THE TRUE NATURE OF SPEECH

With Application to a Voice-Operated Phonographic Alphabet Writing Machine

BY JOHN B. FLOWERS

ABSTRACT OF PAPER

That speech is a rapid variation in intensity of the voice and mouth-tones according to definite sound patterns called letters of the alphabet, is proved, by showing that speech is the result of action of the mouth-parts in varying the intensity of the voice and mouth-tones, and through photographs taken with the string-galvanometer of each letter sound of the alphabet, showing definitely the characteristic variation in intensity of tone for each letter of the alphabet. From the curves, the phonographic alphabet is obtained by measuring the variations in intensity of the main tone of the record.

A design for a voice-operated phonographic alphabet writing machine is described. The object of this device is to record speech automatically in ink on paper in the form of an easily read compact system of natural characters called the phonographic alphabet. Its design comprises a high-power telephone transmitter controlling electric resonator circuits, the intensity of currents in which is measured by the vibration of mirrors reflecting light upon a selenium cell connected to a special recording pen.

DEFINITION OF SPEECH

SPEECH is a rapid variation in amplitude of one or more tones, there being a definite form of variation making a pattern for each letter sound of the alphabet. These patterns are not formed by the superposing of tones of different pitch to give a certain quality to the tone but by a definite form of variation in the amplitude of a single tone or tones. This previously unknown form of variation of amplitude, quite different from the ordinary change from soft to loud, should be called the speech-variation. What is meant is that if a tone of a pitch of say 1000 cycles per second commences in such a manner that the first 10 to 20 cycles vary in intensity after the definite pattern , the sound *b* will be produced and heard.

Manuscript of this paper was received April 26, 1915.

Thus:

20 CYCLES OF A 1000-CYCLE TONE

But no matter what the pitch of the tone, if its intensity is made to vary according to this definite pattern in a *small fraction of a second*, the sound *b* will be produced.

Thus:

5 CYCLES OF A 250-CYCLE TONE

When we speak the letter *b*, the fundamental tone and the overtones of the vocal chords together with the mouth-tones are all varied at the same instant of time according to the definite pattern . The same is true for each letter sound of the alphabet, as will appear in the table of patterns of the letters of the alphabet called the phonographic alphabet on another page:

In other words, if an open reed-organ pipe of, say, middle C pitch were blown and at its large open end were placed a mechanical device like the human mouth for opening, varying the opening, and closing the opening, and trilling and varying the air pressure, the organ pipe would speak like a human being, with wonderful power and majesty.

We quote Helmholtz: "Sensations of Tone," pages 66-67-68. "There has been a general inclination to credit quality with all possible peculiarities of musical tones that were not evidently due to force and pitch. But very slight consideration will suffice to show that many of these peculiarities of musical tones depend upon the way in which they begin and end. The methods of attacking and releasing tones are sometimes so characteristic that for the human voice they have been noted by a series of different letters. To these belong the explosive consonants *B*, *D*, *G*, and *P*, *T*, *K*. The effects of these letters are produced by opening the closed, or closing the open passage through the mouth. For *B* and *P* the closure is made by the lips, for *D* and *T* by the tongue and upper teeth, and *G* and *K*

by the back of the tongue and soft palate. In wind instruments where the tones are maintained by a stream of air, we generally hear more or less whizzing and hissing of the air which breaks against the sharp edges of the mouthpiece. It is well known that most consonants in human speech are characterized by the maintenance of similar noises, as F, V; S, Z; Th in thin and in then; the Scotch and German guttural CH, and Dutch G. For some the tone is made still more irregular by trilling parts of the mouth, as for R and L. In the case of R the stream of air is periodically entirely interrupted by trilling the uvula, or the tip of the tongue; and we thus obtain an intermitting sound to which these interruptions give a peculiar jarring character. In the case of L the soft side edges of the tongue are moved by the stream of air, and, without completely interrupting the tone, produce inequalities in its strength. The formation of M and N in so far resembles that of vowels, that no noise of wind is generated in any part of the cavity of the mouth, which is perfectly closed, and the sound of the voice escapes through the nose."

PROOF THAT A DEFINITE FIXED PITCH IS NOT THE CHIEF
CHARACTERISTIC OF INDIVIDUAL LETTER SOUNDS

Anyone may easily prove this for himself. Place upon the phonograph a talking record. Run the record, first at 60 revolutions per minute, which is below the normal speed, then at 160 revolutions per minute, which is $2\frac{2}{3}$ times the lower speed. The articulation remains intelligible even at the low and high speeds. The pitches of all speech sounds at the high speed of the record are nearly an octave and a half above the pitches at the low speed. Hence if but one fixed pitch was characteristic of each letter sound, the speech would have been absolutely unintelligible. But the speech was entirely intelligible and therefore pitch is not the distinguishing element between letter sounds.

DEFINITION OF VOWELS AND CONSONANTS

A vowel is a definite sound pattern which repeats itself at intervals, while a consonant is a definite sound pattern occurring but once. The consonant sound pattern must exist at least for 0.01 second, in order to be perceived by the brain, 0.01 second being the minimum perception time for speech sounds. The vowel sound patterns proper, are about 0.01 second in duration, but as they repeat themselves again and

again, may be of any duration the speaker desires. Under this definition, the consonants are b, c = k, d, g, j, k, p, t, and the vowels are, a, c = s, e, f, h, i, l, m, n, o, r, s, u, v, w = ūō, y, z.

Written language is made up of consecutive pattern pictures, one pattern or form standing for each letter of the alphabet, and the eye is able to distinguish between them if they approximate the standard form. Likewise, spoken language is made up of consecutive pattern pictures, one pattern or form standing for each letter of the alphabet, and I show in another paper not yet published, how the ear is able to distinguish between them when they approximate a standard form.

In reading a submarine cable message on the tape, the eye must pick out the pattern forms of each letter. That the eye can read these messages as fast as ordinary writing is known to all cable operators. Spoken language is similarly a succession of these different pattern forms, strung along one after the other, and the ear distinguishes each of these pattern forms as it arrives in the consecutive order of utterance, even as the eye distinguishes pattern forms of written or submarine cable language.

CABLE MESSAGE RECEIVED BY THE W. U. TELEGRAPH CO., FEBRUARY 11, 1915, OVER AN ATLANTIC CABLE

Speech may be whispered using no voice (without movement of the vocal chords). This fact was determined by actual observation of quiet vocal chords during a whispered word, "ah."

DISTINCTION BETWEEN SPEECH AND VOICE

Note the distinction between speech and voice and where each is produced, viz., voice by the vocal chords and speech through variation in the intensity of the compressed air forced through the throat, mouth, and nasal cavity. This variation in intensity is accomplished by the action of the muscles of the throat and mouth. Each person has a slightly different pitch of voice, men low and women high. An average man's voice has a pitch lying between 85 and 160 complete vibrations per

second, while the average woman's voice has a pitch lying between 150 and 320 complete vibrations per second. Thus, the pitch of one person's voice is different from another's.

But speech may be produced by whispering, that is, using no voice, and whispered speech always has the same sound, no matter who is the speaker, man or woman. Any person can determine this by his own observation, getting someone to whisper with him and comparing the sound of the whispers. No difference in the sound of the whispers can be heard. In other words, during a whisper, speech is independent of the vocal chord action.

All previous speech records have been taken of spoken or sung words and none of whispered words on account of the lack of instruments sufficiently sensitive to record whispered speech. The resulting spoken records contained the fundamental tones and overtones of the vocal chords as well as the mouth-tones and their variations combined in so complex a curve that no one has as yet succeeded in deciphering them.

SECTION OF ACOUSTICON TRANSMITTER FOR WHISPERED SPEECH

I found that the acousticon transmitter hooked up to the Einthoven string-galvanometer, as shown in the sketch herewith, were very satisfactory instruments for making records of whispered speech, as they are both marvelously sensitive instruments. On account of whispered speech being independent of vocal chord action, it was decided to make photographic records of it from which to determine by careful study the true nature of speech. It is evident that the whispered speech records will not show the voice-tones and overtones but will simply show the mouth-tones and their variations in intensity. It is easy to follow the variation in intensity of one or two tones on the whispered speech records and thus pick out the intensity pattern for each letter sound which I call the phonographic alphabet.

The diaphragm is made of very thin carbon and so covers the recesses in the carbon block, without touching it, as to prevent the tiny carbon balls from falling out. The sketch shows how the acousticon transmitter collects sound from a large area and by a reflecting system concentrates it on the middle of the carbon diaphragm. There is no spring pressing on the diaphragm and it is believed that the diaphragm moves like a piston with the rarefactions and compressions of the sound waves. It is a highly damped instrument due to the weight of the carbon balls pressing against the diaphragm resisting its motion.

DESCRIPTION OF METHOD OF MAKING WHISPERED SPEECH-PATTERN PICTURES

The acousticon transmitter was placed in a large sound-proof telephone booth and supported by suspension free of all vibration. The diagram shows how the speech current flows from the acousticon through the storage battery cell, through the resistance, through the telephone receiver, through the primary side of the induction coil, and back to the acousticon; on the secondary side of the induction coil, the speech current flows through the string of the galvanometer. The string galvanometer is an electrical instrument having a fine silver-plated quartz fiber, 0.0001 in. thick, supported between two poles of a huge electromagnet. Through a hole in the magnet poles, a powerful ray from an arc lamp is focused upon this quartz fiber or string. By means of another system of lenses, the image of the string is focused upon a slot in the rotating drum-camera. When a speech-current flows through this galvanometer string, the string vibrates back and forth in a direction at right angles to the lines of force of the magnet, and its lightest motion is magnified 900 times by the lens system. The shadow of the string vibrates back and forth on the camera-shutter almost exactly in proportion to the speech sounds. The camera photographs the vibrating shadow of the galvanometer string. A word is whispered repeatedly, and the speech is listened to in the telephone receiver as a check on the articulation of the words. When the shadow of the galvanometer string moves back and forth about two inches, a lever is pressed, the electrically operated camera-shutter is held open for one revolution of the wheel and a photographic record, five feet long, of the word whispered is obtained. Extra-rapid film was exposed and good photo-

graphs obtained at rates varying from 1000 feet to one mile per minute. The rate of 1080 ft. was used mostly in this work as it gave a record with good intervals between crests of speech-waves, the best record to analyze. By interrupting the light from the arc lamp, 500 times per second, vertical lines were photographed upon the film at intervals of 0.002 second by a specially constructed time-wheel, for the necessary time record.

For a consideration of the theory of the string galvanometer, I refer to Dr. A. C. Crehore on the "Theory of the String Galvanometer," *Phil. Mag.*, vol. 28, August, 1914. The amount of air and electromagnetic damping of the movement of the galvanometer string was large enough, so that the motion of

the string corresponded almost exactly to every component of the impressed speech currents, and the resulting speech records should be regarded as being as nearly correct records of speech as can be obtained by any recording instrument.

Speech patterns of the same letter sound are almost exactly identical for all persons and independent of age or sex of speaker. Five hundred records were obtained of three men's speech and one woman's speech which prove this.

THE PATTERN FORMS OF THE NATURAL ALPHABET

Each of the letters has its own pattern form. All following records are of whispered speech, except Nos. 1 and 2.

CURVE 1— $\bar{u}$ in $y\bar{u}$

Showing fundamental tone of voice to be 140 to 200 cycles per second. $\bar{u}$ equals 1100 cycles per second superposed upon the lower note in definite pattern form. Male voice, spoken record.

CURVE 2— $\bar{u}$ in $y\bar{u}$

Fundamental tone of voice shows 250 cycles per second. $\bar{u}$ shows 1050 cycles superposed upon 250 cycles per second in definite pattern form. Female voice, spoken record.

CURVE 3— $\bar{a}$ in $\bar{a}\bar{o}$

$\bar{a}$ shows 2500 cycles per second superposed upon 500 cycles per second in definite pattern picture. (Repeating pattern picture.)

CURVE 4—b in $b\bar{a}y$

b shows special pattern picture.
c equals s.

CURVE 5— $\bar{d}$ in $\bar{d}\bar{a}y$
 $\bar{d}$ shows special pattern picture.

CURVE 6— $\bar{e}$ in $\bar{m}\bar{e}$
 $\bar{e}$ in $\bar{m}\bar{e}$, shows 2500 cycles per second superposed on 200 cycles per second in special pattern picture. (Repeating pattern picture).

CURVE 7— $\bar{f}$ in $\bar{f}\bar{e}\bar{e}$
 $\bar{f}$ in $\bar{f}\bar{e}\bar{e}$, shows alternation of 1000 and 2500 cycles, special pattern. (Repeating).

CURVE 8— $\bar{g}$ in $\bar{g}\bar{o}$
 $\bar{g}$ in $\bar{g}\bar{o}$, shows special pattern picture.

CURVE 9— h in $h\check{i}$
 h in hi , shows special pattern picture. (Repeating).

CURVE 10— $\bar{i}$ in $m\bar{i}$
 $\bar{i}$ shows 1700 cycles in special pattern picture. (Repeating).

CURVE 11— $\bar{i}$ in $h\check{i}$
 $\bar{i}$ shows 2600 to 2700 cycles in special pattern picture. (Repeating).

CURVE 12— j in $J\check{o}c$
 j shows 3000 cycles for 0.01 sec., 2500 cycles for 0.05 second, 2000 cycles for 0.01 second in special pattern picture.

CURVE 13—k in kō

k shows special pattern picture.

CURVE 14—l in lō

l shows alternation of 2500 and 1000 cycles per second in special pattern picture. (Repeating).

CURVE 15—m in m̄

m shows 1200 cycles for 0.01 second, silence 0.006 second, 1200 cycles for 0.01 second in special pattern picture. (Repeating).

CURVE 16—n in n̄

n shows 1100 cycles with silence following in special pattern picture. (Repeating).

CURVE 17—ō in tōē

ō shows 1000 cycles in special pattern picture. (Repeating).

CURVE 18—p in po

p shows special pattern picture.

q equals k $\ddot{y}$ ū.

CURVE 19—r in ro

r shows alternation of 1000 and 1250 cycles in special pattern picture. (Repeating).

CURVE 20—s in say

s shows 1000 alternating with 2000 cycles per second superposed upon 1000 cycles per second. (Repeating).

CURVE 21—t in to

t shows special pattern picture.

CURVE 22— $\bar{u}$ in $t\bar{u}$

$\bar{u}$ shows 1100 cycles in special pattern picture. (Repeating).

CURVE 23— v in $v\bar{e}$

v shows 1000 and 2500 cycles alternating in special pattern picture. (Repeating).

CURVE 24— $\bar{u}\bar{o} = w$ in $w\bar{e}$

w in $w\bar{e}$, shows $\bar{u}$ equals 1050 cycles; same special pattern as $\bar{u}$ passing into $\bar{o}$ equals 1000 cycles. (Repeating).

x equals $\tilde{e}ks$

$\bar{y}$ equals $\bar{u} \bar{i}$

$\tilde{y}$ equals $\tilde{i}$

CURVE 25— z in zec

z shows 5000 and 2000 cycles per second alternating over 1000 cycles per second in special pattern picture. (Repeating).

From Helmholtz's formula for resonance of a spherical cavity with a small opening,

$$n = a \sqrt{\frac{3r}{8\pi^3 R^3}},$$

where n equals pitch of resonance of cavity,

r equals radius of opening of cavity,

R equals radius of cavity,

a equals 33,226 centimeters,

we can calculate the natural resonance tones of the mouth while uttering speech. I have done this and find that the resonance tones for the mouth vary from 800 to 3000 cycles per second, depending on the opening of the mouth. From our records, we find the pitches upon which whispered speech is uttered to be also from 1000 to 3000 cycles per second. This only shows, however, that pitch is an unavoidable characteristic of speech sounds. From the known facts, that a phonograph talking record may be run at more than twice the normal speed so that the speech tones must vary through more than a full octave and the speech is still quite intelligible, it is conclusive that pitch is not the main characteristic of speech sounds. That intensity variation is the main characteristic of speech sounds is well shown by the whispered speech records given, where no characteristic pitch is shown for b, d, g, k, p, t, but only definite pattern pictures. When the speed of the phonograph is increased, this pattern is shortened and distorted as we hear, but its main form is retained and we judge it as nearly the same as the sound at the correct speed and therefore are able to understand the speech. The pitch of speech sounds, then, only exists as a carrier of the speech-variations. In another paper not yet published on the Nature of Hearing and Perception, it is shown how pattern forms operate memory cells, and how we thus perceive the letter sounds of speech.

In general, the pattern form persists for at least 0.01 second and if the letter sound is sustained, the pattern repeats itself again and again. See pattern picture of letter $\bar{o}$ sound, curve No. 17. One-hundredth second is the perception time. This is proved in the paper mentioned above. Therefore the mouth is a device for producing patterns of the necessary length and repeating them for sure perception.

Repetitions of pattern pictures of letter sounds (several times to make up a letter sound; see record of vowel $\bar{o}$, curve

No. 17,) are due to simultaneous excitatory and inhibitory action of impulses on same or opposing muscles in the mouth

or throat. At one moment excitatory impulse overbalances inhibitory impulse, then a maximum appears on the speech curve; when the excitatory and inhibitory impulses balance by interference (equal pull on muscle) no or little amplitude is shown on $\bar{o}$ curve. "A slight and rapid muscle tremor is regularly produced by the simultaneous play of excitation and inhibition on one muscle, just as 'progression' and other rhythmic movements are regularly produced by the simultaneous play of excitation and inhibition on antagonistic muscle pairs. One muscle can actually receive excitatory and inhibitory influences simultaneously, and this condition results in a peculiar and characteristic muscle 'tremor'." (Sherrington C. S., "Nervous Rhythm," Proc. Roy. Soc., 1913, Ser. B, 86, 219-232; Forbes, A., "Reflex Rhythm induced by concurrent excitation and inhibition," Proc. Roy. Soc., 1912, Ser. B, 85, 289-298.)

The repeated use of patterns is the physiological basis of memory.

I conclude from an exhaustive search of 500 vowel and con-

sonant curves, that an accumulator or memory cell exists in the brain for each letter sound, detecting a definite picture

FIG.1—GENERAL ARRANGEMENT AND WIRING DIAGRAM OF THE VOICE-OPERATED PHONOGRAPHIC ALPHABET WRITING MACHINE

pattern, such as that of the letter $\bar{o}$ which has a simple character of varying amplitudes, but repeats this character or picture pattern about 100 times a second. Since it takes 0.01 second

to perceive a picture pattern (perception time for sounds), this seems the right length of time (0.01 second) for a definite picture pattern to last and then to have repetition.

Having shown that speech is made up of a set of pattern pictures of sound waves called the phonographic alphabet, it seemed feasible to design and construct a machine which would record speech automatically in ink on paper in the form of this

FIG. 2—MECHANICAL ARRANGEMENT OF MIRROR-MOVING MECHANISM OF THE VOICE-OPERATED PHONOGRAPHIC ALPHABET WRITING MACHINE

easily read compact system of natural characters called the phonographic alphabet. This machine has been carefully designed and will be completed in the near future.

VOICE-OPERATED PHONOGRAPHIC ALPHABET WRITING MACHINE

Fig. 1 shows a wiring diagram and general arrangement of the elements of this machine. Notation on the sketch explains the function of the different parts of the device. Fig. 2, in

plan and elevation, shows the mechanical arrangement of each electrical resonator circuit and how the resistance of the selenium cell is varied by the differences in the amount of light coming from the vibrating mirrors.

Method of Operation. On reference to Fig. 1, it is evident that by talking into the telephone transmitter, speech currents are caused to flow into the electrical resonator circuits. An audion may be used as a relay if desired, to increase the strength of the voice currents. That resonator which is tuned to the main tone of the speech at any instant, will respond, the current existing in it at any moment varying in strength according to the speech-variation. The magnetic strip placed over the poles of the electrical resonator magnet will vibrate powerfully (if the electrical tuning of the electrical circuit and the mechanical tuning of the magnetic strip are of the same tone as the tone of the speech). On reference to Fig. 2, it is seen that the vibrating magnetic strip causes the connecting rod to oscillate the tiny shaft held in jewel bearings. The connecting rod is pinned to the tiny shaft only 0.004 in. from the center line of the shaft. In this way, a motion of the magnetic strip of 0.0001 in. will cause angular motion of the tiny shaft of $1\frac{1}{2}$ degrees. A mirror fixed to this tiny shaft will oscillate with it and cause the beam of light from the tungsten or other lamp to be reflected onto the selenium cell. In the normal position of rest the mirror reflects the light beam onto a blank space at the middle of the selenium cell box where it has no effect on the selenium cell. When the reflected beam of light oscillates, both parts of the selenium cell on each side of the blank space will be illuminated and the resistance of the selenium will be decreased, allowing more current to pass through it from the battery and causing the electromagnetic recording pen to trace a wavy line on the paper sheet fastened to the revolving platen.

To illustrate the method of operation: Let the word "go" be spoken into the telephone transmitter. From our table of the letters of the phonographic alphabet we shall expect the recording pen to write "go" in the form of . It is known from my research that there are present in the sound of a man's speech, principal tones of 100 cycles per second (the fundamental tone of a man's voice), 200 cycles per second (the 1st overtone of the man's voice), and a tone of approximately 1000 cycles per second (the resonance tone of the mouth for the o-position.) These three tones will be moulded or varied in intensity simul-

taneously by the muscles of the throat and mouth, first in the g-form (a sudden explosion of the three tones) and then in the o-form (a quick waxing and waning of the three tones.) Three tuned electrical resonating circuits marked respectively 100, 200 and 1000 cycles will simultaneously respond to the three tones and three beams of light will oscillate over the face of the selenium cell, each vibrating after the same pattern at the same instant of time. The current in the selenium cell and in the recording-pen magnet will vary in strength, first according to the g and then to the o-form, causing the writing in ink on the rotating paper cylinder of . In the same way, all letter sounds are written and we thus have a visible-writing machine recording our thoughts as rapidly as we speak, in the natural alphabet.

This work was conducted by cooperation of the Dept. of Physiology, College of Physicians and Surgeons, New York, and the Underwood Typewriter Co.