

## PART I. THEMATIC ESSAYS

### A SHORT HISTORY OF ACOUSTIC PHONETICS IN THE U.S.

Ignatius G. Mattingly

*Haskins Laboratories and University of Connecticut*

Acoustic phonetics concerns itself with those topics in experimental phonetics that can reasonably be investigated by the analysis, manipulation, and synthesis of speech signals. These topics include the spectrotemporal structures of the signals and the relation of these structures both to the vocal tract configurations that produced them and to the linguistic units intended by speakers and perceived by hearers. In this essay, some account is given of the history of acoustic phonetic research in the U. S. Speech perception research, however, is considered in a separate paper in this volume and will be referred to here only incidentally.

Acoustic phonetics in the U. S. began in 1879 with a paper read by A. G. BELL before the National Academy of Arts and Sciences. By this time, European scientists such as Willis, Wheatstone, and Helmholtz had already defined two questions that were to preoccupy the field for the next sixty years. First, how many vowel "pitches" (i.e., resonances) are there, and how are they related to the cavities of the vocal tract? Helmholtz's (1877-1954) experiments had led him to the conclusion that these cavities acted as simple resonators. There was one pitch for a back vowel, produced by a single resonator, and two pitches for a front vowel, produced by two connected resonators. Second, are the vowel pitches laryngeal partials amplified by the resonators (the Chord-Tone Theory, proposed by Wheatstone, 1837, and favored by Helmholtz) or are they unrelated to the laryngeal partials, resulting from the momentary excitation of the resonators by successive laryngeal pulses (the Inharmonic Theory, proposed by Willis, 1830). Although Rayleigh (1894-1945) argued correctly that both theories were true and consistent with each other, the issue was to remain alive through the 1930s.

BELL was well aware of these questions. In his paper, he claims to have found, by tapping his throat or cheek, two pitches, systematically related to cavity size, for *all* the vowels, but does not report their values. He supports the Chord-Tone Theory, arguing that laryngeal partials are reinforced to some extent even if they are merely close in frequency to the cavity pitch. But he refutes a variant of this theory proposed by the British investigators Jenkin and Ewing (1878), according to which the pitches of a vowel maintain a constant harmonic relationship, by showing that the quality of a recorded vowel does alter when the speed of rotation of a phonograph cylinder is suddenly altered.

BELL's paper did not immediately inspire further work in acoustic phonetics. At this time, there was in the U.S. no strong tradition of experimental phonetic research such as existed in Britain and Germany, and no further significant acoustic work was done until the next century. Bell's real contribution to

acoustic phonetics is surely the insight expressed to his assistant Thomas Watson during their experiments in telegraphy: "If I could make a current of electricity vary in intensity precisely as the air varies in density during the production of a speech sound, I should be able to transmit speech telegraphically" (Watson, 1915, p. 1506). Not only the telephone, but the electrical technology that was to transform acoustic phonetic research and the theories later developed to explain the relation between the vocal tract and the speech signal, depend ultimately on the analogy between sound and electricity, whose practical importance Bell seems to have been the first to recognize.

#### 1900-1920

But electrical technology did not become available for phonetic research for quite a while. In the meantime, Fourier analysis of speech was a laborious and not very rewarding affair. First, an oscillogram of the sound wave had to be prepared, by optically transducing the vibrations of the membrane of a horn receiver, as with D. C. MILLER's Phonodeik (1916-1922); or the varying depth of the grooves of a wax cylinder recording (as in L. Bevier's work (e.g., 1900). It was of course extremely difficult to do this accurately and reliably and the frequency range now known to be important for speech was very poorly preserved. From the oscillogram, Fourier coefficients were computed from series of ordinate values measured for each pitch pulse, either by hand (Bevier) or with a Henrici analyzer (Miller). Using this device, the experimenter traced the waveform of a pitch pulse with a stylus, the movement of the stylus was mechanically analyzed, and the sine and cosine coefficients of five components appeared on dials. To do twenty components and then compute amplitudes and phases must have taken well over an hour for each pitch pulse.

In general, the number of resonances and their values obtained by these early investigators were consistent with those given by Helmholtz. Indeed, Helmholtz's account was so influential that contradictory data were brushed aside. For example, although a back vowel was supposed to have only one resonance, both Bevier and MILLER report cases where [a] has two resonances at frequencies now known to be appropriate for F1 and F2 of this vowel. Of one such case, Miller observes, "The double peak for this vowel is peculiar to certain voices, and probably there is only one resonance, which is separated into two parts by the absence of a particular partial tone from the sound of a particular voice". (1916-1922, pp. 226-227).

MILLER also synthesized some vowels and CV syllables, using wooden organ pipes with adjustable output levels. Each

pipe produced one harmonic component of the vowel being synthesized, ten pipes being required for a baritone [a]. By opening and closing the air supply, he synthesized the word *papa*.

While MILLER and Bevier accepted the Chord-Tone Theory, the psychologist E. W. SCRIPTURE was a fierce partisan of the Inharmonic Theory, as elaborated by the German investigator L. Hermann (e.g., 1893). Scripture (1906) insisted that a standard Fourier analysis gave a false picture of the vowel spectrum, and proposed a method for analyzing the speech waveform into "frictional" (i.e. damped) sinusoids. He also synthesized waveforms, though not actual signals, from such components. Understanding that the vocal resonators were interdependent, he accomplished this by using two pendulums of different lengths linked so as to vibrate together, anticipating the principle of the serial formant synthesizer. He also gave much attention to "qualitative analysis", that is, close scrutiny of waveforms of connected speech, at a time when others looked only at sustained vowels. His commentaries are remarkably insightful, anticipating many ideas not commonplace until much later. For example:

...the cavity tones in the spoken vowels are never constant. This fact, when thoroughly understood and recognized, must effect changes in the prevailing views of sounds found in the books on phonetics and the dictionaries. These are really written with notions of sounds that are derived from typography and not from actual speech; the conclusions often have little relation to the really spoken sounds. (1906, p. 41).

#### 1920-1945

After World War I, the Research Laboratories of the Bell Telephone System undertook an ambitious program of research on speech and hearing (Fletcher, 1929). The Bell electrical engineers found it natural not only to use electrical methods but also to draw on electrical analogies in thinking about phonetic questions.

An early example was Stewart's "Electrical Analogue of the Vocal Organs" (1922). Stewart's device, the first electrical synthesizer, was actually a *terminal* analog (FLANAGAN, 1957) formant synthesizer, consisting of a buzzer exciting two resonant circuits in parallel. Recognizable versions of all the vowels and some consonants were obtained by varying the tuning of the circuits.

Stewart's synthesizer had two resonators, rather than three or four, because he shared the still generally held Helmholtzian notion that "the air in the mouth cavities possesses, as a rule, only one or two important modes of vibration" (1922, pp. 311). But Crandall (1925), another Bell engineer, found evidence against this view with the help of his electro-optical system for making oscillograms, a great advance over previous systems. He observed for back vowels not only the expected "mean low characteristic frequencies" but also, in the case of several speakers, "scattered low frequencies" that are obviously second-resonance frequencies and, for both back and front vowels, "scattered high frequencies" that may represent higher resonances. As recording and analyzing equipment improved, the number of resonances observed in vowel sounds continued to increase, further embarrassing the Helmholtzian account. In Steinberg's (1934) analysis of *Joe took father's shoebench out*,

three resonances were regularly seen and sometimes four; Lewis (1936) found five resonances in each of the vowels he analyzed, and therefore proposed that there were five simple resonators in the vocal tract, without specifically locating them. Moreover, G. OSCAR RUSSELL (1931) found the Helmholtzian account inconsistent with his x-ray data.

Homer Dudley was another Bell engineer intrigued by electrical analogies. In a 1940 paper, he says, "The fundamental processes in human speech production are ... analogous to those of electrical carrier circuits" (p. 504). He embodied this idea in the Vocoder (Dudley, 1939), a device to reduce the channel capacity required for speech communication. At the sending end, the smoothed outputs of a pitch-detector and a bank of filters covering the speech spectrum were coded as slowly-varying signals; at the receiving end, these coded signals controlled the frequency of a buzzer and the input levels of the buzzer signal to another bank of filters, whose summed outputs formed the synthesized speech. The Voder (Dudley, Riesz and Watkins, 1939) was a terminal analog synthesizer similar to the Vocoder's receiving end. The control signals, however, were provided by a human operator by means of a keyboard, wrist bar, and pedal. It took a long time for an operator to become skilled and the synthetic speech, to judge from recordings that still survive, was not highly intelligible.

#### 1945-1970

The impressive accomplishments of the post-World War II years were due mainly to three developments: The invention of the sound spectrograph, the replacement of the Helmholtzian account of vocal resonance by the Acoustic Theory of Speech Production, and the availability of digital computers for phonetic research.

The spectrograph (Potter, 1945; Koenig, Dunn and Lacy, 1946) had been developed by Ralph Potter and his associates at Bell Laboratories just before the War, but was used for military purposes and not made public until 1945. The spectrograph provided, in a few minutes, a time-frequency-intensity display of a previously recorded 2.4 sec signal, by means of repeated analyses with a band pass filter automatically retuned to increasing center frequencies. For the first time, spectrotemporal phonetic events could be readily observed.

In *Visible Speech*, Potter, Kopp and Green's (1947) textbook for reading spectrographic displays, a great many very clear spectrograms are shown and many of the significant acoustic features of speech are pointed out in colorful terminology. Stops are characterized by a "stop gap" followed by a "spike". Formants are "bars". Every sound has a "hub", the explicit or implicit frequency of "bar 2". The hub, a forerunner of the "locus" (DELATRE, LIBERMAN and COOPER, 1955), is said to be "relatively fixed" for many sounds, but Potter et al. were quite aware of coarticulatory effects (pp. 49-51). While little of their terminology has survived, most of their observations have proved in subsequent research to have been accurate and important.

Up to this point, linguists and conventional phoneticians had avoided acoustic phonetics. BLOCH and Trager had observed, "Acoustic terms, for all their precision, are meaningless to nearly every linguist" (1942, p. 12). But the advent of the spectrograph awakened linguists' interest. Martin JOOS, who had worked with the spectrograph for three years in the Army Signal Corps, was probably the first linguist to realize its potential value for phonetic research. His *Acoustic Phonetics* (1948) was extremely

influential and is still worth reading today for its discussions of segmentation and coarticulation. Another linguist, ROMAN JAKOBSON, found support in spectrograms for his theory of phonological distinctive features. In Jakobson, Fant and HALLE (1952), the features are defined in acoustic terms, with spectrographic examples. Unfortunately, the interests of phonologists and experimental phoneticians would soon diverge again, owing to differences in training and scientific goals.

The spectrograph also inspired FRANKLIN COOPER's Pattern Playback synthesizer at Haskins Laboratories (Cooper, LIBERMAN and Borst, 1951). The Playback could convert spectrotemporal patterns, either actual spectrograms or hand-painted creations, to sound. These control patterns were stable, unlike the human manipulations that had controlled earlier synthesizers, so that as many identical repetitions of an utterance as desired could be produced, yet the patterns were also readily modifiable. These features made the Playback the first synthesizer that enabled systematic perceptual experimentation, and it was the main tool for research at Haskins during the next fifteen years.

Many investigations aimed at describing the various phonetic categories were carried out during this period with the help of the spectrograph. DELATTRE (1951) elaborated on the correspondence, to which JOOS (1948) had called attention, between F1-F2 plots and the traditional phonetic vowel quadrilateral based on tongue position. This correspondence received further abundant support from an extensive study of the vowels of 76 speakers at Bell Laboratories (PETERSON and Barney, 1952). This study also revealed consistent formant patterns within speaker but systematic differences across speakers, dependent on sex and age.

Acoustic properties of stop consonants were investigated by Fischer-Jørgensen (1954) and HALLE, Hughes and Radley (1957); stress, by Fry (1955) and LEHISTE and PETERSON (1959); glides and diphthongs, by Lehiste and Peterson (1961); voice onset time, by LISKER and ABRAMSON (1964); and intonation, by Lieberman (1967). In the case of some categories, however, the analysis provided by the spectrograph was inadequate and special filtering was necessary, as in the case of stop bursts (Halle et al., 1957) and fricatives (Hughes and Halle, 1956). Besides these studies of particular categories, the coarticulatory effects were studied by House and FAIRBANKS (1953), Peterson and Lehiste (1960), and House (1961).

H. K. Dunn at Bell Laboratories and Gunnar Fant at MIT were responsible for the final demise of the Helmholtzian account of vocal resonance. Dunn (1950) argued that Helmholtz resonators were an oversimplification, given the frequencies and cavity dimensions involved. A distributed rather than a lumped treatment was called for, taking shape as well as size into consideration.<sup>1</sup> This could be accomplished by regarding the cavities of the vocal tract as a series of connected tubes of different lengths and areas or equivalently, in the most striking electrical analogy so far, as segments of a telephone transmission line. Though in certain cases a particular resonance might be chiefly dependent on a single cavity, the resonant frequencies specified by the transmission characteristic of the tract depended

<sup>1</sup> Chiba and Kajiyama in Japan had made this point in *The vowel - Its nature and structure* (1941/1958). But most copies of this book were lost during the war (Flanagan and Rabiner, 1973, p. 92), and it did not become known to Dunn until after he had submitted his paper for publication (Dunn, 1950, p.741).

in general on the entire configuration. Dunn ignored damping effects to simplify his discussion, but Fant (1950a,b; 1952) demonstrated that if damping is taken into account, a tube model of the transmission characteristic (or "transfer function"), together with appropriate assumptions concerning source and radiation characteristics, could specify the complete acoustic signal, rather than merely its resonant frequencies. This was the essence of the Acoustic Theory of Speech Production, elaborated later in Fant (1960).

To test his calculations, Dunn had constructed a primitive transmission-line synthesizer, in which the vocal tract was modelled as two uniform tubes connected by a constriction of variable location and area. Soon after, STEVENS, Kasowski and Fant (1953) built a more elaborate synthesizer in which the areas of 35.5 cm tube segments in cascade could be independently varied, allowing any desired vocal-tract area function to be approximated. By defining a fairly naturalistic configuration and varying only lip opening and location and degree of tongue constriction, Stevens and House (1955) demonstrated that the connection between tongue position and vowel formant frequency was not straightforward; because of compensatory articulation, the F1-F2 patterns PETERSON and Barney (1952) had reported could be produced by many different area functions. This synthesizer, supplemented with a nasal circuit, was also used to investigate nasalized vowels (House and Stevens, 1956) and nasal consonants (House, 1957).

Computers began to become available to speech researchers in the 1960s, taking over tasks previously done manually, such as the editing of speech signals (COOPER and MATTINGLY, 1969), or electronically, such as terminal analog synthesis (Kelly and Gerstman, 1961; FLANAGAN, Coker and Bird, 1963), vocal-tract synthesis (Kelly and Lochbaum, 1963), and modelling of the voicing source (Flanagan and Landgraf, 1968).

Other tasks performed by the computer had not previously been practical at all. One such was "synthesis by rule". LIBERMAN, Ingemann, LISKER, DELATTRE and COOPER (1959) had presented rules for synthesis from a phonemic input for the Pattern Playback, but hand-painting spectral control patterns by rule was extremely tedious, and only a few sentences had actually been synthesized. Kelly and Gerstman (1961), however, programmed a computer to derive control functions for their formant synthesizer according to rule. Thus ample amounts of speech could be readily produced and evaluated, and the rules could then be improved.

The computer also made possible a research strategy inspired by the Acoustic Theory of Speech Production: "Analysis by synthesis" (Bell, Fujisaki, Heinz, STEVENS and House, 1961). Fant (1950a,b; 1952) had pointed out that the transfer function could be factored into pole and zero components in the complex plane. This suggested that the transfer function of a natural speech sound could be estimated, given reasonable assumptions about the contributions of the excitation source and radiation, by assembling, with the aid of a computer, the best matching spectrum from an inventory of such components. Bell et al. analyzed vowels in this way and the method was applied to voiceless fricatives by Heinz and Stevens (1961) and to nasal consonants by FUJIMURA (1962).

## 1970-PRESENT

In recent times, although the number of investigators and the amount of research in acoustic phonetics have greatly increased, there has been no theoretical development as remarkable as the Acoustic Theory of Speech Production, and no technical advance as dramatic as the sound spectrograph. Nevertheless, substantial progress has been made in a number of different directions, which there is space to sketch only briefly, citing some representative papers.

Sophisticated computer procedures for speech analysis, such as the Fast Fourier Transform (Oppenheim, 1970), and Linear Predictive Coding (Atal and Hanauer, 1971) have been developed, and are now packaged for use on PCs and the Macintosh (e.g., Keller, 1994).

Some topics investigated earlier have been profitably revisited, as exemplified in the studies of stop consonants carried out by KLATT (1975), BLUMSTEIN and STEVENS (1979), and Kewley-Port (1982) and in the studies of the acoustic effects of coarticulation by, e.g., SOLI (1981), REPP and MANN (1982), Manuel (1990) and Magen (1997). Stevens (1989), in an influential paper drawing on the Acoustic Theory of Speech Production, has argued that speech is "quantal": Ranges of articulatory variables in which the signal changes rapidly relative to articulatory change, and ranges of acoustic variables where auditory perception changes rapidly relative to acoustic change, provide the basis for phonetic distinctive features.

Other topics, previously given little consideration, have now received more attention. Thus, substantial effort has been devoted to prosodic features, especially intonation, e.g., Kutik, Cooper and Boyce (1983) and Liberman and Pierrehumbert (1984); duration, e.g. Crystal and House (1988a, b); and the effects of syntactic structure on both (Cooper, 1976; Cooper and Sorensen, 1977). The pronunciation of non-native speakers has been studied by Flege, Munro, and Skelton (1992) and by Crowther and Mann (1992). The speech of infants and children has been considered by several investigators from a developmental standpoint. Thus, Gilbert (1970) has examined vowel formants; KEATING and Buhr (1978), F<sub>0</sub>; Smith (1978), timing; and Raphael, Dorman and Geffner (1980), durational differences conditioned by voicing.

In a laudable effort to bridge the divide that seems to separate phonetics from linguistics (see OHALA, 1990), some investigators have employed the methods of acoustic phonetics to address phonological questions. Labov (1963) had measured formant frequencies to demonstrate a sound change in progress on Martha's Vineyard. Later, Ohala (1974) showed that certain puzzling historical sound changes could be accounted for in acoustic terms. Most recently, the Experimental Phonology movement has demonstrated that experimental phonetics can contribute to both synchronic and diachronic phonology (see the papers in Ohala and Jaeger, 1986; Kingston and Beckman, 1990; and succeeding volumes in the Laboratory Phonology series).

Acoustic phonetics has also been applied to questions of linguistic evolution. Lieberman, Crelin and KLATT (1972) have argued that while the supralaryngeal tract in non-human primates, human neonates and, arguably, Neanderthal man is essentially a single tube with a limited sound repertoire, there has evolved in the adult human a "bent two-tube" supralaryngeal tract whose shape can be varied by tongue movement to produce a far more extensive repertoire. Hauser and FOWLER (1992) have provided

evidence that primate vocalizations have some of the prosodic properties of human speech.

There has been considerable work on the determination of vocal-tract configuration from spectral information. MERMELSTEIN (1967) had already shown that vocal-tract area functions for vowels cannot be recovered unambiguously unless they are required to be anti-symmetric. Atal, Chang, Mathews and Tukey (1978) have generated an inventory of formant-frequency patterns by vocal-tract synthesis and have confirmed the many-to-one articulatory-to-acoustic relation observed by STEVENS and House (1955). But LADEFOGED, Harshman, Goldstein and Rice (1978) have found that for unrounded vowels, formant frequencies are highly correlated with dynamic factors known to specify tongue shape, Papçun et al. (1992) have used a neural network trained on x-ray microbeam data for Cə syllables to determine vocal-tract gestures with considerable success, and Hogden, Rubin and Saltzman (1996) have shown that a computer can be trained to determine articulator positions from synthetic speech by exploiting continuity restrictions on articulator movement.

In speech synthesis work, highly intelligible speech can now be generated from a phonemic input, and indeed from conventionally-spelled text (see KLATT's 1987 review of text-to-speech systems) and attention has turned to the voice source. Ishizaka and FLANAGAN (1972) have synthesized voiced sounds from their "two-mass" model, elaborating on the earlier "one-mass" model (Flanagan and Landgraf, 1968), and Klatt and Klatt (1990) have demonstrated the importance of breathiness for the synthesis of natural-sounding female speech.

## CONCLUSION

Clearly, our understanding of acoustic structures and their relation to vocal-tract configurations has vastly increased since A. G. BELL's 1879 paper. On the other hand, the relation between acoustic structures and linguistic units, despite much research and discussion, remains clouded and controversial. It is especially in this direction that further progress is to be hoped for.

## ACKNOWLEDGMENT

Support from NIH grant DC02717 is gratefully acknowledged.

## REFERENCES

- ATAL, B. S., J. J. CHANG, M. V. MATHEWS, and J. W. TUKEY. 1978. Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique. *Journal of the Acoustical Society of America*, 63, 1535-1555.
- ATAL, B. S. and S. L. HANAUER. 1971. Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, 50, 637-655.
- BELL, A. G. 1879. Vowel theories. *American Journal of Otology*, 1, 163-180. Reprinted in A.G. Bell, *The Mechanism of Speech*, 8th ed., 1916. New York: Funk and Wagnalls. 117-129.
- BELL, C. G., H. FUJISAKI, J. M. HEINZ, K. N. STEVENS, and A. S. HOUSE. 1961. Reduction of speech spectra by analysis-by-synthesis techniques. *Journal of the Acoustical Society of America*, 33, 1725-1736.
- BEVIER, L. 1900. The acoustic analysis of the vowels from the phonographic record. *Physical Review*, 10, 193-201.
- BLOCH, B. and G. L. TRAGER. 1942. *Outline of linguistic analysis*. Baltimore: Linguistic Society of America.
- BLUMSTEIN, S. E. and K. N. STEVENS. 1979. Acoustic invariance in

- speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66, 1001-1117.
- CHIBA, T. and M. KAJIYAMA. 1958. *The vowel - Its nature and structure*. Tokyo: Phonetic Society of Japan. Originally published 1941.
- COOPER, F. S., A. M. LIBERMAN, and J. M. BORST. 1951. The intraconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Science*, 37, 318-325.
- COOPER, F. S. and I. G. MATTINGLY. 1969. Computer-controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 46, 115.
- COOPER, W. E. 1976. Syntactic control of timing in speech production: a study of complement clauses. *Journal of Phonetics*, 4, 151-171.
- COOPER, W. E. and J. M. SORENSEN. 1977. Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America*, 62, 683-692.
- CRANDALL, I. B. 1925. The sounds of speech. *Bell System Technical Journal*, 4, 586-626.
- CROWTHER, C. S. and V. MANN. 1992. Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America*, 92, 711-722.
- CRYSTAL, T. H. and A. S. HOUSE. 1988a. Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, 83, 1553-1573.
- CRYSTAL, T. H. and A. S. HOUSE. 1988b. Segmental durations in connected-speech signals: syllabic stress. *Journal of the Acoustical Society of America*, 83, 1574-1585.
- DELATTRE, P. C. 1951. The physiological interpretation of sound spectrograms. *PMLA*, 66, 864-875.
- DELATTRE, P. C., A. M. LIBERMAN, and F. S. COOPER. 1955. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 27, 769-773.
- DUDLEY, H. 1939. The Vocoder. *Bell Labs Record*, 18, 122-126.
- DUDLEY, H. 1940. The carrier nature of speech. *Bell System Technical Journal*, 19, 495-513.
- DUDLEY, H., R. R. RIESZ, and S. S. A. WATKINS. 1939. A synthetic speaker. *Journal of the Franklin Institute*, 227, 739-764.
- DUNN, H. K. 1950. The calculation of vowel resonances, and an electrical vocal tract. *Journal of the Acoustical Society of America*, 22, 740-753.
- FANT, G. 1950a. Transmission properties of the vocal tract. *Quarterly Progress Report*, July-September, 1950. Acoustics Laboratory, Massachusetts Institute of Technology. 20-23.
- FANT, G. 1950b. Transmission properties of the vocal tract, Part II. *Quarterly Progress Report*, October-December, 1950. Acoustics Laboratory, Massachusetts Institute of Technology. 14-19.
- FANT, G. 1952. *Transmission properties of the vocal tract with application to the acoustic specification of phonemes*. Technical Report 12. Cambridge, MA: Acoustics Laboratory, Massachusetts Institute of Technology.
- FANT, G. 1960. *Acoustic theory of speech production*. The Hague: Mouton.
- FISCHER-JØRGENSEN, E. 1954. Acoustic analysis of stop consonants. *Miscellanea Phonetica*, 2, 42-59.
- FLANAGAN, J. L. 1957. Note on the design of "terminal-analog" speech synthesizers. *Journal of the Acoustical Society of America*, 29, 306-310.
- FLANAGAN, J. L., C. H. COKER, and C. M. BIRD. 1963. Digital computer simulation of a formant-vocoder speech synthesizer. *Fifteenth Annual Meeting of the Audio Engineering Society*. Preprint 307.
- FLANAGAN, J. L. and L. L. LANDGRAF. 1968. Self-oscillating source for vocal tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, AU-16, 57-64.
- FLANAGAN, J. L. and L. R. RABINER (Eds.). 1973. *Speech synthesis*. Stroudsburg, PA: Dowden, Hutchinson and Ross.
- FLEGE, J. E., M. J. MUNRO, and L. SKELTON. 1992. Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America*, 92, 128-143.
- FLETCHER, H. 1929. *Speech and hearing*. New York: Van Nostrand.
- FRY, D. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765-768.
- FUJIMURA, O. 1962. Analysis of nasal consonants. *Journal of the Acoustical Society of America*, 34, 1865-1875.
- GILBERT, J. H. 1970. Formant concentration positions in the speech of children at two levels of linguistic development. *Journal of the Acoustical Society of America*, 48, 1404-1406.
- HALLE, M., G. W. HUGHES, and J.-P. RADLEY. 1957. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, 29, 107-116.
- HAUSER, M. D. and C. A. FOWLER. 1992. Fundamental frequency declination is not unique to human speech. *Journal of the Acoustical Society of America*, 91, 363-369.
- HEINZ, J. M. and K. N. STEVENS. 1961. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 33, 589-596.
- HELMHOLTZ, H. 1954. *On the sensations of tone*. A.J. Ellis, Trans., 1885, New York: Dover. Original work, 4th ed., published 1877.
- HERMANN, L. 1893. Phonophotographische untersuchungen, IV. [Phonophotographic investigations, IV]. *Pflügers Archiv für des gesamte Physiologie des Menschen und der Tiere*, 53, 1-51.
- HODGEN, J., P. RUBIN, and E. SALTZMAN. 1996. An unsupervised method for learning to track tongue position from an acoustic signal. *Bulletin de la communication parlée*, no. 3, 101-116.
- HOUSE, A. S. 1957. Analog studies of nasal consonants. *Journal of Speech and Hearing Disorders*, 22, 190-204.
- HOUSE, A. S. 1961. On vowel duration in English. *Journal of the Acoustical Society of America*, 33, 1174-1178.
- HOUSE, A. S. and G. FAIRBANKS. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25, 105-113.
- HOUSE, A. S. and K. N. STEVENS. 1956. Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders*, 21, 218-232.
- HUGHES, G. W. and M. HALLE. 1956. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, 28, 303-310.
- ISHIZAKA, K. and J. L. FLANAGAN. 1972. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal* 51, 1233-1267.
- JAKOBSON, R., G. FANT, and M. HALLE. 1952. *Preliminaries to speech analysis*. Technical Report 13. Cambridge, MA: Acoustics Laboratory, Massachusetts Institute of Technology.
- JENKIN, F. and J. A. EWING. 1878. Helmholtz's vowel theory and the phonograph. *Nature*, 17, 384.
- JOOS, M. 1948. Acoustic Phonetics. *Language Monograph No. 23*. *Language*, 24, 2, Suppl. Baltimore: Linguistic Society of America.
- KEATING, P. and R. BUHR. 1978. Fundamental frequency in the speech of infants and children. *Journal of the Acoustical Society of America*, 63, 567-571.
- KELLER, E. 1994. *Signalize™/Signal analysis for speech and sound*. Version 3.0. [Computer software] Lausanne: InfoSignal.
- KELLY, J. L., JR. and L. J. GERSTMAN. 1961. Artificial talker driven from a phonetic input. *Journal of the Acoustical Society of America*, 33, 835.
- KELLY, J. L., JR. and C. LOCHBAUM. 1963. Speech synthesis. *Journal of the Acoustical Society of America*, 35, 1115.
- KEWLEY-PORT, D. 1982. Measurement of formant transitions in naturally-produced stop-consonant-vowel syllables. *Journal of the Acoustical Society of America*, 72, 379-389.
- KINGSTON, J. and M. BECKMAN. (Eds.). 1990. *Papers in Laboratory Phonology I*. Cambridge, England: Cambridge University Press.
- KLATT, D. H. 1975. Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*

- 18, 686-706.
- KLATT, D. H. 1987. Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, 82, 737-793.
- KLATT, D. H. and L. C. KLATT. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.
- KOENIG, W., H. K. DUNN, and L. Y. LACY. 1946. The sound spectrograph. *Journal of the Acoustical Society of America*, 18, 19-49.
- KUTIK, E., W. E. COOPER, and S. BOYCE. 1983. Declination of fundamental frequency in speakers' production of parenthetical and main clauses. *Journal of the Acoustical Society of America*, 73, 1731-1738.
- LABOV, W. 1963. The social motivation of a sound change. *Word*, 19, 273-309.
- LADEFOGED, P., R. HARSHMAN, L. GOLDSTEIN, and L. RICE. 1978. Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America*, 64, 1027-1035.
- LEHISTE, I. and G. E. PETERSON. 1959. Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, 31, 428-435.
- LEHISTE, I. and G. E. PETERSON. 1961. Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America*, 33, 268-277.
- LEWIS, D. 1936. Vocal resonance. *Journal of the Acoustical Society of America*, 8, 91-99.
- LIBERMAN, A. M., F. INGEMANN, L. LISKER, P. DELATTRE, and F. S. COOPER. 1959. Minimal rules for synthesizing speech. *Journal of the Acoustical Society of America*, 31, 1490-1499.
- LIBERMAN, M. and J. PIERREHUMBERT. 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle (Eds.). *Language Sound Structure*. Cambridge MA: MIT Press. 157-233.
- LIBERMAN, P. 1967. *Intonation, perception and language*. Cambridge MA: MIT Press.
- LIBERMAN, P., E. S. CRELIN, and D. H. KLATT. 1972. Phonetic ability and related anatomy of the newborn, adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, 74, 287-307.
- LISKER, L. and A. S. ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.
- MAGEN, H. S. 1997. The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 25, 187-205.
- MANUEL, S. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, 88, 1286-1298.
- MERMELSTEIN, P. 1967. Determination of the vocal-tract shape from measured formant frequencies. *Journal of the Acoustical Society of America*, 41, 1283-1294.
- MILLER, D. C. 1922. *Science of musical sounds*. 2nd. ed. New York: Macmillan. Originally published 1916.
- OHALA, J. J. 1974. Experimental historical phonology. In J.M. Anderson and C. Jones (Eds.), *Proceedings of the First International Conference on Historical Linguistics, Edinburgh*. Amsterdam: North-Holland. 353-389.
- OHALA, J. J. 1990. There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics*, 18, 153-171.
- OHALA, J. J. and J. JAEGER. (Eds.). 1986. *Experimental phonology*. New York: Academic Press.
- OPPENHEIM, A. V. 1970. Speech spectrograms using the Fast Fourier Transform. *IEEE Spectrum*, 7, 57-62.
- PAPCUN, G., J. HOCHBERG, T. R. THOMAS, F. LAROCHE, J. ZACHS, and S. LEVY. 1992. Inferring articulation and recognizing gestures from acoustics with a neural network trained on x-ray microbeam data. *Journal of the Acoustical Society of America*, 92, 688-700.
- PETERSON, G. E. and H. L. BARNEY. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
- PETERSON, G. E. and I. LEHISTE. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693-703.
- POTTER, R. K. 1945. Visible patterns of sound. *Science*, 102, 463-470.
- POTTER, R. K., G. A. KOPP, and H. C. GREEN. 1947. *Visible speech*. New York: Van Nostrand.
- RAPHAEL, L. J., M. F. DORMAN, and D. GEFFNER. 1980. Voicing-conditioned durational differences in vowels and consonants in the speech of three- and four-year old children. *Journal of Phonetics*, 8, 335-341.
- RAYLEIGH, J. W. and B. STRUTT. 1945. *The Theory of Sound*. New York: Dover. Original work, 2nd ed., published 1894.
- REPP, B. H. and V. A. MANN. 1982. Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 71, 1562-1567.
- RUSSELL, G. O. 1931. *Speech and voice*. New York: Macmillan.
- SCRIPTURE, E. W. 1906. *Researches in experimental phonetics. The study of speech curves*. Washington: Carnegie Institute of Washington.
- SMITH, B. L. 1978. Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, 6, 37-67.
- SOLI, S. D. 1981. Second formants in fricatives: Acoustic consequences of fricative vowel coarticulation. *Journal of the Acoustical Society of America*, 70, 976-984.
- STEINBERG, J. C. 1934. Application of sound measuring instruments to the study of phonetic problems. *Journal of the Acoustical Society of America*, 6, 16-24.
- STEVENS, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.
- STEVENS, K. N. and A. S. HOUSE. 1955. Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, 27, 484-493.
- STEVENS, K. N., S. KASOWSKI, and G. FANT. 1953. Electrical analog of the vocal tract. *Journal of the Acoustical Society of America*, 25, 734-742.
- STEWART, J. Q. 1922. An electrical analog of the vocal organs. *Nature*, 110, 311-312.
- WATSON, T. A. 1915. How Bell invented the telephone. *Transactions of the American Institute of Electrical Engineers*, 34, 1503-1513.
- WHEATSTONE, C. 1837. Reed organ pipes, speaking machines, etc. *Westminster Review*, 27, 30-37. Reprinted in *Scientific Papers of Sir Charles Wheatstone 1879*. London: Taylor and Francis. 48-367.
- WILLIS, W. 1830. On vowel sounds, and on reed-organ pipes. *Transactions of the Cambridge Philosophical Society*, 3, 231-268.