

April 12, 1966

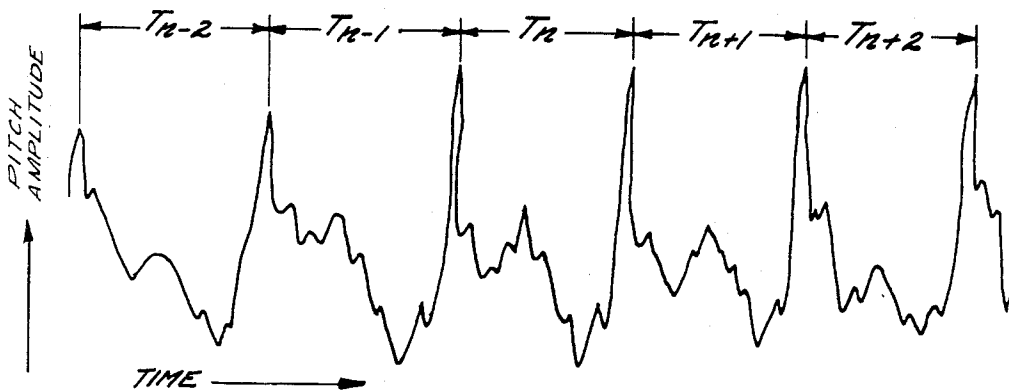
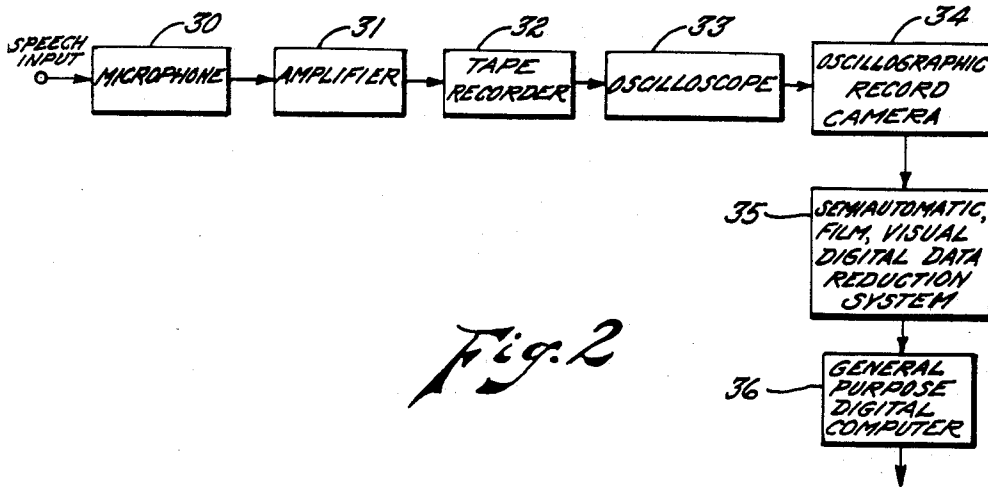
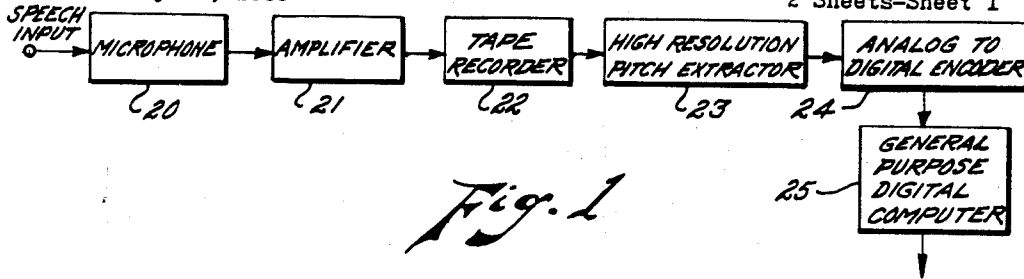
P. LIEBERMAN

3,245,403

SYSTEM FOR ACOUSTIC DETECTION OF PATHOLOGIC LARYNGES

Filed May 22, 1963

2 Sheets-Sheet 1



INVENTOR
PHILIP LIEBERMAN
BY
Wade King
George Lind
ATTORNEYS

April 12, 1966

P. LIEBERMAN

3,245,403

SYSTEM FOR ACOUSTIC DETECTION OF PATHOLOGIC LARYNGES

Filed May 22, 1963

2 Sheets-Sheet 2

--- = NORMAL SPKR ($T_{median} = 9.6$)
 — = SPKR. EIGHTEEN ($T_{median} = 9.1$)

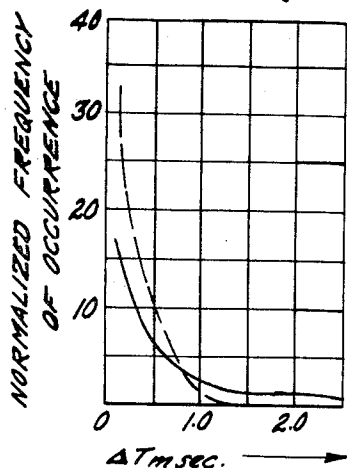


Fig. 4

○ = LARGE MASSES OR CA. ON VOCAL CHORDS
 • = NORMAL CONTROLS

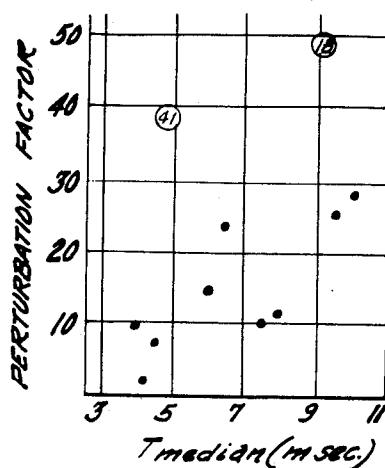


Fig. 5

○ = SMALLER MASSES ON VOCAL CHORDS
 • = NORMAL CONTROLS

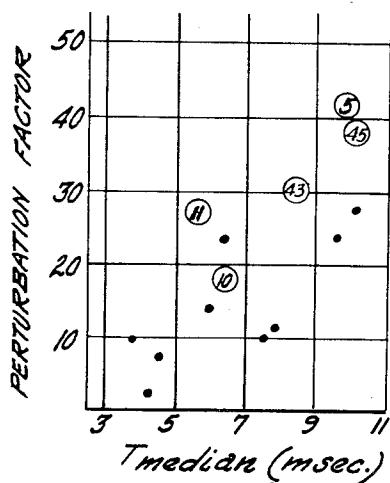


Fig. 6

○ = LARGE MASS OR CA. NEAR VOCAL CHORDS
 • = NORMAL CONTROLS

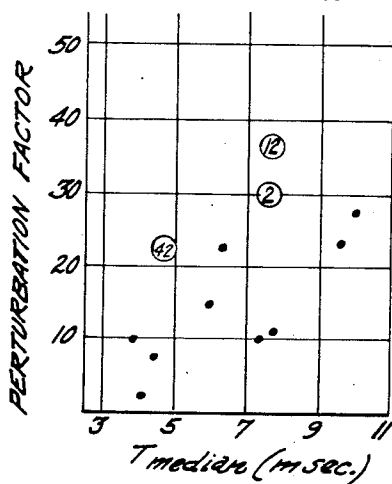


Fig. 7

INVENTOR.
 PHILIP LIEBERMAN

BY

Walter L. Smith
 George F. Line
 ATTORNEYS

1

2

3,245,403 SYSTEM FOR ACOUSTIC DETECTION OF PATHOLOGIC LARYNGES

Philip Lieberman, Cambridge, Mass., assignor to the
United States of America as represented by the Secre-
tary of the Air Force

Filed May 22, 1963, Ser. No. 282,521

1 Claim. (Cl. 128—2)

The invention described herein may be manufactured and used by or for the United States Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to acoustic detection of pathologic larynges, and more particularly, to a system wherein the fundamental period of phonation of human speech is utilized to determine the presence of a pathologic human larynx.

The detection of pathologic human laryngeal conditions is at present effected primarily by visual examinations of the affected organs by a trained Otolaryngologist through the techniques of indirect and direct laryngoscopy. Direct laryngoscopy is often the only certain way of telling what is wrong. These examinations are costly and time consuming and often require hospitalization of the patient in the case of direct laryngoscopy and extensive anesthesia. The patient's physician must first suspect the presence of pathologic laryngeal conditions through his subjective impression of the patient's voice. He must decide whether the patient's voice appears to be especially hoarse or whether the patient's vocal quality has changed. In certain situations when the patient is being treated at a dispensary, e.g. in the armed services, the doctor does not know what the patient's normal voice sounds like so that he is unable to note subtle changes in the patient's vocal quality. Thus many laryngeal conditions develop into permanent disabilities and surgery is ultimately necessary. Cancer of the vocal chords, in particular, causes large masses to develop on the vocal cords.

The invention to be described hereinafter detects comparatively small masses on the vocal cords through an acoustic testing procedure that can be easily administered by relatively unskilled personnel to large groups of people on a routine basis. Thus many laryngeal conditions are detected and treated in their early stages before extensive surgery becomes necessary and permanent disabilities result. A thorough examination by a trained specialist will still be necessary after a potential subject has been detected through this screening procedure. However, the test is of great value in detecting cancer of the vocal cords and other nearby organs in the vocal tract as well as other pathologic laryngeal conditions.

In accordance with the present invention detection of cancer in the larynx at an early stage is made possible by analyzing human speech signals, particularly the pitch thereof. Three factors are involved in pitch generation—the vocal cords, the variation of air pressure within the vocal tract, and muscular tension. Their movements establish pitch periodicity of the acoustic waveform. These movements are regulated by the air pressure against the vocal cords and by muscular tension. Even with normal speakers, i.e. persons whose vocal cords do not have growths thereon, variations in pitch periodicity occur because of small transients in air pressure. However, growths or polyps on or near the vocal cords alter the relative mass of the cords and create an imbalance between the two vocal cord elements. The system's mechanical damping factor is reduced and air pressure transients result in larger and more prolonged variations in pitch periodicity than is the case for normal speakers.

In the present invention fundamental pitch periods are measured with a high degree of resolution. Pitch per-

turbations, that is, small but rapid variations in the fundamental periodicity of normal connected speech, are then computed by measuring the differences between durations of adjacent fundamental periods. The aforesaid pitch perturbations that is the variations in the fundamental periodicity of normal connected speech is well known in the prior art and is described by A. Risberg in "Statistical Studies of Fundamental Frequency Range and Rate of Change" which appeared in "Speech Transmission Laboratory Quarterly Progress and Status Report," Royal Institute of Technology, Stockholm, Sweden, January, 1962, pp. 7-8 and also described by S. Saito, K. Kato, and N. Teranishi in "Statistical Properties of the Fundamental Frequencies of Japanese Speech Voices," appearing in J. Acoust. Soc., Japan, 14: 111, 1958. Perturbations having an absolute value ≥ 0.5 msec. (0.5 milliseconds) are induced by transients in the air pressure drop across the glottis which are in turn caused by changes in vocal tract configurations. The perturbations also depend on the durations of the fundamental periods and speakers, i.e. persons, with longer fundamental periods tend to have larger perturbations. Speakers or persons with pathologic larynges have larger pitch perturbations than do normal speakers with the same median fundamental periods. Growths on the vocal cords appear to reduce the laryngeal damping factor and larger perturbations are more easily induced by air pressure transients. The "Perturbation Factor," which is defined as the percent of the time that perturbations ≥ 0.5 msec. in magnitude occurred, is computed for each speaker. The differences between the Perturbation Factor of speakers with normal and pathologic larynges are proportional to the size of the pathologic growths on or near the speaker's vocal cords.

It is to be noted that the present invention is applicable to people of all nationalities as pitch perturbations are not artifacts of any particular language and are caused by the physical processes of speech production. Since the fundamental periodicity of speech is established by the activity of the larynx, then these pitch perturbations are due to the irregularities in the periodic motions of the vocal cords.

An object of the present invention is to provide a system wherein the human speech is utilized to detect pathologic larynges.

Another object of the present invention is to provide a system wherein the pitch content of human speech is utilized to detect pathologic larynges.

Yet another object of the present invention is to provide a system wherein pitch perturbations of human speech are measured to detect pathologic larynges.

Still another object of the present invention is to provide a system to calculate the Perturbation Factor in human speech to detect pathologic larynges.

The features of this invention, which are believed to be new, are set forth with particularity in the appended claims. The invention itself, however, together with further objects and advantages thereof may best be understood by reference to the following description when taken in conjunction with the accompanying drawings, in which:

FIGURE 1 shows a block diagram of one embodiment of the present invention including tape scanning utilized to extract the durations of fundamental periods from speech waveforms;

FIGURE 2 shows a block diagram of a second embodiment of the present invention including an oscillographic record camera to photograph the speech waveform displayed on an oscilloscope;

FIGURE 3 shows speech waveform typical of those provided by pitch extraction of either of the systems

shown in FIGURES 1 or 2 to be read for measurement purposes in which the duration of each fundamental period is determined from the leading edge of one amplitude peak to the next succeeding one to thus permit the subsequent subtraction of the duration of each period from that period immediately preceding it to provide pitch perturbations;

FIGURE 4 shows two curves provided from the output of computer 25 or 36 of either of the systems shown in FIGURES 1 or 2 which are typically representative of frequency distribution of ΔT for two speakers;

FIGURE 5 shows plots of the Perturbation Factor for normal speakers and speakers with large pathological growths, the Perturbation Factor being provided at the output of computer 25 or 36 of either system shown in FIGURES 1 or 2;

FIGURE 6 shows plots similar to FIGURE 5 but shows speakers with small growths; and

FIGURE 7 is similar to FIGURE 5 but includes speakers who had large growths or carcinoma near their vocal cords.

Now referring in detail to FIGURE 1, there is shown microphone 20 which may be of the wide range condenser or dynamic type. It may also be a telephone handset. The individual to be evaluated first reads a three minute selection into microphone 20 which serves to introduce the test material. An alternate procedure is to engage the individual in conversation. The introductory selections serve to insure that the individual will read the test material in his "usual" voice. Only a short selection is necessary to perform the actual test. The individual reads in an acoustically treated room in which the reverberant sound field has been minimized. Standard audiometric rooms are often suitable. However, it is possible that useful results may be obtained in a standard type of room that remains acoustically untreated.

Microphone 20 serves as a transducer and the output thereof is an electrical signal representative of the input speech thereto. The representative electrical signal is passed through amplifier 21 which provides sufficient gain to properly operate tape recorder 22. Tape recorder 22 is a standard monaural single channel analog tape recorder covering the range 50-10,000 c.p.s. with a minimum of wow and flutter and complies with the NARTB standards at 7.5 inches per second recording speed.

The output from tape recorder 22 is then fed into high resolution pitch extractor 23 which determines the onset and end of phonation from the recorded speech waveform and measures the duration of each fundamental pitch period to the nearest 0.1 msec. The fundamental pitch period is equal, for example, to T_n which is the duration between two amplitude peaks as illustrated in FIGURE 3. Pitch extractor 23 may be such as shown and described in a paper entitled "Pitch Extraction by Computer Processing of High Resolution Fourier Analyses Data" by C. M. Harris and M. R. Weiss published March 1963 in the Journal of Acoustical Society of America, volume 35 at pages 339-343, or may be such as shown and described in the Proceedings of the 4th International Congress of Acoustics, Copenhagen, Denmark, Aug. 21-28, 1962, in paper No. G-26, entitled "Extraction of the Voice Pitch from the Vibration of the Outer Skin of the Trachea" by T. Sugimoto and T. Mirura. Pitch extractor 23 may also be such as described by Gold in the Journal of the Acoustical Society of America, volume 34 at pages 916-921, published in July of 1962.

The output signal from pitch extractor 23 is fed into analog to digital encoder 24 which converts the measurements of the duration of the fundamental pitch periods into a computer format. The encoder, for example, digitizes and punches the duration of each period on IBM cards (or if desired on paper tape, or digital magnetic tapes). The encoder also indicates the beginning and end of voicing and assigns speaker identification numbers to the data. Analog to digital encoder 24 may

be such as shown and described in a doctoral thesis by G. W. Hughes entitled "Recognition of Speech by Machine," published at the Massachusetts Institute of Technology, Cambridge, Massachusetts, in the Department of Electrical Engineering in the year 1959. It is to be noted that analog to digital encoder 24 may be such as is commercially available from Epsco Corporation in Cambridge, Mass.

The output information from analog to digital encoder 24 is fed to general purpose digital computer 25 and is programmed to calculate the Perturbation Factor and median fundamental period of each individual. Digital computer 25 is of the conventional general purpose type and is commercially available for example, Model No. 7090, sold by International Business Machines Corporation.

Perturbation Factor is defined as the integral of the frequency distribution of ΔT for $\Delta T \geq 0.5$ msec.

$$\Delta T = T_n - T_{n-1}$$

where T_n is the duration of period n and T_{n-1} is the duration of period $n-1$. ΔT is the pitch perturbation and is the difference in time between the time duration T_n and T_{n-1} , T_n and T_{n-1} being illustrated in FIGURE 3 which shows the duration of successive fundamental pitch periods for a preselected time as represented by aforesaid spoken short selection. The Perturbation Factor, in other words, is the percent of the time that perturbations are greater than or equal to 0.5 msec. Pitch perturbation, ΔT , being the difference in time in milliseconds (msecs.) between adjacent fundamental pitch periods and where this difference is 0.5 msecs. or greater between successive periods over a preselected time as represented by aforesaid spoken short selection then the percent of time that ΔT is equal to or greater than 0.5 msec., for said preselected time represents the Perturbation Factor. Individuals with pathologic larynges have larger Perturbation Factors than do normal individuals with the same median fundamental period. The differences between the Perturbation Factors of individuals with normal and pathologic larynges is proportional to the size of the pathologic growths when these growths are on the vocal cords. The Perturbation Factor is also sensitive to large growths near the vocal cords. Carcinoma of the larynx generally results in the formation of comparatively large masses on or near the vocal cords. Thus the Perturbation Factor can detect the presence of Carcinoma as well as other pathologic conditions that involve growths on or near the vocal cords. The Perturbation Factor may be adjusted with respect to the individual's median or mean fundamental period if a single index of the likelihood of pathologic conditions is desired.

Now referring in detail to FIGURE 2, there is shown microphone 30 which is of the type having a wide frequency response. A speaker whose larynges are to be evaluated first reads a three minute selection which serves to introduce the test material. The introductory selections serve to insure that the speaker will read test material in his usual "voice." Only a short selection is necessary to perform the actual test. The speaker reads sentences into microphone 30, in an acoustically treated room in which the reverberant sound field has been minimized. Standard audiometric rooms are often suitable. It is also possible to utilize rooms that are acoustically untreated.

Microphone 30 serves as a transducer and converts speech into its representative electrical signal. The representative electrical signal is passed through amplifier 31 to tape recorder 32. Tape recorder 32 is a standard monaural type covering the frequency range of 50 to 10,000 c.p.s. with minimum wow and flutter. Tape recorder 32 also meets NARTB standards at 7.5 inches per second recording speed.

The recorded signal on tape is played back for display on the face of oscilloscope 33 and simultaneously photo-

graphed with oscillographic record camera 34 running at 1,200 inches per minute. Oscillographic record camera 34 is commercially available from Dumont Corporation and is conventional. The 35 mm. film obtained from the aforementioned photographing is processed and then read on semiautomatic film, visual digital data reduction system 35 on which the duration of each pitch period (as indicated in FIGURE 3) is measured and also automatically punched on IBM cards. Semiautomatic film, visual digital data reduction system 35 is commercially available and one type is sold as the Gerber Digital Data Reduction Systems type GDDRS-3B. The overall time resolution of the entire system is 0.05 msec.

Digital computer 36 is programmed to process the pitch periods punched on aforesaid IBM cards by semiautomatic film, visual digital data reduction system 35. Computer 36 first calculates the pitch perturbations by subtracting the duration of each period, T_n , from that of the period preceding it, T_{n-1} , T_n and T_{n-1} being illustrated in FIGURE 3. Computer 36 lists and graphs the sequence of perturbations for each speaker. Digital computer 36 is of the same type shown in FIGURE 1 as digital computer 25 which is in the general purpose category and being commercially available, as for example, Model No. 7090, sold by International Business Machines Corporation.

General purpose computer 36 was programed in one instance to calculate and plot, for each speaker, the frequency distribution of absolute value of the difference in duration between adjacent periods, i.e., the hystogram of $[\Delta T = T_n - T_{n-1}]$ where T_n is the duration of period n , T_{n-1} is the duration of period $n-1$, and ΔT is the pitch perturbation and is the difference between T_n and T_{n-1} . It is to be noted T_n and T_{n-1} are illustrated in FIGURE 3. The two curves of FIGURE 4 are typical frequency distributions of ΔT for two speakers who had approximately the same median periodic duration. Speaker 18 had a large benign mass on his left vocal cord while the other speaker's larynx was normal. It was found that for both normal and pathologic larynges higher frequencies of occurrence of large ΔT 's resulted as T median, the speaker's median periodic duration increased. However, speakers with large masses on their vocal cords had larger frequencies of occurrence of large ΔT 's than did normal speakers with similar fundamental periods.

Since perturbations ≥ 0.5 msec. in magnitude reflected variations in the vocal cords' vibratory cycle, the integral of the frequency distribution of ΔT for $\Delta \geq 0.5$ msec., i.e., the percent of the time that perturbations ≥ 0.5 msec. occurred, was computed for each speaker. This computation is termed the Perturbation Factor. In FIGURE 5 the Perturbation Factor is plotted on the ordinate for a control group of normal speakers, each of whom is marked by a solid dot, and for speakers 4, 18 and 41, who all had comparatively large masses on their vocal cords. The median fundamental period of each speaker is plotted on the abscissa. Speaker 41 has squamous carcinoma with a mass measuring 15 x 5 x 5 mm. on his right vocal cord. Speaker 18 has a fairly large keratotic polyp-like area on the anterior half of his vocal cord.

In FIGURE 6 the same normal group is repeated and compared with a group of speakers who had smaller growths on their vocal cords. Speaker 43 had a recurrent 5 mm. polyps in his right vocal cord. Speaker 11 had a polyp 1 cm. across, extended in the middle third of her left vocal cord. Speaker 45 had a leukoplakia-like lesion on his right vocal cord. Speaker 10 had a laryngeal polyp 5 mm. across on her right vocal cord and speaker 5 had a small singer's nodule on his left vocal cord. Note that these speakers' Perturbation Factor falls into a lower region than the pathologic cases plotted in FIGURE 5.

In FIGURE 7 the normal group is again repeated and compared with speakers who had large growths or carcinoma near their vocal cords. Speaker 2 in this group had a large subglottic mass which filled the area below his left vocal cord anteriorly while speaker 42 had a recurrent cancer to her posterior cricoid area. Speaker 12 had a squamous carcinoma of his epiglottis. Note that these cases also fall into a lower region than those of FIGURE 5. It is to be noted that in each of FIGURES 5, 6, and 7. Each of the speakers having normal larynges is represented by a solid dot. Each of the speakers having a growth is represented by a numeral which is encircled. Accordingly, by a quick glance at FIGURES 5, 6, and 7, it is easily discerned (by noting the encircled numerals) the speakers having growths.

Thus the present invention provides a most important system to evaluate and detect a pathologic condition on or near a speaker's larynges. Speakers with pathologic larynges have larger pitch perturbations than do normal speakers with the same median fundamental period. The Perturbation Factor is mass sensitive. The differences between the Perturbation Factors of the normal and pathologic larynges are proportional to the size of the pathologic growths when these growths are directly on the vocal cords. Speakers with large growths near their vocal cords have smaller Perturbation Factors than speakers with large growths directly on their vocal cords. However, speakers with large growths near their vocal cords had much larger Perturbation Factors than the normal speakers.

What is claimed is:

In a system for detecting growths in the larynx of a human being, the combination of:

- a microphone for transforming the voice waves of a person speaking a preselected sentence, into an electrical signal that is representative of the spoken sentence;
- a recorder for making a record of said signal;
- means for responding to a playback of said record to measure the duration of each fundamental pitch period in said signal;
- means for converting the pitch period measurements into a punched record of said measurements; and
- a digital computer for processing said punched record, said computer being programmed to calculate the Perturbation Factor of said signal, said Perturbation Factor being defined as the percent of the time that perturbations equal to or greater than 0.5 milliseconds occur during the time required to speak said preselected sentence, said perturbations being defined as the differences in the durations of adjacent fundamental pitch periods, whereby the differences between the Perturbation Factors of speakers with normal and pathologic larynges are proportional to the size of the pathologic growths on or near the vocal cords.

References Cited by the Examiner

UNITED STATES PATENTS

2,393,717	1/1946	Speaker	128—2.1
2,442,805	6/1948	Gilson	128—2.1
2,616,415	11/1952	Kirby	128—2
2,678,692	5/1954	Ranseen	128—2
3,024,783	3/1962	Timcke	128—2

OTHER REFERENCES

"Bell Labs Record," Davis, pp. 263-267, June 1950, 128-2.1.

RICHARD A. GAUDET, *Primary Examiner.*

LOUIS R. PRINCE, SIMON BORDER, *Examiners.*