United States Patent [19]

Fuller

[11] **3,855,417**

[45] **Dec. 17, 1974**

[54] **METHOD AND APPARATUS FOR PHONATION ANALYSIS LENDING TO VALID TRUTH/LIE DECISIONS BY SPECTRAL ENERGY REGION COMPARISON**

[76] Inventor: Fred H. Fuller, 4450 S. Park St., Rockville, Md. 20014

[56]   **References Cited**

UNITED STATES PATENTS

2,181,265	11/1939	Dudley	179/1 VS
3,238,303	3/1966	Dersch	179/1 SA
3,509,280	4/1970	Jones	179/1 SB
3,588,363	6/1971	Herscher	179/1 SA
3,679,830	7/1972	Uffelman	179/1 SA
3,752,929	8/1973	Fletcher	179/1 SA

FOREIGN PATENTS OR APPLICATIONS

1,113,225	5/1968	Great Britain	179/1 SA

OTHER PUBLICATIONS

Philip Lieberman, Some Acoustic Correlates of Word Stress in American English, J.A.S.A. 1960, pp. 451–54.

Lieberman & Michaels, Some Aspects of Fundamental Frequency & Envelope Amplitude as Related to the Emotional Content of Speech, J.A.S.A. 1962, pp. 922–27.

Medical Electronics, Electronics, 6/1966, p. 40.

[57]   **ABSTRACT**

A method and apparatus for indicating emotional stress in speech by normalizing the ratio of peak amplitude signals in two or more frequency regions of a single response. Normalization is achieved by comparing all subsequent ratios with a selected stored ratio of the same speaker.

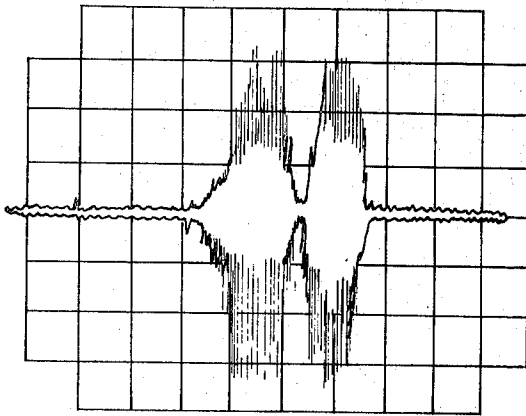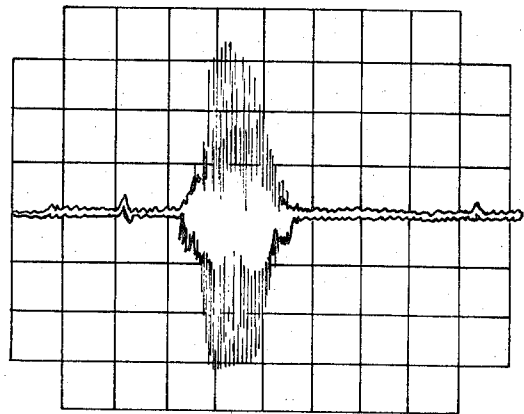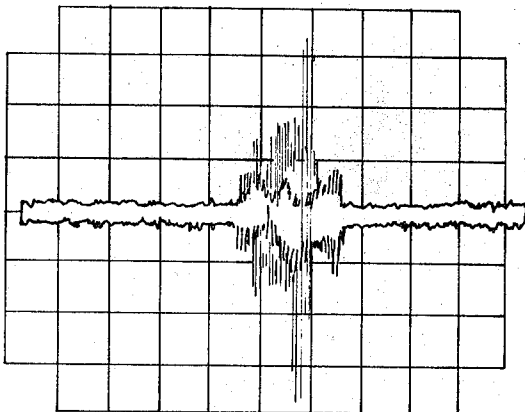**9 Claims, 10 Drawing Figures**
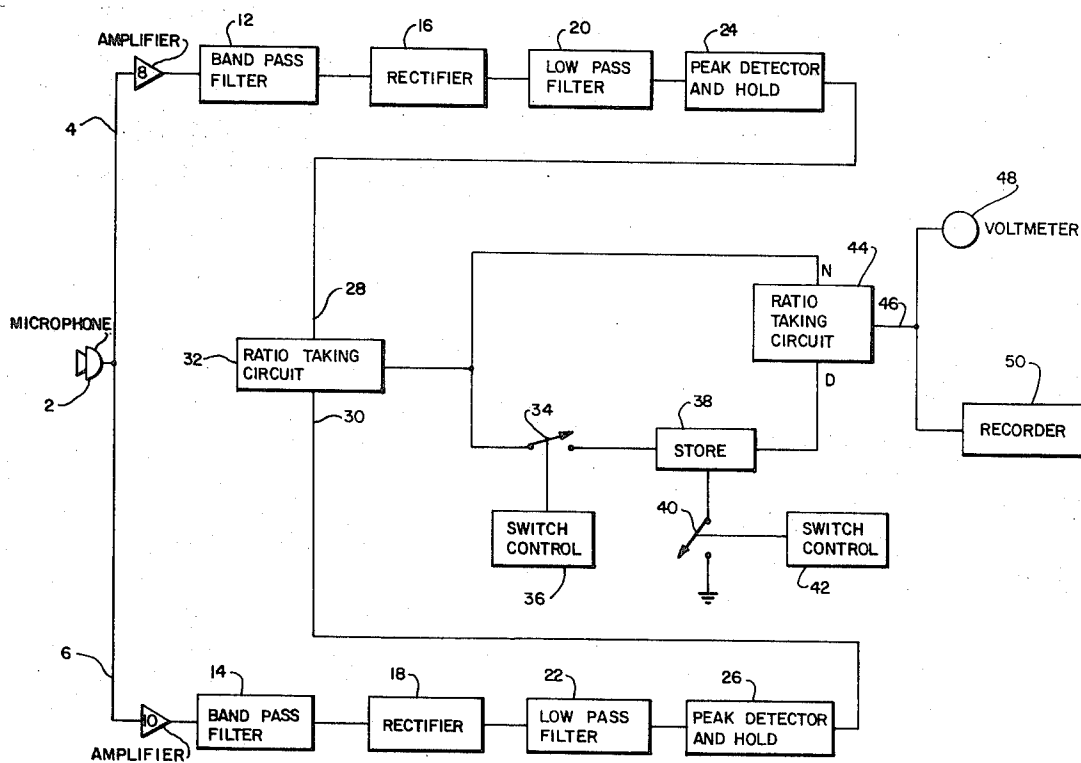
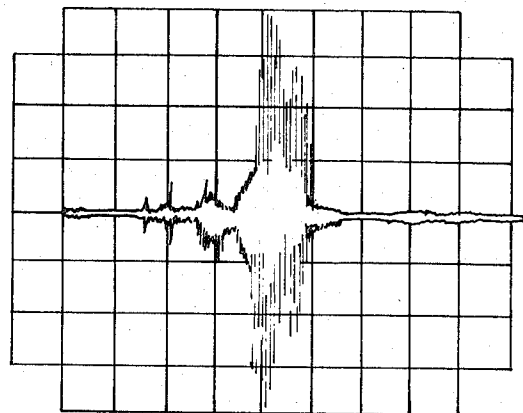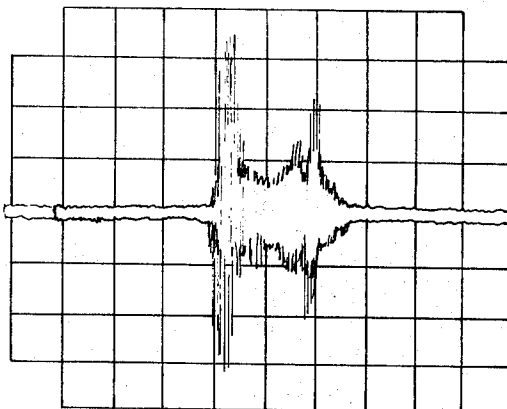FIG. 1   "YES"

FIG. 2   "NO"

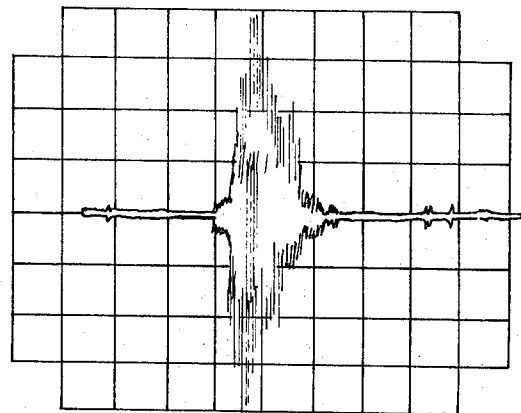FIG. 3A   150-300Hz

FIG. 3B   600-1200Hz

FIG. 4A   150-300Hz

FIG. 4B   600-1200Hz

3,855,417

FIG. 5

FIG. 6

PROB. OF
ACCURATE
ASSESSMENT

RATIO VALUES

FIG. 7

PROB. OF
ACCURATE
ASSESSMENT

NORMALIZED RATIO VALUES
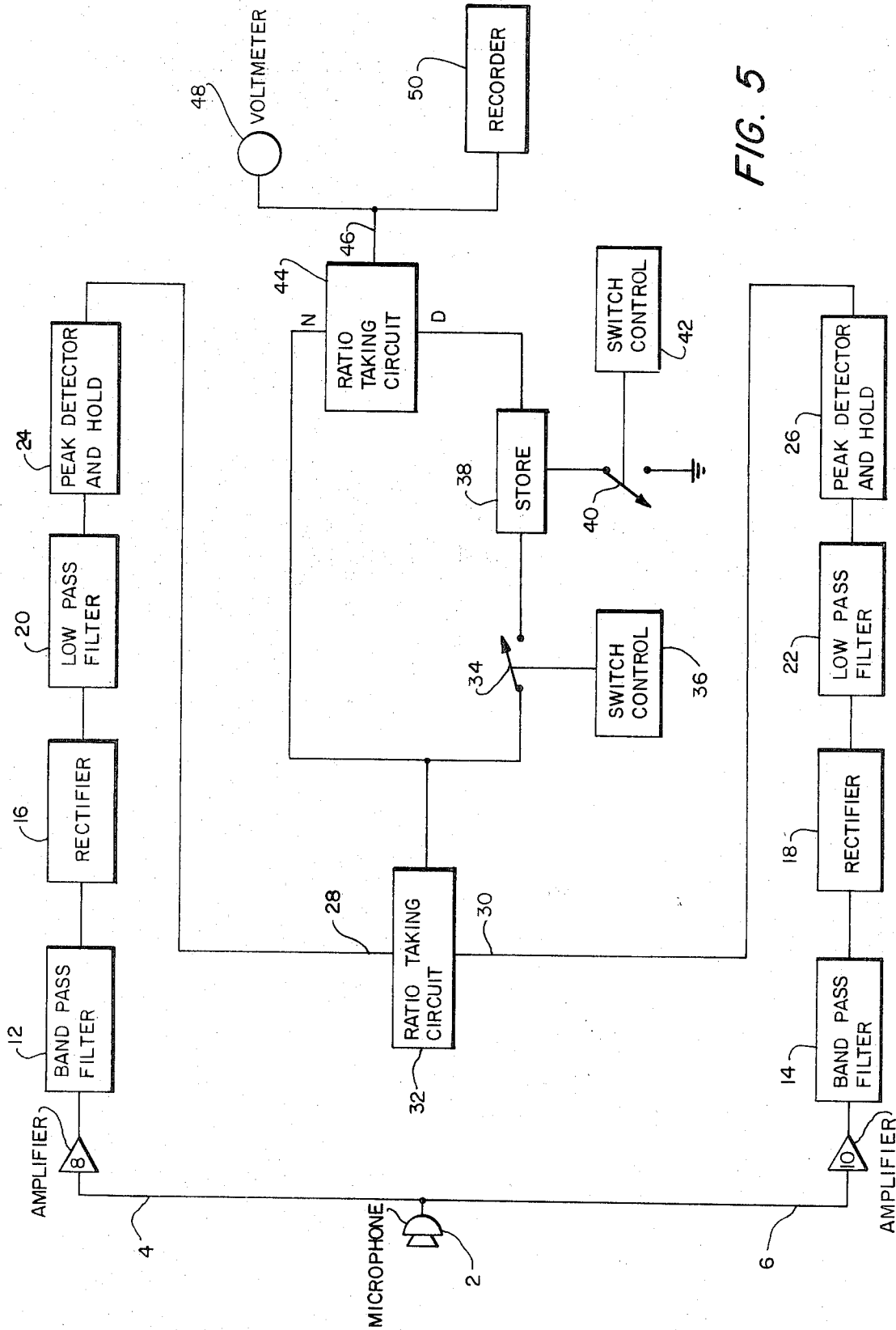
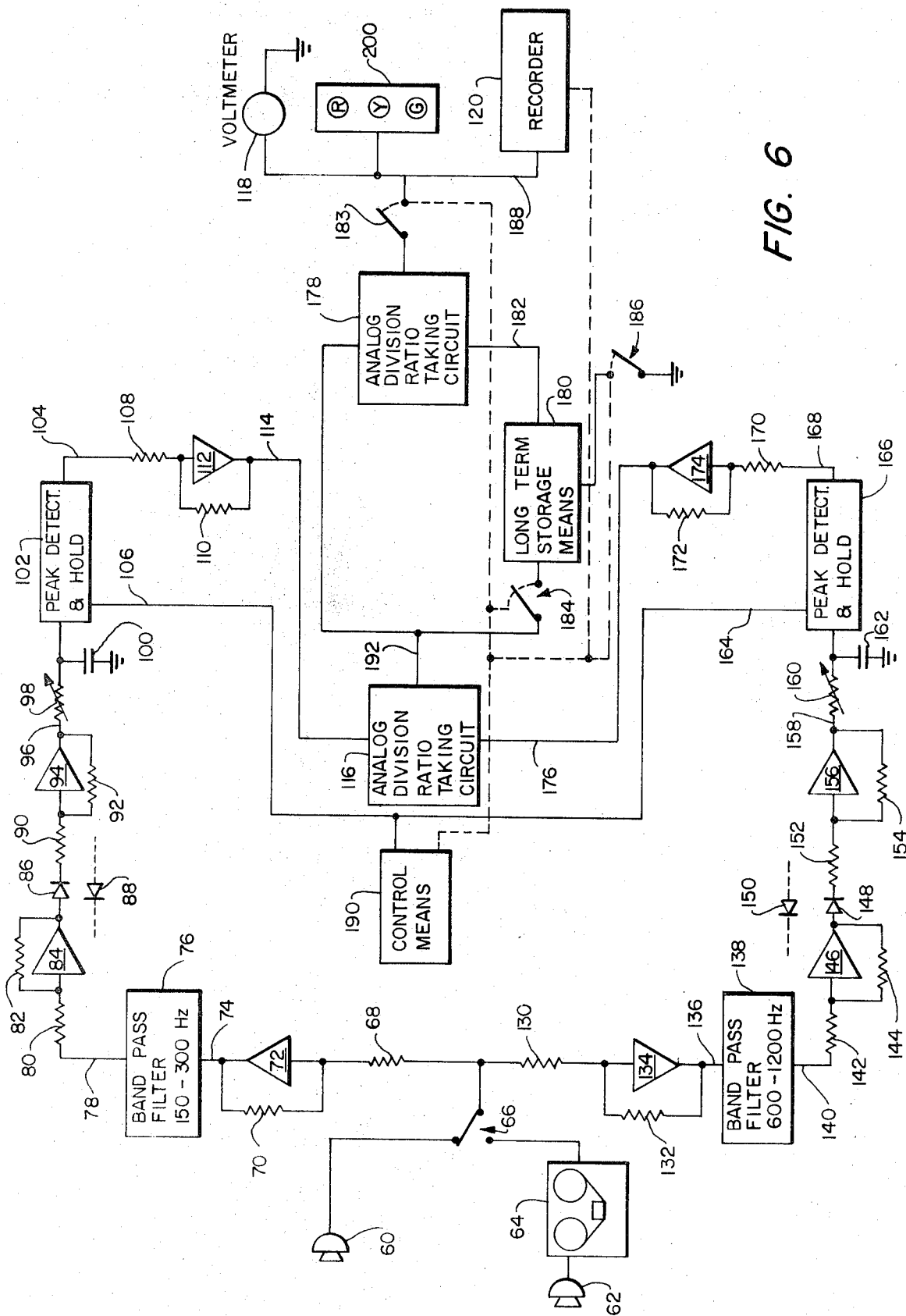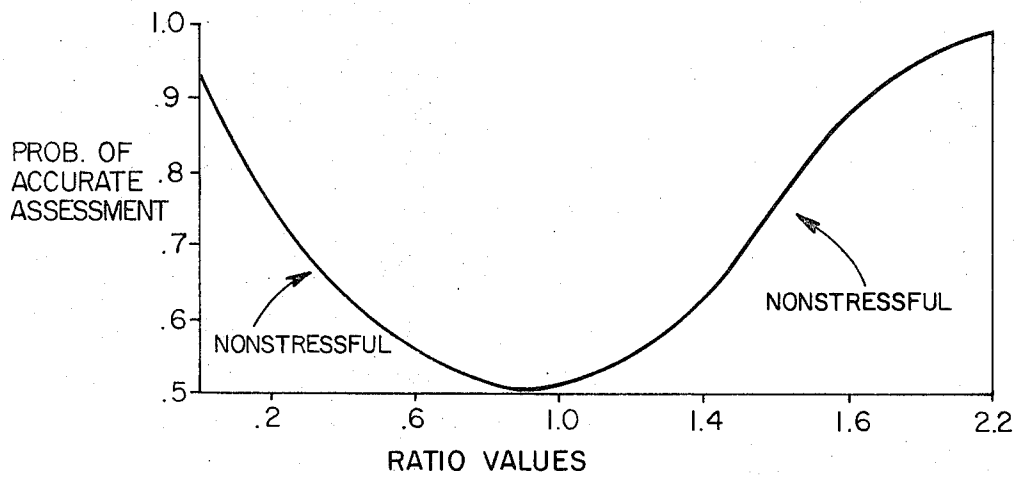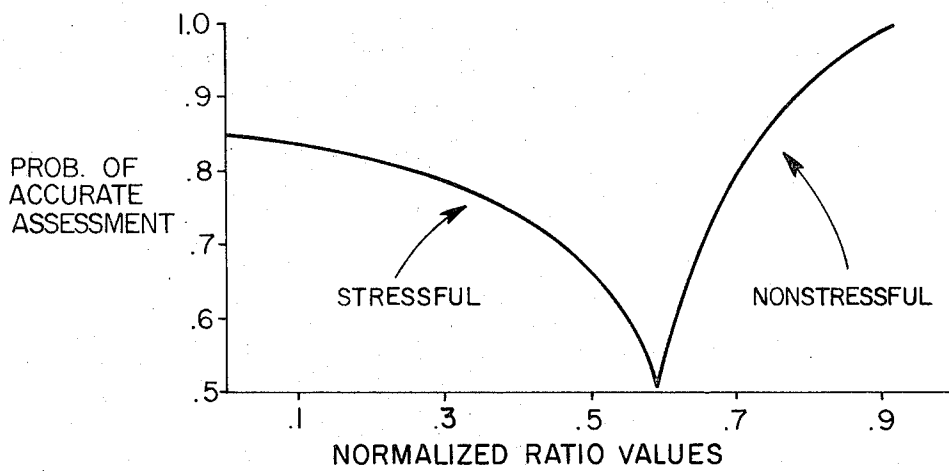*FIG. 8*

METHOD AND APPARATUS FOR PHONATION ANALYSIS LENDING TO VALID TRUTH/LIE DECISIONS BY SPECTRAL ENERGY REGION COMPARISON

BACKGROUND OF THE INVENTION

The present invention relates generally to voice signal analysis systems and more specifically to a method and apparatus for detecting emotional stress within a voice pattern. The presence of an emotional state will be used to determine the truthfulness of a response to questions asked by a skilled interrogator.

DESCRIPTION OF THE PRIOR ART

It has long been known that the voice may be, and often is, used to convey the emotions of the speaker. The emotional state of the speaker produces readily observable variation in the measurable parameters of the voice.

Speech is the acoustic energy response of: a) the voluntary motions of the vocal cords and the vocal tract which consists of the throat, the nose, the mouth, the tongue, the lips and the pharynx, and b) the resonances of the various openings and cavities of the human head. The primary source of speech energy is excess air under pressure, contained in the lungs. This air pressure is allowed to flow out of the mouth and nose under muscular control which produces modulation. This flow is controlled or modulated by the human speaker in a variety of ways.

The major source of modulation is the vibration of the vocal cords. This vibration produces the major component of the voiced speech sounds, such as those required when pronouncing the vowel sounds in a normal manner. These voiced sounds, formed by the buzzing action of the vocal cords, contrast to the voiceless sounds such as the letter "s" or the letter "f" produced by the nose, tongue, and lips. This action of voicing is known as "phonation."

The basic buzz or pitch frequency, which establishes phonation, is different for men and women. The vocal cords of a typical adult male vibrate or buzz at a frequency of about 120Hz, whereas for women this basic rate is approximately an octave higher, near 250Hz. The basic pitch pulses of phonation contain many harmonics and overtones of the fundamental rate in both men and women.

The vocal cords are capable of a variety of shapes and motions. During the process of simple breathing, they are involuntarily held open and during phonation, they are brought together. As air is expelled from the lungs, at the onset of phonation, the vocal cords vibrate back and forth, alternately closing and opening. Current physiological authorities hold that the muscular tension and the effective mass of the cords is varied by the learned muscular action. These changes strongly influence the oscillating or vibrating system.

Certain physiologists consider that phonation is established by or governed by two different structures in the pharynx, i.e. the vocal cord muscles and a mucous membrane called the conus elasticus. These two structures are acoustically coupled together at a mutual edge, within the pharynx and cooperate to produce two different modes of vibration.

In one mode, which seems to be an emotionally stable or non-stressful timbre of voice, the conus elasticus and the vocal cord muscle vibrate as a unit in synchronism. Phonation in this mode sounds "soft" or "mellow" and few overtones are present.

In the second mode, a pitch cycle begins with a subglottal closure of the conus elasticus. This membrane is forced upward toward the coupled edge of the vocal cord muscle in a wave-like fashion, by air pressure being expelled from the lungs. When the closure reaches the coupled edge, a small puff of air "explosively" occurs, giving rise to the "open" phase of vocal cord motion. After the "explosive" puff of air has been released, the subglottal closure is pulled shut by a suction which results from the aspiration of air through the glottis. Shortly after this, the vocal cord muscles also close. Thus in this mode, the two masses tend to vibrate in opposite phase. The result is a relatively long closed time alternated with short sharp air pulses which may produce numerous overtones and harmonics.

The balance of respiratory tract and the nasal and cranial cavities give rise to a variety of resonances, known as "formants" in the physiology of speech. The lowest frequency formant can be approximately identified with the pharyngeal cavity, resonating as a closed pipe. The second formant arises in the mouth cavity. The third formant is often considered related to the second resonance of the pharyngeal cavity. The modes of the higher order formants are too complex to be very simply identified. The frequency of the various formants vary greatly with the production of the various voiced sounds.

One of the acoustic correlates of emotional involvement transmitted through human speech is a measure of the normalized but relative peak energy at low and high frequencies in voiced phonation. Statistical data reveals that the normalized ratio between peak input signal values, measured within specified frequency ranges, corresponds in a significant manner to the degree of emotional stress during the assessed phonation. Other parameters through to be related to the emotional transmission of information include: Phonetic Content, Gross Changes In Fundamental Frequency, the Speech Envelope Amplitude and the Fine Structure of the Fundamental Pitch Frequency. This latter parameter is discussed in my copending patent application, Ser. No. 311,392. These parameters all contribute to the conveyance of emotion or a stressful condition existing in the speaker.

Speech analysis and the equipment for accomplishing the same has been developed for a variety of loosely related purposes. One of the primary concerns is the transmission of speech with a high order of intelligibility and presence over a very reduced bandwidth. The applicability of this particular art becomes obvious in civil and military communications. Other fields in which speech analysis equipment are used are the voice operated printing or recording device, such as a typewriter and systems, equipment and devices that are commanded and controlled by the spoken word or phrase. While these activities are interesting and valuable in themselves, they do not relate to the detection of emotional content of a speech wave nor its use to determine the veracity of the speaker.

SUMMARY OF THE INVENTION

The present invention determines the amount of emotional stress in the voice of a person under interrogation by comparing the peak amplitude in two differ-

ent frequency ranges. The peak amplitudes in a 150–300Hz and a 600–1200Hz frequency band are detected and held after separation by band-pass filters, rectifying and smoothing. The ratio of the peak amplitude in the two frequency regions can indicate emotional stress content after much analysis for the individual subjects. To accent the emotional stress content and provide a quantitative information thereof irrespective of the subject, the present invention stores a peak amplitude ratio from the subject and compares it with subsequent peak amplitude ratios in a second ratio circuit. The second ratio provides a normalization of the peak amplitude ratios. The stored ratio may be updated at the discretion of the interrogator or done automatically at periodic intervals.

OBJECTS OF THE INVENTION

It is an object of the present invention to provide a means for detecting a stressful or emotional condition in a human being who is speaking.

An additional object of this invention is to detect this emotional or stressful condition while the person who is speaking is under direct and skillful interrogation.

A further object of this invention is to provide means whereby a valid Truth/Lie decision can be rendered by direct observations of the data readout of a voice or speech analysis system.

A still further object of this invention is to detect the emotional or stressful condition by analysis of the maximum signal amplitude in two or more frequency regions of a human voice.

Other objects, advantages and novel features of the present invention will become apparent from the following detailed description of the invention when considered in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 is an oscillograph of a male voice responding with the word "yes" in the English language in answer to a direct question in a bandwidth of 5kHz;

FIG. 2 is an oscillograph of a male voice responding with the word "no" in the English language in answer to a direct question in a bandwidth of 5kHz;

FIGS. 3a and 3b are oscillographs of a male voice responding "yes" in the English language as measured in the 150–300Hz frequency region and 600–1200Hz regions, respectively;

FIGS. 4a and 4b are oscillographs of a male voice responding "no" in the English language as measured in the 150–300Hz frequency region and 600–1200Hz regions, respectively;

FIG. 5 is a simplified block diagram of a functional embodiment of the invention;

FIG. 6 is a detailed schematic of the preferred embodiment of the invention;

FIG. 7 is the plot of the results of a statistical analysis of measured ratio values versus the probability of accurate assessment of the given emotional state;

FIG. 8 is the plot of the results of a statistical analysis of measured and normalized ratio values versus the probability of correct assessment of a given emotional state.

DESCRIPTION OF PREFERRED EMBODIMENTS

FIG. 1 shows an oscillograph of a male voice responding with the word "yes" in the English language

in answer to a direct question at a bandwidth of 5kHz. The wave form contains two distinct envelopes, the first being for the voiced "ye" sound and the second being for the voiceless "s" sound. Since the first envelope of the "yes" signal wave form is a voiced sound being produced primarily by the vocal cords and conus elasticus, this envelope will be processed to detect emotional stress content or modulations. The male voice responding with the word "no" in the English language in a bandwidth of 5kHz is shown in FIG. 2. This response has a single envelope which will be analyzed by the present device to detect the presence of "fine structure" i.e. the rapid modulation of the phonation constituent of the speech signal.

FIGS. 3a, 3b, 4a and 4b show oscillographs of the same male voice as in FIGS. 1 and 2 responding "yes" and "no," respectively, in the English language as measured in the 150–300 Hz and 600–1200Hz frequency regions. The electrical speech signal in each of these bands of frequencies is well defined. Thus when this energy is rectified and smoothed, signal outputs are provided whose maximum amplitudes may be readily and accurately determined.

A simplified block diagram of the present embodiment of the invention is depicted in FIG. 5. A transducer 2, a microphone in this case, is used to convert the acoustic utterances or phonation of the subject being interrogated to an electrical signal. The electrical signals from the microphone 2 must have adequate fidelity to tranduce the basic pitch frequency of the voice — about 120Hz for male subjects and 250Hz for female. The voltage output of the microphone 2 is directed into two parallel channels, each having an isolation amplifiers 8 and 10 by shielded cables 4 and 6, respectively. The amplified signals are split into separate frequency regions by the employment of separate band-pass filters 12 and 14. In the particular embodiment, amplifier 8 feeds the full voice bandwidth to band-pass filter 12 which separates out a single frequency region. This region might occupy any portion of the audio spectrum but in a typical example it could occupy the fundamental pitch region from 100 to 200 Hz or 150 to 300Hz. After the speech energy is filtered it must be rendered single-valued. A rectifier 16 is a common method to do this. The rectifier 16 in this case selects only positive (or negative) values of the filtered speech energy. Full wave rectifiers could be employed, at the additional circuit complexity involvement. Following rectification, a low pass filter is employed to smooth out the peak fluctuations of the voice energy. In common use, this filter is variable so that the correct amount of filtering may be obtained for the particular voice in question. The output of the low pass filter 20 is commonly termed the speech envelope. A peak detector and hold circuit 24 stores the peak energy value of the filtered speech envelope. The output of the peak detector and hold circuit 24 is then directed to input port 28 of the ratio determining means 32.

Another channel from the microphone is directed through a different pass band region at input connection 6 into amplifier 10 and band-pass filter 14. This bandpass filter is set to pass a different frequency region than the band-pass filter 12. In general, this region will most likely be a higher frequency region than the first, for example, from 600–1200 Hz or thereabouts. As in the other channel, the speech energy is single valued by rectifier 18 which must be identical to the recti-

5

fier 16 in the other channel. The rectified speech energy is then directed through a low pass filter 22 circuit to provide a smoothed signal envelope as did filter 20. This filter employs different value components, however, since the frequency region of this channel is different from the first channel. An identical peak detector and hold circuit 26 is employed. This latter circuit detects and holds the peak energy value of the speech envelope in this particular frequency region. The output of the circuit 26 is directed to input port 30 of the ratio taking circuit 32, which performs the arithmetic computation of dividing the peak envelope energy of one channel by the other.

The output of circuit 32 is separated into two equal value signals. The first signal forms the input to a long time constant storage circuit 38 which holds the received signal value until it is deliberately discharged or reset by the closing of switch means 40. The switch may be controlled manually or automatically by control means 42. The signal is received by storage circuit 38 when switch means 34 is closed either manually or automatically by control means 36. The second portion of the signal from circuit 32 is received at the numerator port of a dividing or ratio taking circuit 44. This circuit is of the same type as the previous ratio taking circuit 32. The long term stored signal at the output of the storage device 38 is used as the denominator of the second ratio taking circuit 44. The second ratio taking circuit normalizes the signals from the first ratio taking circuit. The output of circuit 46 is directed to a volt meter 48 which reads the value of the ratio for each utterance of the subject and to the analog recorder 50 which records this value for subsequent analysis and comparison.

FIG. 6 is a detailed block schematic of the preferred embodiment of the invention. A detailed discussion of the components and functioning thereof will further serve to explain the behavior and the functioning of the invention.

A microphone 60 is shown as the acoustic/electric transducer which transfers the acoustic energy of the human voice into an electrical signal. The microphone used in this manner is perfectly typical except that the frequency response of the unit must cover the frequency regions of the follow-on filter circuits. Switch means 66 is shown which may be used to alternately select the sonic signal from the microphone 60 or from a combination of another microphone and a conventional tape recorder 64. The behavior of this second combination of microphone and tape recorder must retain the fidelity of the microphone 60. Namely the units must pass the entire frequency region that allows operation of the system.

The switch means 66 is followed by two operational amplifiers 72 and 134 used for isolation purposes. One operational amplifier 72, and its gain determining resistors 68 and 70, feeds speech energy to one channel of the instrument. The other operational amplifier 134, with its gain determining resistors 130 and 132, feeds the second channel of the instrument.

The first channel of the instrument feeds the amplified signal at terminal 74 to a band-pass filter 76. From the broad-band speech energy which the enters the band-pass filter 76 only the region from 150–300Hz is allowed to pass through. This region might change depending upon the voice involved. FIGS. 1 and 2 show the waveforms of a male human voice responding with

6

the word "yes" and the word "no" in the frequency region from 100Hz through 5000Hz. FIGS. 3 and 4 show the waveforms of the same voice after passing through the filters 76 and 138. Filter 76 passes the spectral region from 150–300Hz and filter 138 passes the region from 600–1200Hz.

Another stage of isolation employing operational amplifiers follows both of the band pass filters. The bandpass filter 76 is followed by operational amplifier 84 with its gain determining resistors 80 and 82. The bandpass filter 138 is followed by operational amplifier 146, with its gain determining resistors 142 and 144.

In the preferred embodiment of the invention, the conversion of the dual polarity signals out of each of the isolation amplifiers is rendered single valued or rectified by simple solid state diodes 86 and 148. Reversed polarity diodes are also shown as items 88 and 150. This polarity would allow the instrument to function equally as well. If diodes with a particular characteristic were employed, such as a square law characteristic, the instrument would function upon the measure of speech energy, i.e. power, in the two band-pass regions. In the present device, there is no statistical difference in the behavior of the instrument with a true "power"characteristic in the diodes or with a straight rectification process. In the practical case therefor, a pair of simple diodes have been found to suffice.

In both the low frequency channel and the high frequency channel, this rectification process takes place in the manner described and the both channels are directed again into two operational amplifiers for the purpose of circuit isolation. In the low frequency channel, the operational amplifier 94 with its gain determining resistors 90 and 92 perform this function. In the high frequency channel, operational amplifier 156 with its gain determining resistors 152 and 154 is likewise employed. In the low frequency channel, the signal output appears at point 96 and it passes into a low pass filter network consisting of a variable resistor 98 and a fixed capacitor 100. It can be seen that other types of low pass filters could be employed here to remove the high frequency fluctuations appearing at point 96 and rendering the output of the filter essentially that of the envelope of the speech signal, in the defined pass band. The exact time constant of this filter is adjusted depending upon the pitch of the voice under assessment. This envelope of signal energy then passes into a peak detect and hold circuit 102. Such circuits can be readily fabricated from a variety of components and modules by those skilled in the art. A single module named Infinite Sample Hold which is manufactured by Hybrid Systems Corp. can be used. This particular module has the advantage of non-decay of the peak value until the circuit is reset. The output peak value appears at point 104 while the reset signal is applied at point 106. The peak value is isolated from the follow-on circuit by another operational amplifier 112, with its gain determining resistors 108 and 110. The output at 114 of amplifier 112 appears as the denominator of an analog division circuit 116. Again, there are many ways that the ratio of two voltages could be taken. Two voltages could be read on two volt meters and the ratio determined arithmetically. The two voltages could also be recorded and the recorded values read from a chart and the ratio determined arithmetically. Preferably a modern analog computer module such as Model 107C,

manufactured by Hybrid Systems Corp. could be used.

In the high frequency channel, the circuitry is quite the same. The signal energy appears at the output of the isolation amplifier **156** at point **158** and passes into a similar R/C filter network, where the high frequency components of the speech envelope in the high frequency region are filtered out. The value of the time constant which consists of variable resistor **160** and fixed capacitor **162** is different from that in the low frequency channel, since the frequency is quite different.

Again, the signal passes out of the filter into a peak and detector and hold module **166**. Under control of the reset signal at port **164** and providing output at port **168**, the peak detector and hold module acts to detect and store the peak signal envelope value that occurred during the phonaton of the subject in the selected frequency region. An isolation amplifier **174**, with its gain determining resistors **170** and **172** provides the high frequency channel peak signal level to the analog division module as the numerator of the ratio expression. The quotient appears at port **192** where it divides into two equisignal paths. One path travels to switch means **184** which is under the control of control means **190**. The other path travels to a second analog division circuit or ratio circuit **178** which is identical to circuit **116** and appears at the numerator port of the analog ratio taking circuit **178**.

The switch means **184** connects the analog ratio of the two frequencies at **192** to a long term sample hold storage means **180**. When switch means **184** is closed by control means **190**, the long term storage means will keep holding the peak ratio appearing at **192** until it is reset by reset switch means **186**, also under control of control means **190**. This circuit then feeds this stored value for a selected utterance of the speaker into the ratio taking circuit **178** on the denominator buss **182**. Thus the machine may be calibrated or "normalized" for a particular speaker, with the result that his other variations in subsequent utterances will be highly significant. The reset switch **186** operated by control means **190** will not be reset until a new subject is being interrogated or until the analyst desires to update the normalization. The output of this ratio taking circuit **178** enters switch means **183** which is also operated by control means **190**. When a suitable normalizing denominator has been held or stored and entered into the final ratio taking circuit **178**, the ratios formed for all subsequent utterances will pass to the indicators. The indicators consist of DC volt meter **108** and recording means **120**. The recording means functions only for the second and all subsequent utterances of the subject since the control means **190** closes switch **183** after the first utterance. The control means **190** also provides switching signals to peak detect and hold module **102** and **166** as required. It also switches to signals into ans out of the long term storage **180**, the final ratio taking circuit **178** and it operates the recorder **120**.

FIG. 7 is the plot of the statistical analysis of the ratio values obtained from a number of stressful and non-stressful utterances of a number of different speakers. The plot shows the probability of correctly identifying an utterance as stressful or non-stressful as a function of the ratio value of that utterance. The plot indicates that when either low or high ratio values occur, the utterance can be assessed to be non-stressful. There is no

ratio value range in which an utterance can be assessed, with any confidence, i.e. greater than 50 percent, as being stressful.

On the other hand, FIG. 8 is a plot of an analysis of the data taken from the preferred embodiment using a normalizing circuit. This data has been normalized by assessment of a specific utterance of the subject. All subsequent utterances of the same subject are then normalized with this information. It can be seen that normalized ratio values less than about 0.59 are stressful and ratio values higher in value are non-stressful. To obtain a confidence band of greater than 70 percent, the ratio values may be set as less than 0.46 for stressful and greater than 0.65 for nonstressful. With this circuit behavior it is quite apparent to those skilled in the art that a set of limit lights **200** may be applied. For example, a red light could activate for an established lower limit indicating stress or an untruthful response, an amber light could indicate indecision or no opinion in the middle range of the ratio values, and a green light could indicate non-stress or a truthful response. When the measured normalized ratio value was higher than a given amount, the lights may be controlled by level responsive switches such as relays or transistors or a combination thereof.

With the above description, any person skilled in the art could discern the proper functioning of the invention described here. Statistical data taken with the present instrument has demonstrated that conditions of emotional stress and in particular the Truth/Lie decision can be analyzed and correctly discerned with a high degree of confidence. Although the invention has been described and illustrated in detail, it is to be clearly understood that the same is by way of illustration and example only and is not to be taken by way of limitation, the spirit and scope of the invention being limited only by the terms of the appended claims.

What is claimed is:

1. A method for detecting emotional stress in the utterance of an individual comprising:
   converting said utterance to an electrical signal;
   selecting two different frequency bands of said electrical signal;
   detecting and holding the peak amplitude of each frequency band for the duration of the utterance;
   computing the ratio of the held peak amplitude of one frequency band with the other;
   storing a previously computed ratio of said peak amplitudes;
   comparing subsequent ratios with said stored ratios; and displaying the compared results which would be indicative of emotional stress.

2. A method as in claim **1** wherein selecting comprises amplifying, band-pass filtering, rectifying and smoothing wherein said band-pass filtering and smoothing is different for each selected frequency band.

3. A method as in claim **1** wherein displaying comprises indicating quantitatively stress, non-stress and indecision.

4. A device for indicating emotional stress from the utterances of a human comprising:
   means for converting said utterances into electrical signals;
   first channel means connected to said converting means for detecting and holding a peak amplitude

in a first frequency band for the duration of the utterance;

second channel means connected to said converting means for detecting and holding a peak amplitude in a second frequency band for the duration of the utterance;

first ratio means connected to said first and second channel means for taking the ratio of said detected and held peak amplitudes

means connected to said first ratio means for storing a previous ratio taken by said first ratio means

second ratio means connected to said first ratio means and said storing means for taking the ratio of said stored ratio and subsequent detected peak amplitude ratios; and

means connected to said second ratio means to display said second ratio.

5. A device as in claim 4, said first and second channel means each including:

means for passing electrical signals in a selected frequency band;

means connected to said passing means for rectifying said passed signal;

means connected to said rectifying means for smoothing said rectified signal; and

means connected to said smoothing means for detecting and holding the peak amplitude of said smoothed signal.

6. A device as in claim 5 wherein said first channel passes a frequency of 150–300Hz and said second channel passes a frequency of 600–1200Hz.

7. A device as in claim 4 wherein said display means comprises three indicators each responsive to a select region of second ratio values thereby indicating stress, non-stress and indecision in the utterance.

8. A device as in claim 4 wherein said first channel's frequency band is of a lower frequency than said second channel's frequency band, and wherein said first channel's detected peak amplitude comprises the denominator of said first ratio means.

9. A device as in claim 4 wherein said stored ratio comprises the denominator of said second ratio means and said subsequent peak amplitude ratios comprises the numerators of said second ratio means.

*   *   *   *   *

5

10

15

20

25

30

35

40

45

50

55

60

65