

May 27, 1941.

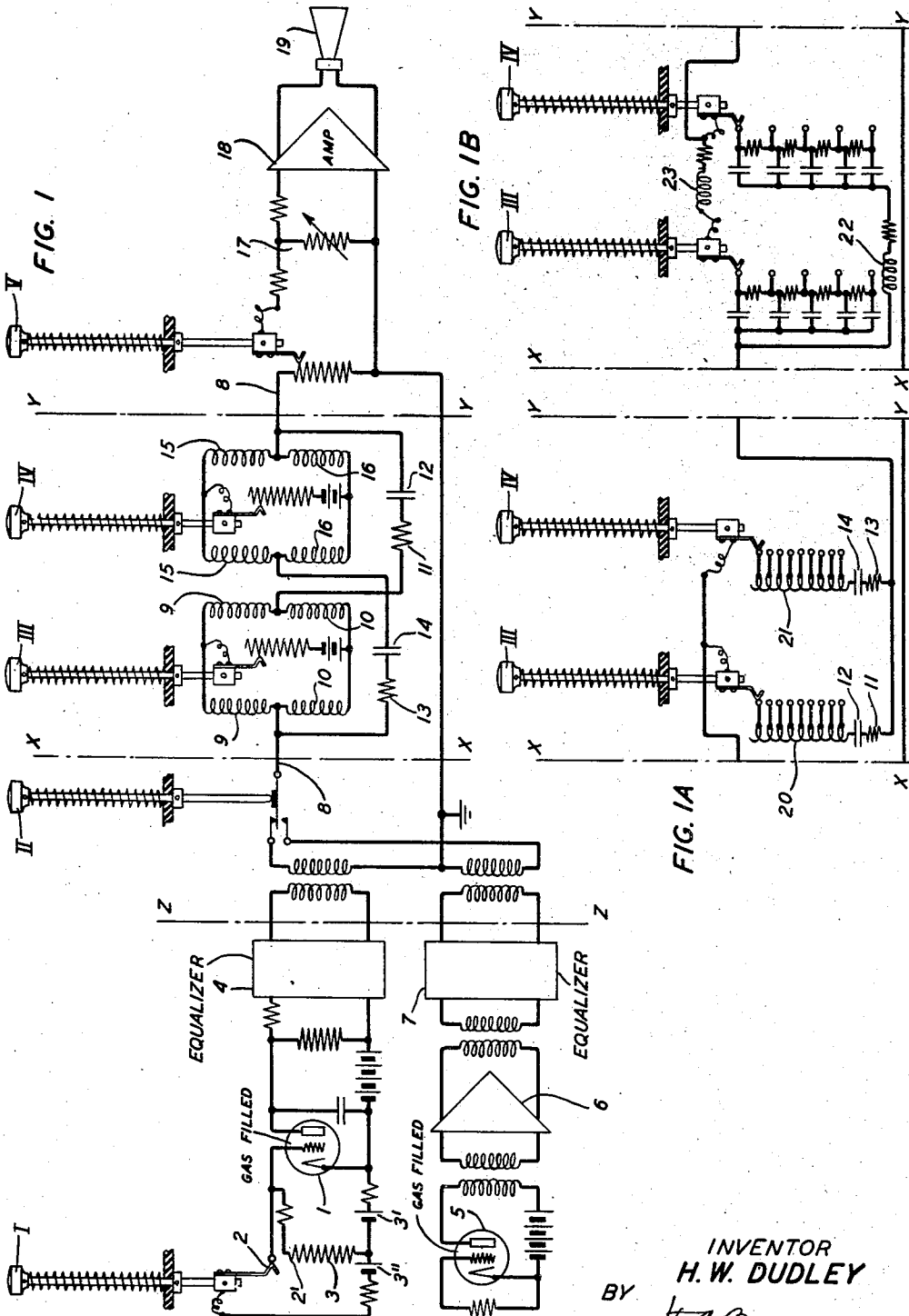
H. W. DUDLEY

2,243,525

PRODUCTION OF ARTIFICIAL SPEECH

Filed March 16, 1940

4 Sheets-Sheet 1



INVENTOR
H. W. DUDLEY
BY *H. A. Rogers*
ATTORNEY

May 27, 1941.

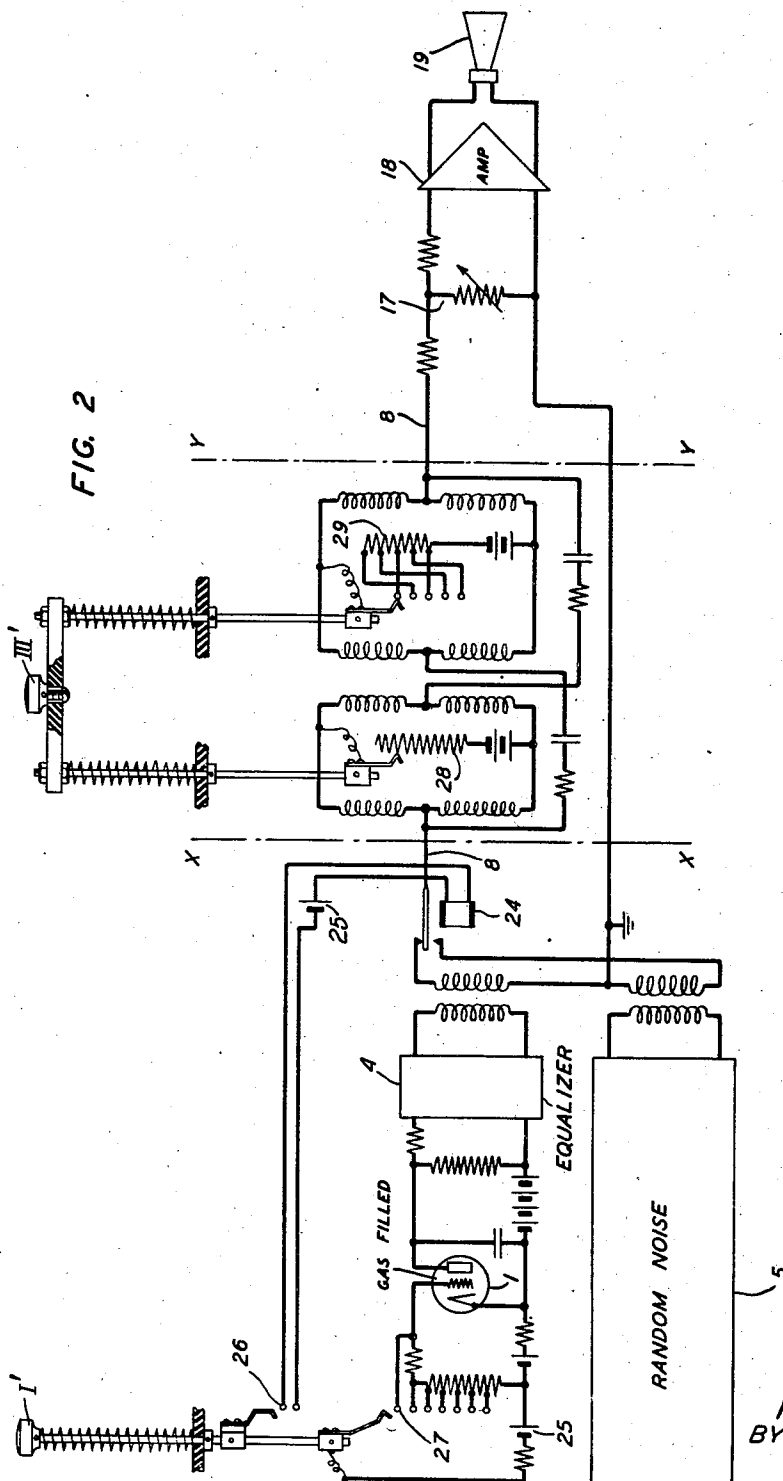
H. W. DUDLEY

2,243,525

PRODUCTION OF ARTIFICIAL SPEECH

Filed March 16, 1940

4 Sheets-Sheet 2



INVENTOR
H. W. DUDLEY
BY *H. A. Burgess*
ATTORNEY

May 27, 1941.

H. W. DUDLEY

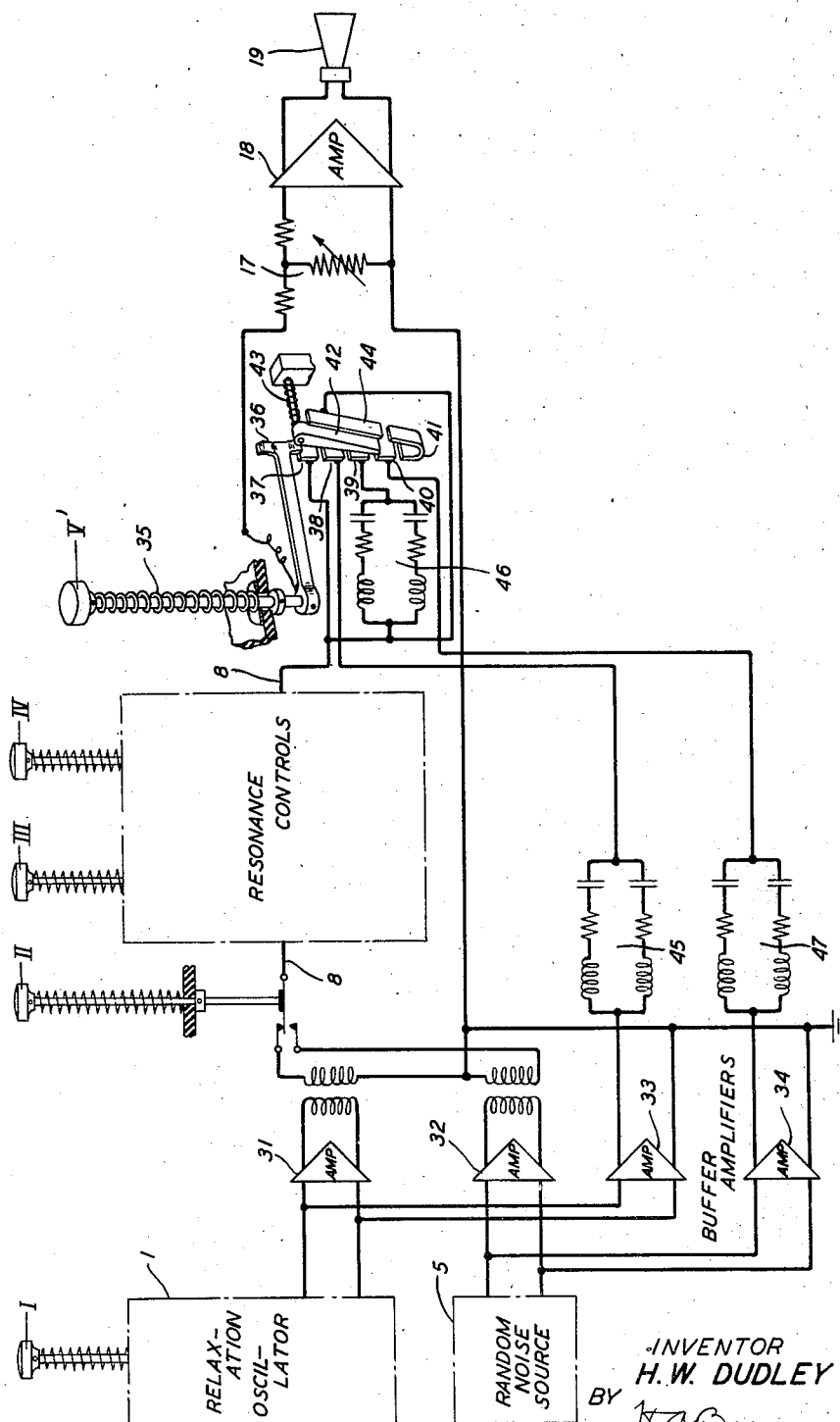
2,243,525

PRODUCTION OF ARTIFICIAL SPEECH

Filed March 16, 1940

4 Sheets-Sheet 3

FIG. 3



INVENTOR
H. W. DUDLEY
BY *H. A. Burger*
ATTORNEY

May 27, 1941.

H. W. DUDLEY

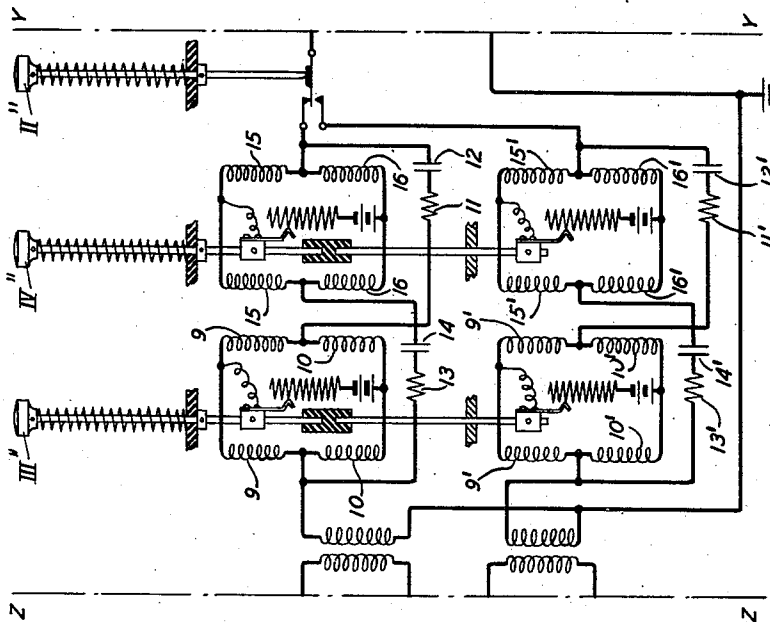
2,243,525

PRODUCTION OF ARTIFICIAL SPEECH

Filed March 16, 1940

4 Sheets-Sheet 4

FIG. 4



INVENTOR
H. W. DUDLEY
BY *H. A. Burgess*
ATTORNEY

UNITED STATES PATENT OFFICE

2,243,525

PRODUCTION OF ARTIFICIAL SPEECH

Homer W. Dudley, Garden City, N. Y., assignor to
Bell Telephone Laboratories, Incorporated, New
York, N. Y., a corporation of New York

Application March 16, 1940, Serial No. 324,286

9 Claims. (Cl. 179-1)

The present invention relates to the artificial production of speech or similar complex waves. The invention has particular reference to, and will be disclosed in connection with, a key-operated or finger-operated mechanism for building up understandable speech waves from waves produced in suitable wave sources and constituting the raw material out of which the speech is to be constructed.

An object of the invention is the artificial production of speech or similar waves in simplified and effective manner.

In my prior Patent 2,121,142, June 21, 1938, I have disclosed a system for artificial speech production from waves of two types, a buzzer-like tone wave and a hiss-like noise wave, which are combined in particular manners under control of keys to build up speech waves. In that system as disclosed a number of narrow band filters, for example, ten filters, are used in the synthesizing operation, these filters having fixed frequency limits and being brought into the circuit in such combinations as needed. Usually only a small number of these filters are needed for any one sound, such as two or three filters, the rest of the group being momentarily unused.

The present invention aims at a simplification in apparatus, especially in the matter of the frequency selective devices used. The present invention achieves such simplification by imitating more closely the process carried out by the human speech mechanism. While the system of my prior patent has operated successfully as an artificial speech producer, the use of a large number of fixed filters which are switched into and out of circuit in groups has no close counterpart in the human speech mechanism. Rather, in human speech production the frequency controls are exercised by the air chamber resonances of the mouth comprising principally two variable resonances, and changes are made mostly by gliding from one sound to another as a result of continuous or gradual, instead of abrupt, changes.

In talking, the size of the total mouth cavity is varied by movement of the lower jaw and the lips while the tongue modifies the mouth resonances in a very important manner by shaping the mouth cavities. For example, the tongue by rising in the middle divides the mouth cavity into two principal chambers with air coupling between them, and the resonances of these two chambers are varied between wide limits by varying the respective air chamber sizes and degree of coupling. The tongue may be drawn back and

the tip raised so as to provide two resonance chambers near the front of the mouth. For some sounds the mouth cavity is used mainly as a single resonance chamber.

While there are many controls exercised in the general ways above indicated, each producing its characteristic sound effect, the number of resonance chambers used for producing any one sound is limited by the nature of the speech mechanism to two principal ones, formed by the action of the tongue in dividing the mouth cavity into two principal portions. Moreover, as indicated above, the positions of jaws, tongue and lips used for producing one sound effect, change by a continuous movement to the positions necessary to produce the next sound effect, resulting in a continuous transition from one sound to another, in a large number of cases.

According to the present invention the conditions of nature are more closely approached by using, in place of the large number of fixed filters disclosed in my prior patent, a small number of resonant circuits, such as two, which are variable and which can be variably controlled to produce sounds that merge, one into the next, as the transition is made from one sound to another.

The nature of the invention and its various objects and features will be made more clear from the following detailed description of particular embodiments as illustrated in the attached drawings.

In the drawings:

Fig. 1 is a schematic circuit diagram of a system for artificially producing speech sounds manually under the control of keys in accordance with one form of the invention;

Figs. 1A and 1B show modifications respectively that may be substituted for the corresponding portion of the circuit of Fig. 1, as indicated by the broken lines;

Fig. 2 is a schematic circuit diagram similar to Fig. 1 but with certain simplifications according to the invention;

Fig. 3 is a diagram generally similar to Fig. 1 but showing a modified system according to the invention; and

Fig. 4 is a fragmentary circuit schematic showing a modification that may be substituted in the system of Fig. 1, as indicated by the broken lines.

Referring to Fig. 1, a system for producing artificial speech is shown in which five keys or controls are used, numbered I, II, III, IV and V in the figure. Key I is the pitch control key, key II is the key for selecting between the type of energy

source ("buzz" or "hiss"), keys III and IV are the resonance controls and key V is the stop consonant key.

The source of energy simulating the vocal cords and producing the buzz or complex tone wave is shown as a relaxation oscillator comprising gas-filled tube 1 and associated circuits. This may be of the type shown in United States patent to R. R. Fiesz 2,183,248, issued December 12, 1939. It produces a wave rich in harmonics, the fundamental of which is readily variable by controlling the negative grid bias. Equalizer 4 reduces all of the harmonics to the same amplitude or is designed to produce some other desired relation between them. When key I is in its upper position it makes contact at 2, giving a bias to the grid equal to the sum of the voltages of batteries 3' and 3'', which is great enough to prevent production of oscillations. As key I is depressed, contact is made at 2'. The two parts of the resistance 3 and the resistance adjacent 3'' act like a potentiometer so that when contact is made at 2' a lesser part of the voltage of battery 3'' is effective on the grid and the relaxation oscillator starts to oscillate at a low fundamental frequency. As the key is depressed further, a proportionately smaller voltage is applied from source 3'' to the grid and the fundamental frequency of the oscillator is caused to increase.

The source of continuous spectrum or random frequency noise is shown as comprising gas tube 5, followed by amplifier 6, and equalizer 7 for making the energy distribution constant over the utilized frequency band or otherwise suitably shaping it.

Energy source key II in its upper position connects oscillator 1 to the conductor 8 and in its lower position it connects the random noise source to conductor 8. The other side of the respective source is shown grounded. For simplicity of wiring, Fig. 1 shows use of a ground return although the grounds could be replaced by a conductor, if desired.

The function of keys III and IV is to control the tuning of the two selective circuit branches shown included in conductor 8. The resonant circuit controlled by key III includes inductances 9 and 10, resistance 11 and capacity 12. The resonant circuit controlled by key IV includes resistance 13, capacity 14 and inductances 15 and 16. While the resonances of these circuits might be controlled in various ways the particular method exemplified in this figure is by varying the saturating current flowing in the windings, which have saturable cores. As either key is moved, it actuates a slider over a resistor as shown, which regulates the battery current flowing through the coils 9, 10 or 15, 16 as the case may be. This current varies the point of operation on the permeability curve.

Key V when in its top position transmits the waves at full volume to the variable volume control pad 17 and amplifier 18 leading to loudspeaker 19 or other output such as a line. When key V is depressed it moves its slider along the shunt resistor shown and this governs the volume similarly to a potentiometer. The decreased volume at the start of producing stop consonants due to closing the vocal tract is imitated here by a corresponding rate of decrease of volume upon depression of key V. The following sudden burst of energy upon opening the constriction of the vocal tract is simulated here by the quick release of the key V.

In considering the operation of the circuit of Fig. 1, it will be convenient to have certain frequency relations in mind. It will be assumed for illustration that the random noise covers the human speech band and that for production of normal speech effects the relaxation oscillator circuit can have its fundamental frequency varied between 40 and 800 cycles, these being not limiting but typical values. Assuming that key III adjusts over a lower range of frequencies and key IV over a higher range, the range of adjustment effected by key III may be from 200 cycles up to a frequency of 1,000 cycles, while key IV may effect resonance adjustments at from 1,000 cycles to 5,000 cycles, again using these values as illustrative and in no wise as limiting.

Moreover, the resonance effect produced by either key III or IV is preferably not a sharp resonance and resistances 11 and 13 may be proportioned to give the desired degree of flatness even to the possibility of having them vary with the inductances under control of extra arms on keys III and IV not shown. To give illustrative figures, when circuit 9, 10, 11, 12 is adjusted to resonance at 1,000 cycles, it may give 6 decibels greater attenuation at 250 cycles away from resonance than it has at the resonant frequency; and when circuit 13, 14, 15, 16 is adjusted to resonance at 3,000 cycles it may give 6 decibels greater attenuation at 450 cycles away from resonance than at the resonant frequency. These values are not critical and in practice can be varied widely to produce desired effects.

With these general values of frequencies in mind, it is seen that manipulation of key I produces rising and falling inflection and determines the normal talking fundamental, while keys III and IV mold the output of either of the two energy sources, as chosen by key II, into recognizable sounds similarly to the way in which the human mouth molds the vocal cord waves and unvoiced waves into speech sounds. It is necessary, of course, to coordinate the movements of these keys for best effects and these can best be determined by trial. Stop consonant effects are introduced by sharp, quick movements of key V.

While in Fig. 1, as well as in the figures to follow, the use of only two resonance controls is shown it is within the invention to use more than two resonance controls, and additional ones may be provided by substantial duplication of the ones shown. The variation of inductance can be made to follow any desired law within considerable limits by use of cores in the different inductances saturating at different rates or in different directions as disclosed more fully in my copending application Serial No. 324,288 filed March 16, 1940.

In Figs. 1A and 1B different types of resonance controls are shown which may be substituted for the type shown in Fig. 1. Key III in Fig. 1A moves a sliding contact along inductive winding 20, while key IV similarly moves its slider along inductive winding 21. The slider may move from turn to turn or at any variation rate desired. As shown taps are brought out at desired intervals to a row of fixed contacts over which the slider moves.

In the form of resonance control shown in Fig. 1B the inductance remains fixed (shown at 22, 23) and the capacity is varied by movement of sliding key contact over taps. Resistance leaks are shown connected across the capacity units. These prevent storage of energy with re-

sultant clicks as the contacts are made or broken. The resistances also give some damping of the circuit resonance.

Fig. 2 illustrates a system in which a simplification in the number of controls has been made. In this figure the sound wave sources 1 and 5 and their associated circuit elements may be the same as previously described. Key II of Fig. 1 has been replaced by a relay 24 adapted to be energized from battery 25 whenever key I' bridges over contacts 26, which is its first operative position, the topmost position being "idle," corresponding to such a large bias on the grid that no oscillations are produced. When relay 24 is energized its armature closes its lower contact and connects the random noise source to conductor 8. Further downward movement of key I' releases relay 24 to connect the tone source 1 to conductor 8 and such further downward movement of key I' also results in controlling the fundamental pitch as previously described.

It is possible to omit the explosive consonant control also. A sudden oncoming of energy of the right frequency distribution will produce a good imitation of the explosive sound desired. This is accomplished in the circuit of Fig. 2 by providing a contact 27 in the path on the slider of key I'. When contact is made at 27 by the rising key I' a silent condition results which can be terminated by lowering the key again resulting in energy changes such as might occur in the sound *g* of the word "ago." Similarly, the "stop" position may be preceded or followed by the random noise source connection above it. By the proper sequence of key I' positions and of adjustment of resonance by key III' any of the stop consonants can be produced.

In general, the resonance controls are not independent. When a sound has a certain resonance in the lower frequency range, it will have an associated resonance in the upper frequency range. This is more or less fundamental in the way the air cavity of the mouth is divided into two parts for producing resonant effects. In any case, a certain lower resonance and a certain upper resonance must go together for a particular sound. Because of these facts there is associated with each resonance A, a single corresponding resonance B for the optimum production of a particular sound. Because of this, it is possible to combine these two resonant controls into one multiple resonance control.

This has been done in Fig. 2 by providing a single key III' with two sliders operating over resistances 28 and 29, respectively, for controlling the direct current in the inductive windings of the respective resonant branches, similarly to Fig. 1. By strapping variously the contacts to points along resistor 29, any type of variation such as increasing or decreasing at various rates can be secured by continuous movement of the slider uniformly in the same direction. The remainder of the system from this point on to the final output may be the same as in Fig. 1. Instead of the particular type of resonance controls shown, other types could be used such as those shown in Figs. 1A and 1B with suitable mechanical linkage between the two sliders and the single control key.

Fig. 3 shows the same system as Fig. 1 except as modified to include a more elaborate stop consonant control. The keys I and II and the two wave energy sources may be as in Fig. 1, and the resonance controls including keys III and IV may be as in Fig. 1, Fig. 1A or Fig. 1B. Branched

outputs from the energy sources are shown, each including a "buffer" amplifier 31, 32, 33 or 34 to isolate the two separate output branches of the same oscillator and prevent undesired reactions.

The stop consonant key V' is similar in construction to that shown in Fig. 38 of my copending application for United States patent Serial No. 181,275, filed December 23, 1937, and is arranged to provide for the making of a series of circuit closures in given sequence and returned to normal position without operating the contacts in reverse order, if such action is desired. The key shank 35 is linked to the slider 36 which in its downward travel engages in succession contacts 37, 38, 39, 40 and 41. If the key is not allowed to rise until it has been depressed the full limit, its end 36 passes under the lower end of guide trigger 42 which is urged to the left by spring 43, so that on its upward stroke the end 36 passes on the right-hand side of member 42 and makes contact with fixed member 44 on such upward stroke. The manner in which it is used in the figure is illustrative of one type of control. In its rest position as shown, wiper 36 connects conductor 8 and contact 37 directly to the outgoing circuit including volume adjusting network 17, amplifier 18 and loud-speaker 19. When moved into engagement with contact 38, wiper 36 connects the output of the relaxation oscillator 1 through buffer amplifier 33 and network 45 to the outgoing circuit. Thus network 45 is substituted for the variable resonance control circuits and by constructing network 45 to have definite characteristics the desired resonance effect is produced when this is brought into the circuit. When moved to engage contact 39, wiper 36 brings network 46 into circuit between the output of the resonance control circuits and the outgoing circuit. Network 46 may be constructed to give its individual characteristic shaping to the waves. When in contact with member 40, wiper 36 connects the output of the random noise source 5 through buffer amplifier 34 and shaping network 47 to the outgoing circuit. This may be used to give a pronounced sibilant or hiss or aspirant or other unvoiced effect and the resonance of network 47 may be chosen to select the particular part of the spectrum to give the desired effect. If depressed only part way the key V' retraces its downward path on returning, but if depressed to the full extent it returns wiper 36 on the opposite side of catch 42 so that 36 makes contact with spring 44 which, as shown, has the same electrical effect as contact between 35 and 37. By variously wiring the contacts of key V', various effects are obtainable, those described being illustrative of some. Moreover, if desired, a plurality of separate keys like key V' may be used in parallel and so wired that each key corresponds to a particular consonant or group of consonants.

In Fig. 4, the circuit of which is to be substituted in Fig. 1 between lines *z-z* and *y-y*, separate resonance control circuits are shown for voice and unvoiced sounds with common control keys III'' and IV''. In this arrangement energy source selecting key II'' is placed after the control keys.

Each resonance control key carries two sliders which operate over separate inductance controls in respective circuits. The resonant branches 9, 10, 11, 12 and 13, 14, 15 and 16 may be the same as in Fig. 1 and controlled in exactly similar manner, as shown. In the present figure, however, these are used only with respect to the

energy from the relaxation oscillator. The energy from the random noise source is similarly controlled by resonant branches 9', 10', 11', 12', and 13', 14', 15', 16' which may be constructed to have resonant properties similar to or, preferably, different from the resonant circuits for the voiced energy. Since the unvoiced sounds usually employ higher frequencies than the voiced sounds, the resonant branches indicated by primed numerals would be resonant to higher frequencies than the others. Such an arrangement makes for ease in manipulation, for a key can be made to correspond, in one position, to one frequency condition in the branch carrying waves representing voiced sounds and to some different frequency condition in the consonant sound branches, so that both of these resonant conditions can be obtained without movement of the control key in question. Thus the transition from voiced to unvoiced sounds and vice versa can be made with less extent of movement of the control keys in many instances.

What is claimed is:

1. In a system for producing artificial speech, a source of electrical wave energy of frequency spectrum resembling vocal cord sounds, a source of electrical wave energy of continuous spectrum for simulating unvoiced sounds, means to select wave energy from either source, means to control variably the waves from the selected source comprising variable resonant circuits simulating the effect of the different resonant cavities of the human mouth.

2. In a system for producing artificial speech, electrical wave generating means for producing electrical wave energy having a spectrum distribution resembling that of vocal cord sounds and other electrical wave energy of continuous spectrum, a switch for selecting either of said types of wave energy at will, a shaping circuit comprising a pair of resonant circuits capable of having their resonances separately varied over the major part of the speech frequency range, individual resonance controls for said circuits, means to apply the wave energy selected by said switch to said resonant circuits, and a sound producer actuated by the energy transmitted through said resonant circuits.

3. In mechanism for producing artificial speech and similar effects, a controlling circuit, means producing waves consisting of a fundamental and a plurality of harmonic components in the speech frequency range, means producing other waves of random frequency distribution within the speech frequency range, selecting means for impressing either of said kinds of waves upon said controlling circuit, said controlling circuit comprising frequency selective branches, each branch being variable as to its frequency transmission range, means to control said selecting means and to vary the frequency transmission ranges of said branches at the muscular rates of

movement of the human speech mechanism, and means to translate the energy in said controlling circuit into sounds simulating speech.

4. A system for the artificial production of speech waves comprising means for generating waves covering the speech frequency range and comprising a fundamental frequency and harmonics thereof, means for generating waves of random frequency distribution, means for producing sound from said waves in sequences simulating the sound sequences in speech and a resonant system connected to said generating and sound producing means having variable resonance means for emphasizing sounds in different portions of the speech frequency band.

5. In a system for producing artificial speech, a source of electrical waves comprising a fundamental frequency and harmonics thereof, a source of waves of random frequency distribution, means for translating waves from either source into sound, a resonant system for transferring waves from said sources to said translating means comprising circuit branches of adjustable resonance, and means for varying the resonant frequencies of said branches.

6. The combination according to claim 5, including manually operated keys for selecting the one of said wave sources to be used and for varying the resonance of said circuit branches.

7. The method of producing speech artificially comprising generating electrical waves simulating the vocal cord vibrations and generating other electrical waves simulating unvoiced sounds, reproducing sounds from said waves and molding the waves which produce said sounds by producing resonance effects upon the electrical waves simulating the action of the resonant air chambers of the mouth and producing continuous variations of said resonance effects to make transitions from one sound to another.

8. In a system for the artificial production of speech, a relaxation oscillator, a random noise source, means for translating electrical waves into sound waves, variable resonance circuits adapted to connect said sources to said translating means and manually operated means for varying the resonant frequencies of said resonance circuits.

9. In a system for the artificial production of speech, a relaxation oscillator, a random noise source, means for translating electrical waves into sound waves, variable resonance circuits adapted to connect said sources to said translating means, manually operated means for varying the resonant frequencies of said resonance circuits and a stop consonant key having a movable contacting element movable over a succession of stationary contacts for connecting the input of said sound reproducing means in rapid succession to a succession of shaping networks.

HOMER W. DUDLEY.