

## The Vocoder

By HOMER DUDLEY  
Circuit Research Department

AT THE World's Fairs in New York and San Francisco great interest was shown in the speech synthesizer in the Bell System exhibits. Known as the Voder, this device creates spoken sounds and combines them into connected speech. Its raw materials are two complex tones, a hiss and a buzz; selection of one or the other and its intensity and tone quality are controlled by an operator through a keyboard.\*

The Voder is an offshoot of a more extensive system, first demonstrated† in its experimental stage some three years ago. That system analyzed

spoken sounds, and then used the information to control the synthesizing circuit. At the time World's Fair displays were under consideration, so it was naturally perceived that the synthesizer, manually controlled, could be made into a dramatic demonstration. Development was for a while concentrated in that field; as a successful Voder became assured, attention was shifted back to the broader and parent system. Shortly thereafter the system was given the name "Vocoder" because it operates on the principle of deriving voice codes to re-create the speech which it analyzes.

Figure 1 shows the over-all circuit for remaking speech; the analyzer is

at the left and the synthesizer at the right. Electrical speech signals from a microphone are analyzed in the top channel and for a group of channels at the bottom.

In the pitch analysis the fundamental frequency, which for voiced sounds will be called the pitch, is determined by a circuit containing a frequency-discriminating network for obtaining this frequency in a pure form; a frequency counter, by more or less pulses, the current reverses, and a filter for eliminating speech frequencies but a slowly changing current is a direct measure of the pitch of the voiced sounds, whether in speech or the unvoiced sounds. The speech, have insufficient energy to operate the frequency counter, the output current of the pitch discriminator then a pitch-defining signal is a current approximately proportional to the pitch of the voiced sounds and equal to zero for the unvoiced sounds.

There are ten spectral

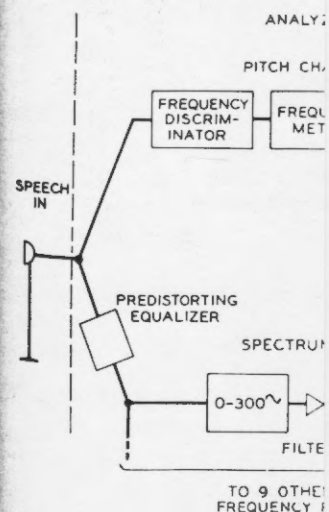
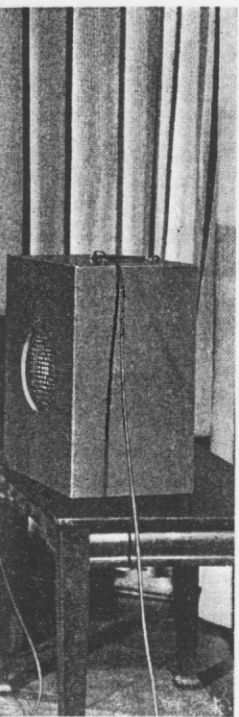


Fig. 1—

\*RECORD, Feb. 1939, p. 170.

†RECORD, Dec. 1938, p. 98.



at the left and the synthesizer at the right. Electrical speech waves from a microphone are analyzed for pitch by the top channel and for spectrum by a group of channels at the bottom.

In the pitch analysis the fundamental frequency, which for simplicity will be called the pitch, is measured by a circuit containing a frequency-discriminating network for obtaining this frequency in reasonably pure form; a frequency meter for counting, by more or less uniform pulses, the current reversals therein; and a filter for eliminating the actual speech frequencies but retaining a slowly changing current that is a direct measure of the pitch. (Unvoiced sounds, whether in whispering or the unvoiced sounds of normal speech, have insufficient power to operate the frequency meter.) The output current of the pitch channel is then a pitch-defining signal with its current approximately proportional to the pitch of the voiced sound and equal to zero for the unvoiced sounds.

There are ten spectrum-analyzing

channels,\* the first handling the frequency range 0-250 cycles and the other nine, the bands, 300 cycles wide, extending from 250 cycles to 2950 cycles, a top frequency which is representative of commercial telephone circuits. Each spectrum-analyzing channel contains the proper band filter followed by a rectifier for measuring the power therein and a 25-cycle low-pass filter for retaining the current indicative of this power but eliminating any of the original speech frequencies.

The operation of the analyzer is illustrated in Figure 2 by a group of oscillograms taken in analyzing the sentence "She saw Mary." To insure that the same speech was analyzed in obtaining the various oscillograms, the sentence was recorded on a high-quality magnetic-tape recorder and reproductions therefrom supplied current to the analyzer. The speech-

\*A 30-channel vocoder covering the wide range of speech frequencies required for high quality has also been built and is being used as a tool in laboratory investigations.

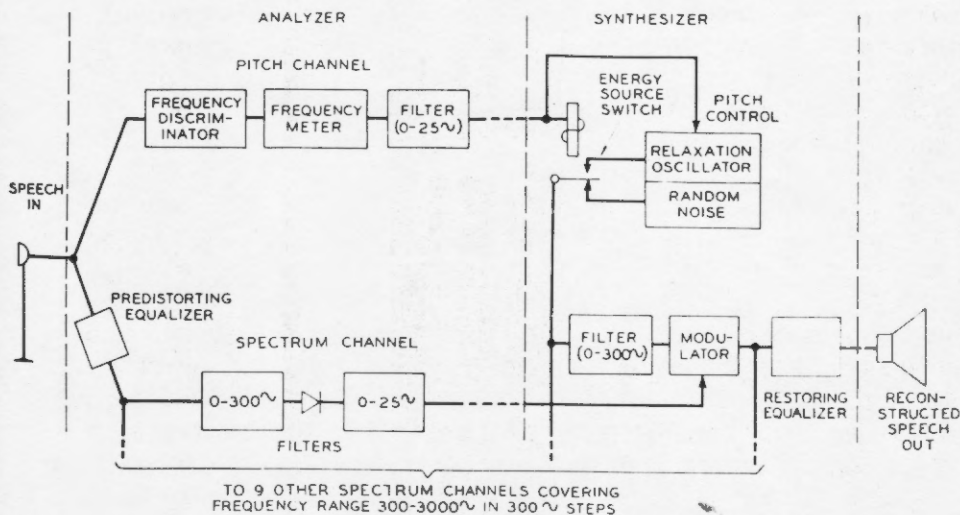


Fig. 1—Simplified schematic of the Vocoder

ds, and then used the in-  
to control the synthesizing  
the time World's Fair dis-  
under consideration, so it  
y perceived that the syn-  
nually controlled, could  
o a dramatic demonstra-  
pment was for a while  
d in that field; as a suc-  
r became assured, atten-  
fted back to the broader  
system. Shortly there-  
stem was given the name  
because it operates on the  
deriving voice codes to  
speech which it analyzes.  
shows the over-all circuit  
g speech; the analyzer is

wave input to the analyzer is shown in the line next to the bottom while the output is shown in the other oscillogram traces; the pitch-defining signal is at the bottom in the figure and the ten spectrum-defining signals in numerical order at the top. For convenient reference the oscillograms are lined up together whereas in the actual circuit the speech-defining signals lag about 17 milliseconds behind the speech-input wave. The inaudible speech-defining output signals contain all the essential speech information as to the input wave, but it is to be noted that they are slow-changing and in this way correspond to lip or tongue motions, as contrasted with the higher audible vibration rates of the rapid-changing speech wave itself. The dropping of the pitch to zero for the unvoiced sounds "sh" and "s" is also readily seen.

Figure 2 gives an idea also as to the synthesizing process. In the analyzer the speech wave is the input and the eleven speech-defining signals are the output; in the synthesizer the eleven speech-defining signals are the input and the speech wave the output.

The steps in speech synthesis are indicated at the right of Figure 1. The relaxation oscillator is the source of the buzz; and the random noise circuit the source of the hiss. The hiss is connected in circuit for unvoiced sounds and for quiet intervals. (In the latter case no sound output from the synthesizer results because there are no currents in the spectrum channels.) When a voiced sound is analyzed a pitch current other than zero is received with the result that the buzz is set for the current pitch by the "pitch control" on the relaxation oscillator; also, the relay marked "energy source switch" operates, switching from the hiss source to the buzz source.

The outputs from the spectrum-analyzing channels are fed to the proper synthesizing spectrum controls with the band filters lined up to correspond. The power derived from the energy sources of the synthesizer in these various bands is then passed through modulators under the control of the spectrum-defining currents. The result is that the power output from the synthesizer is sensibly proportional in each filtered band to that measured by the analyzer in the original speech. From the loudspeaker comes, then, speech approximately the same in pitch and in spectrum as the original. This synthetic speech lags the original speech by about 17 milliseconds due to the inherent delay in electrical circuits of the types used.

In the present models of the Vocoder, control switches have been introduced which permit modifications in the operation of the synthesizer. Through the manipulation of these controls interesting effects are produced. Some of the possibilities of the Vocoder were recently demonstrated by the author and his associate, C. W. Vadersen, before the Acoustical Society of America and before the New York Electrical Society. In those presentations Mr. Vadersen supplied by his own voice the incoming speech which was picked up by a microphone as shown in the headpiece; and at the same time he manipulated the controls to produce desired effects. A remote-control switch was also provided through which, for purposes of comparison, the author could switch the microphone directly to the loudspeaker and so let the audience hear how the speech would sound if it had not been modified by the Vocoder.

In these demonstrations comparison is first made between direct speech

and the best re-creation apparatus could make. The relation of dials and switch modified in various ways speech becomes a throaty buzz. Although the hiss is faint, it is shown to be discrimination as between "and "shirts."

Ordinarily the re-created speech moves up and down with the original. If variation is produced, re-created speech is a monotone or a chant. When the relation of pitch is cut in half, the voice seems dragging; when the swing is normal, the voice seems normal; when four times normal, the voice seems febrile, unnatural. Controls can be reversed so that the pitch comes low: the tune of a song becomes unrecognizable, and speed of the lilting characteristic

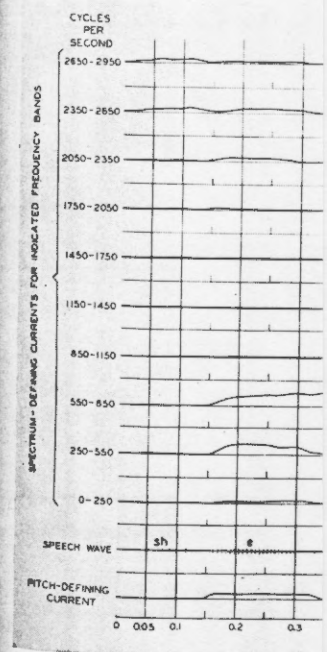


Fig. 2—The original speech wave and the re-created speech wave.



puts from the spectrum-channels are fed to the synthesizing spectrum controls and filters lined up to correct the power derived from the sources of the synthesizer in its bands is then passed through modulators under the control of spectrum-defining currents, so that the power output of the synthesizer is sensibly proportional to each filtered band to that given by the analyzer in the speech. From the loudspeaker, the speech approximately matches the original speech in pitch and in spectrum as well as in timbre. This synthetic speech is about 17% louder than the original speech due to the inherent delay in the circuits of the types used. The present models of the Vocoder have been modified to permit modification of the operation of the synthesizer through the manipulation of several interesting effects are some of the possibilities of which were recently demonstrated by the author and his associate, Vadersen, before the Acoustical Society of America and the New York Electrical Society. Those presentations Mr. Vadersen applied by his own voice to the speech which was picked up by a microphone as shown in the figure and at the same time he manipulated the controls to produce the effects. A remote-control system was also provided through which for purposes of comparison, the speaker could switch the microphone to the loudspeaker and the audience hear how the speech would sound if it had not been processed by the Vocoder.

These demonstrations compared direct speech

and the best re-creation that the apparatus could make. Then by manipulation of dials and switches, speech is modified in various ways. Normal speech becomes a throaty whisper when the hiss is substituted for the buzz. Although the hiss is relatively faint, it is shown to be essential for discrimination as between "church" and "shirts."

Ordinarily the re-created pitch moves up and down with that of the original. If variation is prevented, the re-created speech is a monotone, like a chant. When the relative variation is cut in half, the voice seems flat and dragging; when the swings are twice normal, the voice seems more brilliant; when four times normal it sounds febrile, unnatural. The controls can be reversed so that high becomes low: the tune of a song is then unrecognizable, and speech has some of the lilting characteristics of Scandi-

navian tongues. Another control fixes the basic value of the re-created pitch; if this is "fluttered" by hand, the voice becomes that of an old person. By appropriate setting of the basic pitch, the voice may be anything from a low bass to a high soprano, and several amusing tricks can be performed. In one of these, the basic pitch is set to maintain a constant ratio of 5 to 4 to the original. This is a "major third" higher and harmonizes with the original. In two-part harmony, the demonstrator then sings a duet with himself. Connecting a spare synthesizer set for a 3 to 4 ratio he then sings one part in a trio, the others being taken by his electrical doubles. Finally, with the basic pitch-control of the apparatus, he becomes a father reprimanding his daughter; then the girl herself, and then becomes the grandfather interceding for the youngster.

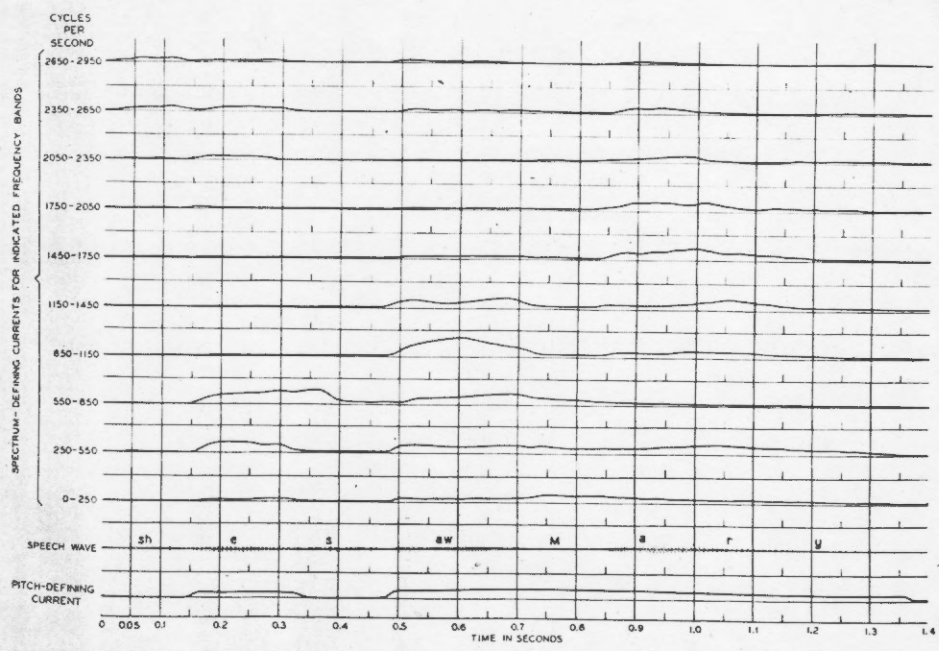


Fig. 2—The original speech wave and an analysis of its components, expressed as the variation of several direct currents

For the vocal-cord tones of the original, the Vocoder substitutes the output of a relaxation oscillator. But any sound rich in harmonics can be used: an automobile horn, an airplane roar, an organ. In some demonstrations, the sound, taken from a phonograph record, replaces the buzz input from the oscillator. Keeping careful time with the puffs of a locomotive, the demonstrator can make the locomotive puff intelligibly "We're - start - ing - slow - ly - faster, faster, faster" as the puffs come closer together. Or a church bell may say "Stop - stop - stop - don't - do - that." A particularly striking effect is that of singing with an organ to supply the tones. Although the words may be spoken, the demonstrator usually sings them to hold the rhythm. It

makes no difference whether his voice is melodious or not; the tonal quality comes only from the musical source.

These tricks and others have suggested uses for the Vocoder in radio and sound pictures. It appears to have possibilities as a tool in the investigation of speech, since by its numerous controls important variables in speech can be isolated for study. The engineering possibilities which may grow out of the application of the principles employed in this device are hard to predict at the present time. The speech-defining currents, however, do have features of simplicity and inaudibility which may open the way to new types of privacy systems or to a reduction in the range required for the transmission of intelligible telephonic speech.

## Contributors to this Issue

J. C. GREENE received the A.B. degree from Harvard in 1921 and spent the following year at Trinity College, Cambridge University, England. He then took two years' graduate study at the Harvard Engineering School, and received the degree of S.B. in Mechanical Engineering in 1924. He at once joined the Technical Staff of the Laboratories, associating with the apparatus analysis group in the Physical Laboratory of the Apparatus Development Department. The following year he transferred to the Equipment Development Department, and for the next ten years was with the current development and toll switchboard equipment groups. In 1935 he transferred to the Toll Facilities Department, where he has since been concerned with development and maintenance problems associated with toll-switching systems.

A. M. ELLIOTT joined the Western Electric Company at Hawthorne in 1923, where he designed equipment and meth-

ods for testing vacuum tubes, thermocouples and switchboard lamps, and later conducted inspection control surveys on carrier and repeater apparatus. In 1925 he transferred to the Installation Department of the Western Electric Company at New York with a group preparing the Handbooks used in testing toll equipment. Two years later he came to the Laboratories as Field Engineer in the Inspection Engineering Department and later was in charge of the Complaint Bureau. In 1938 he transferred to the Toll Facilities Department where he has been concerned with toll switchboards.

W. J. LACERTE joined the Laboratories immediately after receiving his B.S. degree in Electrical Engineering from Kansas University in 1923. For a few years he was engaged in laboratory testing of circuits for dial offices and private branch exchanges, and later conducted studies in the field on trial installations of circuits of these types. Since June of 1937



J. C. Greene

he has been concerned with the development of the crossbar system associated with the first

F. R. STANSEL received the B.S. degree from Union College in 1924, the M.E.E. degree from B.S. in Electrical Engineering from the Technic Institute in 1934. He was assigned to the Whippany Laboratories in 1926 and was later assigned to the Whippany Laboratories where he was engaged in the development of high-power radio transmitters for coast and transatlantic service. He has been associated with the development of oscillators and detection purposes.

HOMER DUDLEY has been associated with the Research Department since his graduation from Penn State where he received the degree of B.S. in Electrical Engineering in 1921. His first work related to the development of transmission features of local telephone systems. Then for several years he supervised work on voice-operated circuits for transatlantic radio, long lines and submarine telephone cables. In 1930 his activities were diverted to research and development problems on coaxial-conductor systems. For the past five years he has been engaged in the design and construction of

ference whether his voice or not; the tonal quality from the musical source. Some and others have suggested for the Vocoder in radio pictures. It appears to be used as a tool in the imitation of speech, since by its controls important variations can be isolated for engineering possibilities. It is now out of the application of the principles employed in this device to predict at the present time the speech-defining currents, which have features of simultaneous audibility which may be used to new types of privacy or to a reduction in the need for the transmission of telephonic speech.

## Issue

vacuum tubes, thermionic switchboard lamps, and later detection control surveys on the heater apparatus. In 1925 he joined the Installation Department of Western Electric Company with a group preparing the manual in testing toll equipments. Later he came to the Western Field Engineer in the Engineering Department and in charge of the Complaint Department. He transferred to the Toll Department where he has been in charge of toll switchboards. He joined the Laboratories after receiving his B.S. in Electrical Engineering from Cornell University in 1923. For a few years he was engaged in laboratory tests for dial offices and private exchanges, and later conducted field work on trial installations of various types. Since June of 1937

December 1939



*J. C. Greene*



*A. M. Elliott*



*W. J. Lacerte*

he has been concerned with field work associated with the first three installations of the crossbar system.

F. R. STANSEL received the B.S. in E.E. degree from Union College in 1926 and the M.E.E. degree from Brooklyn Polytechnic Institute in 1934. He joined the Laboratories in 1926 and until 1936 was assigned to the Whippany Laboratory where he was engaged in the design of high-power radio transmitters for broadcast and transatlantic service. Since 1936 he has been associated with the development of oscillators and detectors for testing purposes.

HOMER DUDLEY has been a member of the Research Department since his graduation from Penn State where he received the degree of B.S. in Electrical Engineering in 1921. His first work related to the transmission features of local telephone systems. Then for several years he supervised work on voice-operated circuits for transatlantic radio, long lines and submarine telephone cables. In 1930 his activities were diverted to research and development problems on coaxial-conductor systems. For the past five years he has been engaged in the design and construction of

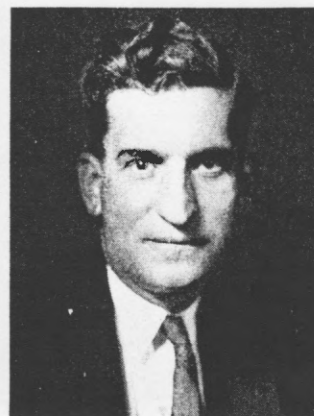
the speech-analyzing and synthesizing circuits which are described in this issue of the RECORD. He received the degree of M.A. in Physics from Columbia in 1924.

H. C. MONTGOMERY received an A.B. at the University of Southern California in 1929 and an M.A. at Columbia in 1933. Since joining the Research Department of the Laboratories in 1929 he has been engaged largely in studies of hearing acuity and related problems in physiological acoustics. More recently his attention has been given to investigations in speech analysis.

J. M. HUDACK joined the Laboratories in 1916 as a member of the Transmission Research Department. Beginning at an early date he assisted in development



*F. R. Stansel*



*Homer Dudley*

December 1939