The Bell System Technical Journal

Vol. XIX *October, 1940* *No. 4*

The Carrier Nature of Speech

By HOMER DUDLEY

Speech synthesizing is here discussed in the terminology of carrier circuits. The speaker is pictured as a sort of radio broadcast transmitter with the message to be sent out originating in the studio of the talker's brain and manifesting itself in muscular wave motions in the vocal tract. Although these motions contain the message, they are inaudible because they occur at syllabic rates. An audible sound is needed to pass the message into the listener's ear. This is provided by the carrier in the form of a group of higher frequency waves in the audible range set up by oscillatory action at the vocal cords or elsewhere in the vocal tract. These carrier waves either in their generation or their transmission are modulated by the message waves to form the speech waves. As the speech waves contain the message information on an audible carrier they are adapted to broadcast reception by receiving sets in the form of listeners' ears. The message is then recovered by the listeners' minds.

SPEECH is like a radio wave in that information is transmitted over a suitably chosen carrier. In fact the modern radio broadcast system is but an electrical analogue of man's acoustic broadcast system supplied by nature. Communication by speech consists in a sending by one mind and the receiving by another of a succession of phonetic symbols with some emotional content added. Such material of itself changes gradually at syllabic rates and so is inaudible. Accordingly, an audible sound stream is interposed between the talker's brain and the listener. On this sound stream there is molded an imprint of the message. The listener receives the molded sound stream and unravels the imprinted message.

In the past this carrier nature has been obscured by the complexity of speech.[1] However, in developing electrical speech synthesizers

[1] Speech-making processes are here explained in the terms of the carrier engineer to give a clearer insight into the physical nature of speech. The point of view is essentially that of the philologist who associates a message of tongue and lip positions with each sound he hears. This aspect also underlies the gesture theory of speech by Paget and others and the visible speech ideas of Alexander Melville Bell. The author has been assisted in expressing speech fundamentals in carrier engineering terms by numerous associates in the Bell Telephone Laboratories experienced in carrier circuit theory. Acknowledgment is made in particular of the contributions of Mr. Lloyd Espenschied.

copying the human mechanism in principle, it was soon apparent that carrier circuits were being set up. Tracing the carrier idea back to the voice mechanism there was unfolded, a little at a time, the carrier nature of speech. Ultimately the speech mechanism was revealed in its simplest terms as a mechanical sender of acoustic waves analogous to the electrical sender of electromagnetic waves in the form of the radio transmitter. Each of these senders embodies a modulating device for molding message information on a carrier wave suitable for propagation of energy through a transmission medium between the sending and receiving points.

The Carrier Elements of Speech

This carrier basis of speech will be illustrated by simple speech examples selected to show separately the three carrier elements of speech, namely, the carrier wave, the message wave, and their combining by a modulating mechanism. These illustrations serve the purpose of broad definitions of the carrier elements in speech.

The illustration chosen for the carrier wave of speech is a talker's sustained tone such as the sound "ah." In the idealized case there is no variation of intensity, spectrum or frequency. This carrier then is audible but contains no information, for information is dynamic,[2] ever changing. The carrier provides the connecting link to the listener's ear over which information can be carried. Thus the talker may pass information over this link by starting and stopping in a prearranged code the vocal tone as in imitating a telegraph buzzer. For transmitting information it is necessary to modulate this carrier with the message to be transmitted.

For the second illustration, message waves are produced as muscular motions in the vocal tract of a "silent talker" as he goes through all the vocal effort of talking except that he holds his breath. The message is inaudible because the motions are at slow syllabic rates limited by the relatively sluggish muscular actions in the vocal tract. Nevertheless these motions contain the dynamic speech information as is proved by their interpretation by lip readers to the extent visibility permits. Another method of demonstrating the information content of certain of these motions is the artificial injection of a sound stream into the back of the mouth for a "carrier" whereby intelligible speech

[2] The information referred to is that in the communication of intelligence. There is, however, static information in the carrier itself. This serves for "station identification" in radio and may similarly help in telling whether it was Uncle Bill or Aunt Sue who said "ah."

can be produced from almost any sound stream.[3] The need of an audible "carrier" to transmit this inaudible "message" is obvious.

The final example, to illustrate the modulating mechanism in speech production, is from a person talking in a normal fashion. In this example are present the message and carrier waves of the previous

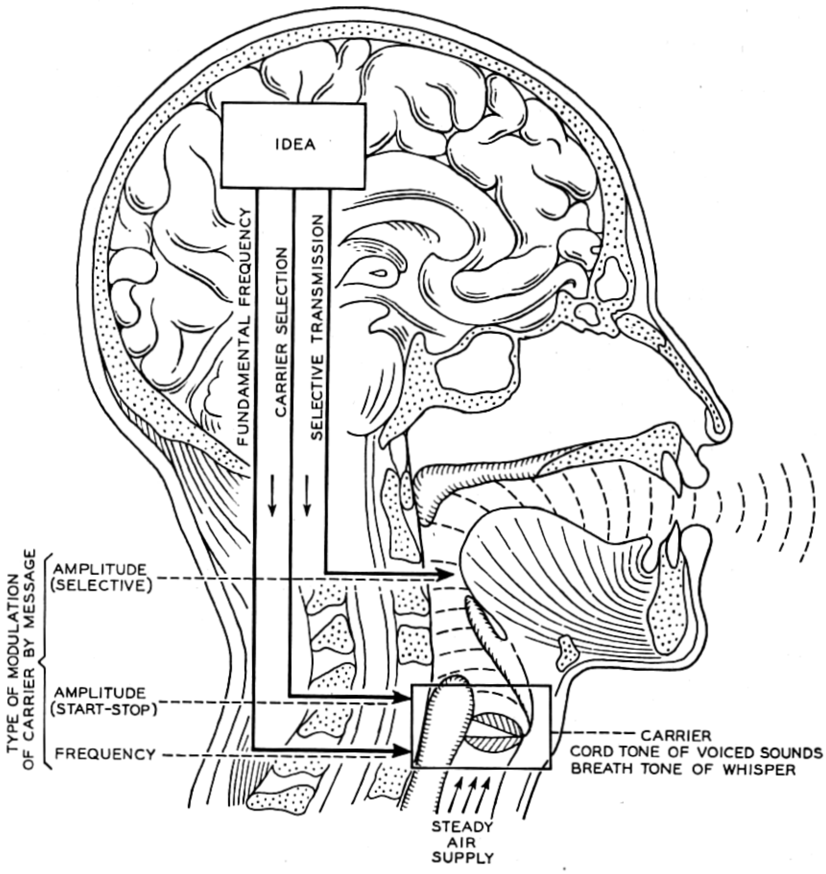

Fig. 1—The vocal system as a carrier circuit.

examples, for both are needed if the former is to modulate the latter. However, the mere presence of the carrier and message waves will not make speech for if they are supplied separately, one by a silent talker and the other by an intoner, no speech is heard but only the audible intoned

[3] R. R. Riesz, "Description and Demonstration of an Artificial Larynx," *Jour. Acous. Soc. Amer.*, Vol. 1, p. 273 (1930); F. A. Firestone, "An Artificial Larynx for Speaking and Choral Singing by One Person," *Jour. Acous. Soc. Amer.*, Vol. 11, p. 357 (1940).

carrier. Ordinary speech results from a single person producing the message waves and the carrier waves simultaneously in his vocal tract, for then the carrier of speech receives an imprint of the message by modulation.

The Speech Mechanism as a Circuit

The foregoing three illustrations by segregating the basic elements in speech production reveal the underlying principles. The present paper treats of these elements as functioning parts of a circuit. In Fig. 1 is shown a cross-section of the vocal system. The idea to be expressed originates in the talker's brain at the left top. Thence, impulses pass through the nerves to the vocal tract with the complete information of the "message," that is to say, what carrier should be used, what fundamental frequency if the carrier is of the voiced type and what transmission through the vocal tract as a function of frequency. The carrier whether voiced or unvoiced is shown for simplicity as arising at the talker's vocal cords. This carrier is modulated to form speech having the complete message imprinted on it preparatory to radiation from the talker's mouth to the ear of the listener, who recognizes the imprinted message.

In discussing the speech mechanism as a circuit, it is clearer to start with a block schematic. Figure 2 has thus been drawn to sketch the

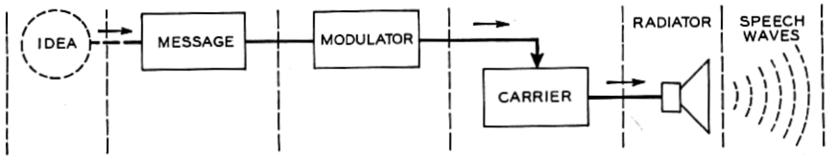

Fig. 2—The basic plan of synthesizing speech.

basic plan of speech synthesizing. As in Fig. 1, the idea gives rise to the message which modulates the voice carrier to produce the speech radiated from the talker's mouth. One can follow the path of the message from its inception in the talker's brain to its radiation from his mouth as an imprint on the issuing sound stream. The progress of the sound stream is also seen from its origin as an oscillatory carrier to its radiation from the talker's mouth carrying the message imprint.[4] The light arrow heads indicate direction of flow while the heavy ones indicate a modulatory control of the carrier by the message. This

[4] Here the carrier path is stressed to show the alteration of the carrier sound stream as it proceeds on its way from the point of origin to the point of radiation. This also accords with the importance of the voice carrier which is received and used by the ear, and thus differs from the treatment of the carrier in simple radio broadcast reception.

modulatory control is exerted on the carrier wave in part as the carrier is generated and in part as it is transmitted after generation.

Relevant Carrier Theory

The heart of the speech-synthesizing circuit of Fig. 2 is the part in which the group of waves making up the message modulate the component waves of the carrier. In any one of these modulations, there is the simple carrier process blocked out in Fig. 3. Here a message [5]

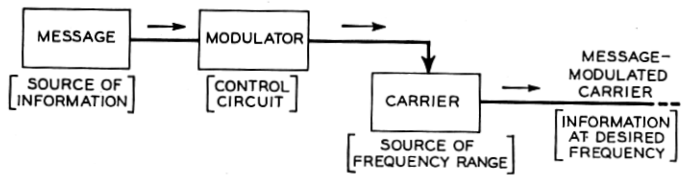

Fig. 3—The elements of a carrier sender.

containing the information modulates a carrier determining the frequency range so that the end product in the form of the message-modulated carrier contains the information of the message translated to frequencies in the neighborhood of the carrier. In this way the carrier sound stream of speech is imprinted with the message.

The prerequisites of the carrier system sender are, as indicated in Fig. 3, first, a carrier wave source; second, a message wave source; and third, a modulating circuit of variable impedance by which the message controls the carrier. The carrier wave is for the simplest case a single sine wave function of time characterized by an amplitude, a frequency and a phase. The message wave as a rule is more complex but may be analyzed as the sum of component sine waves each of which is characterized by its own amplitude, frequency, and phase. In most carrier circuits the frequency range of the message is below that of the carrier. This is true of speech production.

The function of the modulating circuit is supplying a means for the message wave to modify a characteristic of the carrier. If the carrier wave amplitude is modified by the message wave amplitude the process is known as amplitude modulation; if the carrier wave frequency is so modified the process is called frequency modulation while if the carrier wave phase is so modified the process is called phase modulation. No distinction is made as to whether the modification occurs during or

[5] The word "message" has been substituted for the usual carrier term "signal" to avoid confusion since the input signal is commonly speech whereas here the output wave is speech. "Message" seems particularly appropriate with its suggestion of code as in telegraph.

after the generation of the carrier. Modification of the carrier wave characteristics by other than the amplitude of the message need not be considered here. In the voice mechanism significant amplitude and frequency modulations of the carrier occur. Phase modulation takes place also but will not be discussed because the listener's ear is not very sensitive to these phase changes in the carrier.

In attempting to segregate the carrier elements of speech we run into one serious difficulty. In an idealized carrier circuit as shown in Fig. 3 connections can be cut between the two energy sources and the modulator so that each boxed element can be studied independently. With the human flesh of the voice mechanism this is no longer feasible; the use of cadavers would help very little because normal energizing is then impossible. The same difficulty often appears in electrical modulators as, for example, within a modulating vacuum tube where a grid voltage modulates a plate current. In such a case of common parts it is necessary to discuss the action of each of the three elements in the presence of the other two.

With this carrier theory review as a background we are in a position to analyze the three elements making up the carrier transmitting system of the human voice. While the picture presented is over-simplified in details the principles hold and aid in applying carrier methodology to explain the mechanism of speech.

The Voice Carrier

In electrical circuits the carrier is obtained from an oscillatory energy source. The same holds for speech. In the electrical circuit the oscillatory waves (a-c.) are ordinarily generated from a supply of d-c. energy.[6] The same is true in speech with the compressed air in the lungs furnishing the steady supply. Confusion must be avoided, for in speech the conversion of steady to oscillatory energy is often described as *modulation*. Here this conversion of energy form will be considered as an oscillatory action so that the term *modulation* can be reserved for the low-frequency syllabic control of this oscillatory energy to produce the desired speech. *Oscillatory* then will refer to automatic natural responses while *modulatory* will refer to forced responses which are controlled volitionally. This distinction is consistent with carrier terminology.

In the simplest electrical modulating circuits the carrier is a sine

[6] In the usual electrical circuit the carrier is cut off by turning off the output but leaving the carrier oscillator energized as, for example, in voice frequency telegraphy. In the voice mechanism, however, the oscillator is stopped at the source. The difference between the electrical on-off switching and the start-stop switching of speech is not fundamental but results from the use of the most suitable action in each case in view of the conditions prevailing.

wave although this is not true of the damped wave carriers of multi-frequency type once commonly used in spark wave radio telegraphy. The carrier wave in speech is not a simple sine wave. Such a sound would be like a whistle and so too limited for the rich flexibility of speech. Instead the voice carrier is a compound tone having a multiplicity of components of different frequencies which together cover the audible range fairly completely. While these components may be considered as a multiplicity of separate carriers it is simpler to think of the ensemble as a single complex carrier; so this terminology has been used in the earlier carrier illustration and elsewhere in this paper.

Aside from this compound nature of the voice carrier, the voice has two distinct types of carrier, one for voiced and one for unvoiced sounds. Some sounds such as "z" have both types present at the same time but this case may be treated as the superposition of one carrier on the other. For voiced sounds the carrier is the vocal cord tone, an acoustic wave produced by the vibration of the vocal cords consisting of a fundamental frequency component and the upper harmonics thereof. These decrease in amplitude with increasing frequency. For unvoiced sounds the carrier is the breath tone, a complex tone resulting from a constriction formed somewhere in the vocal tract through which the breath is forced turbulently to produce a continuous spectrum of frequency components in the audible range.

These carrier waves must be dissociated from any effects of resonant vocal chambers, for such characterize the speech message rather than the carrier. Furthermore, these carrier waves must be mentally pictured as sustained indefinitely with the starting and stopping of them also characterizing the message wave. Pauses for breath, due to incidental human limitations, do not invalidate the fundamental theory.

The Speech Message

Since a sustained voice carrier has no dynamic flow of information there is need for a source of message waves and a modulating mechanism for imprinting the message on the carrier. Conversely, any variation from the sustained carrier infers the presence of a message wave molding the carrier. The message consists of those articulating, phonating and inflecting motions of the vocal parts which imprint the information on the carrier sound stream. The importance of the message waves cannot be stressed too much. Any impairment of them is an impairment of the message.

The message waves include the motions producing speech changes at infra-syllabic rates, such as the effect of anger when a talker may be high-pitched for many minutes. When the carrier is thus altered over a long period of time the question arises whether to use a long- or short-term value of the carrier. The answer may well be the same as in the analogous radio problem. If weather causes a carrier frequency to be slightly high all day, this higher value is taken as the normal carrier in studying short-term effects such as the degree of modulation. But in long-term studies of carrier stability the deviations from the mean represent a frequency modulation which is observed as a "message" effect.

Due to the inseparability of the message wave motion and its associated wave of impedance change in the modulating mechanism there may be confusion in distinguishing between the modulating elements and the source of the message waves. The rule followed here is simple. From the standpoint of the human flesh lining the vocal tract, the message source is internal, the modulating elements, external. The message consists of those muscular motions (or pressures or displacements) in the vocal tract which are present in the "silent talker" and are volitional in nature. This definition excludes the oscillatory motions which make up the carrier. The modulating elements are acoustic in nature since the carrier starts as a sound stream and ends as a modulated sound stream.

There are three important variations of the voice carrier and so three types of message and of associated modulation. These variations are: first, selecting the carrier; second, setting the fundamental frequency of the voiced carrier; and third, controlling the selective transmission of the vocal tract.[7] The message waves in the three cases will be discussed with the corresponding modulation reserved for consideration under the next heading.

Selecting the carrier appears as a simple start-stop message, complicated somewhat by the presence of two types of carrier and by locating the constriction for the unvoiced type at several places in the vocal tract. We may think of a start-stop type of message for each point where constrictions are formed, including the vocal cords for the voiced type of carrier. A constriction message may be plotted as the opening between vocal parts at the constriction with critical values for the onset of audible carrier. The constrictions are to a certain extent independent. Thus with the vocal cords vibrating, a constriction from the tongue tip to the upper teeth may also be formed, as in making the "z" sound. Again, in whispering, there may be simul-

[7] A fourth message characteristic prescribes the intensity of the speech. This message may be included in the carrier selection if the carrier is selected for intensity as well as type. The matter of intensity is passed over rather lightly here because a comparison is being developed between the human and electrical speech synthesizers with the final intensity in the latter under control of an amplifier setting.

taneous constrictions, both of the unvoiced type, one at the vocal cords and one in the mouth. As the voice has two distinct types of carrier, the vocal cord tone and the breath tone, the selection sets up one of four carrier conditions at any instant: no carrier, vocal cord tone only, breath tone only, or a combination of vocal cord tone and breath tone. This start-stop message resembles the on-off type of telegraph where switching controlled by other muscular motions sets up speech information in another code, that of telegraph. As mentioned earlier a communication system can be made with the vocal system by starting and stopping a voice carrier in a vocal imitation of a telegraph buzzer. While this would be a clumsy way of communicating information it marks the start-stop control of the voice carrier as a speech message and not part of the voice carrier. Another check is that the "silent talker" does form such constrictions.

The second type of message wave specifies the fundamental frequency with any related voice changes for the voiced type of carrier. This message, in a mechanical form, may be the time variation of the tension of the vocal cords. As the frequency of each upper harmonic is changed in the same ratio as the fundamental frequency, a single parameter suffices for all of the carrier components. The unvoiced carrier has no message of this type impressed since the unvoiced sounds are not characterized by pitch.

The third and final type of message wave controls the selective transmission in the vocal tract. By comparison, the first two types of message are simple, with the selecting of carriers ideally changing all components of the carrier by the same amplitude factor and the fundamental frequency control changing them by a uniform frequency factor. The vocal transmission, however, results from a multi-resonance condition with more than one degree of freedom. There follows a selective amplitude modulation with some carrier components decreasing in amplitude at the same instant that others are increasing. Maximum transmission occurs when a component coincides with an overall resonance, minimum transmission when it coincides with an anti-resonance and intermediate transmission for other cases. The voice message for transmission appears in mechanical form as the displacements of lips, teeth, tongue, etc., with as many such displacements considered as are needed for adequately expressing the speech content. This infers finding the simplest lumped impedance structure equivalent to the distributed impedance structure of the vocal tract to the necessary degree of approximation.

All these mechanical displacements of vocal parts that together constitute the voice message lead to corresponding displacements of air in the vocal system, resulting in a set of air waves that likewise contain all the information of speech. These airborne message waves, however, are at syllabic rates and so below the frequency range of audibility.

The Voice Modulators

The three voice modulators associated with the three speech messages are the mechanisms of (*a*) selecting the carrier, (*b*) setting the fundamental frequency and (*c*) controlling the selective transmission. The mechanism for starting and stopping a voice carrier is simple. Assume a sustained carrier of either the voiced or unvoiced type. It can be stopped by opening the constriction at which it is formed. This alters the acoustic impedance of the opening which is then the modulating element in this case.

The modulating mechanism for controlling the fundamental frequency appears in the vibrating portions of air at the glottis. The exact mechanism is of no importance here so long as the message wave at the vocal cords finds means for altering the fundamental frequency under the control of the will.[8] This is a case of frequency modulation of multiple carriers harmonically related.

The modulating mechanism for controlling the transmission through the vocal tract as a function of frequency consists of the masses and stiffnesses of air chambers and openings in the vocal tract. These are varied under control of the message in the form of muscular displacements of vocal tract parts. There is a more complicated modulation in the vocal tract than in the usual electrical circuit for amplitude modulation because the varying impedances are reactive in the voice mechanism but resistive in the electrical circuit and also because several independent modulator elements are used in the voice mechanism as against either a single one or a group functioning as a unit in the simple electrical modulator. The reactive nature of the vocal impedances leads to the selective control of the amplitudes of the various harmonics of the voice carrier. The amplitude modulation of each carrier component by the combined message waves produces an output containing the carrier and sideband frequencies.

Comparison of Speech Synthesizing Circuits

The fundamental processes in human speech production are thus analogous to those of electrical carrier circuits. There is a switching of voice carrier energy comparable to that in voice frequency telegraph;

[8] For a simplified theory of the larynx vibration see R. L. Wegel, *Bell Sys. Tech. Jour.*, Vol. 9, p. 207 (1930) and *Jour. Acous. Soc. Amer.*, Vol. 1, Supp. p. 1, April 1930. The analogy of the larynx to a vacuum tube oscillator is described in an abstract, *Jour. Acous. Soc. Amer.*, Vol. 1, p. 33 (1929).

there is an altering of speech frequencies as in frequency modulating circuits; and finally, there is an amplitude modulation to yield a selective transmission of the various carrier components of the voice. However, the voice mechanism differs from the usual carrier circuit markedly as regards complexity. In the voice mechanism there are two types of carrier each with a multiplicity of partial carrier components. The incoming message has a multiple nature. Finally, several modulations take place including both amplitude and frequency types. This multiplicity of carrier relations indicates the wide range of voice phenomena possible.

Any electrical speech synthesizer must be a functional copy of the human speech synthesizer in providing the essential speech characteristics sketched in the preceding paragraph. There have been developed two such electrical synthesizers referred to in the introduction. A brief description of these will be given followed by some circuit comparisons.

These electrical synthesizers are known as the vocoder and the voder. The vocoder was so named because it handles the speech in a coded form; the voder, because it serves as a Voice Operation DEmonstratoR. Considerable interest has been manifested at the public showings of each of these synthesizers, the vocoder in a limited number of lecture demonstrations and the voder at the San Francisco and New York World's Fairs. Circuit details have been published elsewhere.[9]

Of these two speech synthesizers the vocoder was constructed first. It works on the principle of automatically remaking speech under control of spoken speech instantaneously analyzed to derive the code currents for the control. The vocoder as set up for demonstration is shown in Fig. 4.

The voder was derived from the vocoder by substituting manipulative for automatic controls. The resulting voder as displayed at the New York World's Fair is shown in Fig. 5. In the Fair demonstration, repeated continuously at intervals of about five minutes, the male announcer gives a simple running discussion of the circuit with the girl operator replying to his questions by forming sounds on the voder and connecting them into words and sentences. She does this by manipulating fourteen keys with her fingers, a bar with her left wrist and a pedal with her right foot. This requires considerable skill by the operators. The vocoder, automatic in nature, presents no problem of operating technique.

[9] The vocoder in the *Jour. Acous. Soc. Amer.*, Vol. 11, pp. 169–177, October 1939, "Remaking Speech," Dudley; the voder in the *Journal of the Franklin Institute*, Vol. 227, pp. 739–764, June 1939, "A Synthetic Speaker," Dudley, Riesz and Watkins.

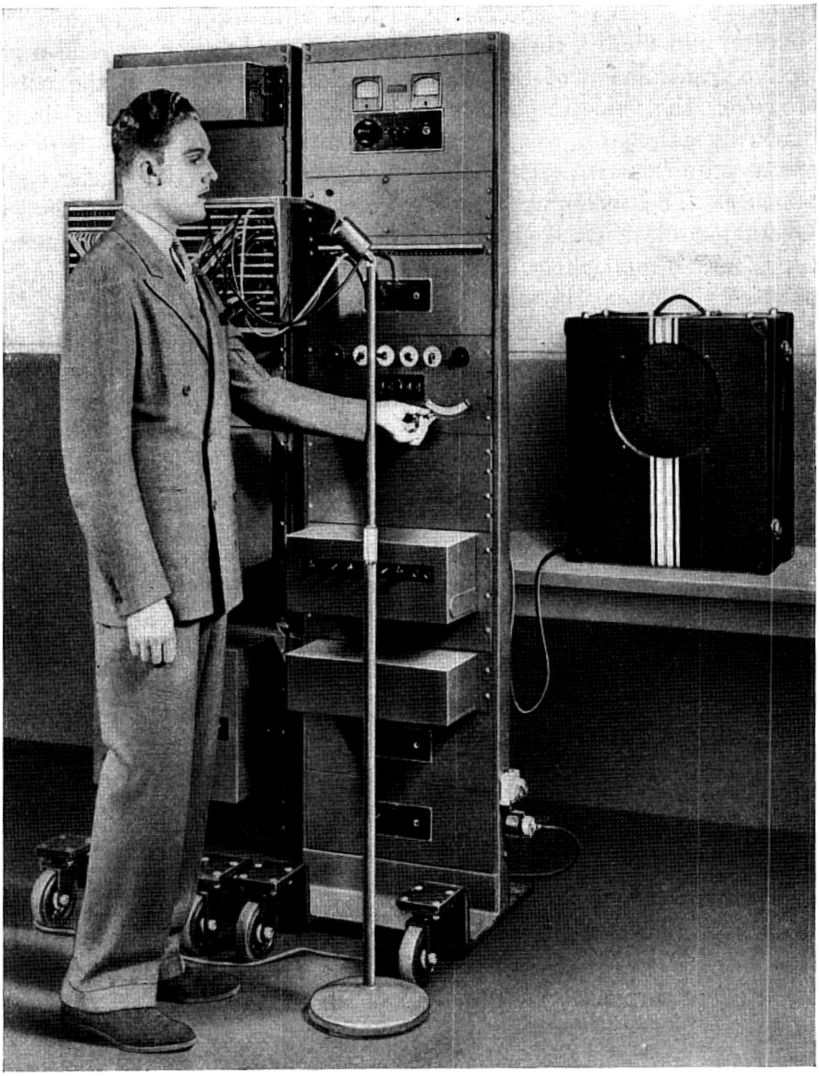

Fig. 4—The vocoder as demonstrated.

Circuit diagrams supply a shorthand for expressing the salient features of electrical circuits. In the next three figures comparative block circuits will be shown for the human and the two electrical speech synthesizers, tracing the communication from the origin of an idea in the communicator's brain to final expression as speech. In each cir-

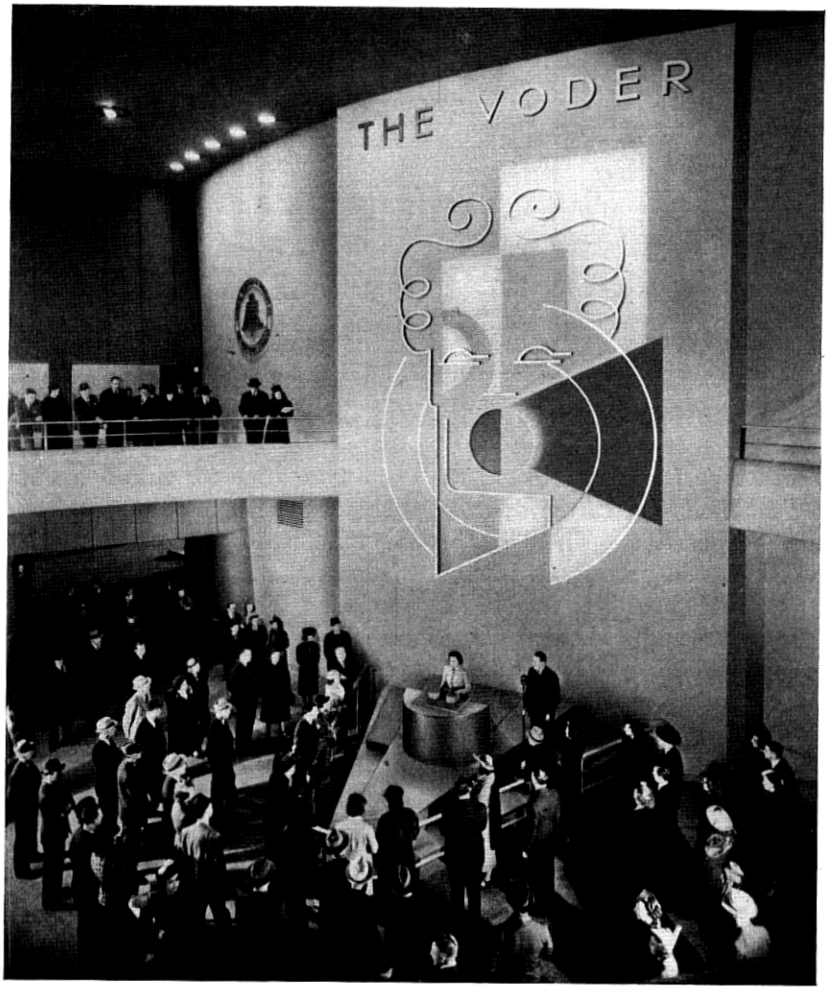

Fig. 5—The voder being demonstrated at the New York World's Fair.

cuit, the arrangement in Fig. 2 will be followed with sufficient detail to show the functional relations of the parts discussed in this paper.

Figure 6 gives a block diagram of the voice mechanism of Fig. 1 with approximating electrical circuit symbols. The same communication paths can be traced. Thus from the talker's brain are sent nerve impulses that set up the message as a set of muscular displacements containing information as to the voice carrier to use, the fundamental frequency for the voiced carrier, and the selective transmission of the vocal tract. The air expelled from the lungs sets up as carriers the

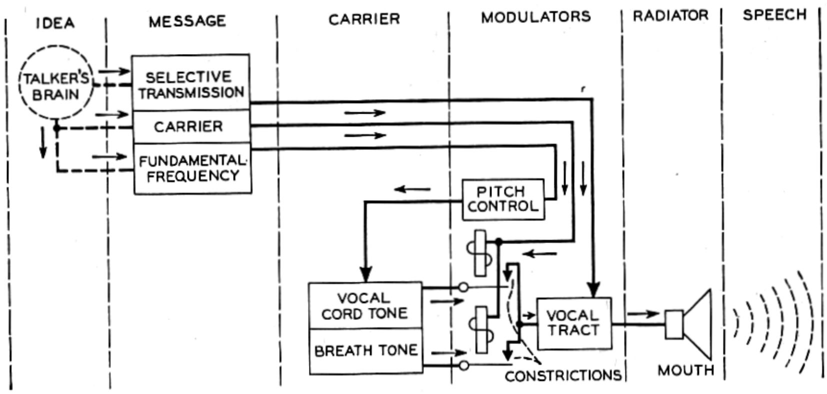

Fig. 6—Block diagram of the voice mechanism.

breath tone for unvoiced and the vocal cord tone for voiced sounds. For simplicity the carrier selection is shown after instead of before the carrier generation. These carriers are modulated by the message wave to produce the output of speech in the form of the message-modulated carrier in the audible range of frequencies.

Figures 7 and 8 show similar block schematics for the vocoder and the voder. The voder circuit has been simplified by the omission of a few controls for easier operation. In these electrical synthesizers, the carrier is provided by a buzzer-like tone from a relaxation oscillator for the voiced sounds and from a hiss-like sound from a gas-filled tube for

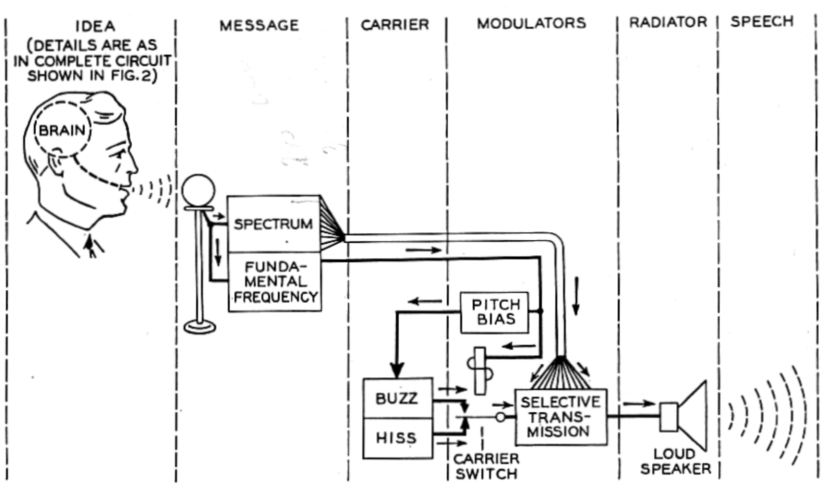

Fig. 7—Schematic circuit of the vocoder.

the unvoiced sounds. In the vocoder, for simplicity's sake, one or the other of these energy sources is used according to whether the sound is voiced or unvoiced, with no provision for the mixed types of sounds found in the human voice. The analyzer of the vocoder derives the

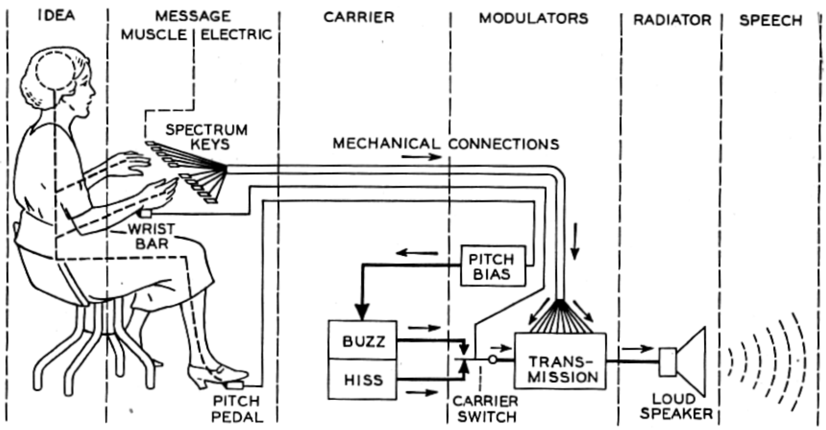

Fig. 8—Schematic circuit of the voder.

original speech message in terms of a modified set of parameters. This analyzer suppresses the original carrier of the talker and so resembles the demodulator in radio reception. The analyzer acts as an electrical ear to tell the artificial vocal system of the vocoder what to say, the whole vocoder acting as a synthetic mimicker.

The basic similarity of the electrical and human speech synthesizers is seen in these figures. In all three cases the message is originated by the brain of the sender of the speech information. There is in each case a transmission of control impulses by the talker's nervous system to the appropriate muscles. The muscles produce displacements of body parts formulating the speech information as a set of mechanical waves. These waves appear in the vocal tract in the case of normal speech; in the fingers, wrist and foot in the case of the voder, but in the case of the vocoder use is made of electrical currents derived from and equivalent to the vocal tract displacements in ordinary speech. In each case the message contains the speech information in syllabic waves. In all cases the message waves control the choice of carrier, the fundamental frequency of the voiced type carrier and the spectrum of power distribution in the speech output. Differences arise in the details rather than in the principles.

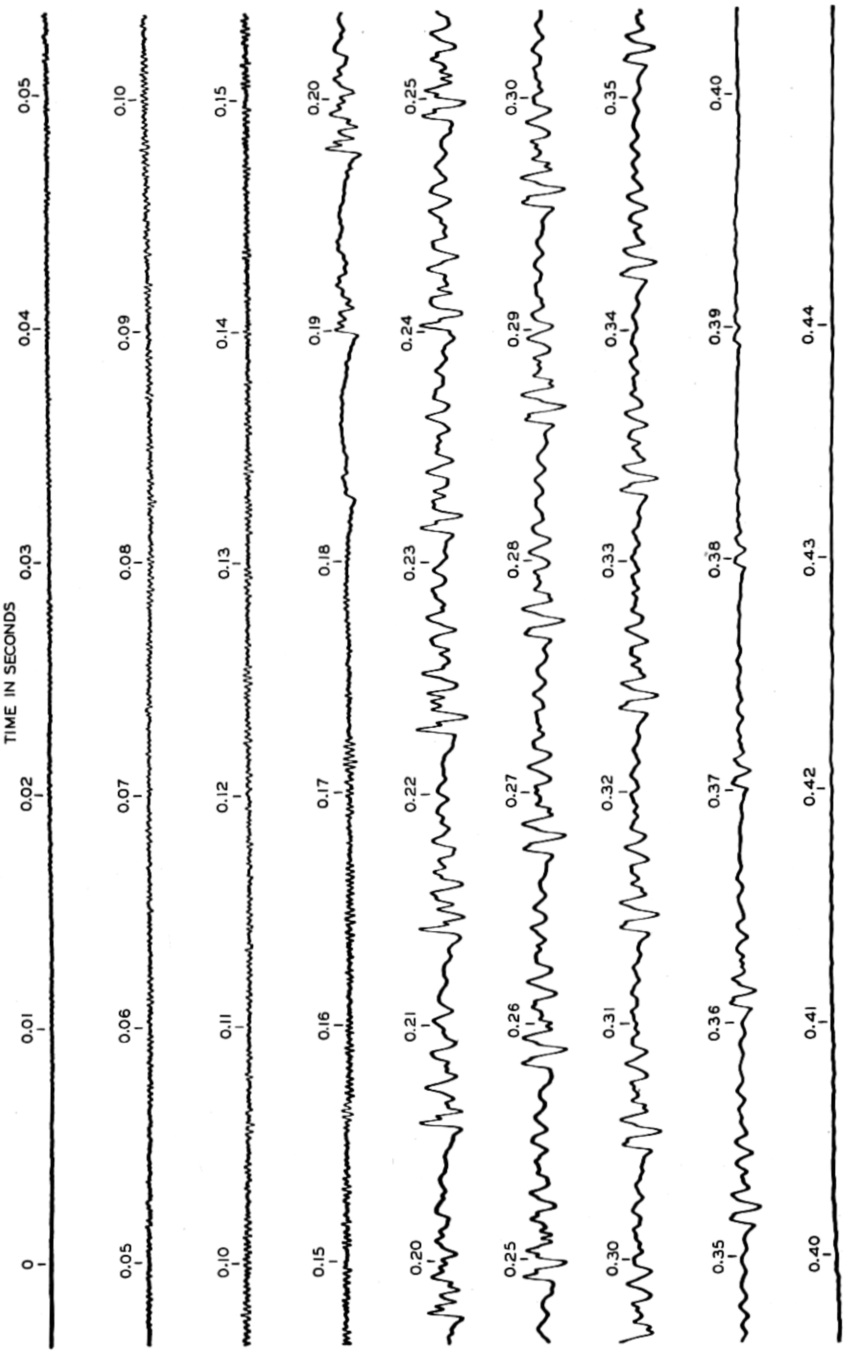

Fig. 9—Oscillogram of the sound "sa."

Speech Characteristics from the Carrier Point of View

Now that the mechanism of speech has been described in carrier terms it is of interest to observe carrier features as they manifest themselves in the characteristics of speech. Some of these can be seen by the eye in speech oscillograms. Some can be demonstrated to the ear with a speech synthesizer such as the vocoder.

For a visual illustration there is shown in Fig. 9 a high quality oscillogram taken from Crandall [10] of the sound "sa" (Plate No. 160. Spoken by M. B.) for a medium-pitched male talker. The carrier shown by the oscillogram is of the unvoiced type for the earlier and of the voiced type for the later part. As one looks at the oscillogram he sees a great mass of the high-frequency components of the carrier. Scrutiny, however, reveals modulated on the carrier the message information in terms of switched energy sources, controlled fundamental frequency and varied transmission characteristic. Shortly after .17 second the switching off of the unvoiced carrier begins. Remnants of the unvoiced carrier can be seen in the voice period just before .19 second and the one starting at about .19 second. The switching on of the voiced carrier appears just after .18 second and seems to be reasonably well completed at the end of the second voice period just before .20 second. This switching was not instantaneous. However, the ear probably does not observe the duration time of the switching. The fundamental frequency falls rapidly at the beginning followed by a leveling out and then a final slight fall in the last few periods. It starts at 140 cycles per second, dropping to around 110 in the level portion, and then to 101 at the end. The resonance conditions cannot be followed too well by eye. However, around .20 second there is a major lower-frequency resonance of about 800 cycles. At .33 second this resonance appears to have increased to 1100 cycles or so. A similar alteration of resonance conditions may be observed if the little shoulder on the rear side of the peak just in front of the .25 second mark is traced in adjacent periods. It can readily be followed back to the third period just before .20 second and can still be seen in the last distinct voicing period starting before .39 second. The dynamic variation of the speech at syllabic rates in accordance with the message content is thus revealed.

For another visual illustration of the speech message Fig. 10 shows a set of oscillograms [11] from the vocoder analyzer for the words "She saw Mary." The oscillogram of the input speech is the trace next to

[10] *Bell Sys. Tech. Jour.*, Vol. 4, p. 586, 1925.
[11] This figure is a copy of Fig. 3 in the paper "The Automatic Synthesis of Speech," Dudley, *Proc. Nat. Acad. Sci.*, Vol. 25, pp. 377–383, July 1939.

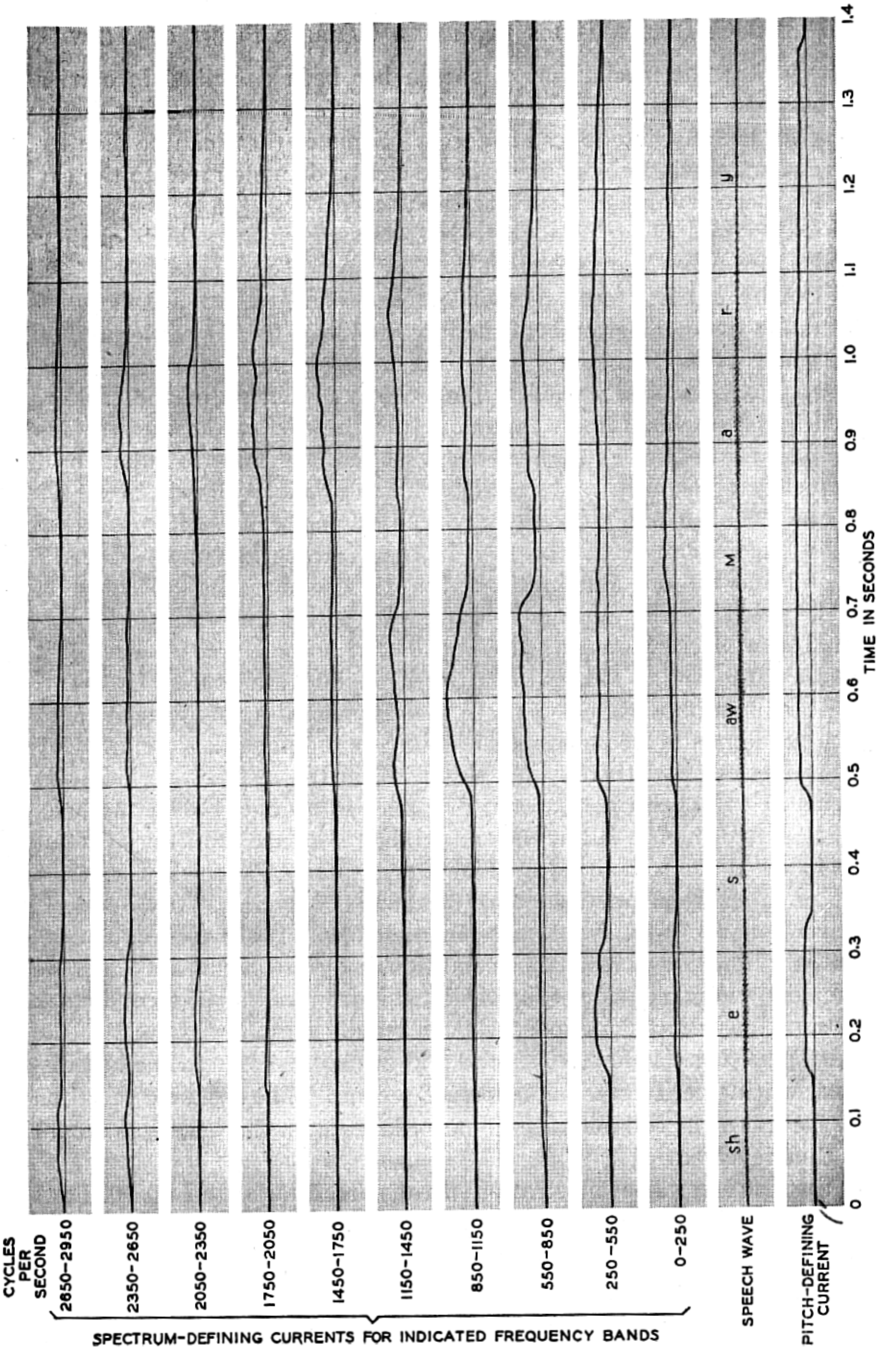

Fig. 10—Coded speech (derived message currents) for the words "She saw Mary."

the bottom. The trace below shows the defining current for the fundamental frequency, while the ten traces above show currents indicating the rectified power in ten frequency bands of 300 cycles width except that the lowest one extends from 0 to 250 cycles. The slow rates of change are noted in the message currents when compared to the original speech wave.

Demonstrations of the vocoder indicate to the ear the carrier nature of speech. Thus the carrier used for remaking speech, whether a monotone or a hiss sound, is observed to have no intelligibility when heard alone. The message currents derived from spoken speech are not audible. However, intelligible "speech" is produced by the modulation of either type of carrier by the message currents of selective transmission. Similarly, there can be used for the carrier a wide variety of sound from the puffs of a locomotive to instrumental music. Upon imprint of the transmission message currents from spoken speech, new forms of odd sounding but nevertheless intelligible "speech" are produced.

The carrier conception of speech reveals what is important and not important in evaluating speech characteristics. An example of interest is the matter of phase. It has long been known that phase was unimportant to the ear at reasonably low listening levels. From the carrier point of view this is natural, for the phase changes referred to are those in the carrier and so, unimportant. When the phases of the message components are altered, there is a very noticeable effect on the ear, for phonetic units are now being shifted.

The great advance in recent years in the application of carrier circuits has been guided by mathematical theory. Since in electrical speech synthesizers the carrier and message currents are separated physically, it is possible to use carrier equations expressing the modulation phenomenon. Similar equations may be written for the voice mechanism as represented by Fig. 6. This has been done in the attached appendix, thus separating speech into syllabic and carrier factors.

APPENDIX

Mathematical Relations

The speech concepts developed in the body of the paper may be expressed in mathematical terms which not only give the fundamental relations in simplest form but also aid in the application of the well-established carrier technique to speech. For voiced sounds, periodic by nature, the carrier C_v may be written as a function of the time t thus:

$$C_v = \sum_{k=1}^{n} A_k \cos (kPt + \theta_k). \tag{1}$$

Here C_v is composed of n audible harmonics of relatively high frequencies with the kth of amplitude A_k, frequency kP radians per second, and phase θ_k. The choice of fundamental frequency P is somewhat arbitrary but may well represent the average of the talker over the period of interest.

By modulation processes, there is molded on to this carrier the total message information at the relatively low syllabic frequencies. The message is divided into three parts: (*a*) the starting and stopping of the carrier; (*b*) the instantaneous fundamental frequency; and (*c*) the selective transmission through the resonant vocal tract.[12] These three message functions as they manifest themselves in varying the carrier will be represented by s, p, and r, respectively. Equation (1) will be modified to indicate the effect on the carrier of each of these modulations separately, after which the equation will be rewritten to show the effect of all three acting simultaneously.

The effect of starting and stopping the carrier is described mathematically as a function of time by multiplying C_v by the switching function $s(t)$, giving:

$$\text{Switched } C_v = s(t) \sum_{k=1}^{n} A_k \cos (kPt + \theta_k). \tag{2}$$

For simple on-off switching, $s(t)$ alternately equals zero and unity, although it may in general represent more gradual changes or even any variations of intensity over the frequency range.

The instantaneous fundamental frequency is obtained by multiplying P by the inflecting factor $p(t)$. The effect of the frequency modulation [13] is represented by substituting for Pt the integrated quantity

$$\int_0^t Pp(t)dt = P \int_0^t p(t)dt.$$

Writing this value for Pt in equation (1) gives the inflected carrier wave:

$$\text{Inflected } C_v = \sum_{k=1}^{n} A_k \cos \left[kP \int_0^t p(t)dt + \theta_k \right]. \tag{3}$$

[12] As in the body of the paper, the effect of phase modulation is neglected here.
[13] "Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency Modulation," J. R. Carson and T. C. Fry, *Bell Sys. Tech. Jour.*, Vol. 16, p. 513 (1937).

The effect of the selective transmission is allowed for by multiplying C_v by the transmitting factor $r(\omega, t)$, ω indicating that the transmitting factor is a function of frequency at any instant. Applying this factor in equation (1) gives:

$$\text{Transmitted } C_v = \sum_{k=1}^{n} r(\omega, t) A_k \cos(kPt + \theta_k). \tag{4}$$

The r factor is placed inside the summation to indicate that as k changes the different frequencies have different values of the multiplying factor r. If a multiplicity of carrier waves is assumed, the transmitting factor would be $r_k(t)$, individual to the kth component.

In normal voiced speech, S_v, these three modulations are all present simultaneously, so that:

$$S_v = s(t) \sum_{k=1}^{n} r(\omega, t) A_k \cos \left[kP \int_0^t p(t)dt + \theta_k \right]. \tag{5}$$

Equation (5) shows how the message in the form of the s, r, and p functions has imprinted its characteristics on the original carrier C_v of equation (1).

The derivation of (5) was for voiced speech. Unvoiced speech, however, is also covered by (5) as a degenerate case. Nevertheless, further information is presented by writing out the unvoiced carrier separately. For unvoiced speech, the frequency P approaches zero and the number of terms, n, approaches infinity, giving an integral instead of a finite sum of components in equations (1) and (5). The unvoiced carrier C_u is then:

$$C_u = \int_{\omega_1}^{\omega_2} A(\omega) \cos [\omega t + \theta(\omega)] d\omega \tag{1'}$$

and the unvoiced speech:

$$S_u = s(t) \int_{\omega_1}^{\omega} r(\omega, t) A(\omega) \cos [\omega t + \theta(\omega)] d\omega \tag{5'}$$

with the continuously variable frequency ω (radians per second) varying over the audible range of energy contribution from ω_1 to ω_2 and the unvoiced carrier spectrum defined by amplitude $A(\omega)$ and phase $\theta(\omega)$. The unvoiced speech has no inflecting factor but does have switching and transmitting factors to make up the message impressed on the carrier.