prepared by the Haskins group in collaboration with the University of Rochester Medical School show clearly how illusory this articulatory uniformity is. The difficulty is, however, not so much inherent in the facts as in the manner in which the question is posed; it stems directly from the bull's eye analogy.

When we speak of hitting the bull's eye we have in mind some ideal event in space and time, namely phoneme x, which we are trying to duplicate as best we can. This is, however, not the only possible model of the speaking process. We can say that in pronouncing a phoneme rather than trying to duplicate an ideal pattern we are attempting to transmit enough information so that the pattern would be recognized. As long as we transmit all the distinctive differences—the features that differentiate phoneme x from all other phonemes of the language—our listeners would recognize it correctly regardless of whether or not the various actualizations of the phoneme resembled each other on the spectrogram.

Implicitly this fact is recognized by our traditional way of categorizing the sounds of speech as stops, fricatives, vowels, voiced, nasalized, etc. For it is these properties that signal the distinctive differences, and it is by knowing which of these properties are actually present in the signal that we can determine which phoneme was uttered. The study of the sounds of speech is in the first instance a search for these distinctive cues; they are the invariants of speech.

Each device employed in the analysis of the speech signal must first be subjected to the test of whether or not it detects the distinctive cue(s) in question. The spectrograph portrays certain of these cues very clearly, for example, voicing and vowel formants. It obscures others, e.g., spectral properties of the stop burst and changes in intensity. It is, therefore, a good instrument for the study of the former attributes of speech and not of much value in the study of the latter.

Similarly, the fact that transitions in the second vowel formant are very striking is an interesting acoustical fact; in the study of speech this fact, however, has interest only after we have shown that it does indeed function as a distinctive cue. Furthermore, the fact that certain acoustic cues can be easily isolated from one another by available equipment is no argument for isolating them. If simpler rules for identification can be stated by treating two or more acoustically separate cues together, they should be treated together and not separately.

It is to be noted especially that the distinctive cues are not always statable simply, in absolute terms. A case in point is the center frequency of the burst in the Haskins experiments on the perception of stops before artificial two formant vowels. It was found, for example, that a burst centered at 1440 cps was judged /k/ when followed by /ɔ/, and /p/ when followed by /i/. In absolute terms there is a clear case of overlap, which has caused a great deal of discomfort to various investigators. If, however, the question were posed in relative terms, e.g., "What is the frequency position of the burst in relation to the second formant of the following vowel?" we would have the simple formula that in the region below 3 kc /k/ judgments were caused by placing the burst somewhat above the second formant of the succeeding vowel, and /p/ judgments anywhere else. Such a statement does, of course, take into consideration the immediate context in which a phoneme is found, but then nobody would expect events which succeed each other as rapidly as do phonemes not to affect one another.

A slightly different problem is touched upon by Hockett when he points out that it is apparently impossible to find a single formula that would describe the data on stop perception obtained by the Haskins group in experiments where the stop was simulated by a burst and in experiments where it was simulated by transitions in the vowel formants. The difficulty here is due to the artificial limitation of the vowels to two formants.

In experiments with three-formant vowels, which were described in the Haskins Laboratories' Final Report No. 2 and Quarterly Progress Report No. 9, though unfortunately never published elsewhere, /k/ judgments were produced by bending formants two and three towards each other, i.e., concentrating the energy in the region above the second formant. On the other hand, /p/ and /t/ judgments were caused by the absence of an energy concentration in the region above the second formant. In this connection the behavior of F_2 (formant two) in the /t/ stimuli is particularly interesting. In the case of the front vowels, where F_2 is close to F_3, F_2 was bent downwards, i.e., away from F_3, so that there was no concentration of energy in the region above F_2. On the other hand, when F_2 and F_3 were far apart, as in the back vowels, F_2 was bent upwards.

These facts obviously are in agreement with observations made with stop bursts mentioned above: /k/ judgments are produced by a concentration of energy in the region somewhat above the second formant; /p/ and /t/ judgments by an absence of such a concentration. The observations with transitions in two-formant vowels can be explained without too much difficulty by taking the three-formant case as a point of departure.

The suggested formula is further supported by the report (Haskins Quarterly Progress Report No. 9, p. 14) that in synthesizing syllables with both burst and transition great difficulties were experienced with /p/ in position next to vowels with low F_2. It was found that in this case omitting the burst altogether, i.e., eliminating the concentration of energy in the F_2 region, increased /p/ judgments considerably.

The point has often been made (and it is made by Hockett too) that speech can be perceived correctly even in the absence of many acoustic cues including phonologically distinctive cues. For acoustic phonetics this remark has less relevance than is usually thought. On the one hand, there are contexts in which the correct identification can be made only if all distinctive features are correctly identified; for example, for English consonants the context "Please pronounce after me the word '—ill'." On the other hand, if certain acoustic cues are omitted in the utterance, no acoustic analysis can show more than that they have been omitted. By knowing the rules of phoneme distribution we may be able to infer what was left out. This, however, is a logical procedure involving our knowledge of phoneme probabilities and not their acoustical properties.

The final problem is connected with indicating exact boundaries for the beginning and ending of each phoneme. This problem has been raised particularly with regard to the so-called imploded (burstless) stops; e.g., /p/ in the word "apt," when /pt/ is pronounced with but a single explosion. Since the cue for this kind of stop clearly is contained in the vowel transition, it was argued that the vowel is a sufficient cue for the entire stop-plus-vowel syllable. Curiously, the fact was omitted that in order for a stop to be perceived at all the transition must be adjacent to a "silence." If the "silence" is filled by something the stop is no longer perceived. The existence of segments belonging to two adjacent phonemes must not obscure the fact that the cues are sequentially ordered; i.e., every phoneme has extension in time (suprasegmental features like stress are not phonemes, but simply phoneme attributes).

Our picture of the process of speech production and perception resembles but slightly Hockett's belt and wringer. We see acoustic signals succeeding each other in time, with adjacent segments exerting an influence on each other. The signals, however, can be interpreted only by the person who is familiar with the code; that is, who knows what questions to pose. Without this information, the signals are noise, gibberish—not human speech, but a mess of broken eggs and shells passed through a wringer.

Morris Halle
Research Laboratory of Electronics
Massachusetts Institute of Technology