ROYAL AIRCRAFT ESTABLISHMENT

*

Technical Report 79056

May 1979

A DEVICE FOR TRACKING THE FUNDAMENTAL FREQUENCY OF SPEECH AND ITS APPLICATION IN THE ASSESSMENT OF 'STRAIN' IN PILOTS AND AIR TRAFFIC CONTROLLERS

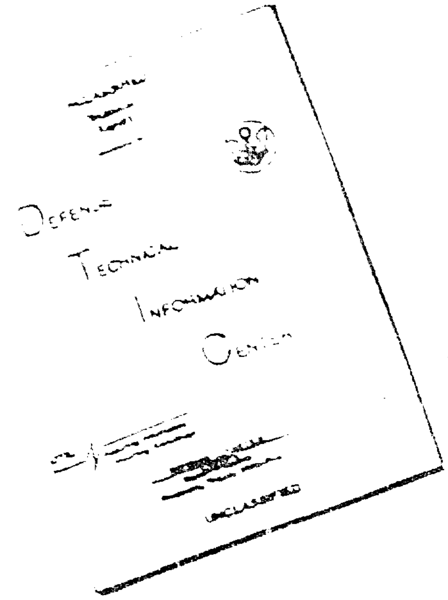
by

J.B. Peckham

*

Procurement Executive, Ministry of Defence
Farnborough, Hants

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

11 May 79

UDC 621.372.54 : 534.63 : 656.7.05 : 159.937 : 358.43 : 159.944 : 613.65 :
534.78 : 612.78

14 RAE-TR-79056

ROYAL AIRCRAFT ESTABLISHMENT

9 Technical Report 79056

Received for printing 22 May 1979

6 A DEVICE FOR TRACKING THE FUNDAMENTAL FREQUENCY OF SPEECH AND ITS APPLICATION IN THE ASSESSMENT OF 'STRAIN' IN PILOTS AND AIR TRAFFIC CONTROLLERS.

by

10 J. B. Peckham

12 57

SUMMARY

The use of the speech signal for assessing physiological and psychological changes resulting from 'strain' in pilots and air traffic controllers is explained and a device is described for tracking one of the parameters of the speech signal, the fundamental frequency, to quantify changes in this parameter due tc 'strain'.

18 DRIC 19 BR-69831

Departmental Reference: FS 100

310450 Gm

LIST OF CONTENTS

056

1 INTRODUCTION

The assessment of the effects upon man of varying mental workloads is an extremely complex problem particularly in view of the many interactions involved between task demands, environmental factors and the individual characteristics of man himself.

A flow diagram indicating the major interactions occurring is shown in Fig 1 and is based upon a model of the working system, in particular of air traffic controllers, due to Laurig[1] and described by Rohmert[2]. For present purposes the term 'stress' is used to describe the input to a 'man at work', of task demands and other environmental factors which affect his performance such as noise, vibration, illumination and climate, and also other demands imposed upon the man such as the responsibilities of the job. 'Stress' on interaction with a man and his individual characteristics, such as his level of competence, results in 'strain'. It is this 'strain' imposed upon a man that may result in changes in certain physiological parameters (Fig 1).

This Report is concerned primarily with the effects of mental stress rather than physical stress and where the term 'workload' is used it refers to mental workload.

Such physiological parameters as heart rate, skin potential, electromyograms, electroencephalograms and the concentration of catecholamines in the urine have been used by many researchers in attempts to provide information about the workload of man, although the measure of heart rate is perhaps the most widely used and easiest parameter to record continuously[3]. However, when using these parameters, particularly heart rate, it is still difficult to distinguish between physical and mental workload. In the case of heart rate, other parameters such as oxygen intake, temperature and catecholamine metabolism must be measured when attempting to distinguish between sources of heart rate changes. A further disadvantage of using these physiological parameters to assess workload is that measurement of them requires some degree of interference with the subject. In some situations it may not be very practicable to instrument the subject to obtain all the measurements required and even if it were there still remains a possibility that the subject's motivation may be affected. Acoustic analysis of the speech signal of a subject at work in an attempt to identify changes in the signal due to 'strain' is attractive since it does not require direct instrumentation of the subject. In the case of pilots and air traffic controllers a communication channel already exists and recordings of their speech may be made for subsequent analysis without interfering in any way with their work; indeed

the subjects themselves need not be aware that anything out of the ordinary is taking place, thus avoiding any modification to their motivation due to awareness of measurements being made on their performance.

The human vocal system is a complex one in which the prime function of the organs making up the system is not that of producing speech at all but one of survival as suggested by O'Connor[4]. The lungs transfer oxygen to the blood, the vocal cords help to prevent foreign bodies entering the trachea whilst the tongue plays a vital role in chewing and swallowing. It is for this reason that organs when fulfilling their secondary function of producing speech may also contribute information to the speech signal which has no linguistic significance but is rather a reflection of the particular physiological state of the organs. This contribution may provide useful information on the 'strain' imposed on a subject as it is reflected in the physiological state of such organs as the respiratory system and the larynx. Such physiological changes which may occur in a subject under '*strain*' are 'uncontrolled' in that they are not effected by direct motor control. In addition to uncontrolled physiological changes resulting in a modification to the speech signal, 'controlled' changes (*ie* under motor control) may take place in the speech signal. In the case of air traffic controllers under high 'stress' a message communicated may be modified due to the requirements of the situation. For example in a situation of high air traffic density it may be necessary for an air traffic controller to communicate specific information to several pilots in a very short space of time resulting in a fast speaking rate but accompanied by a reduction in the number of words used to communicate the information.

At the acoustic level the speech signal may be described by a number of features, referred to as prosodic features, such as the fundamental frequency (or pitch of the voice), formant frequencies, intensity and duration (for a more detailed description of the speech signal, see section 2).

These prosodic features of speech may convey linguistic and extra-linguistic information. However since the same features carry both types of information it is necessary to separate those changes in the prosodic features conveying linguistic information from those conveying extra-linguistic information.

Some physiological changes which may occur in a subject under 'strain' are increased respiration rate and increased muscle tension, both of which could give rise to an overall increase in the voice pitch; these and other physiological changes may also be manifest in some measure in other prosodic features (Fig 2).

In analysing the speech signal to attempt to provide indices of 'strain' arising from the effects of workload and environment we are essentially looking for those prosodic features of the signal which are correlates of the subjects physiological and psychological state. Having found reliable and good correlates there still remains the problem of deciding to which extra linguistic features the changes in prosodics should be ascribed. Since various emotions are reflected in the extra linguistic information this involves, for example, the relationships which may or may not exist between emotional state and strain due to workload.

A number of researchers have studied what they generally refer to as 'the effects of emotions on the voice'. Some of these studies have been concerned purely with measuring the differences between different types of emotion[5-11], whilst others have sought to relate vocal changes to the 'emotional state of pilots during flight'[12], 'pilot stress'[13], 'state of attention'[14], 'the emotional state of man under conditions of space flight'[15], 'task-induced stress'[16] and simply 'stress'[17], 'emotional stimuli'[18] and 'emotional state'[19]. In seeking to find indicators of workload level or task difficulty in the speech signal, caution must be exercised in correlating the changes in certain prosodic features reported by researchers due to various emotions or emotional stimuli with changes due to task or workload level. Whilst the prosodic features which reflect changes in the emotional state of a person may well be the same features which will best give information on the subjects state due to workload level or task difficulty, it is not at all clear that emotional state bears a one to one correspondence with workload state. Certain emotional states such as fear or anxiety may well sometimes be a manifestation of strain resulting from workload imposed upon a subject. However other less obvious 'emotions' may well have nothing to do with strain arising from the workload or task imposed but may be a reflection of the underlying mood of the subject.

Of all the prosodic features, the voice fundamental frequency appears to be potentially one of the best carriers of extra-linguistic information[11,12] and for this reason the 'pitch tracker' described in this Report has been developed to enable a detailed statistical analysis of the voice fundamental frequency of pilots and air traffic controllers to be carried out.

Early techniques for measuring the fundamental frequency consisted of
056 measuring the distance between the periodic epochs in the speech signal by hand thus giving a period by period measurement or making average measurements over consecutive time intervals[20-23]. Whilst such methods render accuracies of around 0.5% they are obviously unsuitable for large quantities of speech and of course

do not permit automatic tracking of the fundamental frequency in real time. Another method which has been used it that of low pass filtering the speech and subjecting the result to some form of frequency measurement. Using fixed cut off frequency filters only a very limited range of fundamental frequency can be measured and a high signal-to-noise ratio signal is required. Variable cut off frequency filters have been tried in the past but in the early days suffered from switching and control transients[24].

The advent of the computer has made possible the implementation of a number of mathematical methods for analysis of the speech signal and the extraction of such parameters as the fundamental frequency. Some of these methods will be discussed in more detail in a later section. Until fairly recently these algorithms could not be implemented in real time on small minicomputers, however the use of microprocessors and specialised digital hardware systems has now enabled real time calculations to be realised. Such hardware can be complicated and expensive to construct and the need for a fairly simple system to track the fundamental frequency of speech often recorded with high background noise has led to the development of the circuitry described in this Report. The method used for extraction of the voice fundamental frequency is based on an algorithm developed by the Joint Speech Research Unit, Cheltenham (JSRU) for a commercial application, but differs in the voiced/unvoiced discussion strategy employed to achieve successful tracking of the fundamental frequency in high levels of background noise. The circuitry used to implement the algorithm is believed to be novel.

2 SPEECH CHARACTERISTICS

2.1 The nature of speech and speech production is a vast and complex subject and so only a brief overview will be given here, the reader is referred to Fant[25] Flanagan[26] and O'Connor[4] for a more detailed discussion of the subject.

Speech may be described as being made up of words each containing one or more sounds which are called phonemes. A particular phoneme string characterises a particular word and in English some 40 different sounds or phonemes have been identified which carry basic linguistic information. At the phonetic level, speech may be studied in two ways, the first being a study of articulatory phonetics, showing new sounds are produced and classifying them according to their method of production. This may then lead onto the way sounds are put together to convey linguistic information. The second way of studying speech is at the acoustic level where the acoustic signal between the mouth and ear is investigated and sounds are described in terms of frequency spectra, intensity and duration.

The varying acoustic features of the speech signal which describe particular sounds and the context and way in which they are spoken are referred to as prosodic features and include voice pitch, intensity, formant frequencies and duration. These features may best be described by first considering the way in which speech sounds are produced. There are two basic components of the speech system which interact in the production of sounds, these are the vocal cords and the vocal tract which is itself composed of the oral cavity, pharynx and nasal cavity (Fig 3).

2.2 Vocal cords

The passage of air over the vocal cords provides a source of energy which excites the vocal tract producing sounds which may be classified into two types according to whether the vocal cords are vibrating or not and are known as voiced and unvoiced sounds. There are further sub-divisions of sound types which are determined by the particular articulatory mechanisms employed.

(a) Voiced sounds

These are produced by the passage of air over the vocal cords (Fig 4) which vibrate in a quasi periodic fashion. The pulses produced as the vocal cords open have a spectrum rich in harmonics (Fig 5), the first harmonic is referred to as the voice fundamental frequency and the period between the pulses as the excitation period*. In voiced sounds the mode of vibration of the vocal cords determines whether the voice is normal, creaky or breathy. Each is distinguished by the amount of air passing over the vocal cords in the open phase. The loudness of the voice is determined by how wide the vocal cords open, together with the amount of lung pressure exerted.

(b) Unvoiced sounds

Such sounds are produced when the vocal cords are not vibrating and an unimpeded breath of air is produced, *eg* as in the sound 'sh', 'ff'. The cords may be partially closed together producing some turbulence of air such as occurs in the sound 'h'. Acoustically these sounds are characterised by a broadband noise spectrum with most energy in the higher frequencies.

The voice fundamental frequency carries intonational information in the speech signal and has a mean value of around 125 Hz in adult male speakers.

---

* When no specific method of measurement is implied the term voice pitch will be used for convenience.

2.3 Vocal tract

The vocal tract which is composed of the pharynx, nasal and oral cavities, is excited by the pulse train from the vocal cords and produces a radiated spectrum at the lips (Fig 7) which is characterised by the particular transfer function of the vocal tract at any instant (Fig 6). The peaks in the radiated spectrum are known as formant frequencies F_1, F_2, F_3 etc and arise as a result of resonances in the vocal tract cavities. It is these formant frequencies which carry the linguistic information in the speech signal, the positions of the formant frequencies varying with the articulation of the vocal tract and tongue. Thus each voiced sound in the English language is characterised by a particular formant frequency pattern which may be seen on a wideband spectrogram which displays the energy, typically in 200Hz frequency bands, over a predetermined frequency range (for example 0-4 kHz) versus time (Fig 8).

The pharynx is tube shaped (Fig 3) and may be varied in length by the raising of the larynx or the soft palate. When the soft palate is raised the nasal cavity is occluded and is therefore not excited by the glottal pulses. Of the three cavities it is the oral cavity which plays the greatest roll in forming the sounds since it is capable of considerable change in shape and size due to the lips, tongue and lower jaw although it is the tongue which contributes most in speech production as the prime articulator. As the phonemes are put together to form words the acoustic pattern resulting is determined by the articulatory requirements to produce a given word and a word may consist of voiced and unvoiced sounds and perhaps gaps of silence within a word, known as 'stop gaps'. The phonemes of the English language may be divided into groups other than voiced or unvoiced according to the manner of articulation producing them, however it is beyond the scope of this Report ot enter into a discussion of this subject.

3 EXTRA-LINGUISTIC CHANGES IN THE PROSODIC FEATURES OF SPEECH

3.1 The existence of extra-linguistic changes in particular acoustic features of speech has already been discussed in section 1. Some of these changes will now be discussed in more detail with particular reference to investigations carried out by other workers into these changes.

3.1.1 Fundamental frequency

Of all the features of the speech signal, fundamental frequency appears to be the most widely studied for extra-linguistic changes and several researchers have found good correlation between changes in the voice fundamental frequency and the subjects emotional state[6,11].

It has been reported that changes in the respiratory pattern frequently occur in situations of fear or anxiety[27] and Williams[11] suggests that "an increase in respiration rate would presumably result in an increased subglottal pressure during speech. This heightened subglottal pressure would give rise to a higher fundamental frequency during voiced sounds in speech". A further result of increased respiration rate could be "shorter durations of speech between breaths". The tension of the larynx muscles would also be expected to have an effect on the fundamental frequency, an increase in muscle tension caused by certain psychological states would give rise to an overall increase in the fundamental frequency and changes in the glottal pulse shape may also result[11]. The work of Williams and Stevens[11,12] shows overall changes in the fundamental frequency between different types of emotion measured for speech samples of several seconds duration, the significant variables between types of emotion were found to be mean fundamental frequency and the fundamental frequency range. Hecker *et al*[16] also report changes in mean fundamental frequency, range and contour between a relaxed subject and the same subject under 'task induced stress'. In some subjects studied the mean fundamental frequency was found to increase under stress whilst in others a decrease was noted. Isao Kuroda *et al*[13] use a measure of the excitation period. By calculating a term called the 'vibration space shift rate' which is a measure of the change of the highest excitation period taken from a phrase spoken by a pilot in a 'normal' situation to one spoken in an 'urgent' or emergency situation the degree of 'stress' was assessed from a nine point classification of 'shift rate'.

From the investigations carried out by these researchers it would appear that changes in fundamental frequency over and above intonational changes do occur under certain emotions and task or workload induced 'strain'. One of the problems associated with assessing changes in the fundamental frequency due to extra-linguistic factors is that of separating out changes due to the linguistic or intonational information being conveyed. Several workers have shown that samples of voiced speed of 20-30 seconds duration must be used in order to obtain estimates of the mean fundamental frequency which are independent of the speech content[28,29].

3.1.2 Formant frequencies

Simonov *et al*[9] report the ability to distinguish between differing emotional states (joy, delight, anxiety and fear) using a measure of fundamental frequency and the average number of zero crossings occurring in the first formant region of the speech signal (an estimate of the first formant frequency F_1). The method

described, plots, for a particular Russian vowel, the ratio of the zero crossing frequency in a state of rest to the frequency in an emotional state versus the ratio of the fundamental frequency in a state of rest to the frequency in the emotional state. Average values for the ratios are determined from three sound impressions. Using a linear determinant function 92-94% discrimination between emotional states and state of rest was obtained for all the cases examined.

It is possible that changes in the formant frequencies may result from increased muscle tension causing a variation in the vocal tract shape above that which is caused by articulation.

3.1.3 Amplitude

Hecker *et al*[16] report changes in the mean amplitude of two phonetically stressed vowels selected from a test phrase spoken under 'control' and 'task induced stress' situations. Only four out of ten subjects however showed 'definite and consistent trends' in the amplitude changes and for one of the four subjects the amplitude change under 'stress' was the reverse of that for the other three. Friedhoff *et al*[17] also report changes in the average amplitude of a spoken word when a subject was presented with emotional stimuli. In this case a 'baseline' amplitude was established with neutral stimuli and masking noise presented to the ears. Some subjects in this experiment showed a decrease in amplitude when presented with emotional stimuli whereas others showed an increase in amplitude over the baseline level. It could perhaps be concluded from these experiments that the direction of change in amplitude is not as important as the fact that a change has taken place. Presumably the changes in amplitude could be related in some way to respiratory changes in the subjects, however in view of the difficulty in establishing constant 'baseline' amplitude levels this technique appears to be very difficult to implement as a method for assessing changes under emotion or 'strain'.

3.1.4 Duration of sounds

If the respiration rate of a subject increases under strain or some emotion it could be expected that the duration of sounds or words between breaths would shorten, there does not however appear to have been any work done to assess this change in the speech pattern.

4 SURVEY OF PRESENT METHODS OF 'PITCH' MEASUREMENT

4.1 The assessment of the pitch of voiced speech may be attempted by one of three methods.

056

(a) Making measurements from the temporal properties of the signal.

(b) Making measurements from the frequency domain properties of the signal.

(c) Utilising both the time and frequency domain properties of the signal.

Examples of time domain pitch detectors are those using zero crossing measurements, peak and valley measurement and autocorrelation measurements. Such methods require the preprocessing of the signal to minimise the effects of formant structure on the waveform. Frequency domain pitch detectors are based on the principle that if the signal is periodic then the frequency spectrum of the signal will contain a fundamental frequency plus harmonics of that frequency. The term 'fundamental frequency' of the voice really refers to the frequency of the first harmonic of the vocal source frequency spectrum (*ie* implies a measurement made in the frequency domain).

An example of a method operating in the frequency domain is cepstrum analysis which will be briefly described later. Hybrid methods of pitch assessment involve measurements in both the frequency and time domain, an example of this method is the simplified inverse filter technique (SIFT) of Markel 1972 in which the speech signal is inverse filtered to produce a spectrally flattened signal which is then autocorrelated.

Early methods of determining the voice pitch involved peak detection or the use of fixed or variable low pass filters, each generally employing some form of non-linear preprocessing of the speech waveform to enhance the fundamental frequency component. Peak detection methods work on the basis of detecting the epochs in the speech waveform corresponding to the vocal cord pulse train produced in voiced sounds. The distance between these epochs or 'pulses' is taken as the excitation or pitch period; however it should be noted that the voiced excitation of the vocal tract is only quasi-periodic. Further difficulties in using the peak detection method of period estimation arise from the fact that the glottal waveform varies in shape and amplitude as well as period and this can make it difficult to decide which epochs on the speech waveform should be chosen for period estimation.

The use of fixed low pass filters to extract the first harmonic of the glottal waveform is limited to high quality speech signals where the fundamental frequency range is small. Variable low pass filters can be used to overcome the problem of fundamental frequency range but early attempts proved unsatisfactory due to switching or control transients[24]. With these problems overcome, this

technique is still only suitable for high quality signals in which the fundamental frequency is present. Non-linear preprocessing methods which have been used to enhance the fundamental include half wave and full wave rectification and squaring[24].

Since the early 1960's much work has been done on the use of algorithms implemented on computers and special hardware to analyse the speech waveform. And there is considerable current activity in this area.

Perhaps one of the best known methods for speech analysis is cepstrum processing. This technique used for the extraction of 'pitch' is described by Noll[31,32]. The term cepstrum refers to the spectrum or Fourier transform of the log-amplitude spectrum and since the reciprocal of frequency is the resultant variable the term 'quefrequency' was coined by the inventors to describe it. The log-taking operation has the property of separating the vocal source characteristics (*ie* harmonic structure) from the vocal system characteristics (*ie* envelope of spectrum), Fig 9. As a result of the separation which occurs, the cepstrum method is suitable for the analysis of formant frequencies as well as larynx excitation frequency. In order to extract automatically the pitch from the resulting cepstrum of consecutive samples of the speech waveform (typically 20 ms) further algorithms must be applied to detect, with appropriate thresholds and boundaries, the peak in the cepstrum representing the quefrequency or period. For a more thorough and detailed description of the method the reader is referred to Refs 31 and 32.

Another approach to the analysis of the speech waveform is the use of inverse filtering in which the excited vocal tract is modelled by a time varying linear filter whose parameters may be determined by linear prediction analysis[33,34]. Numerous other algorithms have been proposed for pitch extraction, see Refs 35-44. An interesting comparison of some of these algorithms is presented by Rabiner *et al*[45].

One of the problems associated with these algorithms is that their implementation on a general purpose minicomputer does not allow analysis in real time. Special purpose hardware can and has been used to implement some of these algorithms allowing analysis in real time, however such hardware can be both complex and expensive. The analysis of bandwidth limited and/or noisy signals, also poses a restriction on the type of algorithm which can be used for analysis, such methods as the cepstrum analysis for example being suited to bandwidth limited signals such as would be obtained from telephone speech. This method can also be used with some success on noisy signal. A digital hardware cepstrum processor

used at the JSRU has proved capable of tracking the pitch of a pilot's voice in fairly high background noise levels.

The pitch extractor to be described here employs a simple frequency domain algorithm similar to that developed at the JSRU and allows implementation in real time using fairly inexpensive analogue hardware. This particular algorithm has also proved to work well with signals containing high background noise levels.

4.2 Description of the algorithm

The basis of the algorithm is the extraction of the envelope of a bandpass filtered speech waveform processed through a full wave rectifier.

If a speech signal is passed through a bandpass filter whose pass band is 300-600 Hz the resulting signal will contain some harmonics of the fundamental frequency and may in the case of a child or female contain the fundamental itself. The harmonics present in the pass band will beat together and the resulting envelope of the signal will correspond to the fundamental frequency. In order to extract this 'envelope' frequency the bandpass filtered signal is full wave rectified, then high pass filtered at 90 Hz to remove the dc component and passed through a variable low pass filter. The cut off frequency of the low pass filter must be such that only the fundamental frequency component or envelope frequency is passed. This is achieved by employing a 'feedback' system in which the output of the variable (tracking) low pass filter is fed to a phase locked loop (PLL) whose loop error voltage is then used to control the tracking filter, Fig 10. Once the PLL has acquired lock the cut off frequency of the tracking filter is moved by the control voltage to a value which admits only the fundamental frequency component. The 'idle' frequency of the voltage controlled oscillator of the PLL (*ie* frequency when no input signal present) must be such that the control voltage to the tracking filter produces a cut off frequency in the filter sufficiently high to allow the PLL to lock onto high voice pitches but not so high as to admit too high a proportion of the harmonics of low fundamentals causing 'pitch doubling' effects. A voiced/unvoiced decision is made simply on the basis of detecting the envelope of the full wave rectified fundamental frequency obtained from the pitch extractor, Fig 11. A variable threshold level at which the voiced/unvoiced decision is made is used to avoid decisions being made on spurious low amplitude signals derived from background noise. This method was found to be superior to the use of energy comparisons in high and low frequency bands for voiced/unvoiced detection when the speech material contained high background noise with a broad spectrum extending up to several kilohertz before falling off in intensity.

5 CIRCUIT DESCRIPTION

5.1 Pitch tracker

5.1.1 Bandpass filter

A commercial (KEMO) four pole variable Butterworth bandpass filter is used in the prototype pitch tracker having an attenuation in the stop bands of 48 dB per octave. The bandpass is set to 300-600 Hz and input signal levels required are of the order of 1-2 volts peak-to-peak.

5.1.2 Full wave rectifier

The output from the bandpass filter is fed to a full wave precision rectifier via a 10× amplifier. Precise rectification is achieved by means of an operational amplifier and diode configuration (Fig 12a) which operates as follows.

For positive inputs the first operational amplifier acts as an inverting amplifier, diode D_1 conducts and D_2 is cut off. The second operational amplifier acts as a unity gain inverter and the output is positive

$$V_o = + V_{in} .$$

Negative inputs cause D_2 to conduct and cut off D_1. For amplifier A_1:

$$\frac{V_{in}}{R} = - \left[\frac{V_o}{3R} + \frac{V_Y}{R}\right] .$$

Since

$$V_Y = V_Z = V_o \frac{2R}{3R}$$

$$\frac{V_{in}}{R} = - \left[\frac{V_o}{3R} + \frac{2}{3}\frac{V_o}{R}\right] = - \frac{V_o}{R}$$

hence

$$V_o = - V_{in} .$$

Thus for negative inputs the output is positive. The envelope of the full wave rectifier output is the fundamental frequency.

5.1.3 90Hz high pass filter

The output from the full wave rectifier is fed to a 90Hz high pass filter with a third order Butterworth response which removes the dc component of the signal and any low frequency syllabic modulations.

5.1.4 Tracking low pass filter

The tracking low pass filter is designed around the so-called 'two integrator loop' (or analogue computing loop) using operational transconductance amplifiers (RCA CA3080) to provide a voltage controllable current source to the integrators of the loop. A 'fourth order' filter system is used in the pitch tracker but since this consists of two identical two integrator loops cascaded together, only one loop will be described in detail.

The transfer operator of a filter may be regarded as a symbolic representation of a linear differential equation[46] which can be solved using standard analogue computing techniques. A second order low pass filter voltage transfer function has the form (Ref 46, p 69).

$$\frac{V_o(s)}{V_i} = \frac{A_0}{1 + b\frac{s}{\omega_0} + \frac{s^2}{\omega_0^2}} \qquad (1)$$

where s = complex variable (frequency domain)

b = coefficient

A_0 = scaling factor

ω_0 = cut off frequency.

If the complex variable is replaced by the differential operator d/dt and the equation rearranged, a differential equation results which can be solved by using two integrators and an adder (Fig 12b).

$$\frac{d^2}{dt^2}\frac{1}{\omega_0^2}V_o = A_0V_i - \frac{b}{\omega_0}\frac{d}{dt}V_o - V_o \quad . \qquad (2)$$

The operation of the 'two integrator loop' may be described as follows, referring to Fig 12b.

$$V_{o_2} = -RC\frac{d}{dt}V_o = -\frac{d}{dt}\frac{V_o}{\omega_0} \qquad \text{since } \frac{1}{RC} = \omega_0$$

$$V_{o_1} = -RC\frac{d}{dt}V_{o_2} = +(RC)^2\frac{d^2}{dt}V_o = \frac{d^2}{dt}\frac{V_o}{\omega_0^2} \quad .$$

It can be shown that

$$V_{o_1} = \frac{2}{R_2 + R_1}\left[V_i R_2 + V_{o_2} R_1\right] - V_o \quad .$$

Substituting for V_{o_1} , V_{o_2}

$$\frac{d^2}{dt^2}\frac{V_o}{\omega_0^2} = \frac{2}{R_2 + R_1}\left[V_i R_2 - R_1 \frac{d}{dt}\frac{V_o}{\omega_0}\right] - V_o \quad . \tag{3}$$

which is the equation we have in (2)

$$\text{where} \quad A_0 = \frac{2R_2}{R_2 + R_1} , \qquad b = \frac{2R_1}{R_1 + R_2} \quad .$$

It can also be shown that V_o is the low pass response by arranging (3) in the form of the general low pass transfer function (1).

$$\frac{V_o}{V_i}\left(\frac{d}{dt}\right) = \frac{V_{o_3}}{V_i}\left(\frac{d}{dt}\right) = \frac{\dfrac{2R_2}{R_2 + R_1}}{1 + \dfrac{2R_1}{R_2 + R_1}\dfrac{d}{dt}\dfrac{1}{\omega_0} + \dfrac{d^2}{dt^2}\dfrac{1}{\omega_0^2}} \quad .$$

The Q of the filter may be adjusted by R_2 where $(R_2)/(R_1) = 2Q - 1$. A 'true' fourth order low pass filter should be obtained by using the fourth order transfer function with a differential operator and solving the equation with the appropriate analogue computing elements in the manner described for the second order response.

Using fixed values of c it can readily be seen that varying R (Fig 12b) allows tuning of the cut off frequency (ω_0) of the filter. In order to allow the value of R to be controlled electrically, operational transconductance amplifiers (OTA) have been utilised, providing voltage controllable resistors to the integrators of the two integrator loop (Fig 13). Only a brief description of the OTA will be given here to provide sufficient understanding of its operation in the two integrator loop, for a more detailed description the reader is referred to Ref 47.

The output current of the OTA (I_{out}) is proportional to the differential input voltage (ε_{in}) and the transconductance (g_m) of the amplifier determined by the amplifier bias current (I_{ABC}) and the differential input voltage

$$g_m = \frac{\Delta I_{out}}{\Delta \epsilon_{in}}$$

also

$$g_m = 19.2 I_{ABC}$$

hence

$$I_{out} = (19.2 I_{ABC})\epsilon_{in} \quad .$$

This equation is analogous to the Ohms law equation

$$I = \frac{V}{R}$$

where $\frac{1}{R} = 19.2 I_{ABC}$.

From these equations it can be seen that variation of the amplifier bias current I_{ABC} provides a current controllable resistor which by the use of a resistive input to the amplifier bias input may be made voltage controllable. The control voltage is derived from the phase locked loop error voltage with appropriate scaling (this is discussed in section 5.1.6) and may be related to the cut off frequency in the following manner. Referring to Fig 13

$$I_1 = \frac{dQ}{dt} = c\frac{dV_o}{dt}$$

where Q = charge on the capacitor

and

$$I_2 = \frac{R_b}{R_a + R_b} V_{o_2} g_m$$

$$= \frac{R_b}{R_a + R_b} V_o 19.2 I_{ABC}$$

now

$$I_1 + I_2 = 0$$

(By Kirchoffs Law) assuming 'virtual earth' at the input of the amplifier.

$$\therefore \; -c\frac{dV_o}{dt} = \frac{R_b}{R_a + R_b} V_{o_2} 19.2 I_{ABC} \quad .$$

056

$$V_{o_2} = - \frac{C}{19.2 I_{ABC} \alpha} \frac{dV_o}{dt} \qquad (4)$$

where $\alpha = \dfrac{R_b}{R_a + R_b}$.

From the discussion on the simple two integrator loop using a resistor in the integrator we found that:

$$V_{o_2} = - RC \frac{dV_o}{dt}$$

$$= - \frac{dV_o}{dt} \frac{1}{\omega_0}$$

where $\omega_0 = \dfrac{1}{RC}$ = the cut off frequency of the filter.

In the expression (4) for V_{o_2} where the resistor is replaced by an OTA the cut off frequency of the filter is thus defined as:

$$\omega_0 = \frac{C}{19.2 I_{ABC} \alpha} .$$

For the particular circuit described here:

$$I_{ABC} = \frac{-V_{s-} - V_C}{R_C}$$

where V_s = negative supply voltage to OTA.

The measured low pass filter response versus control voltage for two cascaded second order filters is shown in Fig 14.

5.1.5 Dc decoupling

Dc decoupling is achieved with the circuit of Fig 15a which is an ac coupled non inverting follower.

The high frequency gain is unity and the dc gain zero with the low frequency cut off being determined by the values of R and C

$$f_c = \frac{1}{2\pi RC} .$$

Values of 27 kΩ for R and 0.1 μF for C give a low frequency cut off of around 60 Hz.

5.1.6 Phase locked loop

The phase locked loop (PLL) used is a SIGNETICS 565 with additional circuitry added to extend its lock range (Ref 46, p 239). Details of the operation of the 565 PLL are given in Ref 46, Chapter 7 and Ref 48 but will not be given here. To enable the loop to acquire lock over a bandwidth extending from 60 Hz to 400 Hz with a free running voltage controlled oscillator (VCO) frequency of around 150 Hz an extended lock range is required since the lock range of the 565 in a standard configuration is only around ±60% of the free running VCO frequency. ?

(a) Extended range

Under normal circuit configuration (Fig 15b) the free running frequency of the VCO is determined by R_1 and C_1

where $f_0 = \frac{1.2}{4R_1C_1}$ Hz (Manufacturer's design information).

The free running frequency is defined as the oscillation frequency of the VCO without input signal and both inputs grounded. Capture range is the range of frequencies either side of the f_0 over which the PLL will acquire lock with an input signal initially starting out of lock. The lock range or tracking range is defined as the range of frequencies either side of f_0 that the VCO will remain locked to once locked on to an input signal.

To extend the lock range of the PLL to greater than ±60% of f_0 the circuitry described in Ref 46, pp 238-239 is used (Fig 16) in which increased loop gain is achieved by supplying all the VCO charging current into pin 8 using an external current source (two BCY 71's) controlled by the loop error voltage. With this circuit configuration the frequency of the VCO now becomes a function of the charging current I_R and C1 (Fig 16) where:

$$f_{VCO} = \frac{5}{2}\frac{I_R}{V_s C_1}$$

056

and

$$I_R = \frac{V_o}{R_B} .$$

The lock range can be controlled by the resistor R_B which sets the gain of the 741 operational amplifier and the centre frequency may be adjusted by altering the voltage to the non inverting terminal of the operational amplifier.

(b) Loop error voltage and filter control voltage

The free running VCO frequency is set at 150 Hz using a timing capacitor C_1 of 0.47 μF and adjusting the voltage to the non inverting terminal of the operational amplifier with a 50kΩ potentiometer connected as a potential divider between +15 volts and ground (Fig 16). At this frequency the output voltage V_o from the operational amplifier is 10 volts, this voltage is used to establish a cut off frequency for the tracking low pass filter of 150 Hz with no input signal to the filter or PLL. A graph of amplifier output voltage (loop error voltage) versus the VCO frequency is shown in Fig 17. The cut off frequency of the tracking filter when no input signal is present is crucial to determining what range of voice fundamental frequencies may be tracked by the filter. Selection of too high a cut off frequency may result in the PLL locking onto the second harmonic of the glottal source spectrum from low pitched voices whilst too low a cut off frequency would prevent the PLL from locking onto the fundamental frequency of high pitched voices such as females or children's voices. With a bandpassed speech signal input to the full wave rectifier (as described in earlier sections) the tracking low pass filter produces the envelope frequency at its output. With this signal input to the PLL, the loop error voltage changes causing the VCO to oscillate at the same frequency as the input signal (providing the input signal is within the capture range of the PLL). The loop error voltage is also used to control the cut off frequency of the tracking filter so that as the PLL locks onto the output from the filter its cut off frequency is optimised for the envelope frequency. As the envelope or fundamental frequency varies, so the loop error voltage varies controlling the tracking filter and thus the filter tracks the fundamental frequency. Using the tracking filter response plot of cut off frequency versus control voltage shown in Fig 14 and the plot of loop error voltage versus VCO frequency for the PLL shown in Fig 17, a suitable transfer function may be developed to use the loop error voltage to control the cut off frequency of the filter and set optimum cut off frequencies for the fundamental frequency being tracked. By plotting the filter control voltage required to extract

particular fundamental frequencies, versus the loop error voltage and VCO frequency, the transfer function is obtained and is found to be a simple linear relationship (Fig 18). This transfer function is implemented using a single operational amplifier with the gain set to the value of the reciprocal of the slope of the line and an 'offset' voltage applied to the non-inverting terminal equal to the constant in the equation

$$V_o = mV_C + C$$

$$V_C = \frac{1}{M}(V_o - C)$$

where V_o = loop error voltage (output from operational amplifier, see Fig 16)
V_C = filter control voltage
m = slope
C = constant.

From Fig 18 it can be seen that the required values of m and C are -0.66 and 5 volts respectively.

The loop error voltage (V_o) is smoothed using an RC low pass filter of time constant 0.1 second and buffered by a voltage follower before being passed to the scaling circuit (Fig 19a) which achieves the transfer function described previously.

5.2 Voiced/unvoiced decision circuitry

5.2.1 Full wave rectifier

The fundamental frequency derived from the tracking low pass filter output described in section 4.1.4 is amplified by a factor of 10 and then full wave rectified using a similar circuit to that described in section 4.1.2. This signal is then low pass filtered in order to obtain its envelope.

5.2.2 Low pass filter

A second order filter with a Butterworth response and a cut off frequency of 45 Hz is used to derive the envelope of the full wave rectifier fundamental frequency (Fig 19b). The general equation for the transfer function of a second order low pass filter is shown in equation (1).

For an active filter A_0 is the gain of the operational amplifier (Ref 46, pp 54-59)

$$A_0 = 1 + \frac{R_b}{R_a} = K \qquad \text{(Fig 19b)}$$

$$\omega_0 = \sqrt{\frac{1}{R_1 R_2 C_1 C_2}}$$

$$b = \sqrt{\frac{R_2 C_2}{R_1 C_1}} + \sqrt{\frac{R_1 C_1}{R_2 C_1}} + \sqrt{\frac{R_1 C_1}{R_2 C_2}} - K \sqrt{\frac{R_1 C_1}{R_2 C_2}} \; .$$

If for simplicity of design $R_1 = R_2 = R$ and $C_1 = C_2 = C$ then ω_0 reduces to

$$\omega_0 = \frac{1}{RC}$$

and the expression for b reduces to:

$$b = 3 - K = 3 - A_0 \; .$$

For a Butterworth response the value of b in the Butterworth polynominal may be found in standard filter design tables and for a second order filter is 1.4142. Using this value the gain of the amplifier may be obtained

where $A_0 = 3 - b \cong 1.6$

also

$$A_0 = 1 + \frac{R_b}{R_a}$$

hence

$$\frac{R_b}{R_a} = 0.6$$

the values for R_a and R_b in this case chosen to be 5.6 kΩ and 3.3 kΩ respectively.

The envelope of the fundamental frequency thus derived is then fed to a comparator which compares this low frequency varying voltage with another preset voltage.

5.2.3 Voltage comparator

The circuit consists basically of a biassed crossing detector (Fig 20) formed around a single operational amplifier whose output voltage goes negative when the input voltage slightly exceeds the reference voltage of opposite polarity. When this criterion is not met the output voltage is a fraction of a volt. With no input signal to the comparator but with a negative reference current i_{ref} flowing from the non-inverting terminal of the amplifier due to the reference voltage V_{ref}, the diode D1 conducts when the current exceeds a few microamperes and an output voltage of 0.5 volt to 0.6 volt is obtained. When an input current i_{in} fractionally greater and of opposite sign to the reference current present, the output voltage swings negative to the value of the zener voltage thus keeping the sum of the currents at the inverting terminal zero.

By adjusting the reference voltage, the threshold at which the comparator switches to its negative output state may be altered. Adjustment of the reference voltage allows for a threshold to be set just above the envelope voltage of any spurious signals derived from the tracking filter due for example to noise. The reference voltage is set by a ten turn potentiometer which, with a multiturn dial, allows precise setting of the threshold and the required settings for particular speech material to be noted.

The output from the comparator, which indicates the presence of voicing in the speech signal, is used to gate the output from a frequency to voltage converter used to convert the fundamental frequency, derived in this case from the VCO of the phase locked loop, to a voltage level.

5.2.4 Frequency to voltage converter

For display purposes and to provide a suitable signal for feeding into an analogue to digital converter, the fundamental frequency is converted into a voltage level by a frequency to voltage converter shown in Fig 21.

The square wave output from the VCO of the phase locked loop is used for convenience in the frequency to voltage conversion and is fed to the base of transistor 1 (Fig 21) via a capacitor C_1. Transistors Tr1 and Tr2 comprise a monostable oscillator and on the positive going edge of an input square wave a negative going pulse is obtained on the collector of Tr1 and a positive going pulse on the collector of Tr2. The negative going pulse is differentiated by a capacitor C_3 and fed to the base of Tr4 which is supplied with a constant current from Tr3, Tr3 and Tr4 together form a ramp generator. A positive spike is produced at the base of Tr4 by C_3 on the positive edge of the negative pulse

causing C_4 to discharge. Before C_4 discharges however, the leading edge of the positive pulse derived from the monostable oscillator switches on the transistor Tr5 allowing C_5 to charge to the value on C_4. C_5 is buffered by an LM308 operational amplifier which has an input impedance of 40 MΩ thus minimising leakage, the voltage on C_5 then appears at the output of the amplifier which is connected for unity gain. The output of the LM308 is fed to the source terminal of a field effect transistor (FET) controlled by the comparator output voltage, when the comparator output is negative (*ie* a voiced decision is made) the FET has a high resistance and the LM308 output is available at the source terminal of the FET. When the comparator is in the off state (*ie* an unvoiced decision is made) the output from the LM308 is grounded by the FET.

6 CONSTRUCTIONAL DETAILS

The fundamental frequency tracking circuitry is assembled on one 114 × 203 mm circuit board and the voiced/unvoiced circuitry on another. The two circuit boards are of the 'plug in' type (43 way) and are incorporated into a 19 inch rack mounting system with 2 inch circuit board fronts on which are mounted BNC input and output sockets and a ten turn precision dial controlling the voiced/unvoiced decision threshold (Fig 24).

A ±15 volt power supply (ITT Powercard PC250A) mounted in a 4 inch wide rack mounting module is fitted into the 19 inch rack and provides power for the two circuit boards.

7 DISCUSSION

7.1 Of all the acoustic features of the speech signal the fundamental frequency appears to be potentially the best candidate for revealing extra linguistic changes and hence giving some indication of the physiological and psychological state of a person. For this reason the fundamental frequency tracker described in this Report has been made to enable extra linguistic changes to be quantified. The realisation of the fundamental frequency tracker with real time operation avoids the use of a minicomputer to implement one of the algorithms described in section 3 which would in all probability not run in real time on a general purpose minicomputer. However a minicomputer can be used, with the fundamental frequency tracker interfaced to it, to perform statistical analysis on the fundamental frequency data derived from speech samples. The tracker will be interfaced to a PDP11/34 minicomputer system and analysis of pilots and air traffic controllers speech will be carried out. By taking around 20 seconds of voiced speech as samples such parameters as mean fundamental frequency, variance and standard

deviation will be calculated and some form of classification statistics applied to investigate whether significant separation in classes can be obtained for speech samples taken under different levels of 'strain'. If such separation can be obtained then it may be possible to correlate changes in fundamental frequency statistics with workload level.

The accuracy of the fundamental frequency tracker has yet to be checked but there are a number of ways in which this could be done and it is hoped that a future memorandum will report on the accuracy found using the various methods. There are three basic methods which could be used and these will be described briefly.

7.2 Subjective tests

The human ear and sound perception mechanism is very sensitive and may be used to assess the accuracy of a fundamental frequency tracker by outputing the frequency, preferably modified as a square wave or sinusoidal wave with some harmonics, in synchronism with the speech from which it is derived. Discontinuities in the tracked fundamental frequency will then be noticed as the extracted 'pitch' is heard alongside the actual voice 'pitch' of the speech.

7.3 Comparison with laryngograph measurements

The laryngograph gives an accurate record of the points of opening and closure of the vocal cords by an impedance measurement method across the larynx in the region of the vocal cords. Such equipment has been pioneered by University College London and arrangements have been made to make a recording of speech and laryngograph measurements simultaneously so that the same speech can be processed by the fundamental frequency tracker and the data derived from it compared with the excitation period data derived from the larynogograph measurements.

7.4 Comparison with mathematical algorithms

Implementation of some of the algorithms described in section 3 for speech processing and particularly cepstrum analysis will allow comparisons to be made between fundamental frequency derived from these methods and the pitch tracker described in this Report.

8 CONCLUSIONS

A low cost device for tracking the fundamental frequency of speech in real time has been achieved. It is particularly suited to processing speech embedded

in noise and will allow extensive statistical analysis of fundamental frequency data derived from pilots and air traffic controllers speech to be carried out to attempt to establish correlation between changes in fundamental frequency statistics and 'strain' in a subject, which may be caused by workload. Other investigations may also be carried out into the effects of vibration, noise and acceleration on the fundamental frequency of speech.

REFERENCES

No.	Author	Title, etc
1	J.H. Kirchner W. Laurig	The human operator in air traffic control systems. Ergonomics 14(5) pp 549-556 (1971)
2	I.W. Rohmert	Determination of stress and strain of air traffic control officers. AGARD Conference preprint No.216 (1977). Methods to assess workload
3	A.H. Roscoe	Use of pilot heart rate measurement in flight evaluation. Aviation, Space and Environ. Med., January 1976
4	J.D. O'Connor	Phonetics. Pelican Books (1973)
5	P. Lieberman S.B. Michaels	Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. J. Acoust. Soc. Am., Vol 34, 922-927 (1962)
6	G. Fairbanks	Recent experimental investigations of vocal pitch in speech. J. Acoust. Soc. Am., Vol 11, 457-466 (1940)
7	G. Fairbanks L. Hoaglin	An experimental study of the durational characteristics of the voice during the expression of emotion. Speech Monographs, Vol 8, pp 85-90 (1941)
8	G.L. Huttar	Relations between prosodic variables and emotions in normal American-English utterances. J. Speech and Hearing Res., 11(3), 481-487 (1968)
9	P.V. Simonov *et al*	Use of the invariant method of speech analysis to discern the emotional state of announcers. Aviation, Space and Environ. Med., pp 1014-1016, August 1975
10	P.V. Simonov M.V. Frolov	Analysis of the human voice as a method of controlling emotional state: achievements and goals. Aviation, Space and Environ. Med., pp 23-25, January 1977

REFERENCES (continued)

No.	Author	Title, etc
11	C.F. Williams K.N. Stevens	Emotion and speech - some acoustical correlates. J. Acoust. Soc. Am., Vol 52, 1238-1249 (1972)
12	C.F. Williams K.N. Stevens	On determining the emotional state of pilots during flight: an exploratory study. Aerospace Med., 40(12), pp 1369-1372 (1969)
13	I. Kuroda *et al*	Method for determining pilot stress through analysis of voice communication. Aviation, Space and Environ. Med., pp 528-533, May 1976
14	P.V. Simonov M.V. Frolov	Utilisation of human voice for estimation of mans emotional stress and state of attention. Aerospace Medicine, pp 256-258, March 1973
15	V. Popov *et al*	Analysis of intonational characteristics of speech as a criterion of the emotional state of man under conditions of space flight. USA Joint Publications Research Service TT 67-30554 (1967)
16	H.M.L. Hecker *et al*	Manifestations of task induced stress in the acoustic speech signal. J. Acoust. Soc. Am., 44(4), 993-1001 (1968)
17	A. Friedhoff *et al*	An electro-acoustic analysis of the effects of stress on the voice. J. Neuropsychiatry, 5(5), 266-272 (1964)
18	M. Alpert *et al*	Transient voice changes associated with emotional stimuli. Arch. Gen. Psychiatry 8, pp 362-365 (1963)
19	E.N. Whitman D.J. Flicker	A potential new measure of emotional state. J. Newark. Beth. Israel Hosp., 17, 167-172 (1966)
20	M. Metfessel	Technique for objective studies of the vocal art. Psychol. Monograph, 36, pp 1-40 (1926)

056

REFERENCES (continued)

No.	Author	Title, etc
21	C.T. Simon	The variability of consecutive wavelengths in vocal and instrumental sounds. Psychol. Monograph, 36, pp 41-81 (1926)
22	D. Lewis J. Tiffin	A psychophysical study of individual differences in speaking ability. Arch. Speech, 1, pp 43-60 (1934)
23	J.M. Cowan	Pitch and intensity characteristics of stage speech. Arch. Speech, 1, Suppl., pp 1-92 (1936)
24	A. Risberg	Fundamental frequency tracking. Proc. of 4th International Cong. of Phonetic Sciences, The Hague: Mouton (1962)
25	G. Fant	Acoustic theory of speech production. s'Gravonhage: Mouton and Co. (1960)
26	J.L. Flanagan	Speech analysis, synthesis and perception. New York: Springer-Verlag (1972)
27	D.B. Lindsey	Emotion. Handbook of Experimental Psychology. S.S. Stevens, ed., Wiley, New York (1951)
28	Y. Horü	Some statistical characteristics of voice frequency. J. Speech and Hearing Research, 18, 192-201 (1975)
29	J.D. Markel *et al*	Long term feature averaging for speaker recognition. IEEE Trans. Acoust. Speech and Sig. Proc., Vol ASSP-25, No.4 (1977)
30	J.D. Markel	The SIFT algorithm for fundamental frequency estimation. IEEE Trans. Audio and Electroacoustics, Vol AU-20, No.5 (1972)
31	A.M. Noll	Cepstrum pitch determination. J. Acoust. Soc. Amer., Vol 41, No.2 (1967)
32	A.M. Noll M.R. Schroeder	Short-time 'cepstrum' pitch detection. J. Acoust. Soc. Am., 36 (1964)

REFERENCES (continued)

No.	Author	Title, etc
33	B.S. Atal S.L. Hanauer	Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am., Vol 50, No.2, pt.2 (1971)
34	J.D. Markel A.H. Gray Jr	Linear prediction of speech. Springer-Verlag (1976)
35	M.J. Ross *et al*	Average magnitude difference function pitch extractor. IEEE Trans. Acoust. Speech and Sig. Proc., Vol ASSP 22 No.5 (1974)
36	Gold L. Rabiner	Parallel processing techniques for estimating pitch periods of speech in the time domain. J. Acoust. Soc. Am., Vol 46, No.2, pt.2 (1969)
37	J.J. Dubnowski *et al*	Real time digital hardware pitch detector. IEEE Trans. Acoust. Speech and Sig. Proc., Vol ASSP, No.1 (1976)
38	J.N. Maksym	Real-time pitch extraction by adaptive prediction of the speech waveform. IEEE Trans. Audio and Electroacoustics. Vol AU-21, No.3 (1973)
39	N.J. Miller	Pitch detection by data reduction. IEEE Trans. Acoust. Speech and Sig. Proc., Vol ASSP-23 No.1 (1975)
40	D.H. Friedman	Pseudo-maximum-likelihood speech pitch extraction. IEEE Trans. Acoust. Speech and Sig. Proc., Vol ASSP-25, No.3 (1977)
41	J.D. Wise *et al*	Maximum likelihood pitch estimation. IEEE Trans. Acoust. Speech and Sig. Proc., Vol ASSP-24, No.5 (1976)
42	C.A. McGonegal *et al*	A semi automatic pitch detector (SAPD). IEEE Trans, Acoust. Speech and Sig. Proc., Vol ASSP-23, No.6 (1975)

REFERENCES (concluded)

No.	Author	Title, etc
43	R.L. Miller	Peformance characteristics of an experimental harmonic identification pitch extraction (HIPEX) system. J. Acoust. Soc. Am., Vol 47 No.6, pt.2 (1970)
44	M.R. Schroeder	Period histogram and product spectrum: new methods for fundamental frequency measurement. J. Acoust. Soc. Am., Vol 43, No.4 (1968)
45	L. Rabiner *et al*	A comparative performance study of several pitch detection algorithms. IEEE Trans. Acoust. Speech and Sig. Proc., Vol ASSP-24, No.5 (1976)
46	G.B. Clayton	Linear integrated circuit applications. Macmillan (1975)
47	RCA	Solid state data book on linear integrated circuits. pp 63-78 (1974)
48	Signetics	Phase locked loops. Applications and data sheets.

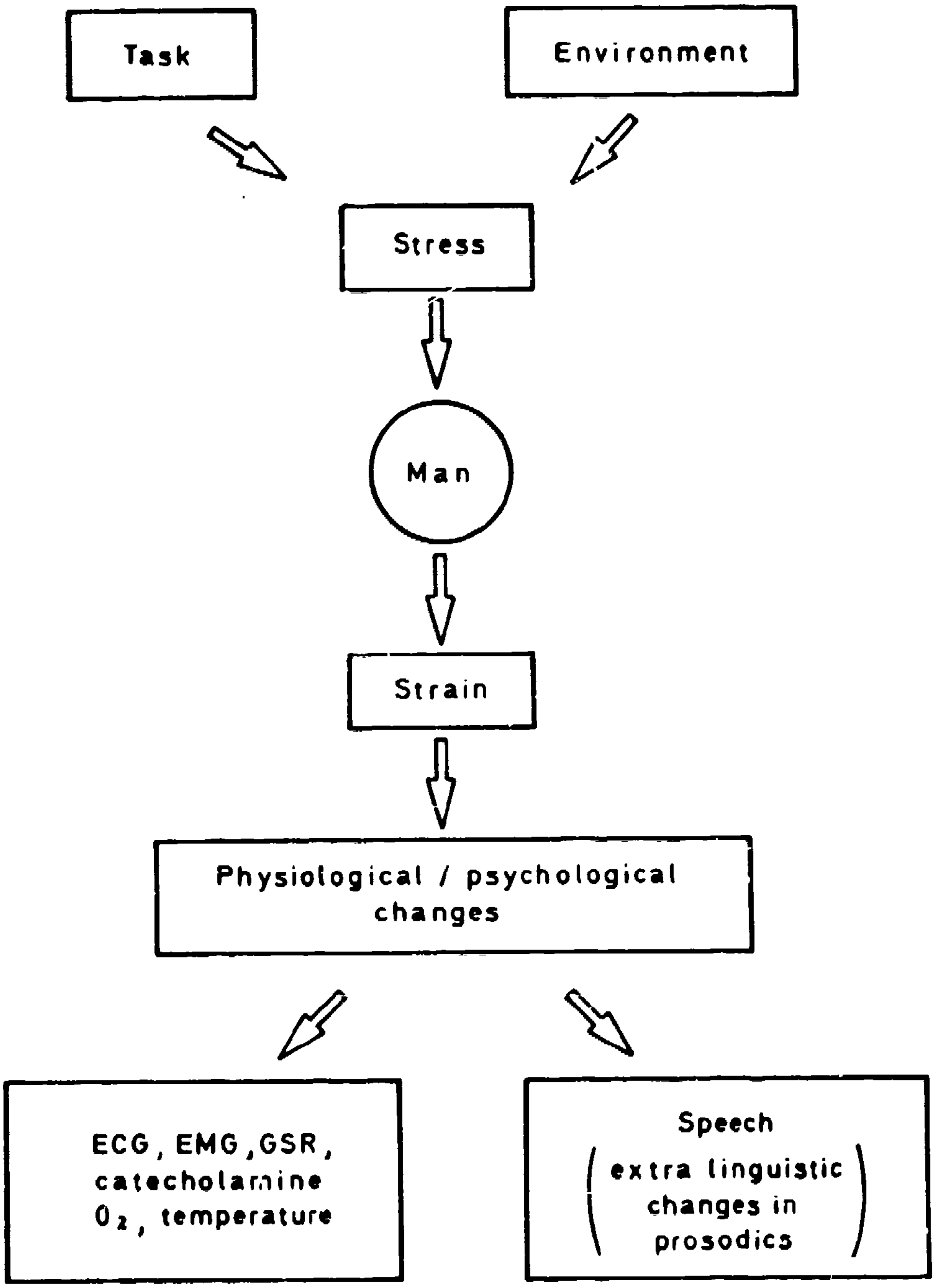


Fig 1 Diagram showing relationship between 'task' and environment and physiological changes in a 'man at work'

Fig 2

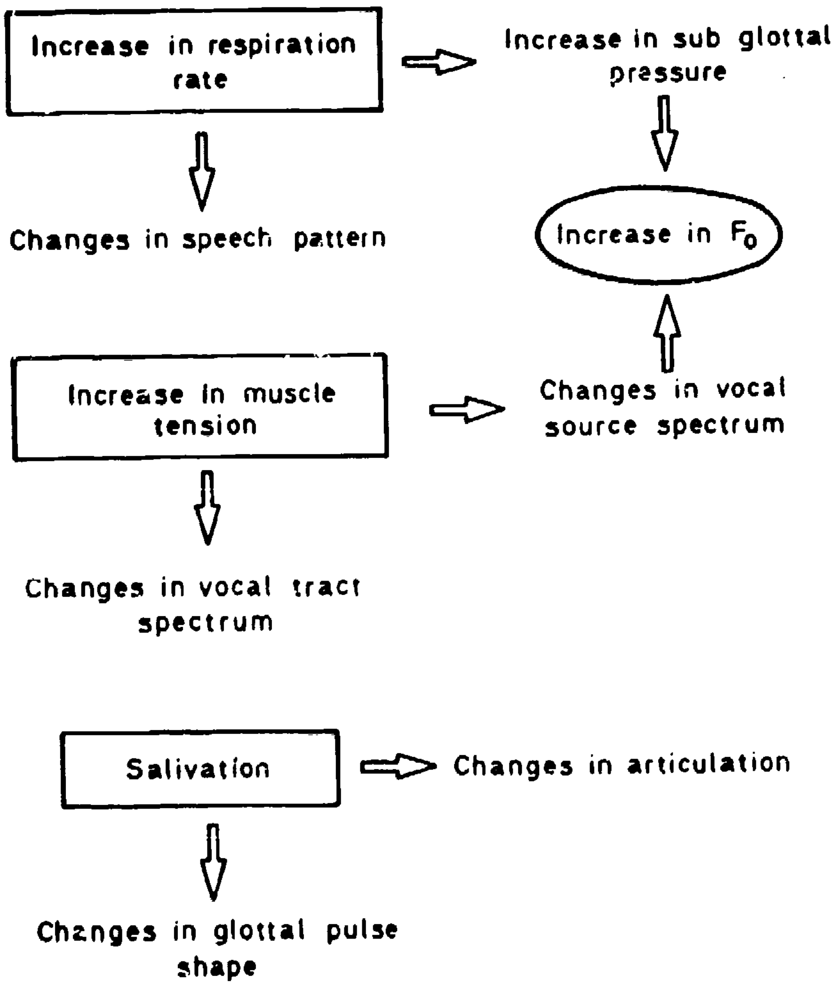


Fig 2 Physiological changes and their possible effect on speech

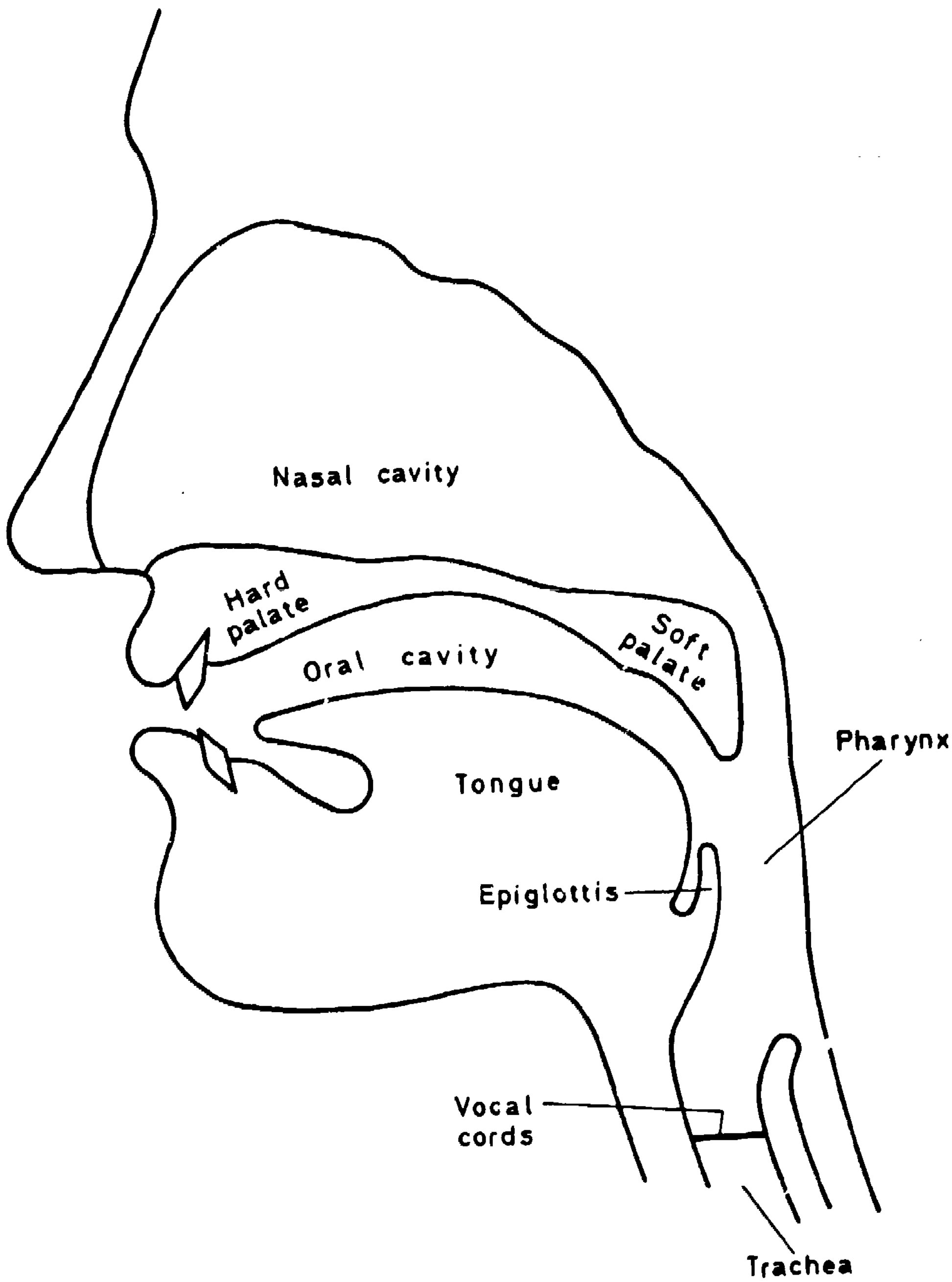


Fig 3 Supra glottal speech organs

TR 79056

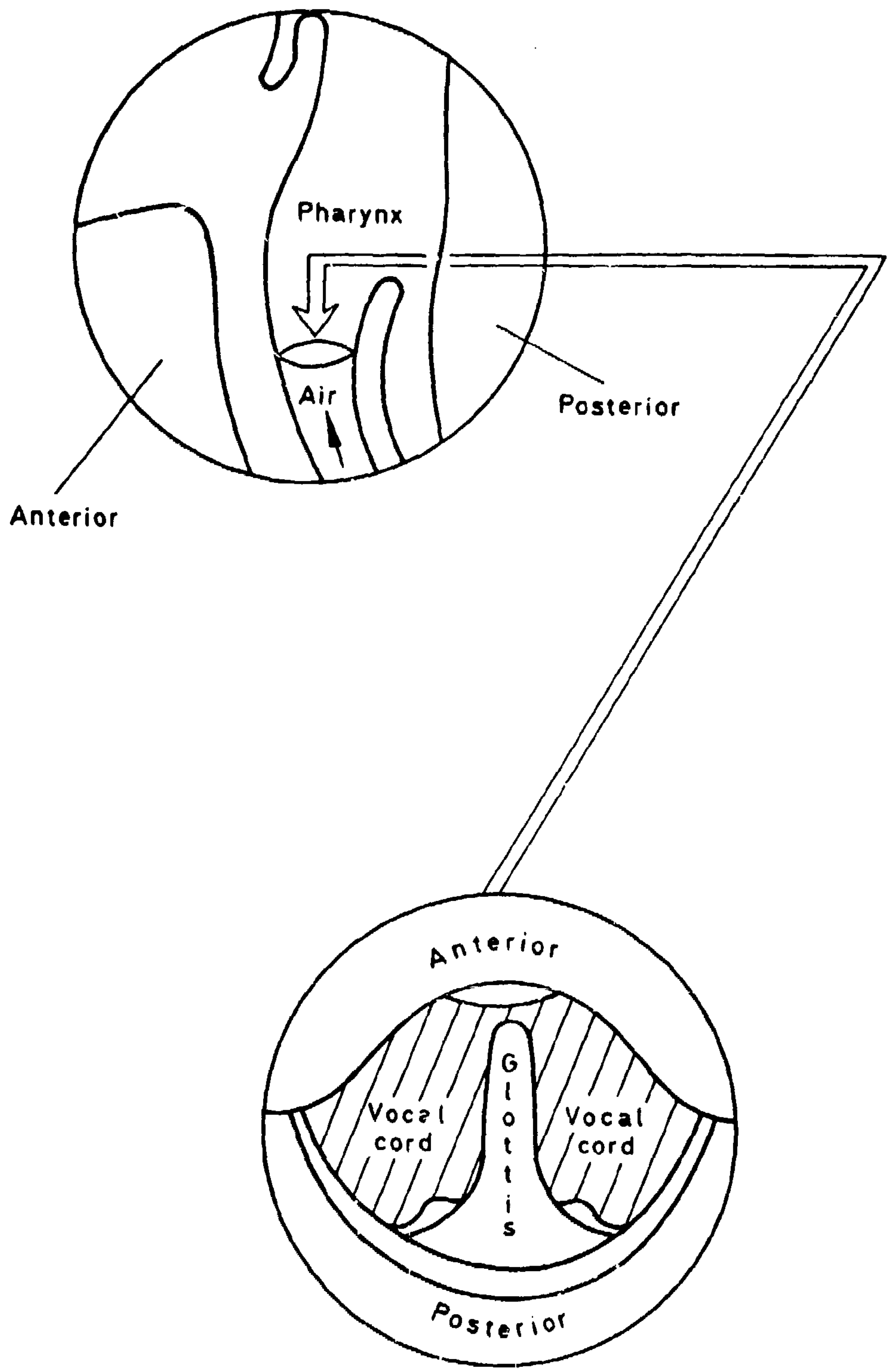


Fig 4 Cross-section of vocal cords open

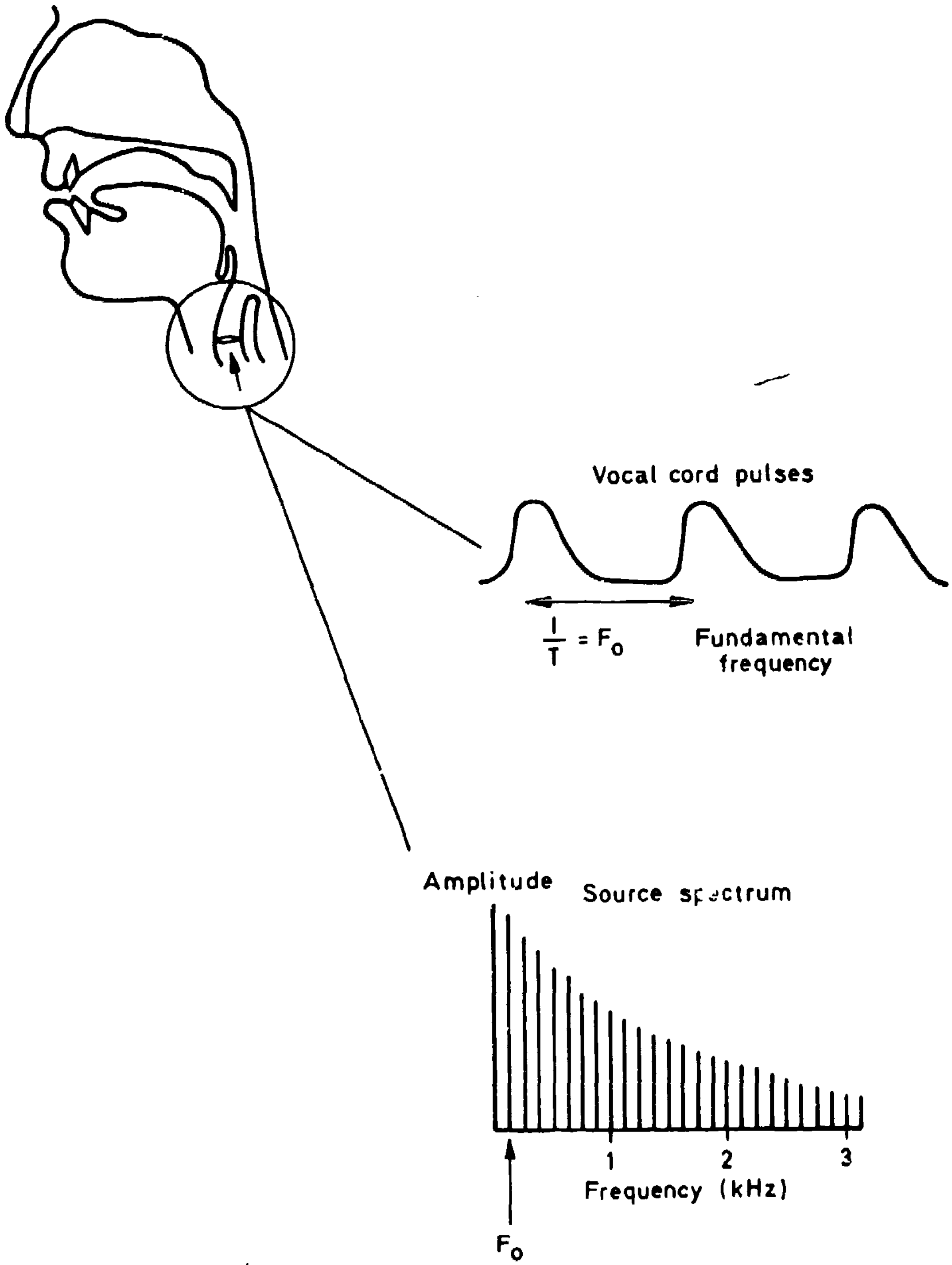


Fig 5 Vocal cord pulses and spectrum for voiced sounds

TR 79056

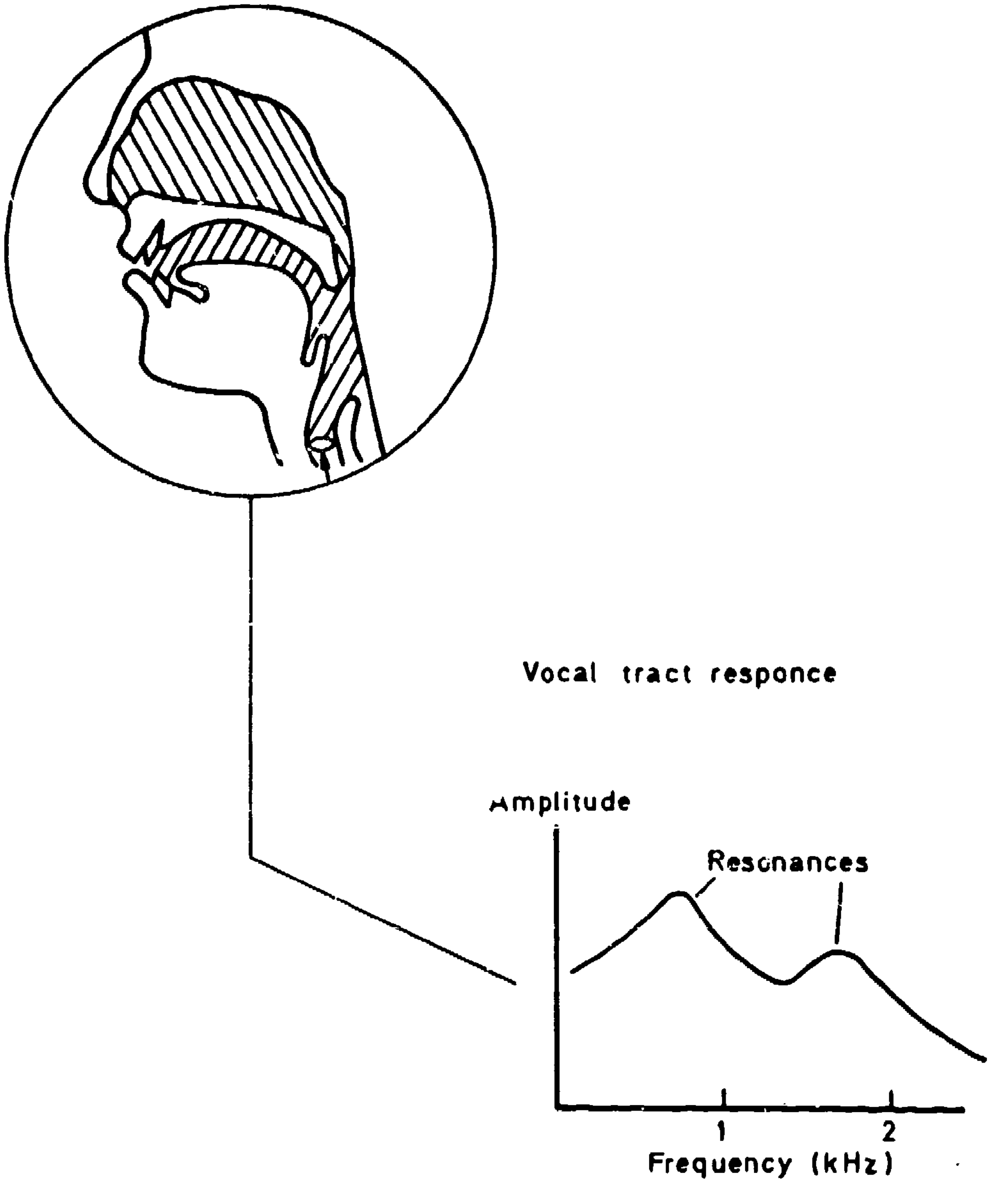


Fig 6 Vocal tract response

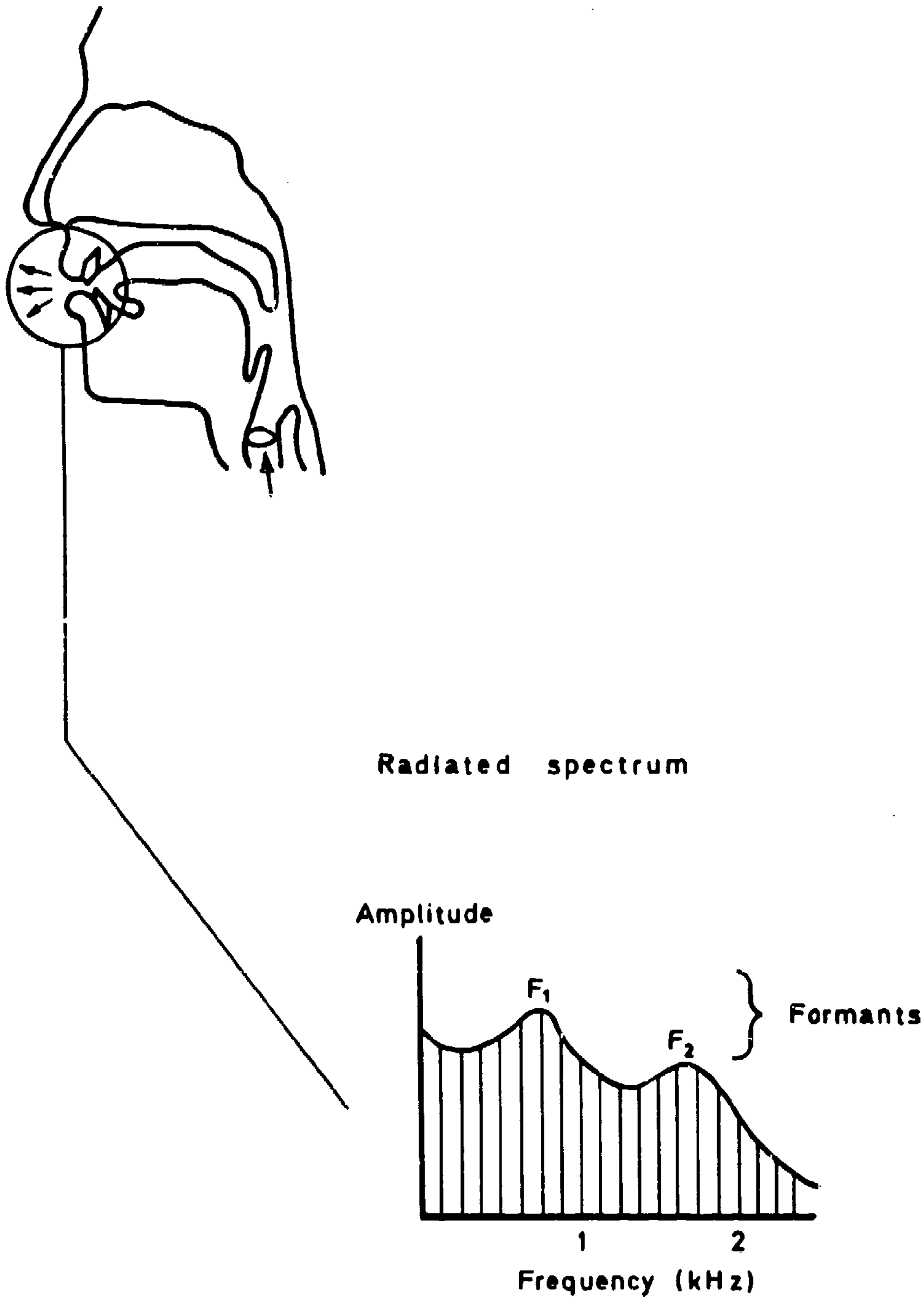


Fig 7 Radiated spectrum

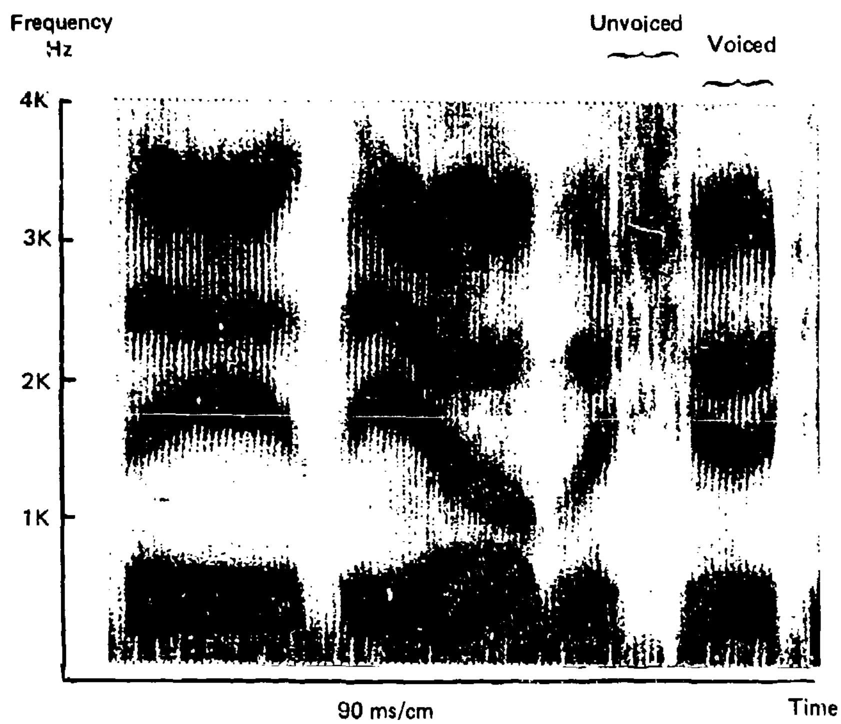


Fig 8 Wideband spectrogram

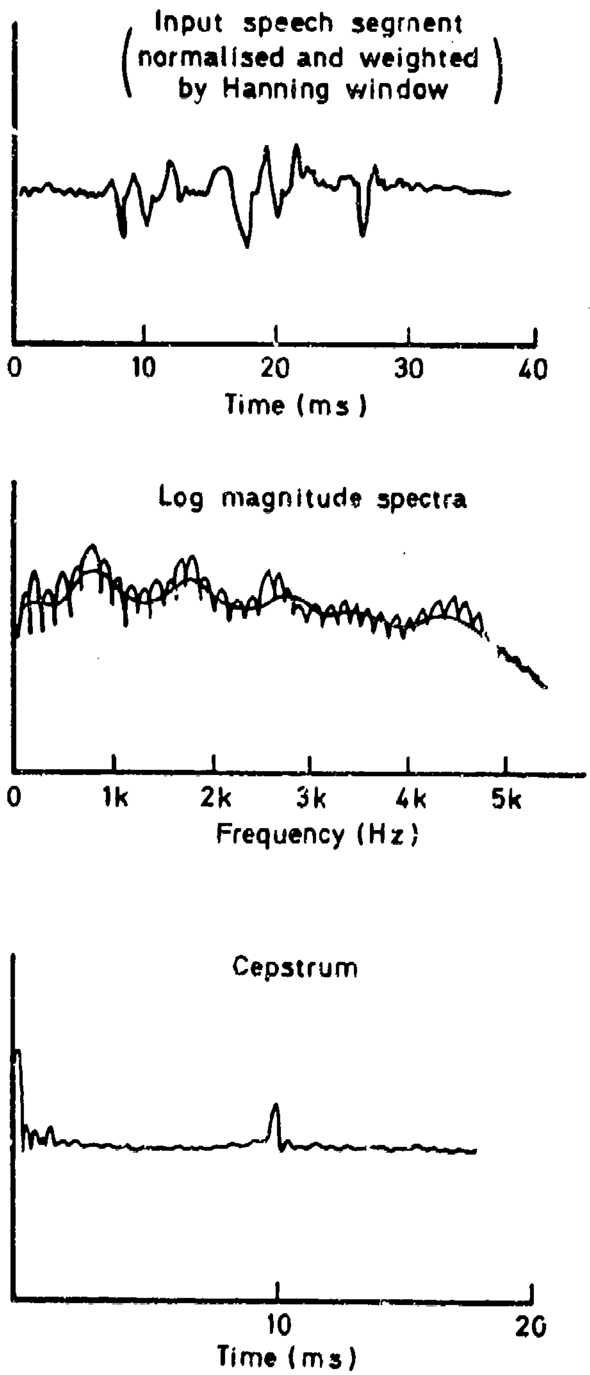


Fig 9 Cepstrum analysis of voiced sound

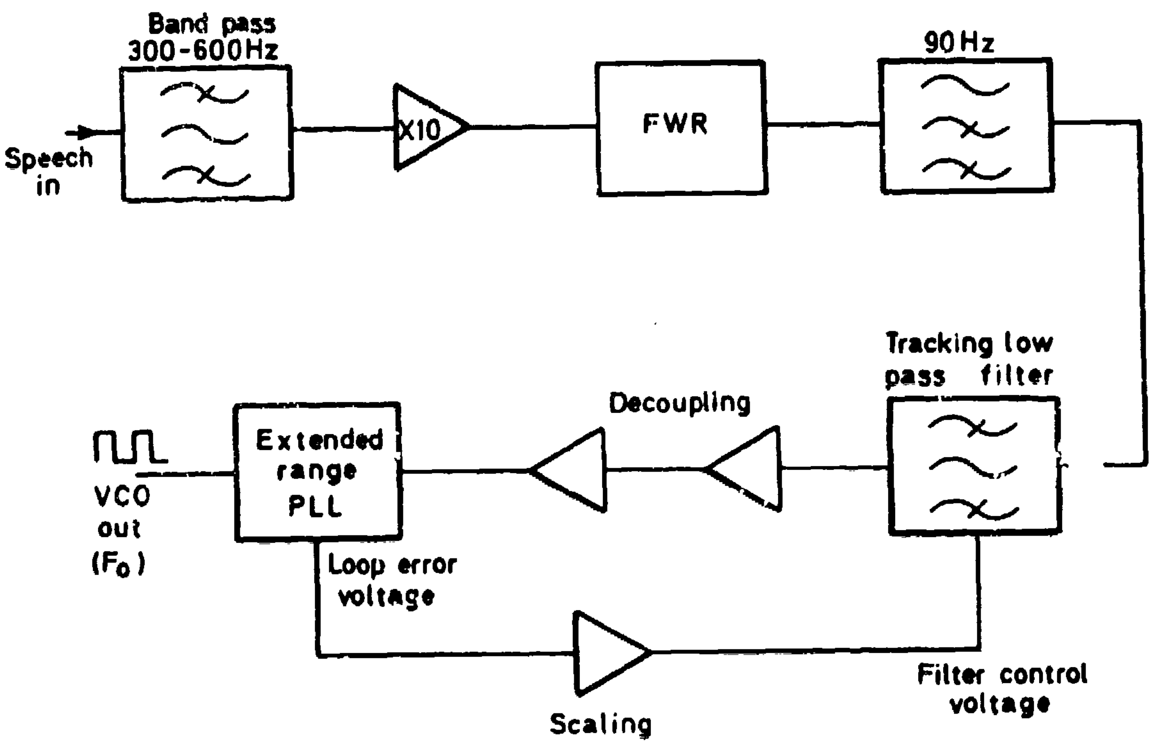


Fig 10 Pitch tracking module — block diagram

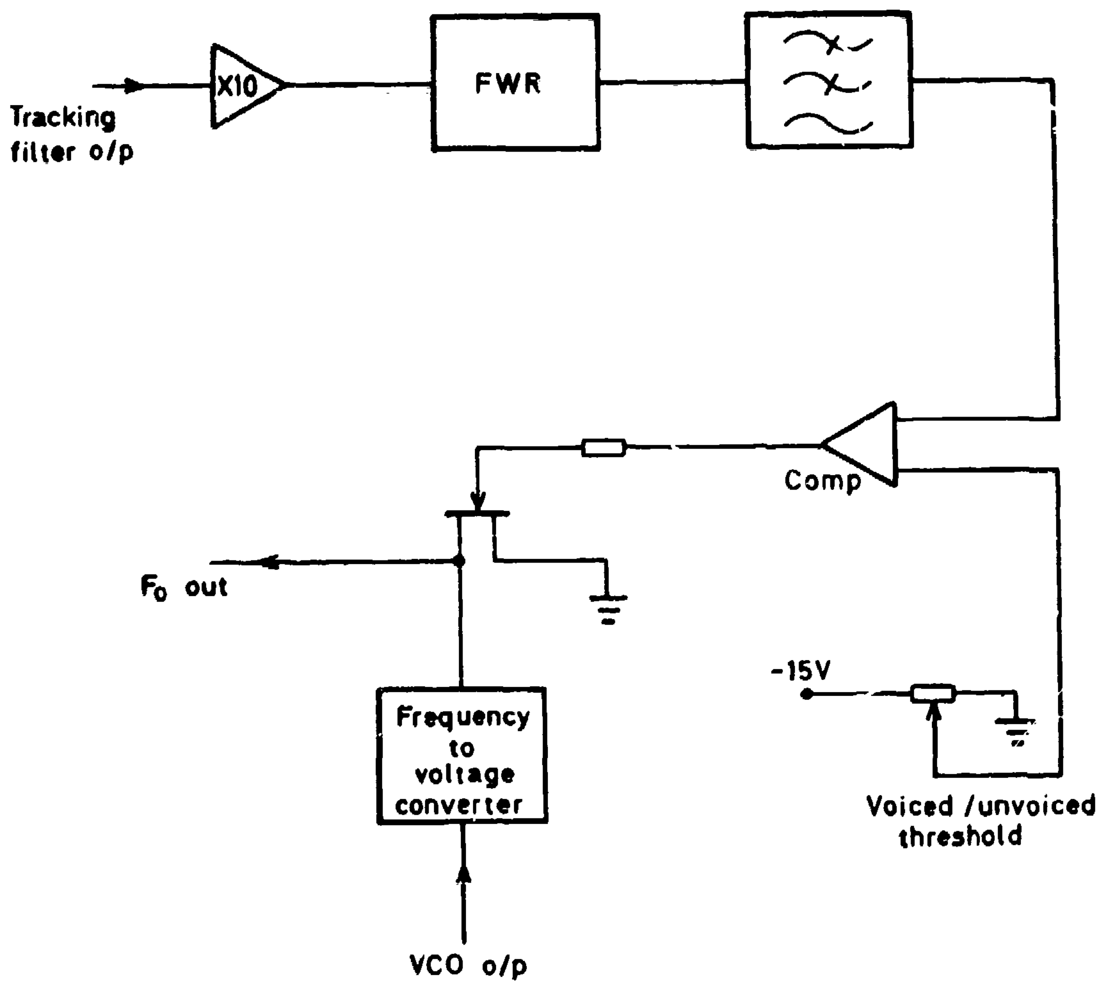
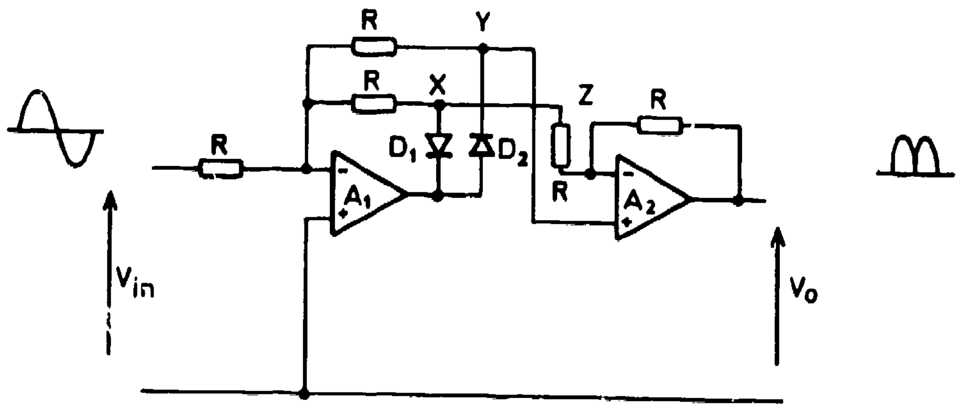
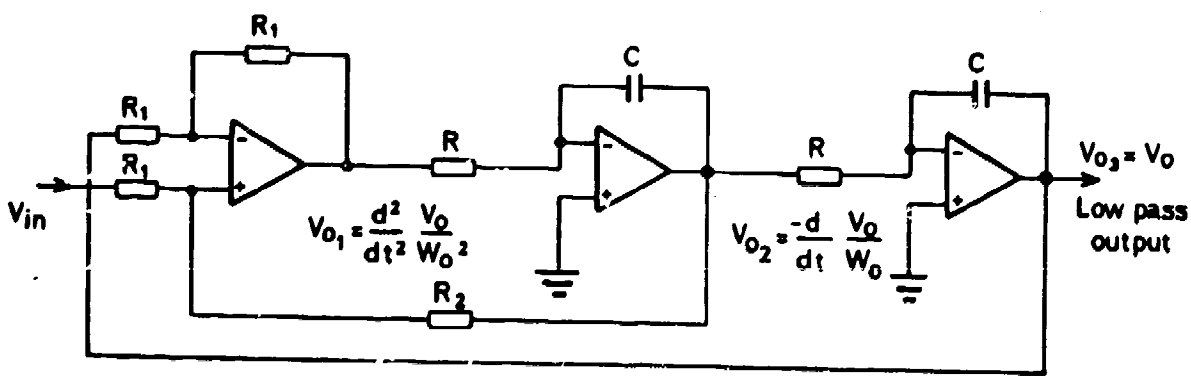


Fig 11 Voiced/unvoiced module – block diagram



a Precision rectifier



b Low pass filter using analogue computing techniques

Fig 12a&b Precision rectifier and low pass filter

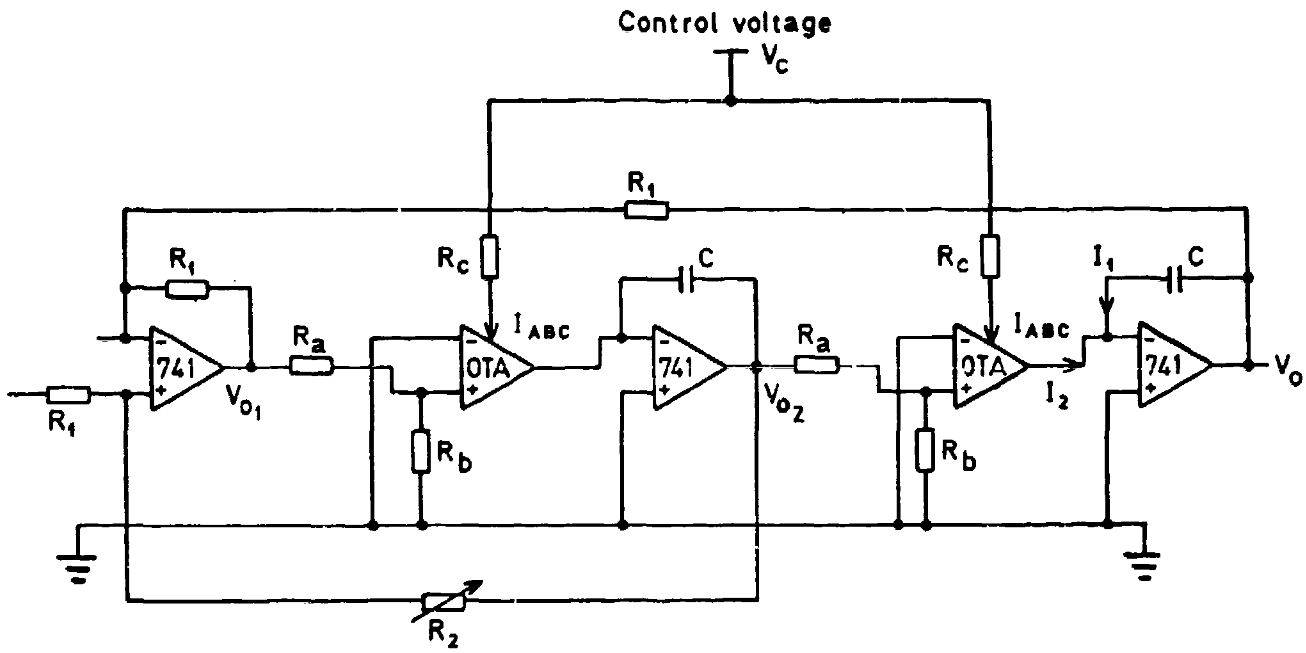


Fig 13 Second order low pass filter voltage controllable ω_o

TR 79056

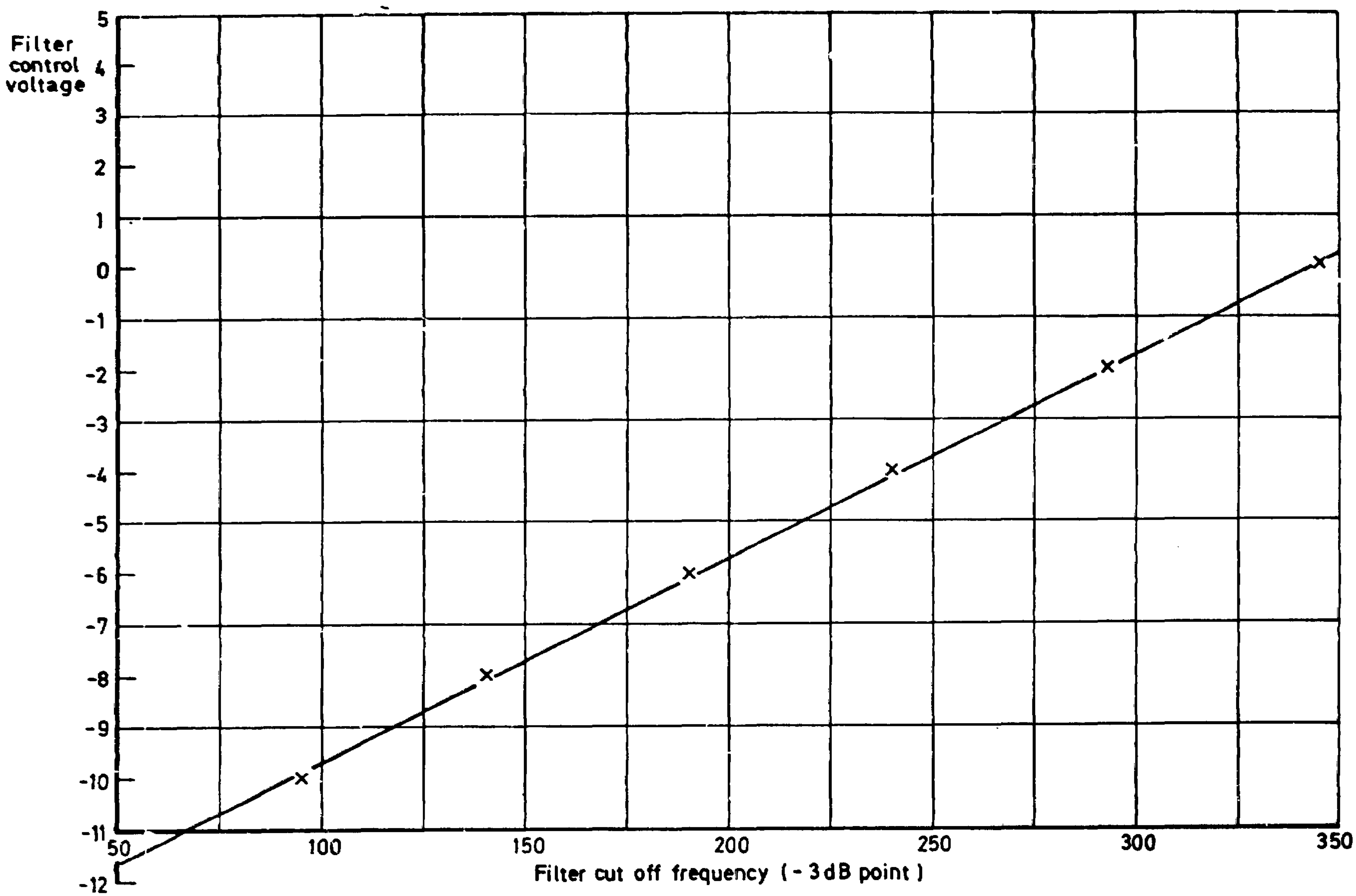
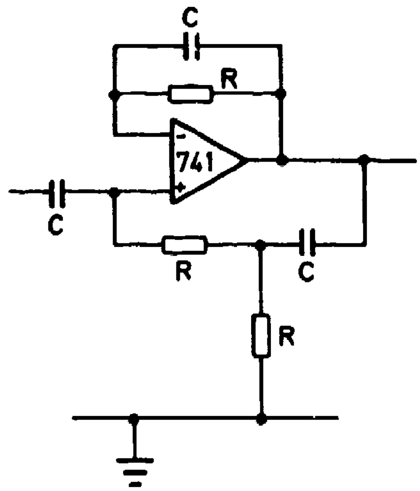
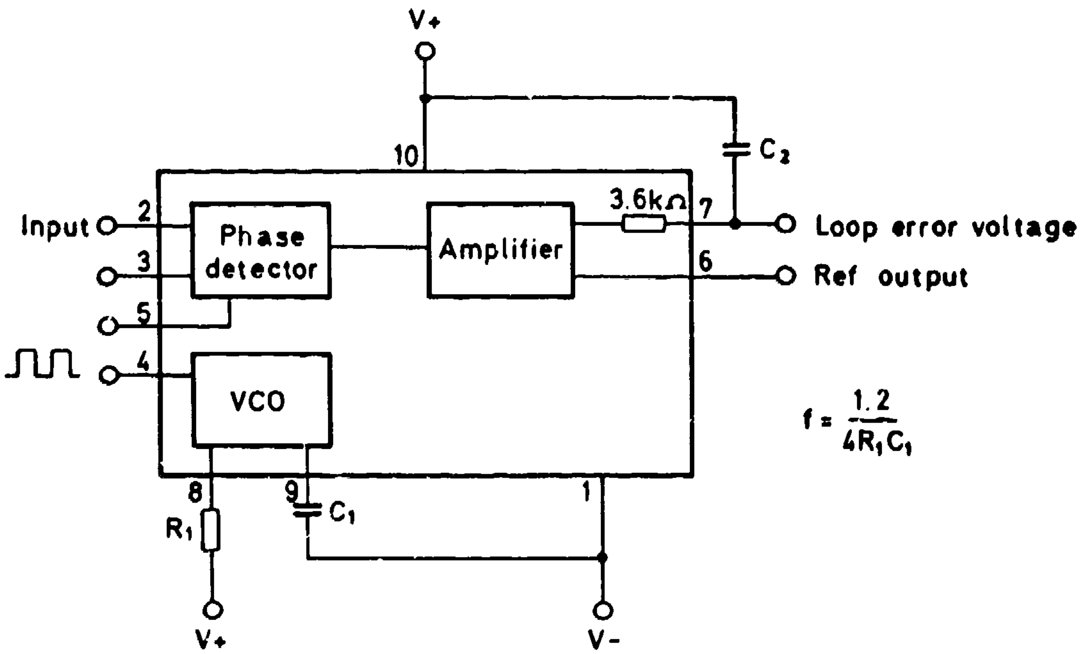


Fig 14 Tracking filter characteristics



a AC coupled follower



b Signetics 565 phase locked loop

Fig 15a&b Alternating current follower and phase locked loop

TR 79056

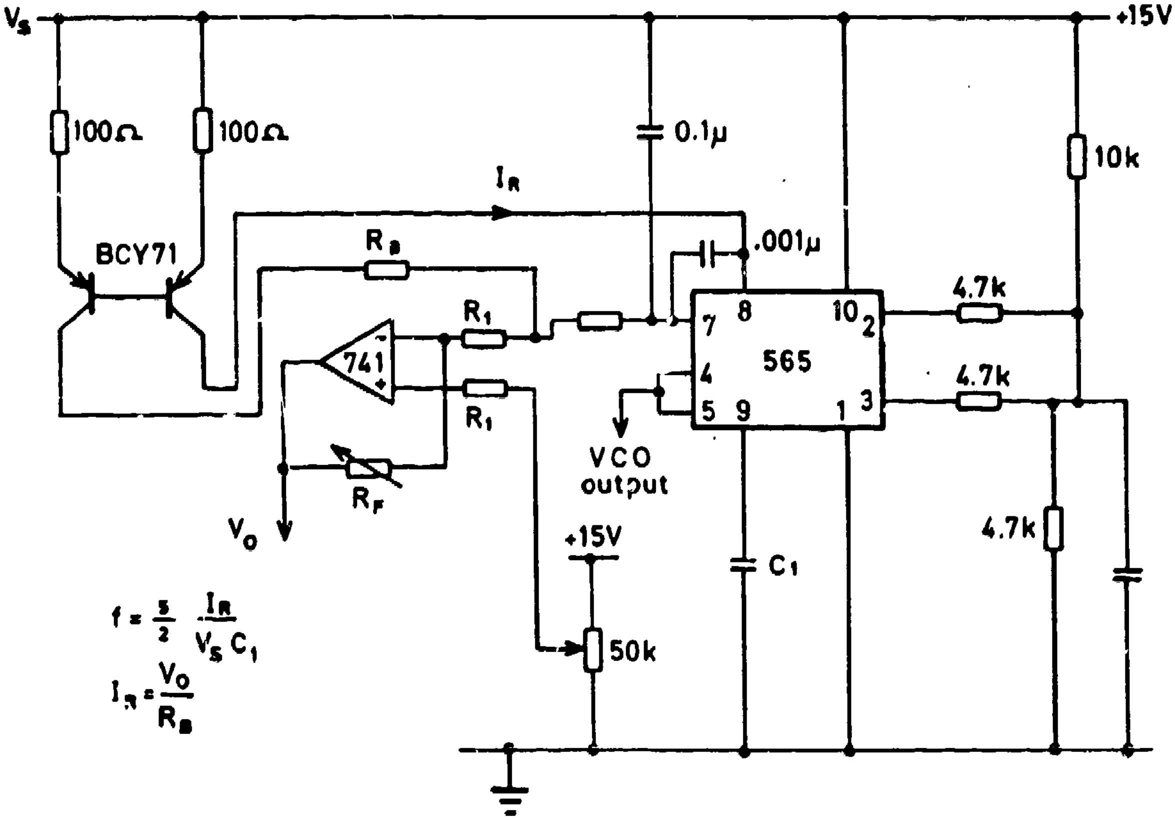


Fig 16 Signetics 565 with increased lock range

TR 79056

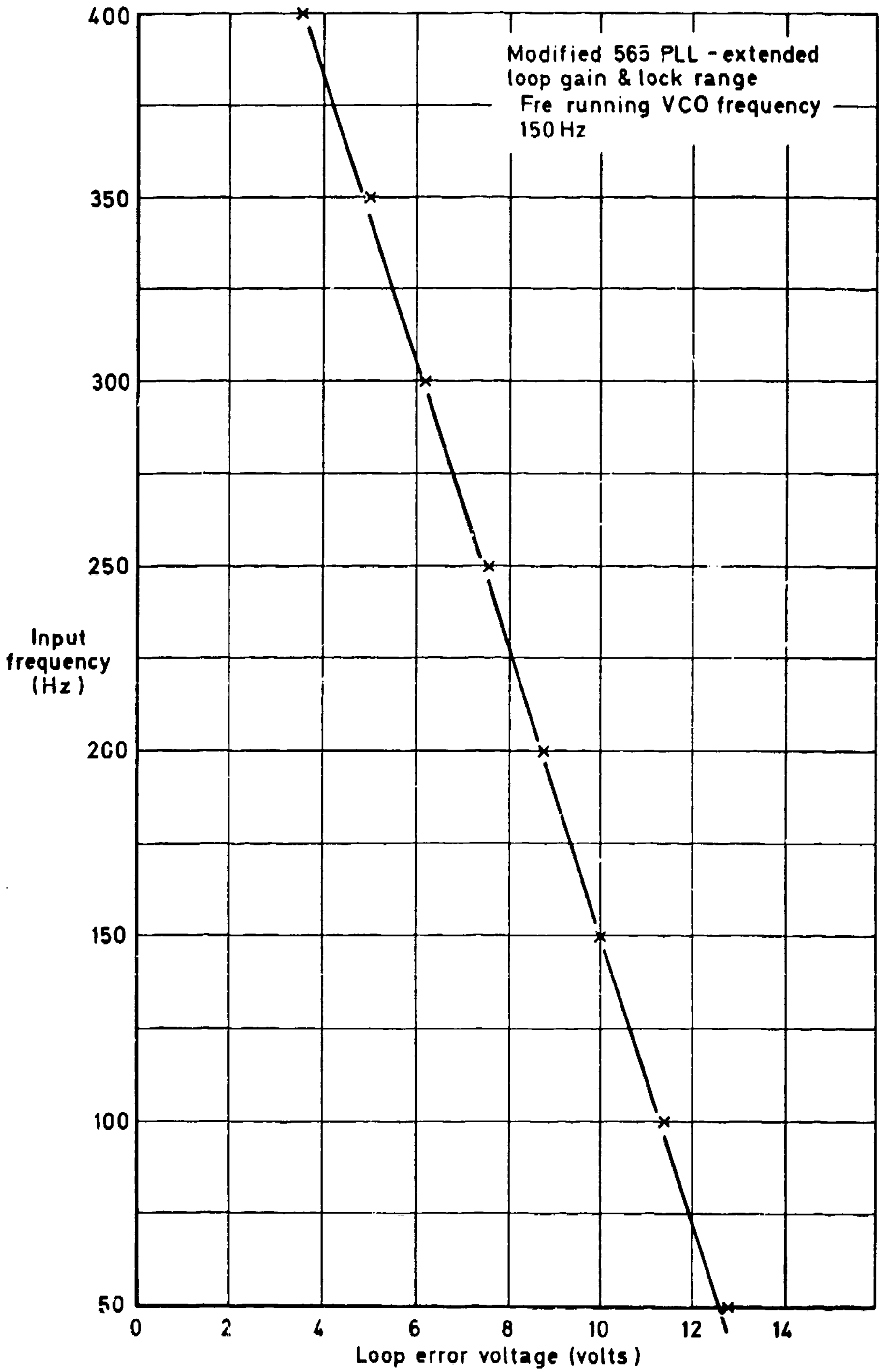


Fig 17 Loop error voltage vs input frequency

Fig 18

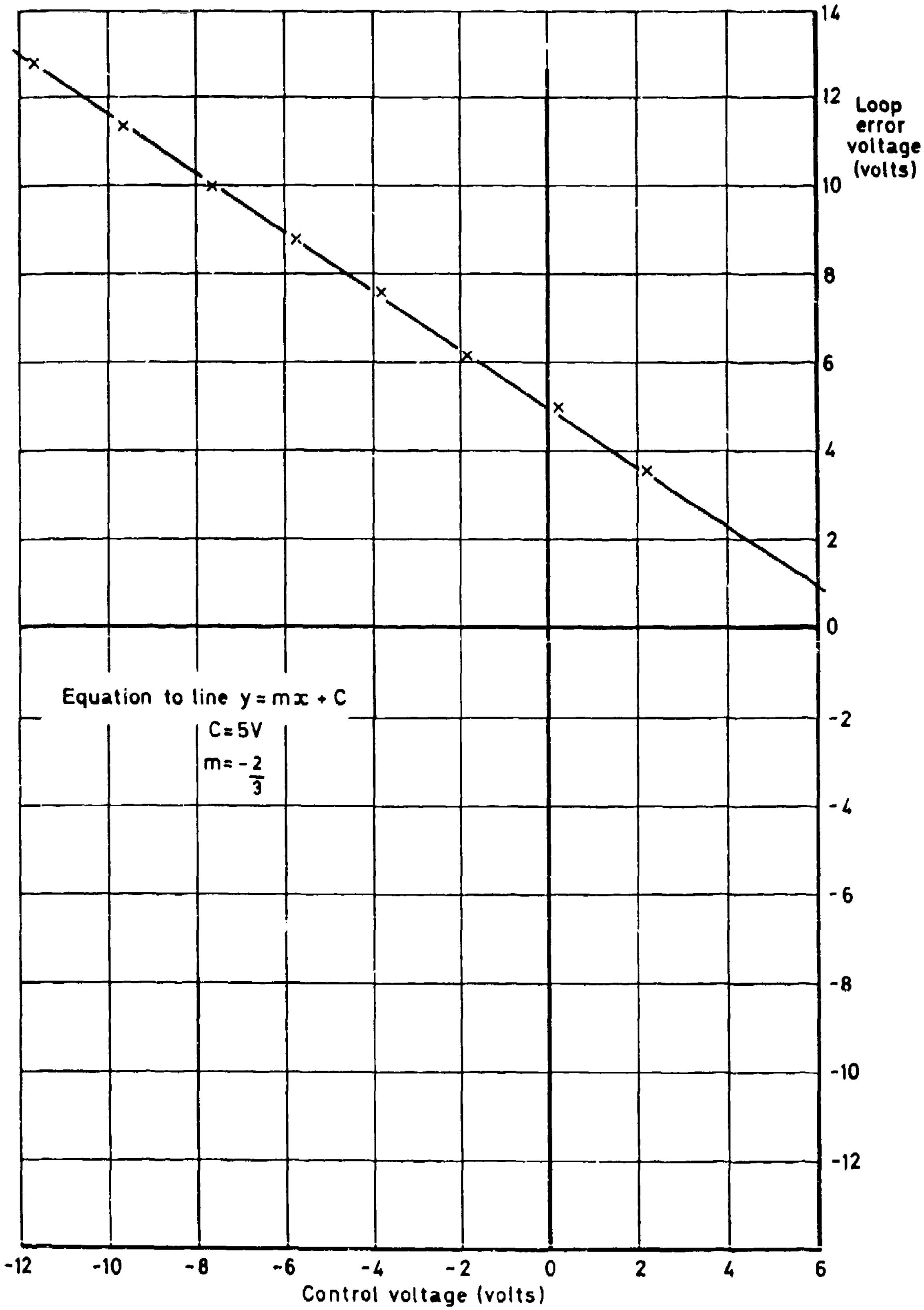
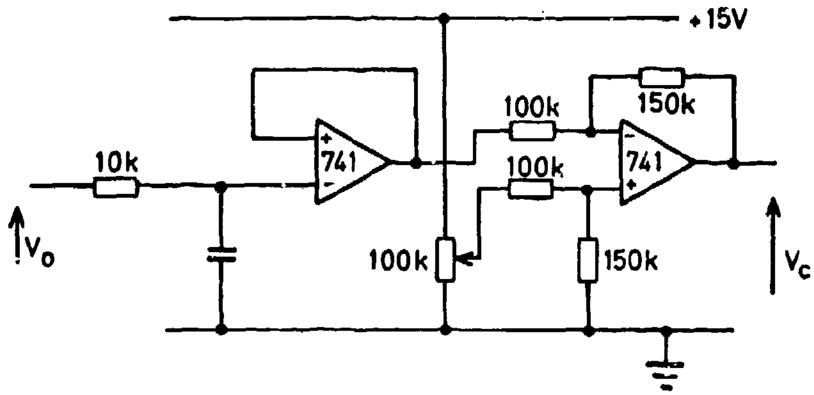
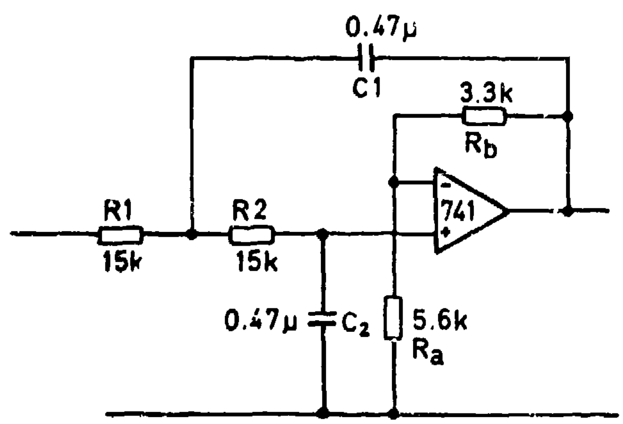


Fig 18 Required transfer function for loop error to control voltage



a Tracking filter control voltage scaling



b Second order low pass filter

Fig 19a&b Tracking filter control voltage scaling and low pass filter

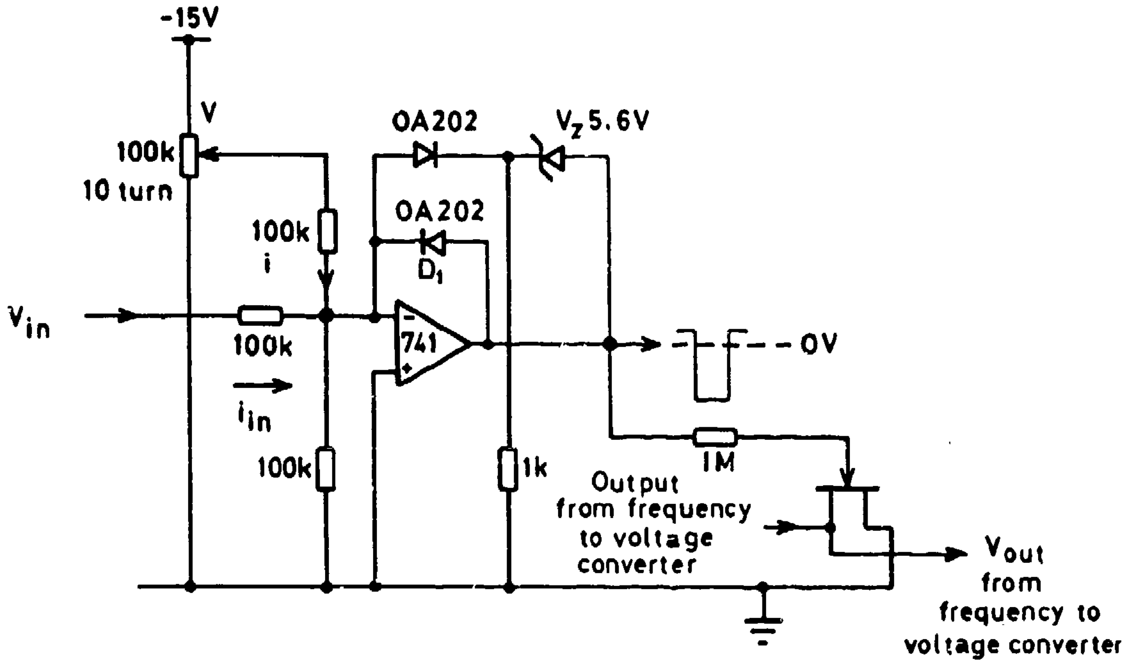


Fig 20 Voltage comparator and gate

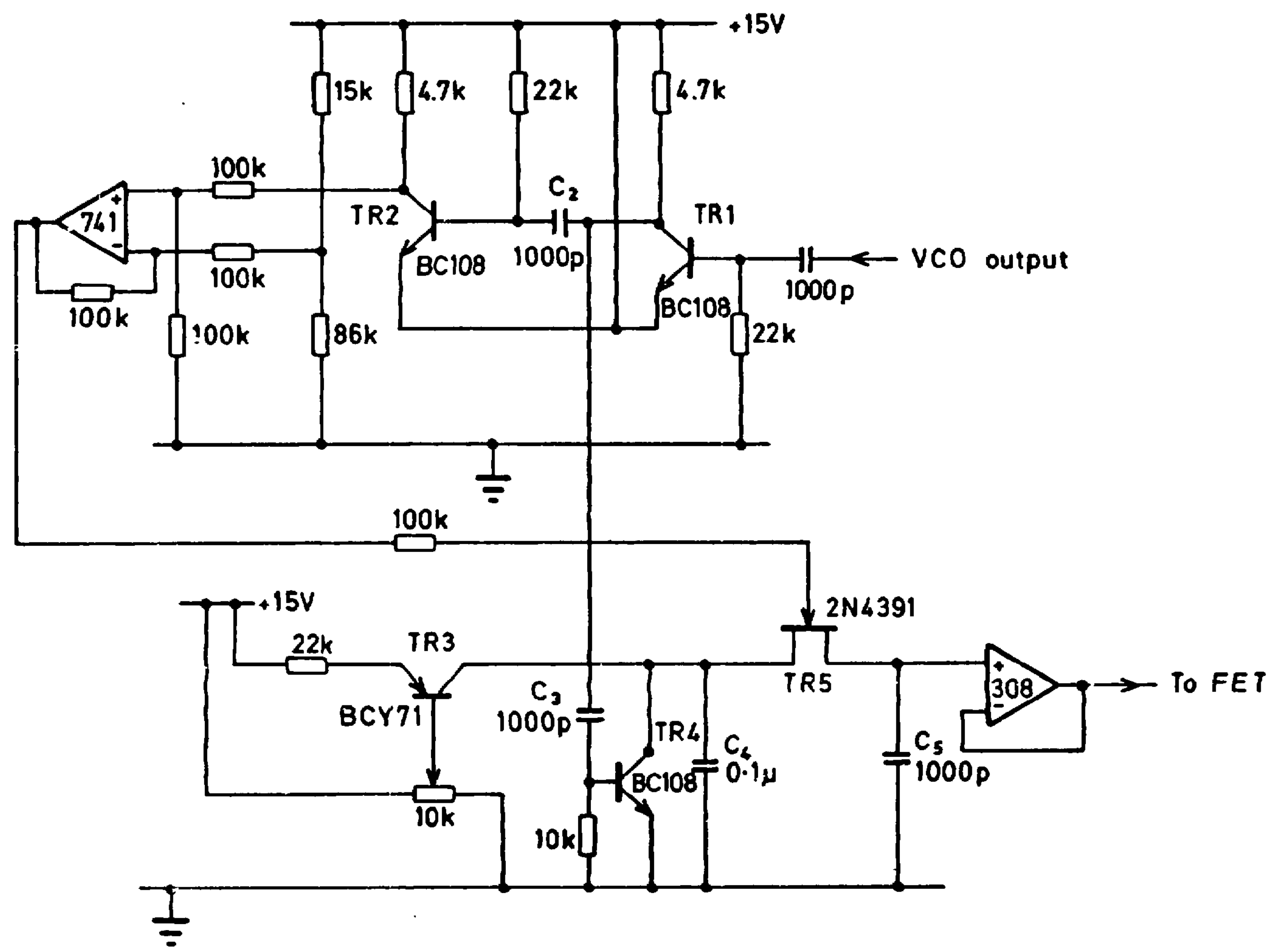


Fig 21 Frequency to voltage converter

Fig 22

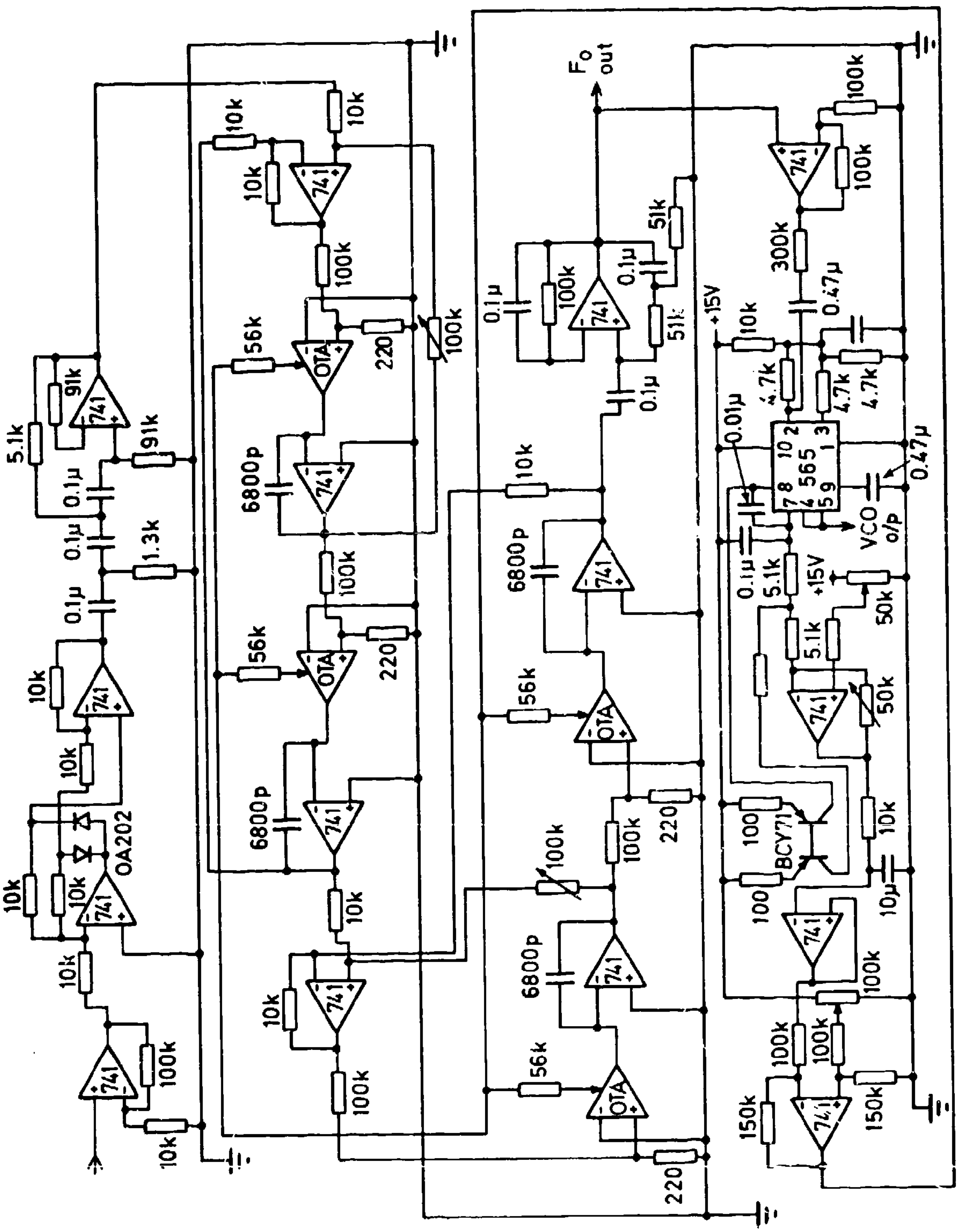


Fig 22 Pitch tracking module — circuit diagram

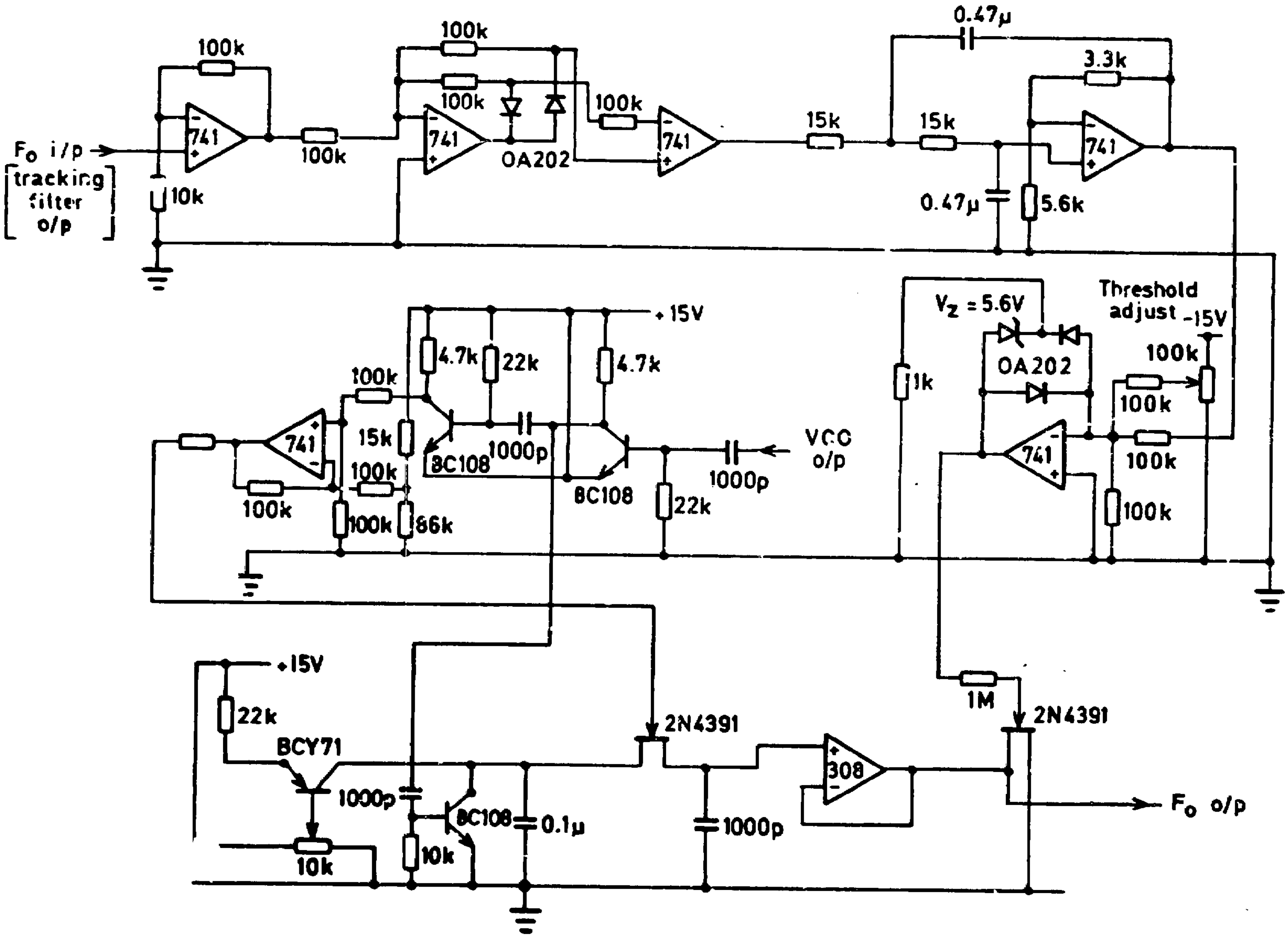


Fig 23 Voiced/unvoiced module — circuit diagram

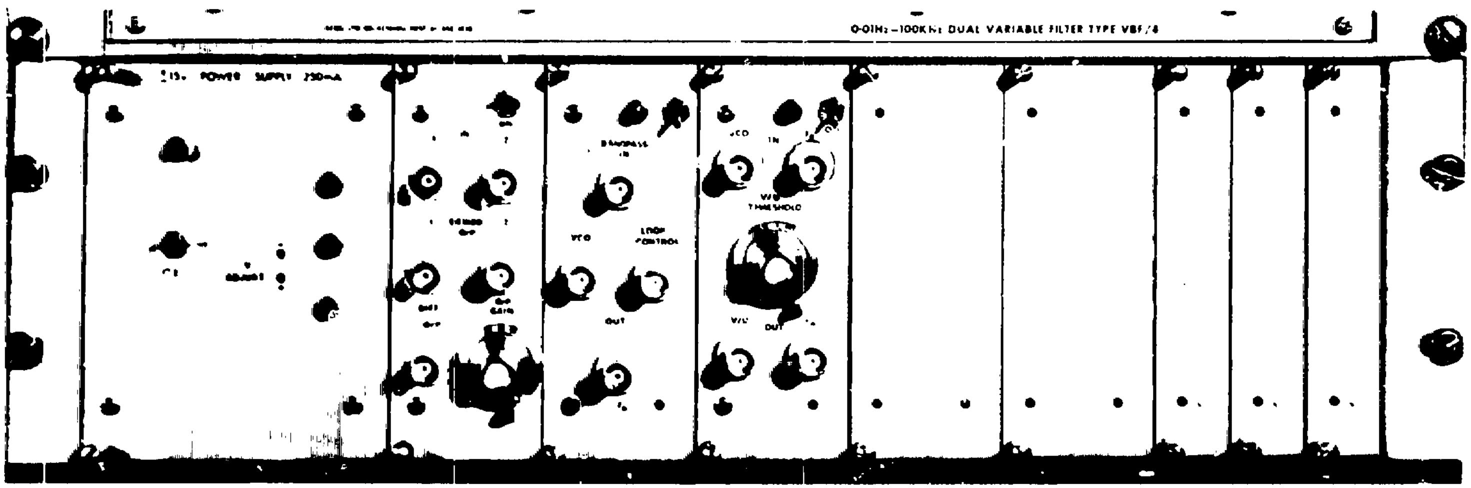


Fig 24 'Pitch' tracker

Overall security classification of this page

UNLIMITED

As far as possible this page should contain only unclassified information. If it is necessary to enter classified information, the box above must be marked to indicate the classification, e.g. Restricted, Confidential or Secret.

1. DRIC Reference (to be added by DRIC)	2. Originator's Reference RAE TR 79036	3. Agency Reference N/A	4. Report Security Classification/Marking UNLIMITED
5. DRIC Code for Originator 7673000W	6. Originator (Corporate Author) Name and Location Royal Aircraft Establishment, Farnborough, Hants, UK		
5a. Sponsoring Agency's Code N/A	6a. Sponsoring Agency (Contract Authority) Name and Location N/A		

7. Title A device for tracking the fundamental frequency of speech and its application in the assessment of 'strain' in pilots and air traffic controllers

7a. (For Translations) Title in Foreign Language

7b. (For Conference Papers) Title, Place and Date of Conference

8. Author 1. Surname, Initials	9a. Author 2	9b. Authors 3, 4	10. Date	Pages	Refs.
Peckham, J.B.			May 1979	55	48
11. Contract Number N/A	12. Period N/A	13. Project	14. Other Reference Nos. FS 100		

15. Distribution statement
(a) Controlled by –
(b) Special limitations (if any) –

16. Descriptors (Keywords) (Descriptors marked * are selected from TEST)

Speech analysis. Pitch tracker. Stress. Workload. Pilots. Air traffic controllers.

17. Abstract

The use of the speech signal for assessing physiological and psychological changes resulting from 'strain' in pilots and air traffic controllers is explained and a device is described for tracking one of the parameters of the speech signal, the fundamental frequency, to quantify changes in this parameter due to 'strain'.

F5910/1

RAE Form A143