July 16, 1946. W. KOENIG, JR 2,403,985

SOUND REPRODUCTION

Filed April 3, 1945 2 Sheets-Sheet 1

FIG. 1

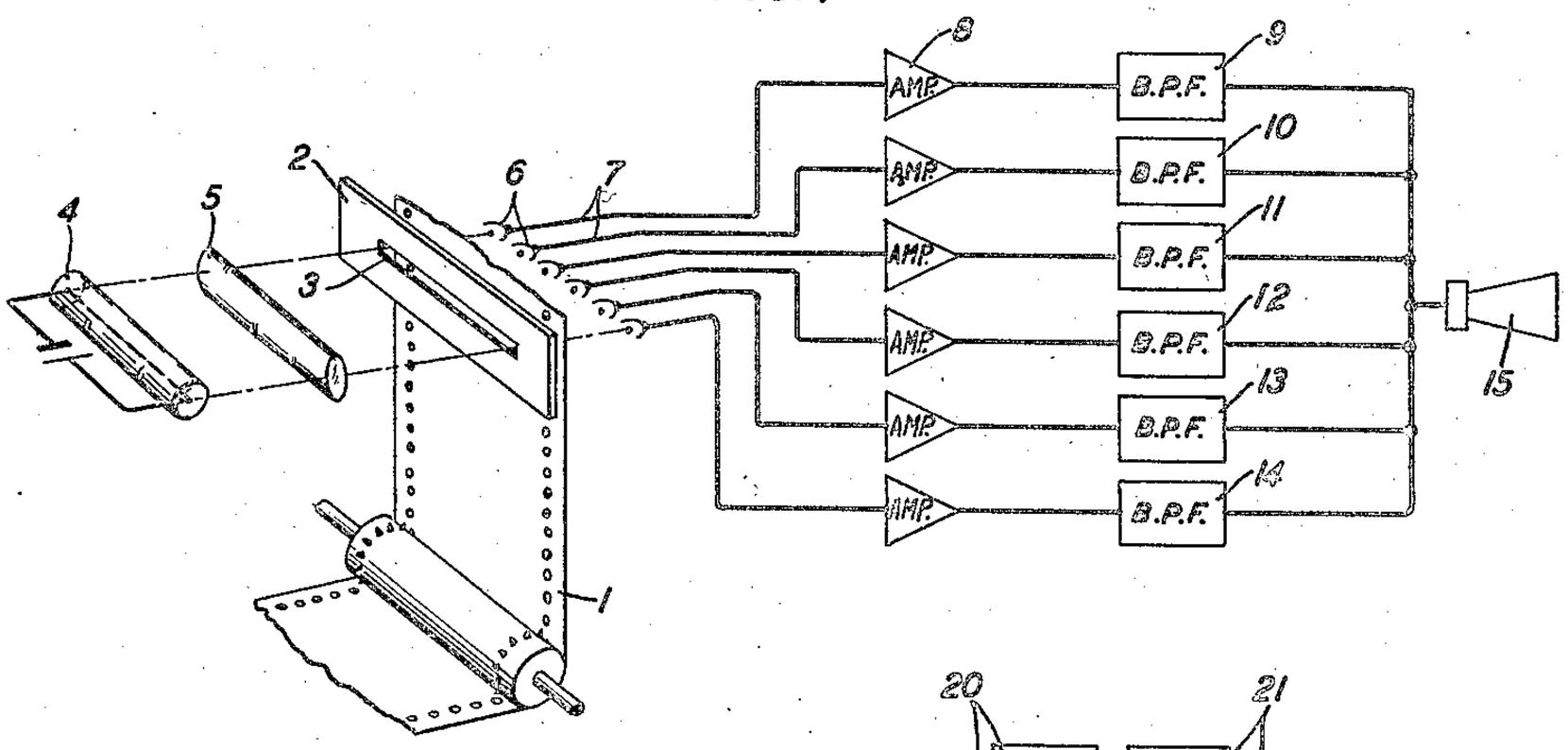
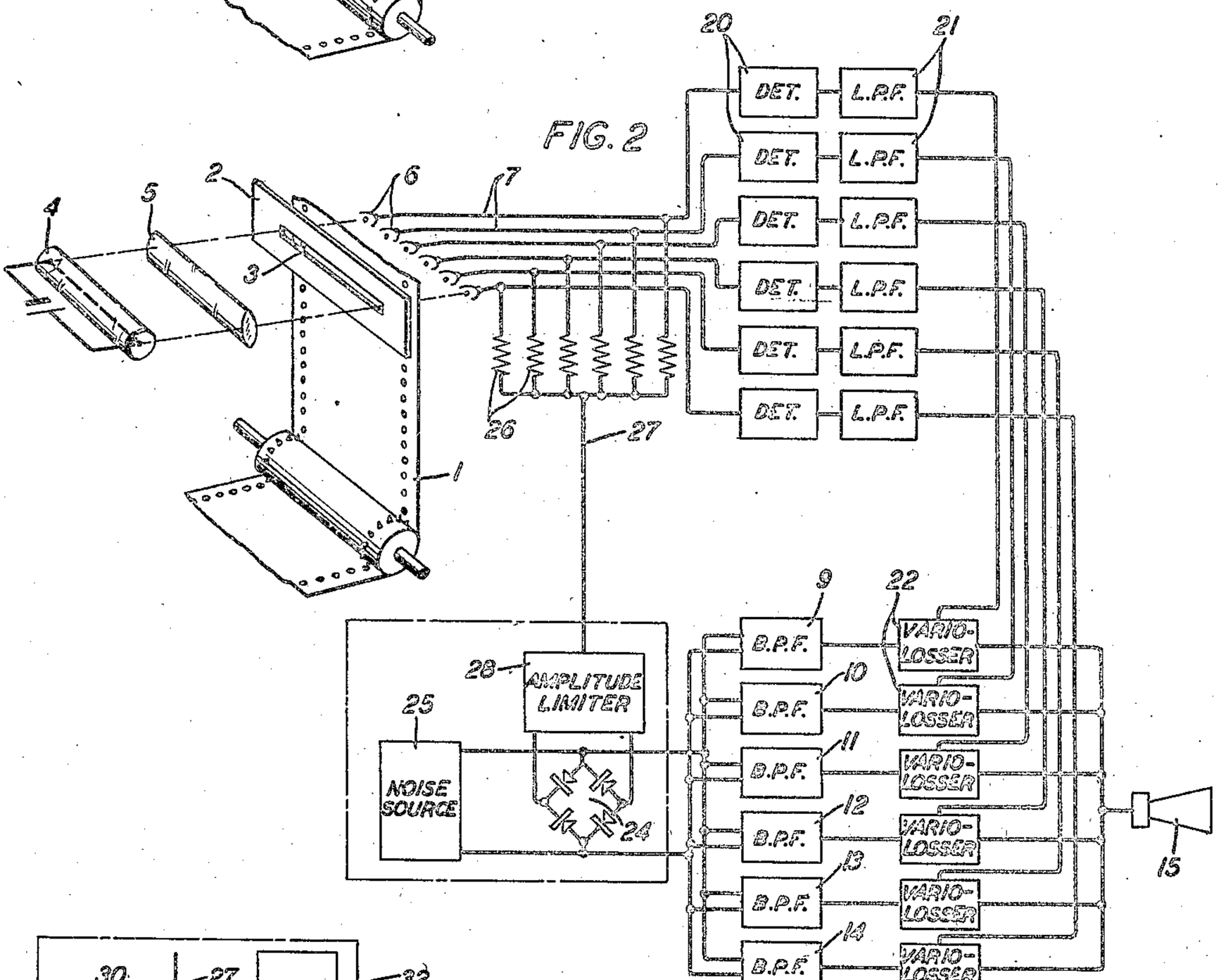


FIG. 2



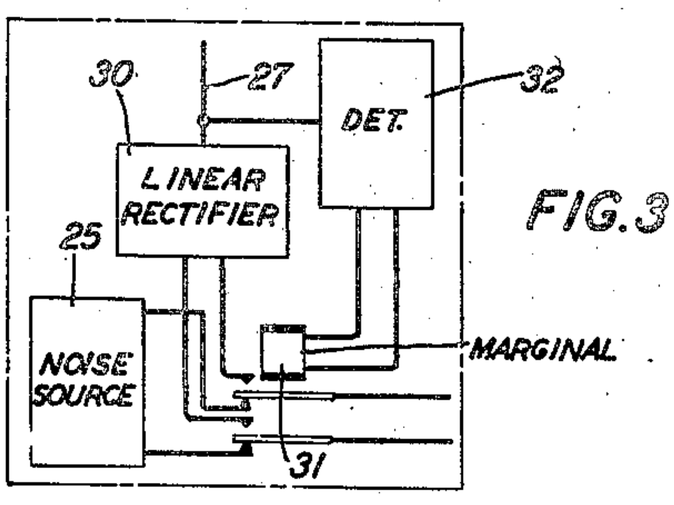


FIG. 3

INVENTOR
W. KOENIG, JR.

BY N. D. Ewing

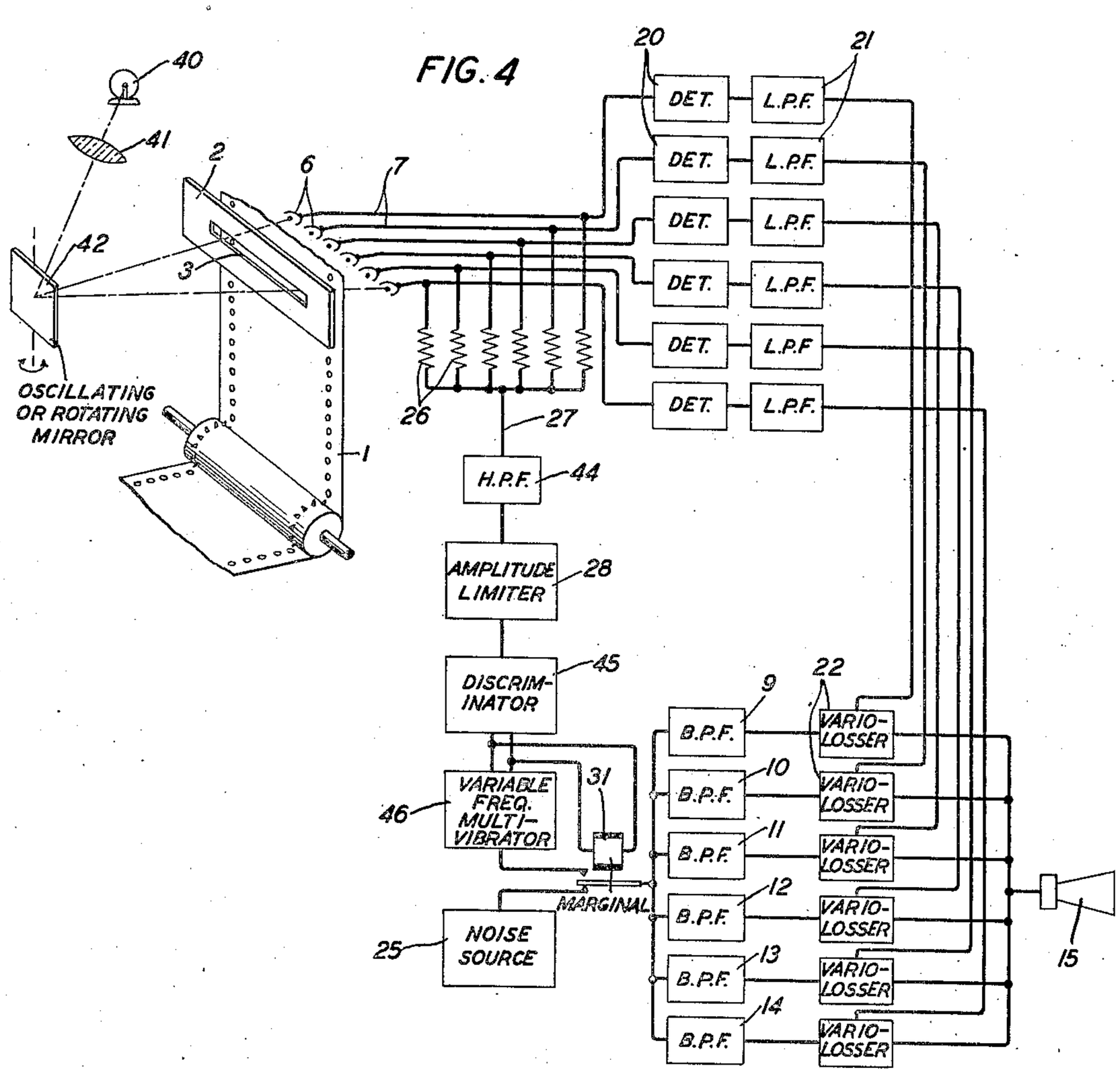
ATTORNEY

July 16, 1946. W. KOENIG, JR. 2,403,985

SOUND REPRODUCTION

Filed April 3, 1945 2 Sheets-Sheet 2.

FIG. 4



INVENTOR
W. KOENIG, JR.

BY N. D. Ewing

ATTORNEY

Patented July 16, 1946

2,403,985

UNITED STATES PATENT OFFICE

2,403,985

SOUND REPRODUCTION

Walter Koenig, Jr., Clifton, N. J., assignor to Bell Telephone Laboratories, Incorporated, New York, N. Y., a corporation of New York

Application April 3, 1945, Serial No. 586,310

13 Claims. (Cl. 179—1)

1

This invention relates to the synthesis of complex sound waves represented in a spectrographic recording, and more particularly to the reproduction of speech waves from a speech spectrogram.

For the recordation of complex sound waves, such as speech waves, it has been proposed heretofore to assign the several component frequency bands to respectively corresponding collateral lines or strips extending longitudinally of a record surface, and to vary the density or darkness of the recording along such lines or strips in conformity with the time variations in the envelope amplitude, or effective intensity, of the wave components appearing in the respectively corresponding frequency bands. The manner in which the total wave power is distributed across the frequency range at any time is indicated directly by the manner in which the density or darkness of the recording varies across the record surface at a corresponding point along its length. A record of this kind is herein designated a sound spectrogram or, specifically, a speech spectrogram. It is to be noted that the sound spectrogram, unlike other sound records, does not contain a record of the variations in instantaneous amplitude of either the complex sound waves or the components thereof.

A principal object of the present invention is to provide improved and simplified methods and means for reproducing from a sound spectrogram, and more particularly from a speech spectrogram, the sound waves represented therein. Another object is to improve the clarity and naturalness with which voiced sounds in general, and inflected sounds in particular, are reproduced from a speech spectrogram.

For the purposes of the present invention speech spectrograms may be divided into two classes, narrow-band and wide-band. In the first class the aforementioned component frequency bands are each so narrow as to embrace only one harmonic of the fundamental voice frequency, and the frequency definition of the spectrogram is accordingly great enough that the several harmonics comprising a vowel sound appear as distinct bars or striations. In the second class the component frequency bands are each wide enough to embrace at least two successive harmonics of the fundamental voice frequency, and only the broad resonance regions defined by the vocal cavities and not the individual harmonics of voiced sounds are represented in the spectrogram. This second class of speech spectrogram is characterized further by regularly spaced transverse

2

striations that are due to the concurrent recordation of two or more successive harmonic components of voiced sounds along each longitudinal line or strip. Methods and means for producing these two classes of speech spectrogram have been disclosed heretofore, as for example in my copending application Serial No. 568,880, filed December 19, 1944, and in that of R. K. Potter, Serial No. 586,769, filed April 5, 1945.

In embodiments of the invention hereinafter described in detail electric waves having a multiplicity of different frequency components corresponding to those found in speech sounds are generated and the different components are concurrently varied in strength in conformity with and under the control of the density or darkness variations appearing along corresponding different parts of a speech spectrogram. These varying components are applied concurrently to a loudspeaker or the like to generate corresponding sound waves. In certain cases the reproduced sound waves have the quality of unvoiced or whispered speech; and in other cases the quality is more nearly that of normal speech, the synthesized vowels and other voiced sounds having the multiplicity of harmonically related tones that is characteristic of such speech sounds.

In accordance with a feature of the invention the transverse striations that appear in the wide-band spectrogram and the generally longitudinal striations that appear in the narrow-band spectogram are utilized to generate or to control the generation of the aforesaid components of different frequency for the synthesis of voiced sounds. More especially, use is made of the fact that the spacing of the striations is definitely related to the fundamental voice frequency of the recorded speech waves, that is, to the fundamental frequency of vibration of the vocal cords. The latter frequency, it should be appreciated, varies continually in normal inflected speech, and the spacing of the striations likewise varies continually as an inverse function thereof.

The nature of the present invention and its various features, objects and advantages will appear more fully upon consideration of the embodiments illustrated in the accompanying drawings and the following description thereof. In the drawings, Figs. 1, 2 and 3 illustrate embodiments of the invention in which the transverse striations of a wide-band speech spectrogram are utilized; and Fig. 4 illustrates an embodiment utilizing a narrow-band speech spectrogram.

Referring to Fig. 1, there is shown diagrammatically a simple system in accordance with the

invention for reproducing or synthesizing speech waves that are recorded on film in the form of a wide-band spectrogram. The film 1 is arranged to be drawn at constant speed past an optical slit which is symbolized by a mask 2 that has an elongated aperture or slit 3 extending across the spectrogram substantially parallel with the transverse striations that appear therein. By means of an optical system represented by an incandescent lamp 4 and a condensing lens 5, a wide beam of light is passed through the slit 3, and through the portion of film exposed therein, to a bank of photoelectric cells 6. The latter are optically shielded from each other and aligned with the slit 3 to receive the light passing through respectively corresponding different portions of the slit and film. Each photocell 6 is identified with a definite speech frequency band, viz., the band embracing all of the speech wave components that are recorded in the portion of spectrogram through which the particular photocell is illuminated.

The fluctuation in the quantity of light incident on any photocell 6 due to the movement of film 1 gives rise to corresponding electrical current fluctuations or waves in its individual output circuit 7. The waves in the several circuits 7 are passed through individual amplifiers 8 and through individual different band-pass filters 9 to 14 to a loudspeaker 15 or other electroacoustic transducer. Each of the filters 9 to 14 is designed to selectively transmit any applied wave components that lie within the particular speech frequency band with which its associated photocell 6 is identified. Although six photocells and six filters are illustrated by way of example, it is contemplated that many more may be employed if desired.

The spectrogram on film 1 is a photographic negative of the usual type of spectrogram recorded on facsimile paper, that is, the greater the envelope amplitude of a recorded wave component the lesser is the opacity of the corresponding portion of the film. Sections of the film that represent pauses between words are accordingly of uniform opacity, and they may be quite opaque; in either case the light, if any, reaching the photocells 6 is unmodulated and no sound is produced by loudspeaker 15. For best results the relative degree of modulation, or variation in opacity, appearing along any longitudinal line or strip in the spectrogram should be proportional to the relative envelope amplitude of the component recorded therein. Slit 3 is made fine enough to resolve the structure of hiss sounds as they appear in the spectrogram.

In considering the operation of the Fig. 1 system, assume first that the spectrographic record of a vowel sound is being scanned by the electro-optical elements. In such case the aforementioned transverse striations are present, and as they pass the slit 3 they modulate the light transmitted to the several photocells, that is, they cause the quantity of transmitted light to vary periodically at a rate depending on the spacing of the striations and the rate of movement of the film. The spacing of the striations is inversely proportional to the fundamental voice frequency represented in the vowel being scanned and hence, if the film is advanced at the proper rate, the modulation frequency will coincide with the fundamental voice frequency and vary as the latter varies. The modulated light gives rise in the affected photocells 6 and their connected circuits 7, to electric current or wave components of the

same fundamental frequency, or pitch, and also to a multiplicity of components harmonically related to the fundamental.

Each of the filters 9 to 14 freely transmits only those generated components that lie within its pass-band. Thus, filter 9 may pass the fundamental and one or more of the lowermost harmonics, if the associated photocell 6 embraces the portion of the spectrogram in which are recorded the fundamental component and the corresponding one or more harmonics of the recorded speech waves. Likewise, filter 10 passes the next higher group of generated harmonics, and the successively higher groups of harmonics are passed through respective filters 11, 12, etc.

The intensity or envelope amplitude of the components transmitted through any one filter varies with the degree of modulation appearing in the portion of the spectrogram identified with that filter, and the degree of modulation is in turn more or less proportional to the envelope amplitude of the speech wave components recorded in that portion of the spectrogram. The wave output of the entire bank of filters therefore comprises a multiplicity of harmonically related components, which may coincide exactly in frequency with corresponding components of the recorded speech waves, and which vary in envelope amplitude in approximate conformity with the variations in envelope amplitude of the respectively corresponding recorded components. The sound produced by loudspeaker 15 accordingly simulates the recorded vowel sound.

When the recording of a hiss sound or other unvoiced consonant is being scanned, the transverse striations are absent but inasmuch as the slit 3 is fine enough to resolve the closely spaced density variations characteristic of such sounds, noise currents are produced in the various circuits 7. The noise currents in any circuit 7 comprise an indefinitely large number of components of different frequency with the power distributed more or less continuously over a wide frequency range, and the intensity of all these generated components varies in substantial conformity with the variations in the intensity of the recorded components identified with the associated photocell 6. The associated filter, as before, selects the generated components that are to pass to the loudspeaker 15.

When voiced consonants appear in the portion of spectrogram being scanned, there may be produced in the circuits 7 both the aforementioned noise currents and the harmonically related components due to the transverse striations. Whatever the character of the recorded speech sound, then, each of the photocells generates a multiplicity of noise components, or of harmonically related components, or both, if and so long as speech components appear in the associated portion of the spectrogram. The filter connected to each photocell suppresses all of the generated components excepting those lying within the frequency band identified with the particular photocell; and the intensity of the transmitted components varies approximately in conformity with the variations in intensity of the corresponding components of the recorded speech sound.

The system illustrated in Fig. 2 comprises elements 1 to 7 of Fig. 1 arranged in the manner described hereinbefore, and like the Fig. 1 system it is adapted for reproduction from wide-band speech spectrograms. The optical slit 3 may be made somewhat wider in this case for a sepa-

rate source of noise currents is provided at 25. Corresponding elements in the several figures are assigned the same reference numbers.

Each of the circuits 7 in Fig. 2 includes an individual detector 20 followed by a low-pass filter 21, which together function to produce at the output terminals of each filter 21 a unidirectional control voltage that fluctuates in conformity with the variations in envelope amplitude that are recorded in the respectively corresponding portion of the spectrogram. These control voltages are applied to individual vario-lossers 22 which are interposed between the bank of filters 9 to 14 and the loudspeaker 15 and which vary the transmission loss or gain in the several band-pass filter circuits in conformity with the variations in the respectively corresponding control voltages.

Noise current source 25, which may be a thermal noise generator, is connected to the input terminals of all of the filters 9 to 14 through a balanced modulator 24 which in its balanced condition allows the noise currents to pass substantially unmodified. The circuits 7 are connected through respective resistance pads 26 to a circuit 27 that includes an amplitude limiter 28. The latter is connected to control the transmission of the noise current components through modulator 24. Whenever the transverse striations appear in the portions of spectrogram being scanned, that is, whenever a voiced sound is to be reproduced, currents of the fundamental frequency are generated by the electro-optical system and applied to limiter 28. These currents, limited in amplitude to a constant value, are impressed on modulator 24, thereby periodically interrupting, or modulating, the noise currents passing through modulator 24. The noise currents delivered to the bank of filters are accordingly chopped at the fundamental voice frequency; they have a harmonic structure and a certain pitch which is that of the original sound. Each of the filters 9 to 14 selects the generated components that fall within its pass-band, and the selected components delivered by any of these filters are varied in strength, by means of the associated vario-losser 22, to simulate the corresponding components of the recorded speech waves.

In the absence of the transverse striations each of the filters 9 to 14 selects the unmodulated noise components that fall within its pass-band, and the selected components are blocked or varied in strength according to the varying intensity of the several control voltages applied to the vario-lossers 22.

The amplitude limiter, modulator and vario-losser are devices well known in the art and any of various forms of them may be used. The vario-losser, for example, may comprise an amplifying vacuum tube the gain of which is varied by applying the varying control voltage to a grid electrode. The modulator may comprise a bridge of rectifying elements as shown.

In the modification of Fig. 2 that is illustrated in Fig. 3 the noise current source 25 is normally connected to the bank of filters 9 to 14 through the back contacts of a marginal relay 31. The currents in circuit 27 are applied to a linear rectifier 30 and also to a detector 32. When currents of the fundamental frequency are applied to rectifier 30, the fundamental and its harmonics appear in the output circuit of the rectifier. The applied currents operate also on detector 32 and, being relatively strong, they cause relay 31 to operate. Source 25 is thereby disconnected and the output circuit of rectifier 30 is connected in its

stead, through the front contacts of relay 31, to the bank of filters 9 to 14. Since the generated components delivered to the filter bank are all in harmonic relation, and the pitch is that of the original sound, the quality of reproduced vowel sounds simulates more closely that of normal, rather than whispered speech.

The system illustrated diagrammatically in Fig. 4 is adapted primarily for reproduction from narrow-band speech spectrograms, and it makes use of the relation between the fundamental voice frequency and the spacing of the striations that appear in such spectrograms. This system differs from that described with reference to Fig. 3 in the optical scanning elements and in the means provided for generating the harmonically related components of voiced sounds. As in Fig. 3, the noise current source 25 is normally connected to the bank of filters 9 to 14 through marginal relay 31, and the generated components selected by the filters are varied by vario-lossers 22 responsive to variations in the control voltages derived from the respectively corresponding circuits 7.

The optical system in Fig. 4 includes a lamp 40, condensing lens 41 and a mirror 42 that is caused to rotate or oscillate continuously. These elements are so proportioned and arranged as to direct a fine beam of light to the slit 3 and to cause the beam to sweep longitudinally of the slit many times a second. The current produced in each of the photocell circuits 7 is thereby interrupted many times a second so long as any light reaches the cells. The effective intensity of the interrupted current depends on and is a measure of the average envelope amplitude represented in the corresponding portion of the spectrogram. The process of detection in detector 20 involves an integration over a period of time dependent on the time constant of the detector, and the latter is designed to smooth out the interruptions and yield an output current that varies according to the variations in envelope amplitude. The filters 21 suppress high frequency variations due to the movement of the light beam; they may have a cut-off frequency of twenty-five cycles per second, for specific example. The vario-lossers 22 thus receive respective control currents that vary in substantially the same manner as those appearing in the systems illustrated in Figs. 2 and 3.

When vowel striations appear in the portion of spectrogram being scanned, the light received by the entire bank of photocells 6 is interrupted at a high rate dependent on the spacing of the striations and the rate of movement of the scanning beam. The latter is so fixed in relation to the frequency scale of the spectrogram that the interruptions occur periodically. Thus, if the spectrogram has a linear frequency scale, the vowel striations at any given point along the film are equally spaced across the film, and a constant rate of movement of the light beam will result in periodic interruptions. Suppose for specific example that the light beam traverses the length of slit 3 in a hundredth of a second, that the fundamental voice frequency represented in the scanned portion is 150 cycles per second, and that the spectrogram represents a frequency range of 4500 cycles. In such case the light beam would traverse 4500/150 striations in a hundredth of a second, and the light reaching the bank of cells 6 would be modulated at 3,000 cycles per second. Corresponding currents of this frequency accordingly appear in circuit 27.

A high pass filter 44 interposed in circuit 27 is

designed to pass the modulated currents derived from the vowel striations and to suppress currents of lower frequency. It will be noted that the modulation frequency is inversely proportional to the fundamental voice frequency and that it therefore varies with inflection of the recorded speech waves. The modulated currents passed by filter **44** are adjusted to constant amplitude, by means of amplitude limiter **28**, and applied to a so-called slope circuit or frequency discriminator **45** which operates in the usual manner to produce a uni-directional voltage that varies according to the frequency of the currents applied to it. The time constant of the discriminator **45** is designed to prevent any substantial change in its output voltage during the interval, if any, between successive sweeps of the light beam. The voltage output of discriminator **45** is applied to a multivibrator **46** to control the operating frequency thereof. The latter is variable over the normal range of the fundamental frequency, and as the applied control voltage varies in conformity with variations in the fundamental frequency of the recorded speech waves, the multi-vibrator frequency varies likewise to reproduce the original pitch at all times.

The output circuit of multivibrator **46** is normally disconnected from the bank of filters **9** to **14**, but whenever a voiced sound appears in the spectrogram, relay **31** operates to connect it to the filter bank. The operating winding of relay **31** may be connected to the output circuit of discriminator **45** for this purpose, as shown. The oscillations produced by multivibrator **46** comprise a fundamental frequency component which is or may be of substantially the same frequency as the fundamental frequency of the recorded speech waves, and also the harmonics of the fundamental frequency. The wave output is substantially free of inharmonically related components and in this respect it closely simulates the components that are produced by the vocal cords. The generated components are separated into groups by the filters **9** to **14** and varied in strength by vario-lossers **22** in the manner described with reference to Figs. 2 and 3 and then applied concurrently to loudspeaker **15**.

Although the embodiments selected for presentation herein involve transmission of light through the spectrographic record, it will be evident that light reflected from the record could be utilized instead. In this respect and in others that will occur to those skilled in the art, one may vary from the disclosed embodiments within the spirit and scope of the appended claims.

What is claimed is:

1. The method of synthesizing speech bearing waves represented in a speech spectrogram which comprises detecting substantially simultaneously the variations in envelope amplitude that are recorded in different portions of the spectrogram respective to different parts of the speech frequency range, continually scanning said spectrogram transversely of striations appearing in portions thereof representing voiced sounds to derive a measure of the varying fundamental voice frequency, generating electric wave components having a multiplicity of different frequencies, varying the strength of the different generated components in conformity with the variations detected in corresponding different portions of the spectrogram, varying the pitch of said generated components under the control of said derived measure, and translating the said components of varying strength and pitch into speech bearing sound waves.

2. The method in accordance with claim 1 in which said components are both generated and varied in pitch by said scanning step.

3. The method in accordance with claim 1 in which said components are harmonically related to each other and are generated independently of said scanning step.

4. The method in accordance with claim 1 in which said detection of variations in envelope amplitude, said generation of components, said variation in strength and said variation in pitch are effected by scanning said spectrogram.

5. A combination for playing-back a speech spectrogram, comprising electro-optical scanning means for deriving from said spectrogram individual measures of the variations in average envelope amplitude indicated for the several parts of the speech frequency range, a multiplicity of electrical circuits each adapted to selectively transmit currents of a frequency lying within an individually corresponding one of said parts of the frequency range, means for supplying each of said circuits with currents of a frequency lying within its said individually corresponding part of the frequency range, means individual to each said circuit for varying the intensity of the currents supplied thereto in conformity with the variations in the corresponding derived measure, and a sound reproducer connected to receive all of said currents of varying intensity.

6. In a combination for reproducing speech waves from a wide-band speech spectrogram, electro-optical scanning means including an optical slit that extends across said spectrogram substantially parallel to the transverse striations that appear in areas of the spectrogram representing voiced sounds, said slit being fine enough to resolve the represented structure of unvoiced sounds, a bank of photoelectric devices each individual to a different part of the speech frequency range, each said device being responsive to light modulated by the variations appearing in the particular portion of the spectrogram in which are recorded any components lying within the part of the frequency range to which it is individual, a multiplicity of wave filters each connected to a different one of said devices and each adapted to selectively transmit electric wave components lying within the part of the frequency range to which the connected device is individual, and an electroacoustic transducer connected to receive concurrently the wave components transmitted by said filters.

7. In a combination for reproducing speech waves from a wide-band spectrogram, electro-optical scanning means for translating into varying electric currents the variations appearing in each of a multiplicity of portions of the spectrogram that are respective to corresponding different parts of the speech frequency range, said scanning means including an optical slit that is at least fine enough to resolve the striations that appear in areas of the spectrogram representing voiced sounds, and a bank of photoelectric devices respective to the said different portions of the spectrogram and responsive to light modulated by the striations therein, a multiplicity of frequency selectors individual to said photoelectric devices and connected to receive therefrom the harmonically related current components produced by the scanning of said areas, each of said selectors being adapted to selectively transmit the components that fall within the part of

the frequency range identified with the connected photoelectric device, and an electroacoustic transducer actuated by the components transmitted by said frequency selectors.

8. A combination in accordance with claim 7 in which said slit is fine enough to produce noise currents in the absence of said striations.

9. In a combination for reproducing speech waves from a speech spectrogram, electro-optical scanning means for deriving from each of a multiplicity of different portions of said spectrogram that are respective to corresponding different parts of the speech frequency range, an individual control current that varies in substantial conformity with the variations in envelope amplitude recorded therein; means including said scanning means for deriving from the striations that appear in areas of the spectrogram representing voiced sounds, an electric current the frequency of which is a function of the spacing of said striations; means for generating a multiplicity of harmonically related electric wave components including means varying the frequency of said components responsive to variations in the frequency of said electric current; means responsive to the variations of the several said control currents for varying the strength of the said components that lie in respectively corresponding different parts of the frequency range; and means for translating said components of varying frequency and strength into a complex sound wave.

10. A combination in accordance with claim 9 in which said generating means comprises a harmonic generator operative on said electric current.

11. A combination in accordance with claim 9 in which said generating means comprises a modulator, a generator of noise currents and means for simultaneously impressing said electric current and said noise currents on said modulator.

12. In a combination for reproducing speech waves from a narrow-band speech spectrogram, electro-optical scanning means including means for sweeping a beam of light repeatedly across said spectrogram transversely of striations appearing therein in areas of the spectrogram representing voiced sounds, a multiplicity of photoelectric devices responsive to the modulated light emanating from respectively corresponding different portions of the spectrogram that are individual to corresponding different parts of the speech frequency range, means individual to the several said devices and responsive to the electric currents produced thereby for deriving a multiplicity of control currents that vary in substantial conformity with the variations in envelope amplitude recorded in the corresponding portions of said spectrogram, means common to a plurality of said devices and responsive to the current component resulting from the modulation of said light by said striations for deriving a measure of the varying fundamental voice frequency of the recorded waves, oscillator means for generating a multiplicity of harmonically related current components including means for varying the operating frequency of said oscillator means in conformity with variations in said derived measure, frequency selective means for separating said generated components, means responsive to said control currents for independently varying the strengths of said separated components, and a sound reproducer actuated by said varying separated current components.

13. The method of synthesizing speech bearing waves recorded in a narrow-band speech spectrogram which includes the steps of repeatedly scanning the spectrogram transversely of the striations that appear in areas of the spectrogram representing voiced sounds to derive a measure of the varying fundamental voice frequency of such sounds, generating a multiplicity of harmonically related current components, varying the frequency of said components in conformity with variations in said derived measure, varying the strength of said components differently in accordance with the variations in envelope amplitude recorded in corresponding different portions of the spectrogram that are individual to different parts of the speech frequency range, and concurrently translating said components of varying strength into sound waves.

WALTER KOENIG, JR.