A Dynamical Study of the Vowel Sounds

By I. B. CRANDALL and C. F. SACIA

Introduction

THE study of the vowel sounds presents a problem which has interested scientists and scholars in varied fields. A knowledge of their nature is of fundamental importance not only in communication engineering but also in acoustic science, phonetics and vocal music. From the earliest theories and the rough experiments of Willis (1829) and Helmholtz (1859) to the later measurements of D. C. Miller (1916) steady progress has been made toward the accurate determination of their characteristics.

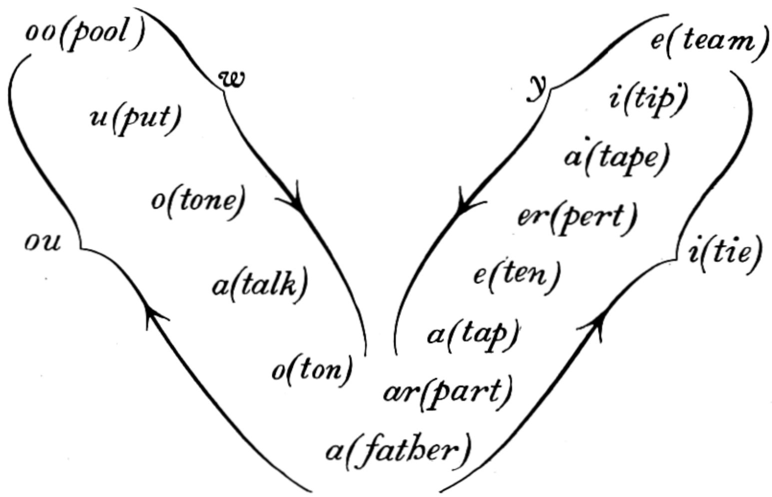
Further progress in this study has been made possible with improved facilities now available in the telephone research laboratory. It has been felt that there was need for more accurate records of the spoken sounds and the development of improved transmitters, amplifiers and other devices has made possible recording apparatus of greater accuracy, range and power than any heretofore used.

In this paper will be given the results of an analysis of spoken vowel sounds based on a set of accurate oscillographic records. The recording apparatus was designed to record the wave forms of the different speech sounds practically free from distortion over the frequency range from 100 to 5000 cycles. A brief description of this apparatus is given in the appendix. The emphasis in the present paper is placed on the composite frequency characteristics of the sounds as revealed by a particular method of analyzing the records so obtained.

Analysis of the Data

The thirteen vowel sounds investigated are shown arranged in a triangle in Fig. 1. The diphthongs *ou*, *w*, *y* and long *i* are not included. Eight records of each sound were taken, four by male and four by female speakers. In speaking these sounds the only constraint imposed on the speakers was that the sound should be completely uttered within an interval of one second. The recording mechanism was so arranged that the whole of the sound from beginning to end was recorded in one continuous graph. In practice the average duration of these sounds was about 0.30 second. Each record shows a sequence of growth and decay in amplitude somewhat as follows: first a period of rapid growth in amplitude lasting about .04 second during which all components are quickly produced and rise nearly to maximum amplitude; second a middle period in which the general amplitude is nearly constant but with varying phase relations between the different components and lasting about 0.17 second; and finally a period of gradual decay lasting about .09 second in which all the components disappear. A typical record so obtained is shown in Fig. 2.

A brief description of the method of mechanically analyzing such a record is given in the appendix. The essential point of the analysis is that the whole record from start to finish is taken as the unit for analysis and the data obtained are therefore the average characteristics of the sounds throughout their duration.



It is usual to exhibit the properties of a vowel sound in a spectrum diagram showing the amplitude of the component vibrations as a function of their pitches or frequencies. For each vowel sound there are, in addition to fundamental tones, certain characteristic regions of resonance which may be at high or low frequencies. It would be possible from the results of this analysis to present the sound spectra of each vowel showing the relative amplitudes for the different frequencies as present in the original air vibration[1] but this treatment has been modified to take into account the relative importance of the various pitches in hearing. Using the data available

[1] In previous publications (*Phys. Rev.* XIX, 1922, p. 228, Fig. 7, and *Bell System Technical Journal*, Vol. 1, No. 1, p. 124,) data have been given showing the actual distribution of energy in average speech. The tremendous concentration of energy in the lower frequencies is somewhat misleading unless account is also taken of the much reduced sensitivity of the ear in this region.

on the relative sensitivity of the ear at different frequencies[2] we have multiplied the acoustic amplitude at each frequency by the corresponding ear sensitivity factor and the results obtained are taken to be the effective amplitude frequency relations which are characteristic of these sounds.

The data from the four male records and from the four female records of each sound are separately composited and the resulting curves are shown in the diagram (Fig. 3). This compositing process was somewhat laborious because the analyses of the separate records were made not with reference to predetermined frequency settings, but rather for those critical frequencies which best determined the shapes of the spectrum curves. The individual curves were therefore plotted, and the average ordinates were then read off for small intervals of pitch. These ordinates were then averaged for each group of four analyses. These average ordinates (after being corrected for the calibration of the recording apparatus) were then multiplied by the ear sensitivity factors for the corresponding frequencies, and the curves so obtained were plotted on the musical pitch scale according to the usual practice. The final spectrum diagram thus shows the relative importance of the amplitudes of all the components of each vowel for male and female speakers.

The amplitude units are entirely arbitrary; it is only the shapes, not the sizes of these curves which have any significance. The order in which these curves are arranged is based upon the vowel triangle in Fig. 1.

Characteristics of the Vowel Sounds

The results of the analyses, as given in Fig. 3 show the essential dynamical properties of these sounds. Consider first the sounds numbered I to VI, which include those vowels usually designated as having single regions of resonance. Progressing through the sequence from I to VI this region of resonance rises in average frequency and becomes narrower in range. The rise in average frequency is of course a well known characteristic. There is also, at least with the male voices, a somewhat scattered and less well defined high frequency range of resonance, perhaps not essential in speech but more highly developed in well-trained singing voices.

The sound *a* (No. VI) is as it were the center of gravity of the vowel diagram and occupies the key position in the phonetics of

[2] See this Journal Vol. II, No. 4, October, 1923. The paper on audition, by H. Fletcher shows a cut of the "Threshold of Audibility" curve from which these data were obtained.

most languages. Now consider the sequence from this sound to No. XIII at the end of the diagram; these sounds include most of those which are known to have two characteristic regions of resonance. The main region of resonance now divides into two parts which gradually recede from each other as we follow the diagram downwards. (Sound X (*er*) is difficult to fit into the diagram in an exact position, but it is evident that it belongs in the series of doubly-resonant vowels.)

Contour lines (nearly vertical) have been drawn on the diagram to indicate the progressive changes in regions of resonance. Viewing the diagram as a whole it is important to consider not only the location of the resonant ranges but also their extent, and their relative separation from other resonant ranges in order to arrive at the essential characteristics of the vowel sound. In other words the individual vowel characteristic depends not only on the absolute pitch but on the relative pitches in case there is more than one region of resonance. It is only in this way that we can explain what is a matter of universal experience in using the phonograph; namely that moderate variations from normal speed in recording and reproducing speech leave the vowel sounds still intelligible.

It is expected to deal in a later publication with the semi-vowel sounds *l*, *ng*, *n*, *m* which seem to be related to the general diagram of the vowel sounds, and on which a preliminary report has already been made[3].

The more interesting features of the original records as such will also be dealt with in a subsequent publication.

APPENDIX

Recording and Analysis of Vowel Sounds

Recording Apparatus

The apparatus used in recording consisted of a condenser transmitter, an amplifier, and an oscillograph, in which important modifications were made. The vibrator was given great stiffness and damping so that the frequency response of the vibrator was nearly uniform up to 5000 cycles. Instead of the usual 12 inch film, special film 51 inches in length was used. This necessitated a much larger film drum. Furthermore the desired length of the record was about four times the circumference of the film drum, so the shutter was arranged to stay open during four revolutions while the vibrator was

[3] *Phys. Rev.* 23, 1924, p. 309—"Preliminary Analysis of Four Semi-Vowel Sounds."

given a slow uniform rotation about its vertical axis. With the film on the drum, the record thus had a helical form. In this way records of the requisite length were obtained.

The condenser transmitter was of the type developed by E. C. Wente, its characteristics combining with those of the amplifier and oscillograph vibrator in such a way that the combined amplitude response for the whole system was fairly uniform up to 5000 cycles, while the phase lag was approximately a linear function of frequency over the same range. This apparatus was therefore well adapted to the production of faithful records of the vowel sounds. The photographic equipment permitted the use of a time scale as great as six meters per second on the record (i.e. 2 inches = 0.01 sec.)

Transformation of Records for Analysis[4]

The oscillograms taken with the above apparatus were line records; in order to analyze these wave forms by the photo-mechanical method outlined below, it was necessary to transform the line record into a black profile. This was accomplished in the following steps:

(1) A positive print of the wave form on the original record was made on motion picture film.

(2) The emulsion of the positive print was then cut through to the base along the line of the wave by means of a stylus.

(3) The entire strip was blackened (on the emulsion side) with printer's ink.

(4) The emulsion on one side of the wave was stripped from the base, thus leaving the profile.

(5) The beginning and end were joined to form an endless belt.

Photo-Mechanical Analysis of the Prepared Records[4]

The principle of the photo-mechanical analysis is as follows: The motion of the strip past the image of an illuminated slit causes fluctuations in a beam of transmitted light which in turn, produce voltage fluctuations in the circuit containing a selenium or photo-electric cell. This voltage is then analyzed by means of a tuned circuit, an amplifier and a rectifier. The frequency of any component selected in this manner is determined by the tuning frequency divided by the ratio of speed transformation (analysis speed divided by the original speed of recording). The measured amplitude of the selected

[4] *Phys. Rev.* 23, 1924, p. 309. It is planned to publish a more detailed description of this apparatus later.

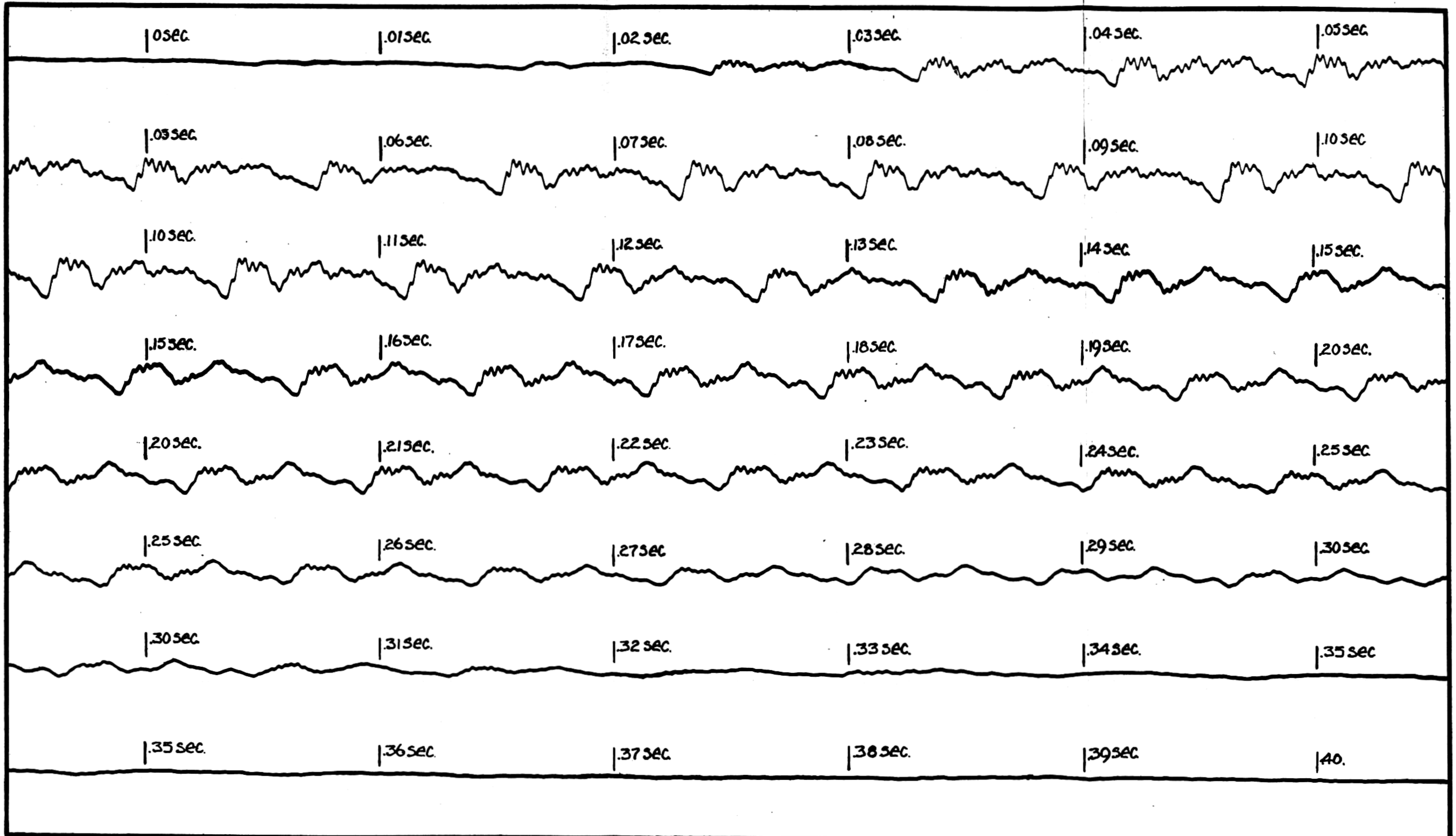
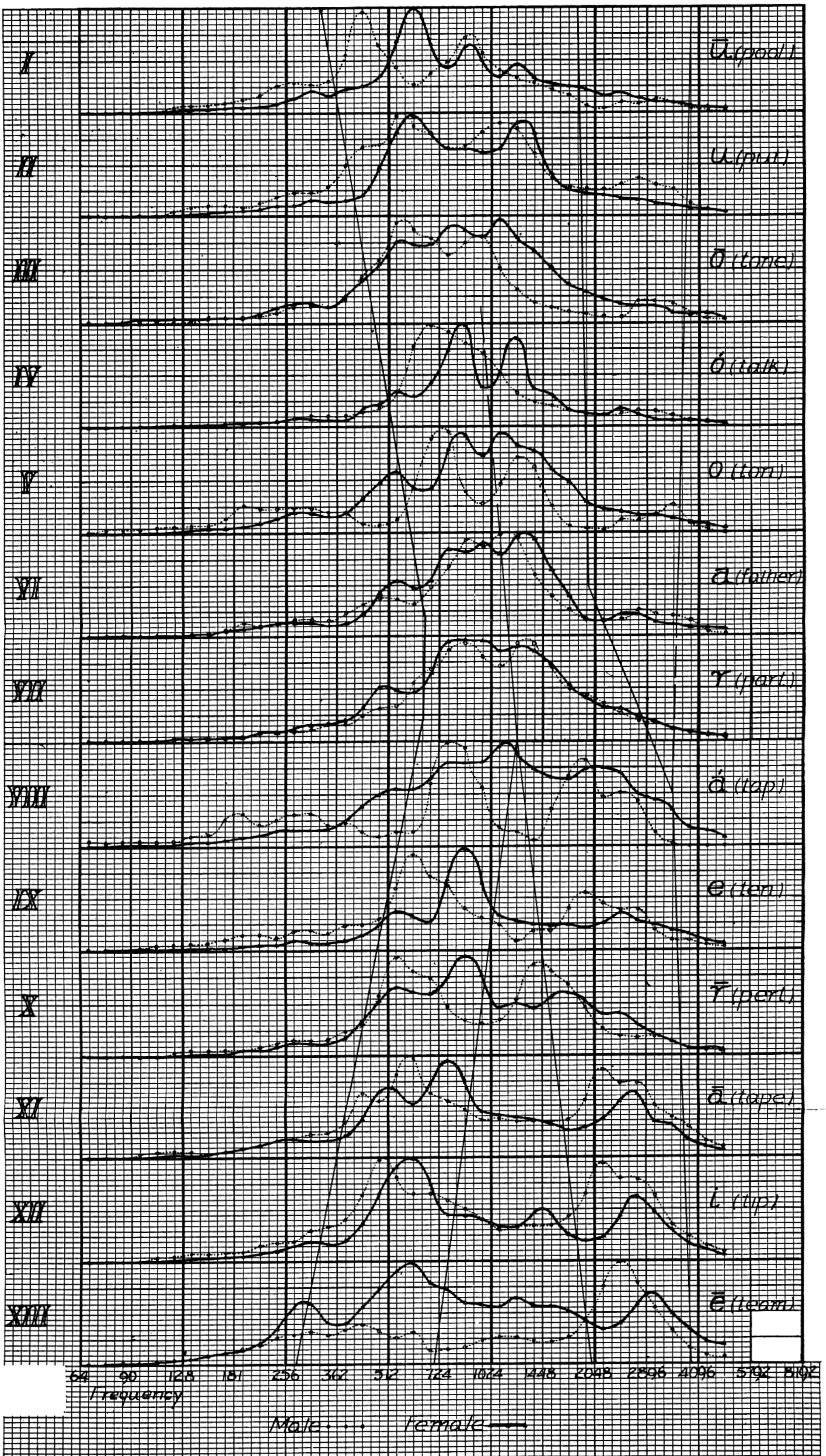


Fig. 2—*oo* as in p*oo*l; spoken by M. B.—male, low pitched. Plate No. 2: made from film record No. 157-A.



I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
ū (pool)
u (put)
ō (tone)
ó (talk)
o (ton)
a (father)
r (part)
á (tap)
e (ten)
r̄ (pert)
ā (tape)
i (tip)
ē (team)
64
90
128
181
256
362
512
724
1024
1448
2048
2896
4096
5792
8192
Frequency
Male
Female

Fig. 3

component is determined by the rectifier output, the sensitivity factor of the selenium cell and the area of the frequency response curve of the tuning apparatus.

Since the wave form of a vowel sound is not a true periodic function, it is represented analytically by a Fourier Integral, not by a Fourier Series. The continued repetition of the motion of the wave past the slit, however, builds up a periodic function consisting of a fundamental and a series of harmonics. The magnitudes of these components bear a simple relation to those of the infinitesimal components of corresponding frequencies in the Fourier Integral. It is this series of harmonics which is measured by the above method, hence the problem of analyzing the aperiodic function represented in the record is solved by means of the related periodic function.