U N I V E R S I T Y   O F   S O U T H A M P T O N

F A C U L T Y   O F   E N G I N E E R I N G

E L E C T R O N I C S

Master of Philosophy

AN ARTICULATION TEACHING AID FOR THE DEAF

by Peter Michael Holland

March 1977

CONTENTS

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING

ELECTRONICS

Master of Philosophy

AN ARTICULATION TEACHING AID FOR THE DEAF

by Peter Michael Holland

This thesis discusses the problems in teaching the profoundly deaf to produce intelligible speech, and describes the design, construction and testing of an electronic aid to assist in this task. The aid is designed to give the deaf user some feedback from his own articulator system in a visual form.

A model of the speech mechanism is constructed and used, together with the results of a literature survey, to select for display the physiological parameters of sub-glottal pressure and vocal cord tension. The difficulties of measuring these parameters non-invasively cause great problems, the attempted solution of which uses the output of a device called a Laryngograph. The Laryngograph measures the variations in electrical conductance through the neck as the vocal cords vibrate, and there is evidence that the mark/space ratio of this waveform is dependent solely on the sub-glottal pressure. It can also be shown that vocal cord tension can be determined from sub-glottal pressure and pitch.

When tested the constructed teaching aid was found to have an unexpectedly erratic response which was also dependent on pitch. More consistent results were obtained by altering the method of testing. This brings into doubt some of the data reported in the literature on basic speech research and upon which the work in this thesis is based. It is proposed that further research into the mechanism of speech production is required before aids like that attempted will be possible.

List of figures

ACKNOWLEDGEMENTS

I wish to acknowledge the assistance that my supervisor Dr. A.F. Newell gave me while I did the work that is reported in this thesis, also his patience in awaiting this written presentation. I also wish to thank Dr. A.J. Fourcin of University College London for data concerning the Laryngograph, and the Institute for Sound and Vibration Research at Southampton University for the loan of equipment.

1. Introduction

The aim of this project was to produce a speech training aid for the deaf; this aid being a machine, the use of which in some way will help the deaf individual to produce a better standard of speech production than would have been possible without it. In order to reach this objective one must provide the deaf person with some vocal standards to aim for, and methods of comparing their own voices with these standards. This can be considered as a feedback situation.

Much work has already been done in this field (see the following literature survey), there is, however, still only one aid which is widely used, this being the hearing aid. (The purpose of a hearing aid is to amplify the sound, thus increasing the amount of aural feedback available to the subject via his residual hearing.) These aids have a long history but until the emergence of electronic technology the only available aid was the ear trumpet, and this although providing some help to a deaf person is a device with a fairly low acoustic gain. It was thus only of use to the group that would now be termed 'hard of hearing', and the profoundly deaf person was considered beyond help, and treated as a severely subnormal being. (For an attempt to define 'hard of hearing' and 'profoundly deaf' see section 2.1)

With high gain electronic aids now readily available the situation has greatly improved. There is, however, a limiting factor for progress in this direction: a deaf person can suffer pain due to high intensity sound without actually hearing anything. A hearing aid is thus useless in these circumstances.

Therefore it has become apparent over the last 20 years that a new variety of aid is necessary to help the profoundly deaf person who can get no aural monitor of their own or anyone else's voice sounds. This project is directed towards providing such an aid.

2. The background to the problem.

2.1 Deafness

Deafness is a very general term and is generally applied to those with any noticeable hearing loss, however this can cause confusion due to the wide range of impairment encompassed by one term. Hearing loss is measured in dB (relative to normal hearing sensitivity) and in order to be meaningful a loss in dB should be quoted at a number of frequencies over the audio band, usually 500 Hz, 1 kHz and 2kHz. Ideally a curve of hearing loss against frequency should be plotted as shown in fig 1. (This curve incidentally shows a very common characteristic: much greater loss at high frequencies than at low.) Thus to say that a subject has a 40 dB loss is meaningless. For convenience a figure called the B.T.A. (Best Two Average) hearing loss is defined. This is the average of the best two figures from the three taken at 500 Hz, 1 kHz, and 2 kHz. This is the measure that will be used, unless otherwise stated, during the remainder of this report.

Generally a loss of more than 80 dB is termed deaf, or profoundly deaf, and between 20 and 80 dB is hard of hearing or partially deaf.

Surveys have been made to discover the prevalence of the problem in this country. However it is not very clear just how many people are involved and how they are distributed by age. This is partly due to the fact that different surveys have used different standards of measurement and classification. They do tend to indicate however that approximately 5% of the population has a significant hearing impairment. In the UK in the late 1960's the Central Office of Information published the following figures: Deaf mutes; 15,000, Totally deaf; 30,000, Deaf to all natural speech; 70,000, Hard of hearing; 1,650,000. Although not accurately defined these figures give some indication of the size of the problem.

Deafness can be caused by many different malfunctions, four categories of which are now generally recognised:

1. Conductive deafness caused by blockage in the outer or middle ear which prevents the vibrations from reaching the cochlea.

2. Receptive deafness caused by damage to the cochlea, so vibrations in the middle ear are not coded into nerve impulses.

3. Transmissive deafness caused by an interruption in the auditory nerve between the ear and the brain.

4. Perceptive deafness caused by damage actually in the brain.

Conductive deafness is less serious than the other types as it can usually be cleared by operation, and even if this is not possible some sound can be transmitted to it through the bone using a bone conductor type hearing aid provided the cochlea is intact.

Those who are born deaf, or become deaf before reaching about two years of age (ie. before attaining speech proficiency) are in greater need of help in the way of a speech training aid than those who have become deaf later in life. This group of people will find the production of intelligible speech a very hard task indeed as they have no foundations upon which to build. Anyone who becomes deaf after having already developed speech will retain intelligibility for a long while even though the speech quality will degenerate.

Thus the group at which this proposed aid is aimed can be described as those with:

1. B.T.A. hearing loss of greater than 80dB.
2. Receptive, Transmissive or Perceptive deafness.
3. Prelingual deafness.

2.2 Speech production

Speech sounds can be subdivided into two sections: voiced and unvoiced, ie. very approximately.... vowels and consonants This is true for the majority of languages. Voiced sounds are caused by the vocal cords vibrating in the flow of air expelled from the lungs. Unvoiced sounds are caused by forcing air from the lungs through a constriction so causing turbulence in the air flow and thus producing a hissing noise. This constriction is typically formed by positioning the tongue against palate or teeth. Both types of sound are modified by the acoustic properties of the larynx, mouth and nasal cavities.

Considering voiced sounds in more detail: The vocal cords are a pair of muscular flaps at the base of the larynx, a much more detailed description of their action is to be found in Appendix I. The characteristics of the vibration depend on the forcing air pressure, the size and elasticity of the vibrating matter, and to a lesser extent the initial open area and the acoustic loading of the upper vocal tract. The pressure waveform produced by this vibration is a series of pulses typically like those shown in fig2. its frequency is called the fundamental frequency, or pitch - f_0 - and lies between 100 and 400Hz.

A fourier analysis of this pressure wave reveals a line spectrum decaying at about 20dB/decade. This can be considered as the input to the acoustic filter formed by the upper tract. Positioning of the tongue, jaw and velum causes resonances the frequencies of which are called the formant frequencies - f_1 f_2 etc. - The overall effect is demonstrated in fig 3. The first two formants are the most significant for speech intelligibility, f_1 lying between 200Hz. and 1kHz. and f_2 lying between 800Hz. and 2.5kHz. These can be used to differentiate between the various vowel sounds, however the accurate determination of them electronically is not easy. (see later sections)

In order to try and understand the overall system involved in speech production I drew a flow chart model. (for a description and discussion of the sort of mental imagery that these diagrams imply refer to : Richardson A. (1969)) A chart for a normally hearing person is shown in fig 4. It can be seen that the sound output is controlled by two feedback loops: an overall aural one and an internal sensory one.

In the learning mode the memory stores a mutully related structure of words, sounds and motor nerve command patterns, and when one is speaking one would expect a continual interplay between the memorised patterns and the patterns that are fed back via the feedback loops.

In a profoundly deaf person the situation would be somewhat

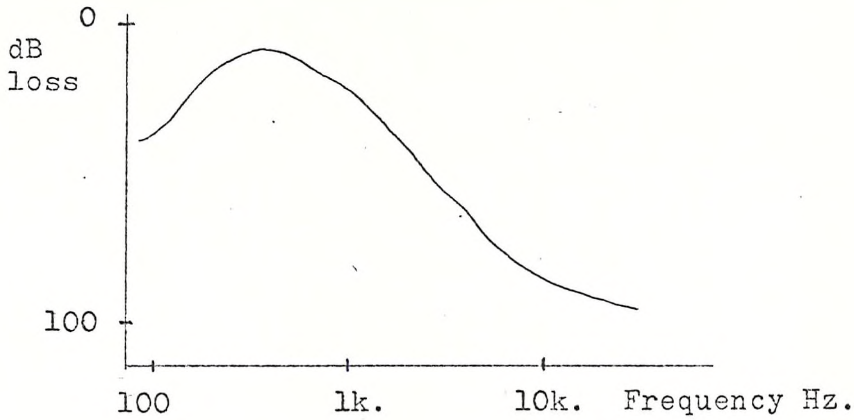


fig. 1 A typical hearing loss characteristic.

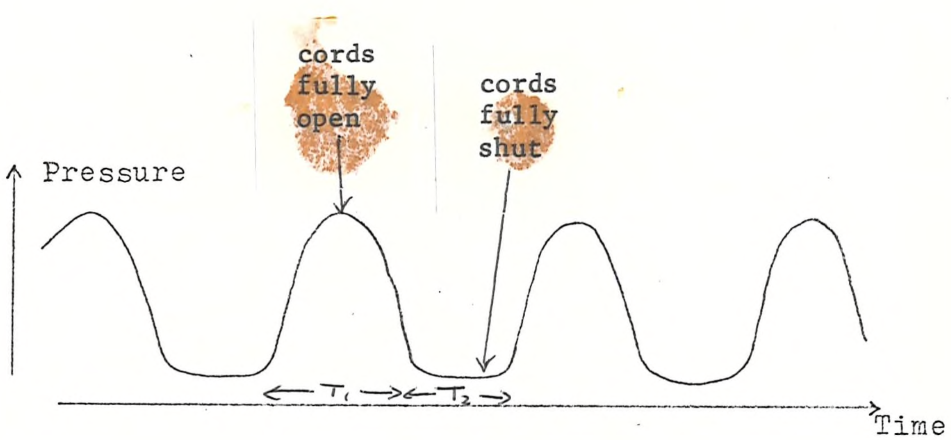


fig. 2 The glottal pressure waveform.

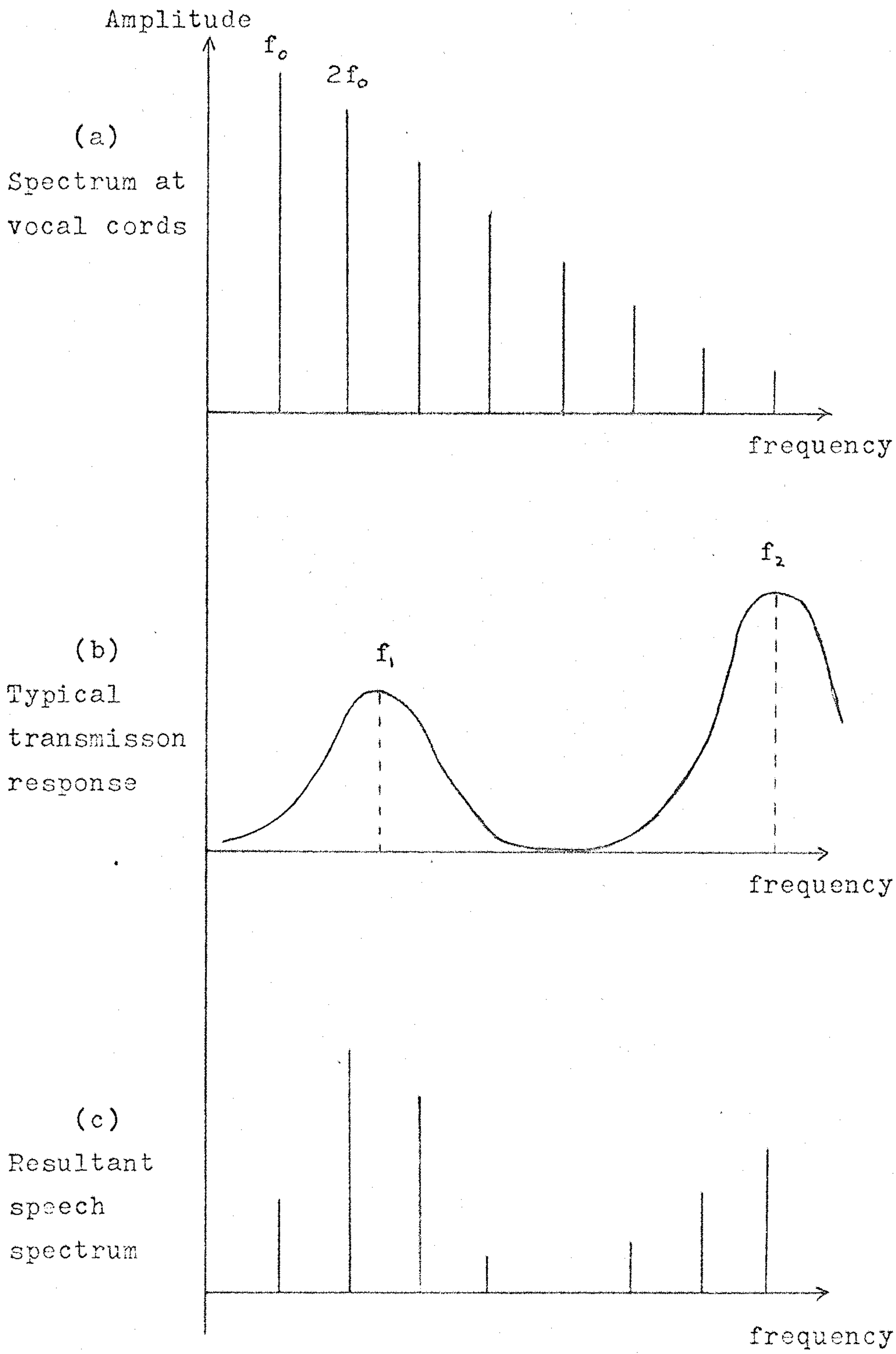


fig. 3 Formation of the speech spectrum

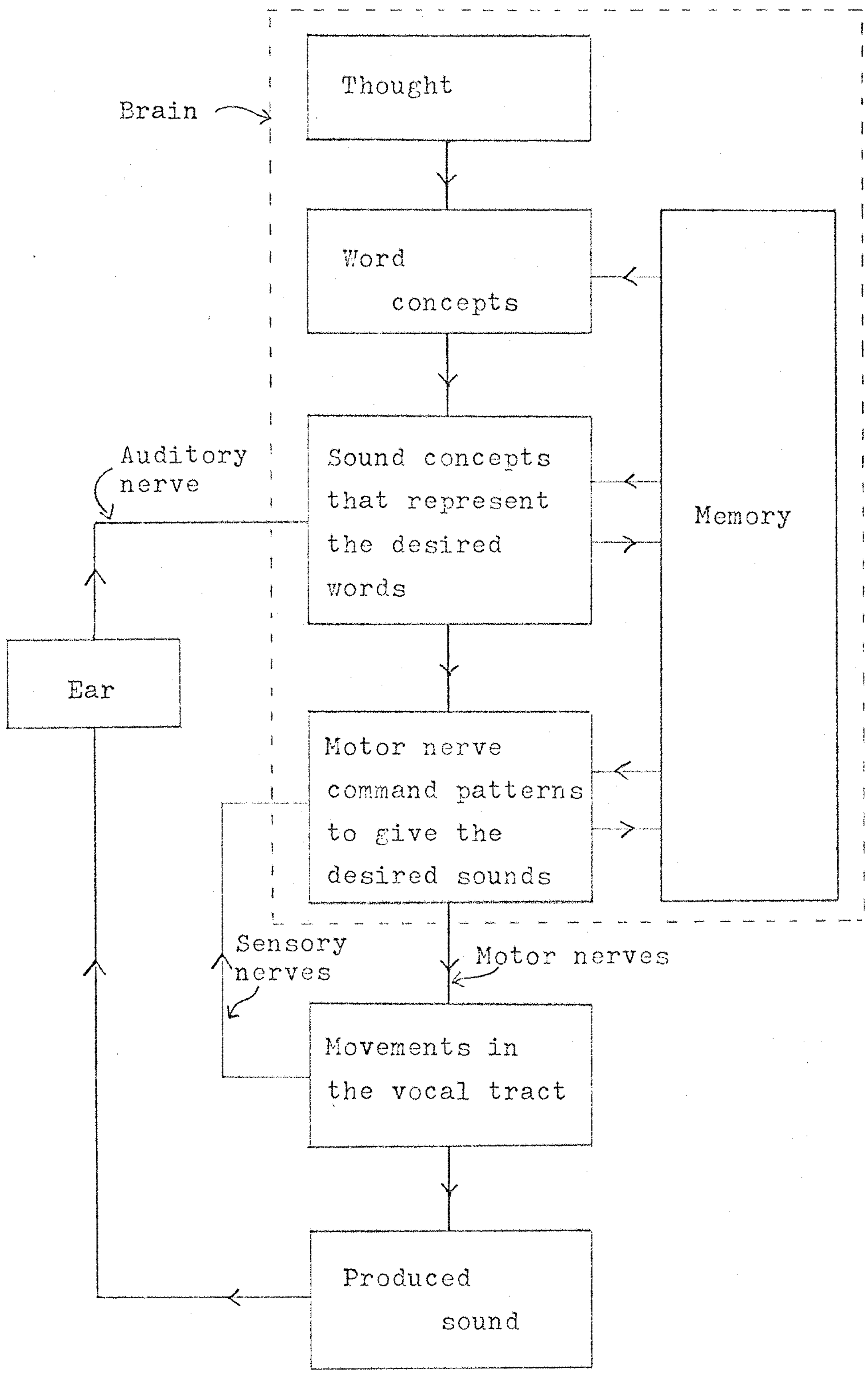


fig. 4 A model of speech production in a hearing person

different. A corresponding chart for a deaf person is depicted in fig 5.

In drawing this chart not only the box labelled 'ear' had to be omitted but also the box labelled 'sounds'. Thus the deaf person has to translate from word concepts direct to motor nerve patterns, this would seem to be an unnaturally large gap, and so the diagrams show that the deaf person is not only handicapped by the loss of aural feedback as is normally quoted, but also by a discontinuity in the forward loop of the production process.

In including the box labelled 'word concepts' in fig 5. an important assumption has been made; namely that the subject has acquired a reasonable level of language proficiency. Without this assumption one would not be justified in assuming that a deaf person thinks in words. Furth H.G. (1966) is entirely devoted to the problem of the psychological effect of deafness, and in particular he tries to answer the question "what do the deaf think in?" Although not in fact answering this question he does come to the conclusion that: "The major significance of the reported findings for theories of thinking is the demonstration that logical intelligent thinking does not need the support of a symbolic system as it exists in the living language of society. This is undoubtedly an internal system, a hierarchical ordering within the person of his interaction with the world. The symbolic system of language mirrors and in a certain way expresses that internal organisation" (Furth H.G. (1966) p.228) This quotation thus validates the separate representation of thoughts and words in the above

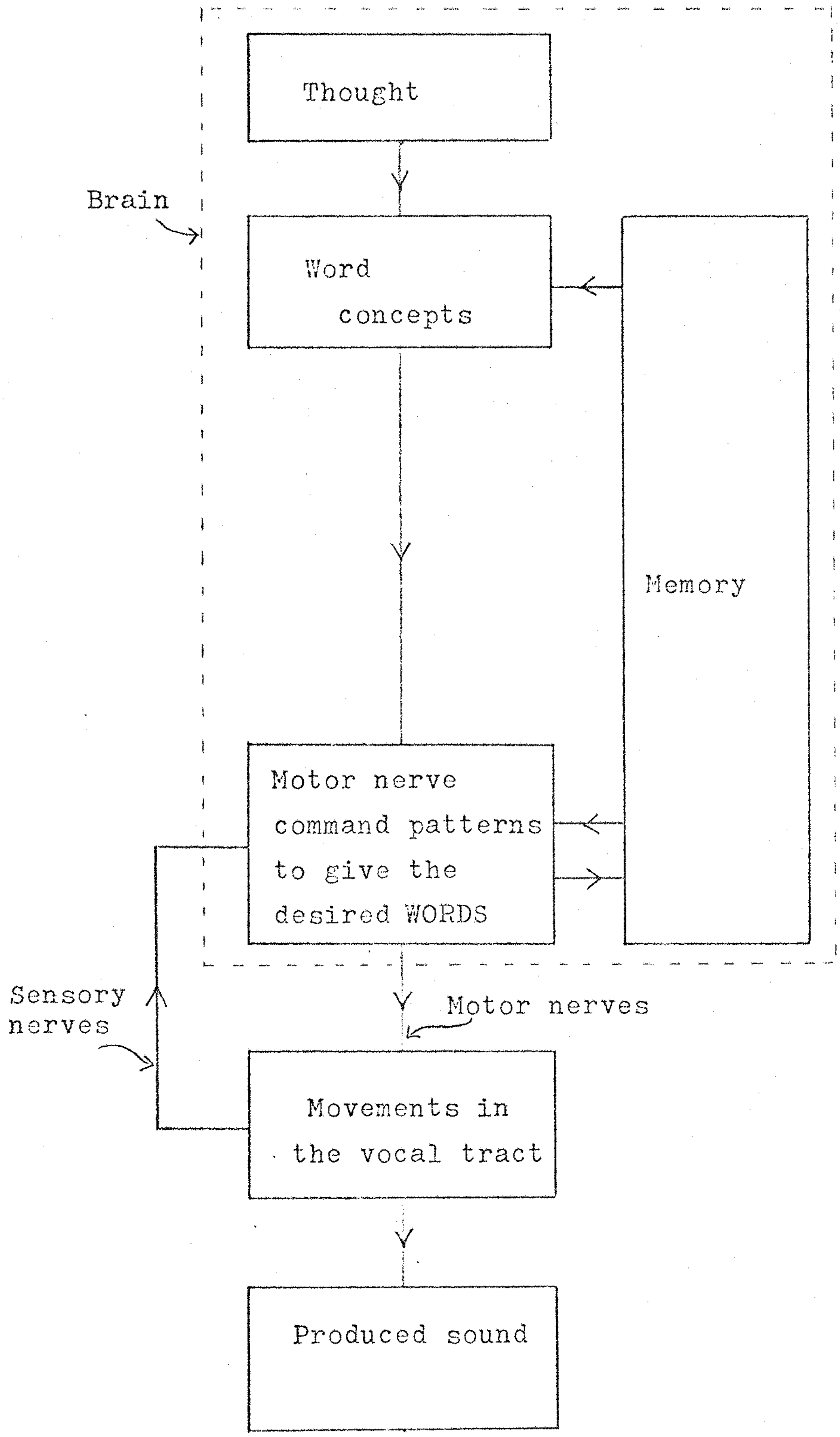


fig. 5 A model of speech production in a deaf person

models, and also indicates that the 'words' box may well be absent in a profoundly deaf person.

In this report I shall always assume that the subjects have sufficient language proficiency, and concentrate on methods of helping them express the words that they desire, as this is the field where I feel that I may be able to offer more.

These models are, at best,only one possible representation of the processes involved and, as explained in the following paragraphs, variations could easily be proposed in the light of further psychological evidence.

Several works, eg. Terwilliger R.F. (1968) and Sokolov A.N. (1972), approach the problem from another view point. It appears, for instance, that a hearing person when listening to speech, or even verbalising internally (so-called inner speech, as when solving a problem in logic, or reading a complicated passage in a book,) makes certain micromovements with his articulators. These movements have been detected electromyographically. When this phenomenom was discovered it was thought that this meant that the understanding of words (received aurally or visually, or used as concepts) depended on, or referred to, the process of articulation: ie. one spoke the words to oneself, actually articulating them, in miniature during the understanding process. This is called the 'motor theory' of hearing, and might be represented as in fig 6. This shows the understanding process to be using not only information from the direct perceptual channels but also via the imitative action of the vocal tract. The evidence for this,

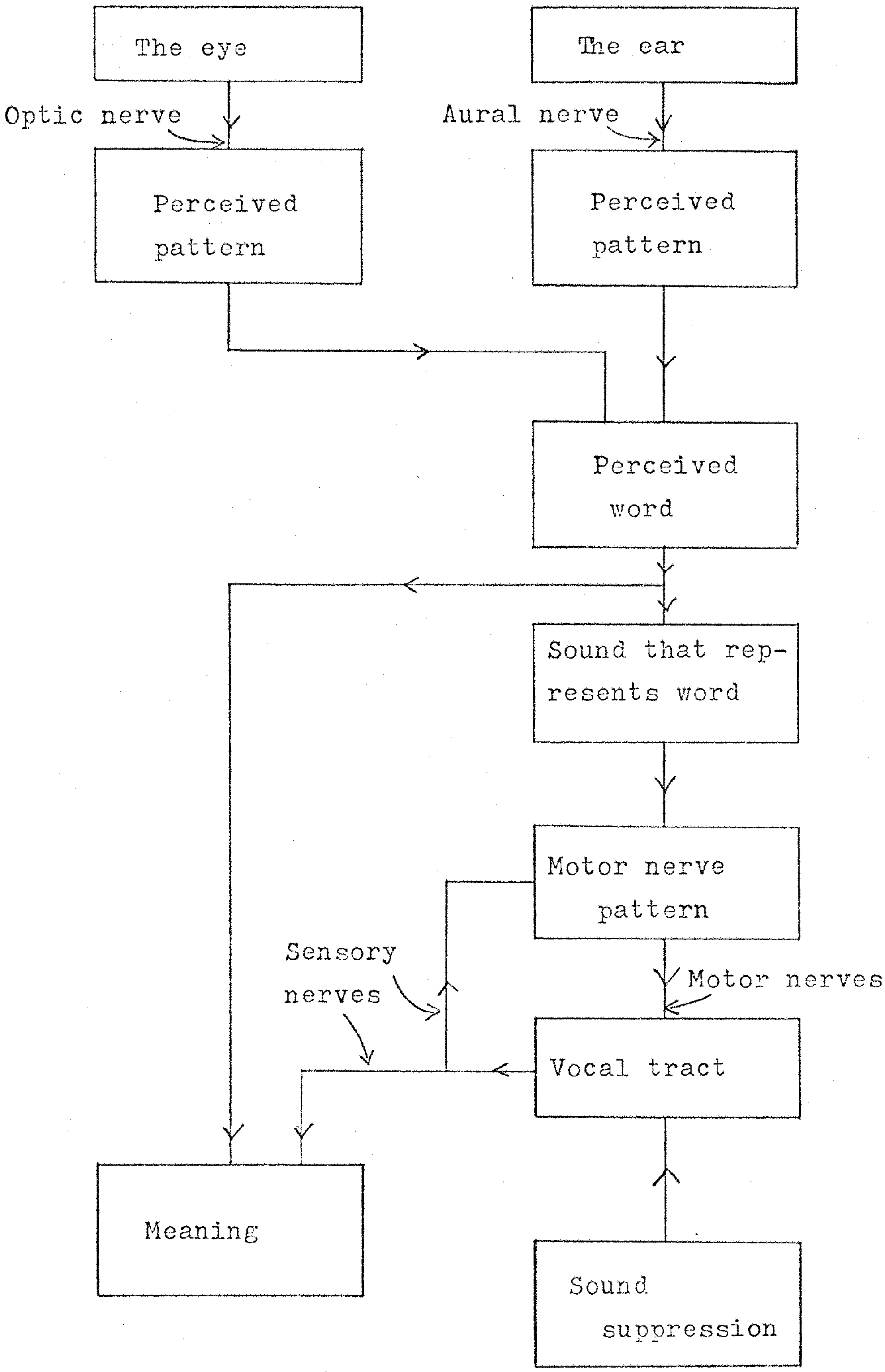


fig. 6 A possible understanding process

however, is far from conclusive as the electromyographic results taken during inner speech of a set word has been found to differ considerably from the results taken during actual speech of that word (see Ballantyne J. and Groves J. (1971)). This may be due to the suppression of the sound output. There are many unknowns associated with this phenomenon and I do not believe that the evidence implies changes in figs 4 and 5 that would have significant results for this study. The above information is included in this report merely to demonstrate that the models drawn in figs 4 and 5 do not claim to encompass all current knowledge on the subject.

2.3 The speech of the deaf

Any written description of the speech of the deaf is inevitably inadequate as words cannot accurately describe sounds. Very few congenitally profoundly deaf people ever reach a level of proficiency such that they can be understood with ease by a stranger.

The faults found in deaf speech are numerous. The most common are: poor timing and pitch control, omission or substitution of consonants and a loss of discrimination between different vowel sounds. These faults compound to produce monotonous indistinct speech. Levitt H. et al (1974) contains a comprehensive survey of the speech of 40 deaf children (age 8-15 years). The general description they give is as follows 'The pattern of errors in the childrens' speech indicates that correct placement of the articulators is not a major problem, but rather incorrect timing, improper velum control and voicing problems are the source of the most serious errors affecting intelligibility'. These deaf children had received much speech training and yet they still produced many language errors. For example, 37 out of the 40 children were classified as having errors in pitch control.

2.4 Speech training methods

The policy in this country has been to educate deaf and partially deaf children in schools apart from normal-hearing children so that they can receive the specialised attention necessary. At the moment there are about 70 schools for the deaf catering for roughly 6,000 pupils. For the profoundly deaf child it is recommended that schooling should start at the age of two. This early start is a very crucial factor in their education.

In the normal-hearing child speech developes gradually. Initially when only a few months old the child 'plays' with his voice seeming to randomly vocalise for quite long periods. In this babbling period the child is learning the relationships between sounds and the motor nerve commands necessary to produce them. Later, at the age of 12-18 months the child starts to produce simple words - using the information already gathered to imitate speech sounds. Thus by the time he reaches about two years the child can speak understandably with a limited vocabulary.

With the deaf child however the babbling stage, although still present, does not perform the same function. The sensations caused by moving the articulators and producing a sound are not complemented by an aural impression of that sound. So the babbling stage does not lead naturally to speech.

Deafness can be diagnosed usually by the age of 6-9 months and it is the policy of most teachers to give a hearing aid as soon as the child will accept it. The teaching of speech relies enormously on whatever residual hearing the child might have. While at home he relies on his parents to expose him continually to as much speech as possible. The parents are encouraged to talk to the deaf child as much, if not more, than they would to a hearing child. Once at school devices called speech trainers are usually available. These can be thought of as high quality hearing aids. The child will receive specialised speech training lessons using this device, and in this manner the majority of deaf children will make some attempt at speaking.

In conjunction with this purely aural training the child will be taught, as much as possible, to use visual, kinaesthetic and tactile cues for differentiating between speech sounds. Articulator positioning (eg. tongue height and lip opening) can be demonstrated quite well visually by a teacher for a child to imitate. This might in some degree explain the findings reported by Levitt H. et al as quoted previously where the major faults in the childrens' speech were found to be other than faults in articulator positioning. There are some children who have to rely entirely on such indirect teaching methods.This project is aimed mainly at this group of children, as were many of those described in the following section.

3. A survey of hearing aids for the deaf

3.1 Introduction

Many electronic aids for the deaf have been designed and constructed in the last 30 years. They have been made possible by the immense growth in electronic technology that has taken place in that time. Some of the first work that was done in this field was performed at the Bell Laboratories in the mid 1940s following on from the work of A.G. Bell on speech coding for line transmission. Since then the field has expanded and progressed greatly, but the 'ideal aid' is still elusive.

I shall group the following description of aids under headings according to the physical sensation they use as a communication channel to the deaf person.

The aids fall into 3 broad groups,

- i Aural
- ii Visual
- iii Tactual.

3.2 Aural aids

3.2.1 Conventional amplification aids

As stated previously these are very widely used. Many sufferers of partial deafness can live quite normal lives when equipped with a suitable aid, but it is not generally realised how much help an aid can give to a profoundly deaf person.

The standard Medresco aid gives 52 dB gain at 750 Hz and a slope of about 8 dB/octave over its whole range of 300 Hz to 3 kHz, and it is possible to get aids that give considerably more gain than this.

The speech trainers referred to earlier have a considerably greater frequency response, much less distortion and a little more gain. They also have facilities for changing the frequency response to suit each particular deaf person. A recurrent problem with hearing aids is acoustic feedback which can cause oscillation in the system and is usually caused by an ill-fitting aid. Speechtrainers attempt to overcome this by using a pair of headphones with large foam pads that seal against the head around the ears.(For more information on conventional hearing aids see John J.E.J. (1964).)

3.2.2 Frequency transposing hearing aids

A deaf person's residual hearing usually displays an increasing loss with increasing frequency. For this reason there have been several attempts to convert some of the high frequency components of speech into a low frequency band. The following methods have all been tried:

1. Modulation. The high frequency components are extracted and modulated by a fixed frequency carrier. The lower sideband of this signal is then added to the original signal. This is described in Biondi E.L. (1967).

2. Distortion. The high frequency components are again separated but with this method they are put through a distorting network to obtain lower frequency components which are then added back to the original signal. See Risberg A. and Spens K.E. (1967).

3. Vocoder System. This is a more complex system: the envelopes of the outputs of two or more high frequency band pass filters are used to modulate the amplitude of two low frequency

bands of filtered noise. This is then added to the original signal; see Ling D. and Druz W.S.(1967).

4. Frequency Dividing. For this the high frequency components are again filtered out and then subjected to infinite limiting, thus creating a pulse train. This is put through binary dividers until of a suitable frequency to present to the subject ie. less than 800 Hz; Guttman N. and Nelson J.R. (1968).

It has also been shown, by Erber N.P. (1972), that the total speech envelope contains considerable information; here the envelope was modulated with band limited noise and it proved to be some assistance during lip-reading.

The degree of success enjoyed by these systems is apparently not very high, they all require a great deal of practice before becoming at all meaningful.

3.3 Visual aids

3.3.1 Introduction

It has been shown that a visual aid will stimulate a child to vocalise even if that display is fairly randomly related to the speech. In Fineman K.R. et al (1969) and Blake P. and Moss T. (1967) a device is used with just two bulbs - corresponding to frequencies above and below a set reference - and these modulated a bright colourful kaleidoscope type of display. This was quite successful in motivating the children (who were electively mute, not deaf) to talk, but little teaching capacity was indicated. These papers do show, however, that children are greatly motivated by a colourful mobile display.

There have been many attempts to provide displays which might prove more successful as a teaching aid, ie. provide information to a deaf user which may help him to improve his speech. These displays can be classified descriptively by the type of information displayed:

1. spectral pattern vs. time
2. frequency vs. amplitude
3. f_1 vs. f_2
4. measure of quality or quantity of single parameter
5. single speech parameter vs. time
6. vocal tract shape.

3.3.2 Spectral pattern vs. time

The Visible Speech Translator (see Stark R.E. et al (1968)) built by the Bell Laboratories produces a picture on a T.V. screen of the speech sprectrum as it varies with time, so the output looks not unlike a spectrogram. One limitation of this equipment is the lack of intensity information: at each point on the screen there are only two possible states - black or white, there is no grey.

3.3.3 Frequency vs. amplitude

L.U.C.I.A. (see Risberg A.(1968)) is a 10 x 20 discrete light display. The abcissa is quantised frequency in 20 steps of 200 Hz and the ordinate is quantised amplitude in 10 steps of 3dB. At each instant this produces an approximate spectrum of the speech input. The display can be frozen in time or run continuously, but in the latter state Risberg describes it as changing too fast during normal speech to be followed by eye. A similar device is described by Borrild K. (1968), this seems a less sophisticated version of L.U.C.I.A.

At the other extreme, Nickerson R.S. and Stevens K.N. (1973) use a computer to generate several displays on a T.V. screen, one of these is a spectrum vs. amplitude display. However, in common with many of these reports, little actual testing of the effectiveness of the system is related.

3.3.4 f_1 vs. f_2

A display which is mentioned in several reports is a plot of second formant, f_2, against first formant, f_1 (see Person S.C. (1973); Georgiou et al (1973); Pickett (1969)). This display can be quite useful for discriminating between vowels and can be used for teaching a deaf person to produce steady vowel sounds. However, it is of little use with connected speech as it is meaningless for all unvoiced speech. Georgiou V.J. describes an interesting modification of the basic $f_1.f_2$ display. Here the output is a dot on a colour T.V. screen which moves as dictated by f_1 vs. f_2 but its size corresponds to the speech amplitude and the colour is varied with the pitch. No test of its effectiveness is reported.

3.3.5 Quality or quantity of a single speech feature

This group includes measures of pitch and the mean frequency of fricatives. In Borrild K. (1968) an 'S' indicator is described that has been in use for many years with good results. This simply lights a bulb with a brightness which increases as the average frequency of the fricative input increases, ie. a bright light indicates a good 'S', In Martony J. (1968) the pitch of the subject's voice was also displayed on a meter, but it is generally accepted that such qualities are best displayed against time.

3.3.6 Single speech feature vs. time

The most common display in this group is the pitch display. In Martony J. (1968); Boothroyd A. (1973); Mermelstein P. (1967) Gopinath G. et al (1967); Wakita H.(1973); Stansfield E.V. et al (1973) pitch is displayed against time on a storage oscilloscope. The papers differ in the methods of obtaining the pitch from the voice signal, but the displays are much the same apart from this. This display is well tested and has been found to be quite useful as it can make teaching intonation and rhythm patterns much easier.

In Stark R.E. (1970) an amplitude contour display was projected on to a storage oscilloscope with the aim of teaching rhythm and stress, but this is not evaluated to the same degree.

3.3.7 Vocal tract shape

This group of displays are all aiming to produce the same output: a graph on an oscilloscope screen, the abcissa of which is distance along the vocal tract, lips inwards, plotted against the cross-sectional area at each point (see Willemain T.R. (1972); Miller J.D. et al (1974); Anderson A.B. et al (1951); Mason J.L. et al (1973); Dobelle W. et al (1973)). The display thus shows movements of tongue and lips as they are happening. This would seem to be quite useful but only in limited circumstances: only voiced non-nasal sounds give a meaningful output. It also has the disadvantage of involving a fairly complicated computation, which must make this an expensive device at present.

3.4 Tactual aids

3.4.1 Introduction

The argument for using the tactual channel to give a deaf person extra information is that it is free of any information concerned with verbal communication, whereas the visual sense and the residual hearing are both already burdened with information. In the literature there seem to have been two methods used to stimulate this sense: vibration and electric current.

3.4.2 Vibration

One problem of using any form of tactile stimulation is that the skin is rather insensitive to the position of stimulation, ie. it is difficult to sense which of two closely spaced vibrators is activated. However this is overcome by using very widely spaced vibrators; Willemain T.R. (1972) contains a report on a pitch display indicator using vibrators on the fingertips. Pitch was quantised into eight steps, but the subject used only three vibrators choosing those of the eight which were of most use to him. In this way they obtained some quite good results in controlling the pitch of deaf subjects.

The tactile display used by Miller J.D. et al (1974) was somewhat different, again using three vibrators on the fingertips, but this time they corresponded to nasalised sound, voiced sound and total acoustic output. Thus it has a much more generalised objective than the one above, and it gave significant improvement in lip-reading ability.

Spectral tactile displays have also been attempted, see Pickett J.M. (1969), the spectrum being quantised into steps, the amplitude in each section modulating the drive to the vibrators (one on each fingertip). This was not found to be very successful.

Possible reasons for the limited success of tactual aids are discussed in Kirman J.H. (1973), the positioning and type of vibrators and the information displayed are criticised and proposals are made for an aid that might have greater success. (No report of any follow-up work on this has been found).

3.4.3 Electrical stimulation

This has generally been rejected for use as an aid to the deaf, probably for safety reasons. However it does have some advantages: the skin differentiates frequencies and intensities much better by electric current variation than by vibration (see Anderson A.B. et al (1951)). Its use also eliminates the noise caused by the vibrators which would be disturbing for any normal hearing person conversing with the deaf subject.

A much more ambitious aid for the deaf is proposed by Dobelle W. et al (1973) who describe how electric pulses were put directly into the brain to stimulate signals in the aural nerve. Such projects can only be expected to produce something useful in the far distant future because of the complexity of the nervous system.

4. Discussion of display objectives

4.1 Derivation of display objectives

Having decided to display some speech varying parameters to the deaf person one can follow one of two path-ways: either towards displaying acoustic parameters such as pitch, formant frequencies etc, or towards displaying the physiological parameters involved in moving the articulators. There are arguments for and against both of these approaches.

Appraisal of the previous section shows that all previously tried aids, except the vocal tract shape indicators (see 3.3.7) have followed the first path - without success in the majority of cases. This could be regarded as a preliminary argument for trying the second approach.

There is one fairly basic argument for employing the first method. The purpose of a teaching aid is to help in the production of natural sounding intelligible speech. Such speech can be produced from many greatly differing articulator positions, as is illustrated in the extreme by ventriloquists' speech. Believers in the acoustic parameter approach reason that a person may use different ways to produce similar speech sounds at different times. Thus teaching a deaf person to put his articulators in one particular position to produce a given sound is unrealistic. However it is undeniable that one articulator position does produce one unique sound, so if one teaches this fact to the deaf person then he will be able to repeatably produce that particular sound. This is certainly a step in the right direction.

One can also indicate the validity of the second method by reference to the flow chart, fig 5. Looked at from the first point of view the desired transfer function - sound/word concept - can be considered as an open loop control system with two unknown functions: motor nerve pattern/word concept, and sound/vocal tract movements. From the second viewpoint the transfer function is vocal tract movements/word concepts and there is only one unknown function - motor nerve patterns/ word concepts. This is a simpler system and the feedback loop has more control over the desired output.

Assuming the use of a display fig 5 becomes fig 7. This represents the use of a generalised display: the input transducers operating on both the vocal tract and the acoustic output, and the display via any sense organ with an undefined function in between. For this to be effective an 'image' box will have to be proposed in the forward path of the feedback loop. Thus for each word the memory will produce a suitable image in terms of these newly formed display parameters, and then a motor nerve pattern will be produced to match the image. This of course represents the 'ideal aid' which is completly replacing the sense of hearing (compare fig 7 with fig 4).

Having stated what an ideal aid would do one must modify the model to allow for the practical limitations of any available system. Using present knowledge and technology it is not possible to analyse speech in such a way that all aurally differentiable sounds are differentiated, (as an example of an attempt to do this see Kung pu li et al (1973)). Thus a more likely configuration is as in fig 8. Here the feedback loop provided by the display and the mental imagary created by it

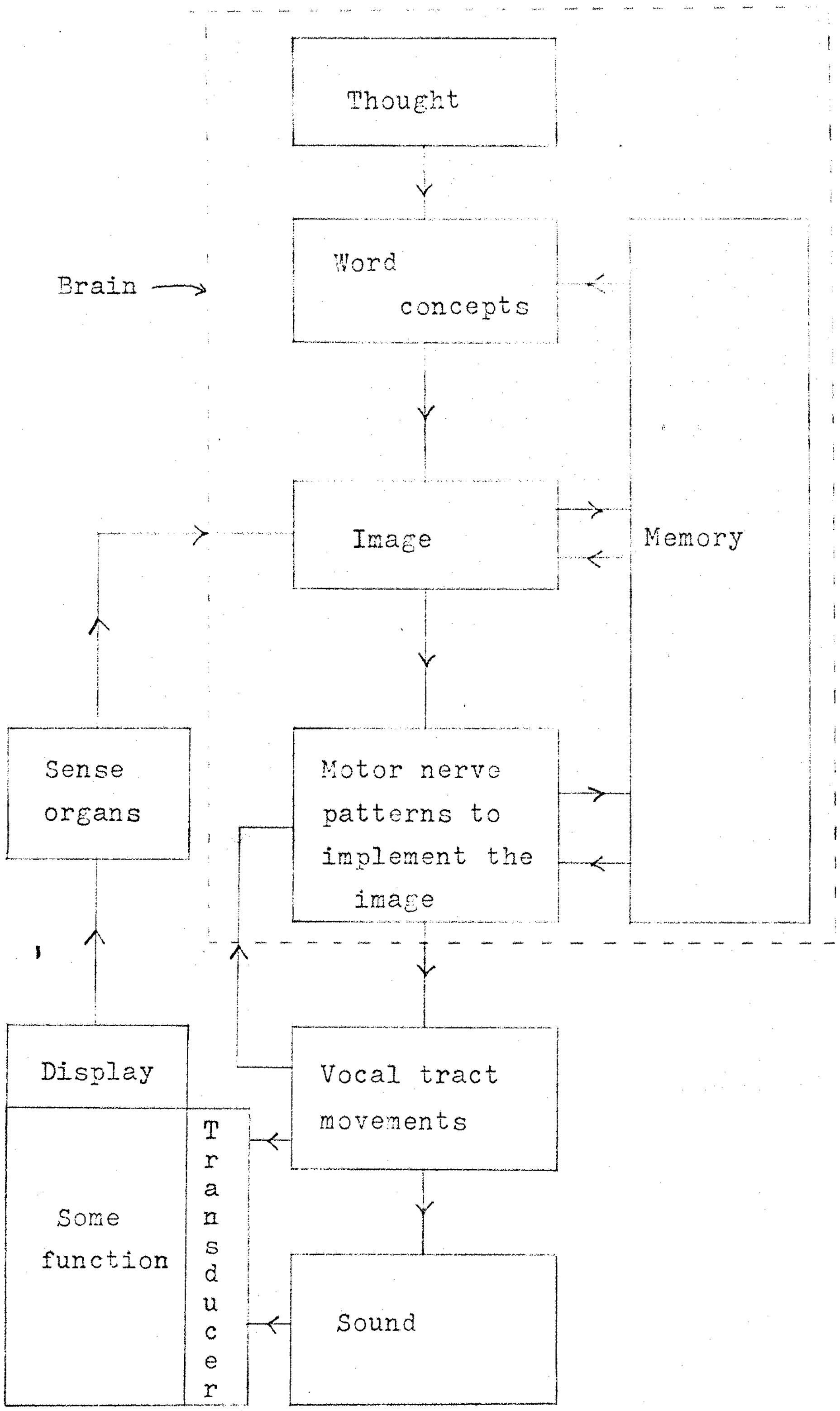


fig. 7 Speech production using a 'generalised' aid

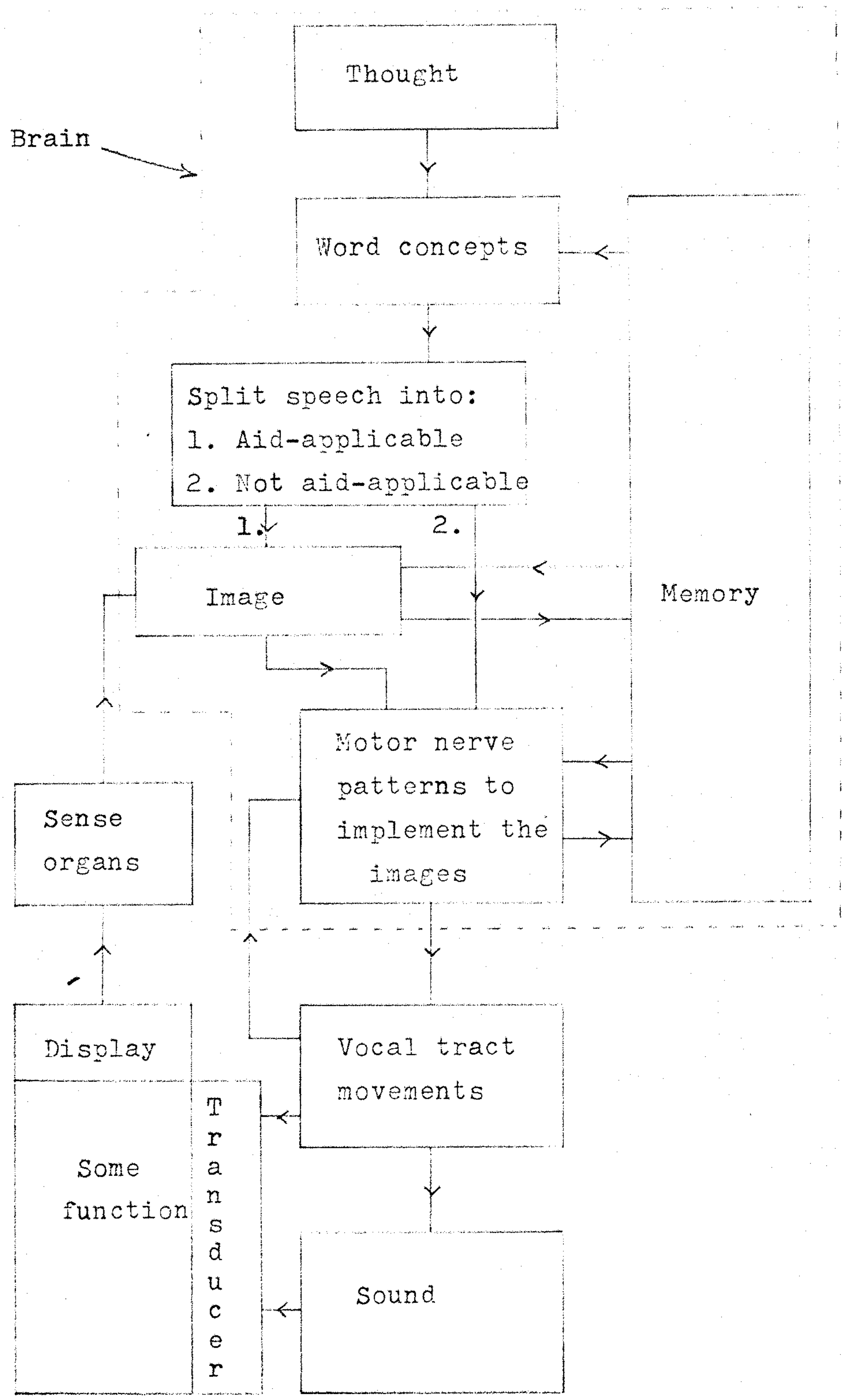


fig. 8 Speech production using a practical aid

are only used for obtaining certain restricted clues about the speech patterns. For example, if the display is a pitch indicator, for a particular word or phrase the memory will produce a required pitch pattern and the display feedback loop will control the pitch during the phrase to fit that pattern; the remainder of the motor nerve pattern controlling articulation will depend on memorised patterns and the inner feedback loop.

The type of aid that is proposed in this report is a teaching machine , this implies that it will be used for short periods of time, lessons, and thus its teaching capabilities ought to be studied. This means that at this stage consideration should be given to the factors which effect the remembering of learnt patterns. According to Cermak L.S.(1972) long term memory has unlimited capacity and is affected mainly by interference of similarly encoded patterns which block access and retrieval of wanted items. There is also a small unlearning (decaying with time) effect. (As with most pyschological phenomena there are many ways of looking at this process. The above statement must also inevitably be a great simplification From the above statement one can draw the conclusion that the less numerous are the patterns to be remembered the more reliably they will be remembered.

According to fig 8 the memory is storing two sets of patterns for each sound: an image related pattern and a motor nerve activity pattern. However this could be simplified, and the strain on the memory reduced, if the image caused by the feedback loop of the aid was of the same form as the motor nerve activity pattern. This logic suggests a teaching aid that displays motor nerve patterns. In fact the only conscious image that exists is the one caused by the movement of the articulators and not the one causing it (these two may be of different forms: generalised tactile sensation as against pulses going to each individual muscle, but they both contain the same infomation).

So a practical aid based on these principles would display the tactile sensations involved in speech, and the teaching process would involve the imitation of correct movements/ sensations (probably provided by a teacher). Thus the patterns that the deaf person is being taught are the actual sensations that he will feel when articulating correctly, he is therefore using fully the one control loop at his disposal.

4.2 Constructing a list of display parameters

Speech involves the use of a considerable portion of the body: all the muscles of the mouth - tongue, jaw, lips, velum and larynx, and the muscles that control the lung volume ie. those of the chest and abdomen. These present a continuum of tactile sensation, but in order to simplify matters parameters must be chosen that represent activity in the area of interest.

The following paragraphs contain a list of parameters that are considered to be the absolute essentials, in due course it may be found that more are needed to represent articulation accurately.

Two parameters are needed to represent the position of the tongue: height of highest point, and distance from the highest point to a fixed reference point, for example the front teeth.

Two parameters are also needed to represent the lip opening: width and height.

The jaw opening and velum opening might each be represented by one parameter.

The effects of the larynx muscles are a little complex. The assumption is made that their action can be represented by a single parameter called vocal cord tension (for anatomical details and justification of this assumption see Appendix I).

Variations in lung volume cause variations in lung pressure, and in particular sub-glottal pressure. This is taken as the final parameter.

Thus we have eight parameters. In order to display these it is first necessary to measure them, a considerable problem. To simplify the problem it was decided to concentrate initially on just two of these parameters.

The most difficult skills to teach a deaf person are breath and laryngeal muscle control, because of this their pitch, timing and stress are generally wrong. It was therefore decided to tackle just this problem. So the two parameters that were chosen were sub-glottal pressure and vocal cord tension.

4.3 The measurement problem

4.3.1 Introduction

The two required parameters (sub-glottal pressure and vocal cord tension) are both indicators of the state of some internal body function, and in order to make the proposed aid acceptable to the user they must be measured entirely from outside the body, with as little interference to the subject as possible. There are several possible media one could work in.

The two most obvious methods are firstly, direct measurement of the muscle potentials or nerve impulses in the relevant area thus giving a direct indication of the amount of effort being exerted in that area, and secondly, at the other extreme, to analyse the acoustic waveform in such a way as to obtain the required outputs.

The direct method has the obvious disadvantage that most of the muscles involved are well below the skin surface, so it is very difficult to measure their potentials in a non-invasive manner. There are also a large number of muscles involved in varying both of these parameters (see Appendix I) so measuring just one muscle's action in each case will not suffice. (Unless one muscle action could be found to represent the action of the whole group.) Thus if the direct method is used a large number of muscle potentials might have to be investigated, involving the use of many electrodes, and because of this the aid would almost certainly be unacceptable to the user.

The second method relies entirely on acoustic analysis, and has the great advantage that the only transducer required is a microphone. It should therefore be very acceptable to the user. The problem lies with the analysis that would have to be done on the acoustic waveform in order to produce the required outputs. As a literature survey of speech analysis techniques gave no great reason for hope in this field (see for example: Kung pu li et al (1973), Hyde S.R. (1968), I.E.E.E. ASSP - 23 no.1 1975) attention was turned to more specialised techniques.

ie. techniques specific to the measurement of just one parameter.

4.3.2 Parameter-specific measurement methods

Initially I shall concentrate on methods of measuring the sub-glottal pressure (P_S) as this will be shown later to be the first priority.

Several methods are available for the measurement of P_S. Firstly it can be measured directly via a hypodermic syringe inserted through the tracheal wall and connected to a pressure transducer (as used in Kunze L.H. (1964)). This is accurate but very obviously it could not be considered in this context.

Secondly a catheter (tube) can be passed down the vocal tract and through the glottis, so that the end is sufficiently below the vocal cords to measure the pressure unaffected by the disturbance caused by their vibration. This also causes considerable discomfort, but does not 'significantly' effect phonation (according to Van den Berg J.W. (1956)). Inserting the catheter in this position was apparently quite difficult, needing medical supervision.

Van den Berg's 1956 paper also contains a description and results of an 'indirect' method of measuring P_S. At a level just below the glottis the oesophagus is sandwiched between the trachea and the vertebral column, so pressure changes in the trachea are transmitted to the oesophagus. The method involves inserting a balloon type catheter in the oesophagus at this level and inflating it till it is pressing on both the trachea and vertebra. In this paper measurements were taken under the condition of abruptly stopping phonation and keeping the thorax and diaphragm for some time in the same position with an open glottis. This is claimed to eliminate an error due to the volume of the lungs varying during measurement, but errors of between 10% and 20% were still recorded. The gross limitation on the measurement procedure means that it is quite inapplicable.

The same instrumentation has been used during continuous speech, see Leiberman P. (1968) and Kunze L.H. (1964). Kunze shows that oesophageal pressure is very dependent on lung volume and it is stated that 'under no experimental conditions did this method provide a valid estimate of subglottal pressure'. Lieberman disagrees with these findings. He applied a simple linear correction factor (with respect to time) and obtained favourable results.

Apart from the disagreement about the accuracy of the method, putting a tube down a subject's throat is not going to be a popular method of approaching the problem of teaching a deaf person to speak. It has therefore been rejected.

An alternative method uses a combination of two measurements: the volume of the lungs and the flow rate into/out of the lungs. From these two the pressure inside the lungs can be computed. Assume for the moment that the temperature of the air and lung tissue is constant, and that both the air and the lung tissue obey the law $P_1V_1 = P_2V_2$ (most materials do obey this law for small changes in P and V).

Take V_{A1} = vol. of air in lungs at 1 atmosphere, at beginning of expiration

V_{L1} = vol. of lung tissue at 1 atmosphere, at beginning of expiration

V_{AP} = vol. of air in lungs at pressure P (atm.)

V_{LP} = vol. of lung tissue at pressure P (atm.)

I = air flow rate at P= 1 out of lungs

V_P = total vol. of lungs and air at pressure P_S

evidently $V_P = V_{AP} + V_{LP}$,

but at time T after start of expiration

$$V_{AP} = \frac{(V_{A1} - \int_0^T I dt)}{P_S}$$

and $$V_{LP} = \frac{V_{L1}}{P_S}$$

? V
? V_1

therefore $$P_S = \frac{V_{A1} + V_{L1} - \int_0^T I dt}{V}$$

$$= \frac{V_1 - \int_0^T I dt}{V_P}$$

Fig 9 shows how this method could be implemented. The integrator is reset and the value of V_1 updated once per intake of breath, so it gives an output for P_S during expiration.

It is desired that pressure differences as low as 1 cm of water should be registered, (speech generally involves pressures between 2 and 20 cm of water) this is a pressure variation of just 0.1% of atmospheric pressure. So small a variation in P_S will cause a correspondingly small change in V in accordance with the ideal gas law.

$$\frac{PV}{T} = \text{constant, } K .$$

If P increases to $P(1+\Delta)$ causing T to increase to $T(1+\delta)$ and V to decrease to $V(1+\gamma)$, then

$$V(1+\gamma) = \frac{K(T + \delta T)}{P(1+\Delta)}$$

$$\approx \frac{KT(1 - \Delta)(1 + \gamma)}{P}$$

therefore $\gamma = \delta - \Delta$ (assuming all increments are small)

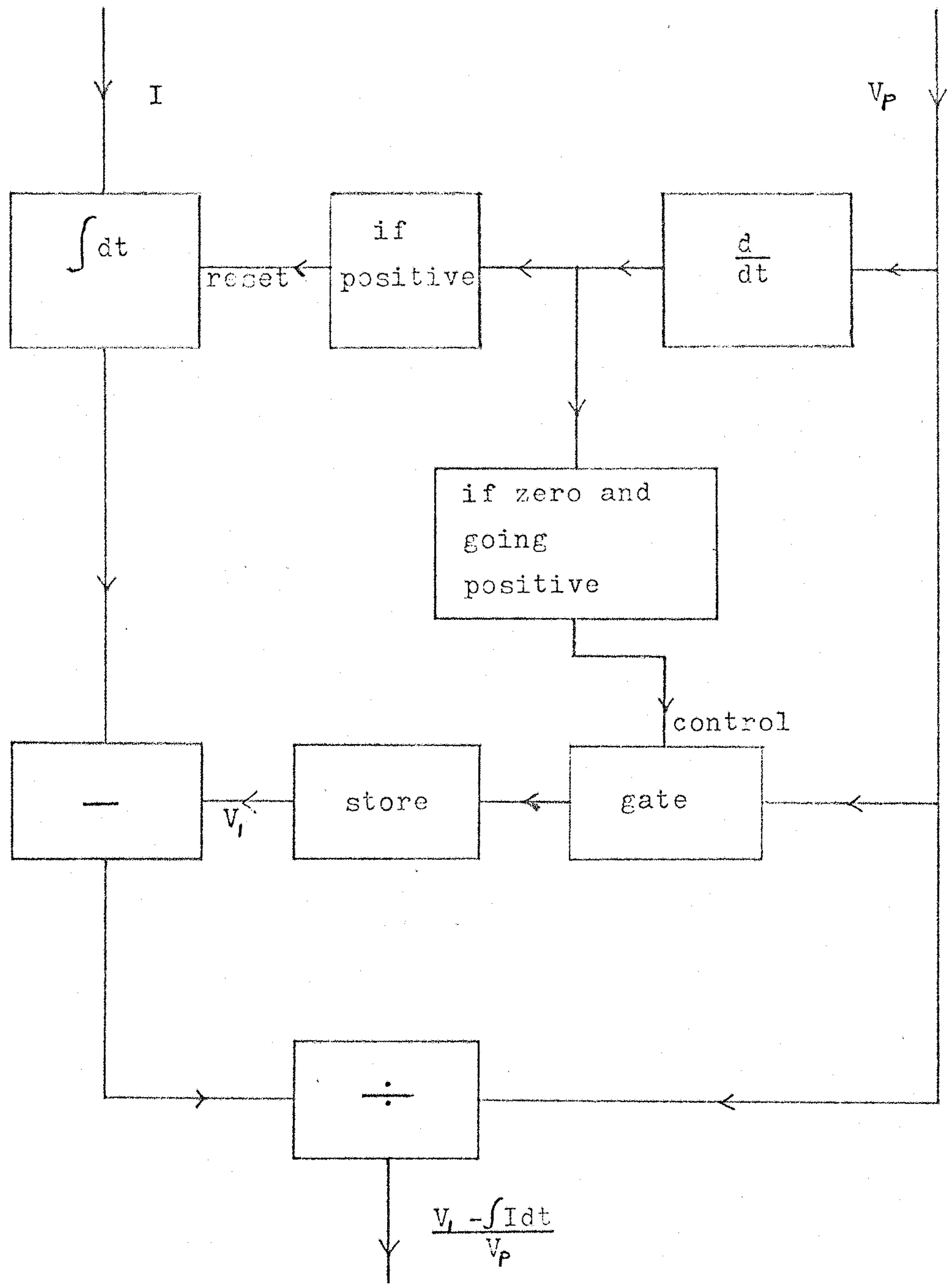


fig. 9 Implementation of P_s calculation

So even assuming that $\delta = 0$, (it is in practice going to be positive) so that γ has its maximum value, we must be able to detect changes in lung volume of the order of 0.1%. Having stated this rather demanding criterion let us consider methods for measuring V.

1. Hixon T.J. (1973) used electromagnetic transducers to measure the distance, front to back, through the chest and abdomen simultaneously. He did not plot any relationship between these measures and the actual lung volume so no facts are available about the accuracy of the system. Intuitively it seems most unlikely that such a simple measure, indicating only the distance between two pairs of transducers could give the volume of the intermediate mass accurately.

2. A well accepted way of measuring breathing rate and depth is by impedance plethysmography. There are many papers describing this technique, for instance Pacela A.F. (1966), Allison R.D. et al (1964), and Baker L.E. (1972). It involves the measurement of the electrical impedance through the chest at a frequency of between 50 and 100 kHz. However experiments have shown an error of about 5% between the lung volume predicted from the impedance and the lung volume measured by total encapsulation of the subject. (Total encapsulation gives the only reliable measure of lung volume but it cannot be used outside the laboratory. The subject is put in a rigid container that comes up to, and is sealed around, his neck. The total body volume and thus the variation in lung volume is determined by measuring the air flow into and out of the container.)

Had either method 1 or 2 been applicable the flow rate could have been measured by a mask, which would have been very accurate but restricting, or by an ultrasonic flowmeter system on the trachea as in Kalmus H.P. (1954).

Another possible and initially very promising method for obtaining P_s is from the mark to space ratio of the glottal pressure waveform during voicing (ie. T_1/T_2 in fig 2). The graph of closed/total vs. P_s is shown in fig 10 (this graph was obtained from two mathematical models (Flanagan J.L. et al (1968) and Ishizaka et al (1972)) and it seems to be fairly accurate in practice (see Lindquist J. (1970)). This curve is claimed by Ishizaka et al (1972) to be invariant with the other parameters (pitch, vocal cord tension, articulator configuration). There is however an unsupported statement in Van den Burg J.W. (1956) that this parameter is dependent on pitch.

Thus if one assumes that Van den Berg's statement was unfounded, taking an accurate measure of the closed/total ratio of the pressure waveform in the larynx before it is degraded by the acoustic filtering of the upper tracts and passing it through a suitable nonlinear network should give a measure of P_s.

There are several ways of trying to obtain the required waveform:

1. A throat microphone certainly reduces the effect of the upper tract resonances, but it does not eliminate them entirely. It also introduces distortions due to the signal passing through the tissue of the neck.

2. Passing a transducer into the larynx has many practical problems and would probably still produce some upper tract features.

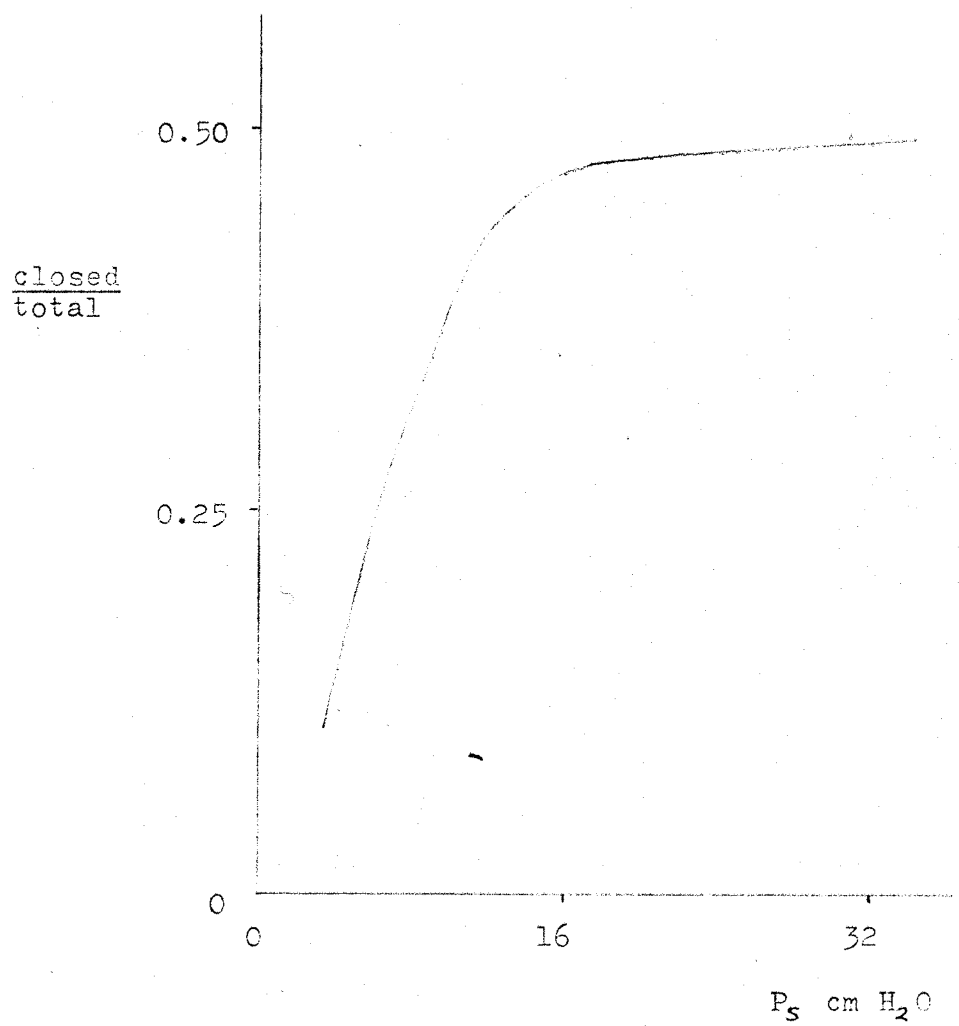


fig. 10 Relationship between $\frac{\text{closed period}}{\text{total period}}$ of the glottal waveform and P_S

3. Inverse filtering. This would seem to offer the best chance of obtaining the required waveform. The speech is analysed into its several resonances - defined by their centre frequency and 'Q factor' - these parameters are then used to control filters through which the speech is passed. Thus the effects of the upper tract are cancelled. Reports of this method (Lindqvist J. (1970), Miller R.L. (1959) and Rothenburg M. (1973)) use continuous vowel sounds as their inputs, presumably because of the errors that must be caused by the time delay in the analysis of the speech wave if the waveform changes. Its application to continuous speech is therefore likely to be of dubious success. This method was used by Lindqvist (1970) and the erratic results obtained could be an indication of the difficulty of measuring this parameter.

4. A method described in Sondhi M.M. (1975) is simple but not very effective or practical in continuous speech. The subject talks into a long reflectionless pipe, with a microphone at a set distance along its length. This provides a high acoustical impedance at the lips, and not a low one as is usual, so the effect of the tract is greatly reduced.

It should be noted that during most speech P_s lies between 5 and 20 cm. of water (see Van den Berg (1956) for example) and some of this range lies above the 'knee' on the curve in fig 10 thus any error in the measurement of closed/total ratio could produce very disproportionate errors in the calculated P_s.

Having considered methods of measuring P_s I shall now move on to the vocal cord tension parameter (Q).

A survey of the literature has revealed no significant attempt to non-invasively measure the muscular activity in the laryngeal muscules during speech. I found no information which would be useful in trying to correlate muscle potentials measured on the surface (which is all that would be practically possible) to the action of the muscles in the larynx. The next possible alternative for getting the relationship between Q and the measurable parameters seems to be a return to a mathematical model of the operation of the cords as no relationship between a simple measure of muscle activity and a general tensing of the vocal cords seems to be available.

Several mathematical models exist, see Flanagan J.L. et al (1968), Ishizaka et al (1972), Titze I.R. (1973) and Obenour J.L. et al (1972). The most informative of these is Ishizaka but the results of all of the models are fairly consistent. One can plot a family of curves on the $f_o.P_s$ plane for varying Q as shown in fig 11. These are all linear relationships to a reasonable approximation. There is a certain amount of agreement between this graph and results plotted in Ladefoged P. (1967) and Lieberman P. (1969) from readings taken from live human subjects. Thus given an accurate indication of P_s, together with f_o, values of Q can be obtained. The numerical accuracy of the results is a little doubtful as the various models and practical experiments predict different scales on the axes on this graph. However this is probably not critical as the absolute values are unnecessary in the context of the design of a teaching machine. Relative values are quite sufficient.

This method of obtaining Q seems to be the only one available and so measurement of P_s is of primary importance.

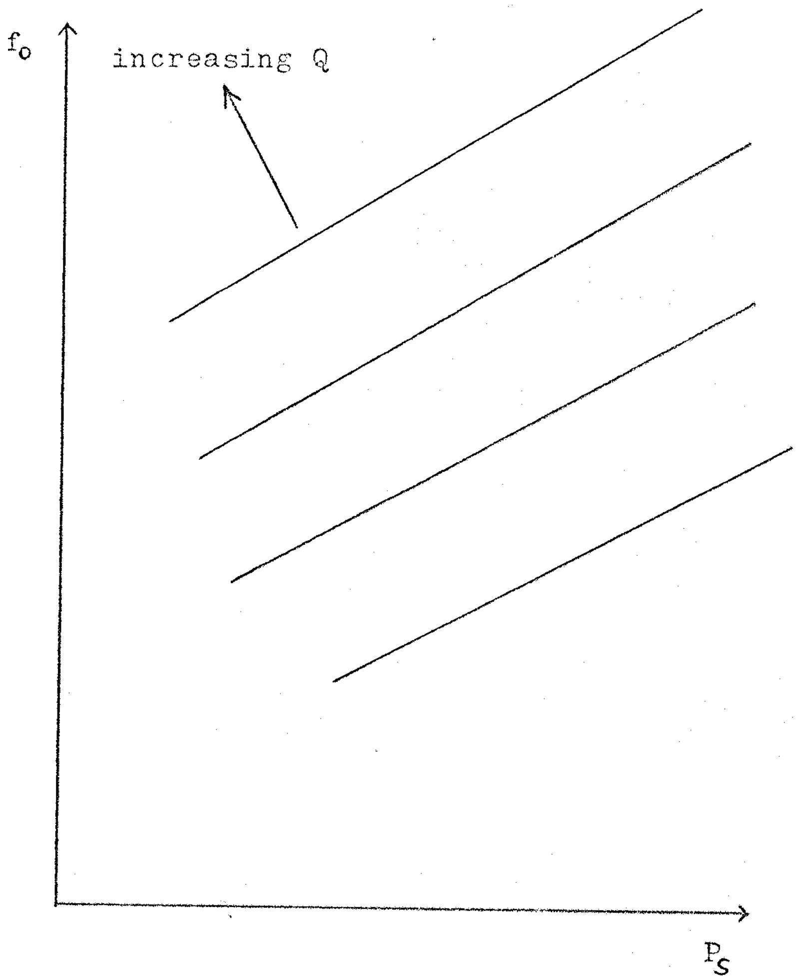


fig. 11 Relationship between pitch f_o, sub-glottal pressure, P_s, and vocal cord tension, Q

5. The measurement method decided upon

5.1 Introduction

After due consideration of the principles outlined in the previous section it was decided to attempt to obtain P_s by a method related to, but not necessarily directly related to, the measurement of the closed/total ratio of the sub-glottal pressure waveform.

This approach does have one very basic disadvantage in that an output is only obtained during periods of voicing so valuable information concerning consonants is lost.

As described in the previous section the methods of obtaining the pressure waveform are unsuitable for use in a teaching aid. There are, however, other ways of obtaining a measure of the vibration of the cords. These other methods must obviously give waveforms which are different to that of the pressure wave but it is thought that the required information should still be present in all the following measurements. This point will be discussed at length once the most practical technique has been chosen.

There are three methods available for obtaining a glottal waveform.

1. Optics - giving an open area function
2. Ultrasonics - proportion of cords in mutual contact
3. Electrical conductance - proportion of cords in mutual contact.

The optical method of monitoring vocal cord movement is by actually viewing the folds while they vibrate. This has generally been achieved by fixing a small mirror at the back of the mouth and filming via this. Unfortunately this makes normal speech impossible. A better method (see Lindqvist J. (1969)) uses an optic fibre inserted via the nose. Under these conditions normal speech is quite possible but this method is not suitable for the continuous on line monitoring that is required for a teaching aid for the deaf.

A literature survey revealed one report (Hamlet S.L. et al (1972)) of using the transmission of ultra-sound through the larynx to obtain an indication of the degree of vocal fold contact. This seems to have considerable problems due to critical transducer placement and erratic inter-personal differences in the results obtained.

One can use a device called a 'Laryngograph' (Fourcin A.J. et al (1969)) to measure the electrical conductivity across the larynx. This is said to give an output that indicates the degree of vocal fold contact. This has several practical points in its favour: it does not use a microphone and so is virtually independent of the upper tract shape (but not completely due to the slight loading effect of the upper tract on the vocal cords) and it is quite free of background noise interference and can only be stimulated by speech. In fact the output is very clean and free from artefacts. It is also a fairly simple device; it is easy to operate and it is claimed that the positioning of the electrodes is uncritical.

For these reasons it was decided to attempt to obtain a signal corresponding to P_s from a Laryngograph waveform. (The Laryngograph was supplied by Dr. A.J. Fourcin of the Department of Phonetics and Linguistics, University College London.)

5.2 The Laryngograph

Fig. 12 shows the operation of the Laryngograph. The oscillator generates a 1 MHz signal and is a voltage source. It feeds into the neck which can be regarded as a resistance network, as shown in fig 13. In this figure $R_{C1,2,3,4}$ are contact resistances between electrodes and skin, and $R_{S1,2}$ are the resistances across the skin surface and $R_{N1,2}$ are the desired resistances through the neck. As can be seen the resistance across the neck is embedded in a network of unknown resistances.

An attempt was made to analyse a simple circuit model to find how dependent the output (ie. the current flowing through the Laryngograph input) is on the various resistances, see Appendix II. However the form was found to be too complex for analysis.

Appendix II also contains an account of a direct implementation of this circuit. It is shown that this simple circuit model must be improved in order to account for the physiology of the situation. Thus a great deal more work would have to be put into this modelling in order to get a useful result.

This was not thought necessary as the Laryngograph has a lower frequency cut-off and will therefore tend to eliminate the effects of changes that happen below about 10 Hz. Spurious resistance changes due to pressure on the electrodes and sweating are all at a frequency below 10 Hz. Thus the output will show only the variation in impedance caused by glottal oscillation, as desired.

Due to its unconventional nature a short description of the box in fig 12 labelled A.G.C. is necessary. The signal entering this box is a large carrier signal with a very small amount of amplitude modulation. In order to amplify this signal before

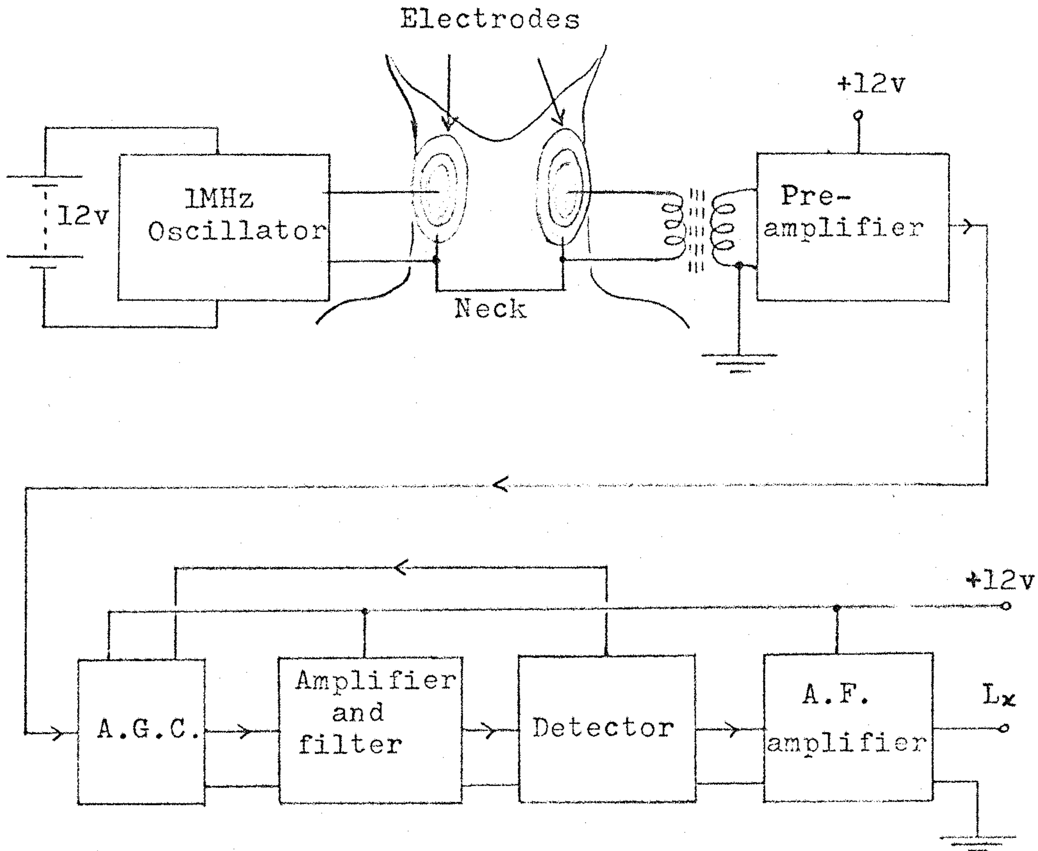
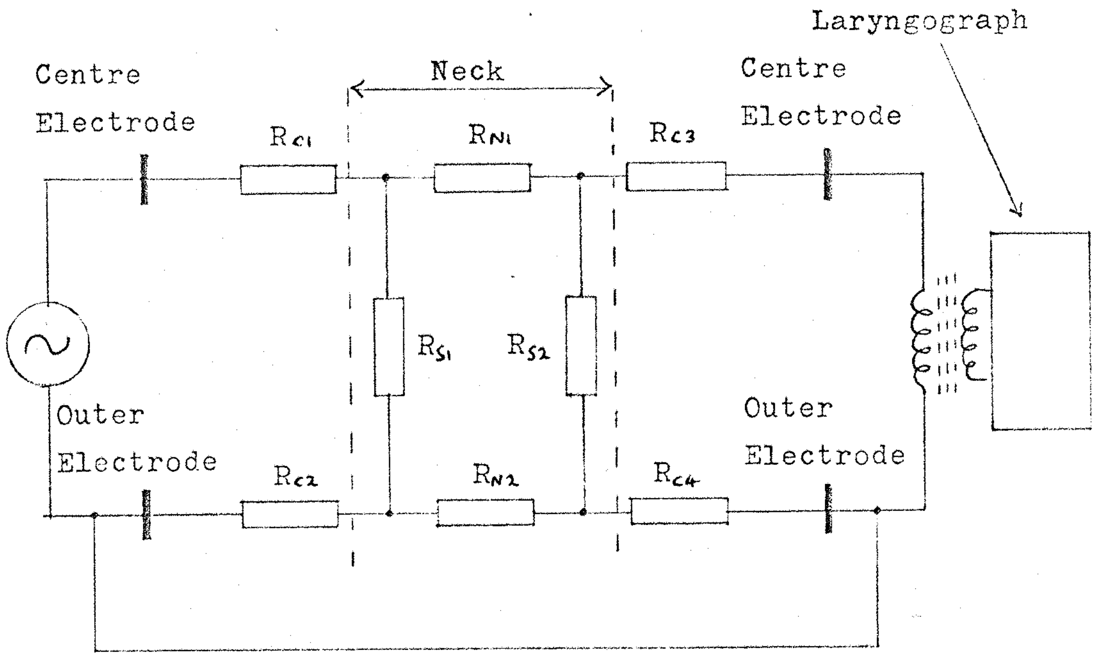


fig.12 Laryngograph schematic



R_c : Contact resistance

R_S : Surface resistance

R_N : Neck resistance

fig. 13 Resistance model of neck.

extracting the audio component the majority of the carrier is sliced off. The feedback from the detector provides a D.C. level below which no signal will pass the A.G.C. circuit. This is demonstrated in fig 14. **(This figure follows 15 (b))**

Thus the relative amplitude of the low frequency to the high frequency has been increased. Without this, amplification would cause very large carrier voltages before the required signal was of reasonable size. So this is not an A.G.C. in the normal sense and one can demonstrate that the overall circuit acts linearly, ie. that doubling the variation of the input current will cause twice the variation in the output. Fig 15 shows a circuit to test this and the results that were obtained.

The signal that the Laryngograph produces during speech is typically like that shown in fig 16. Comparing this with fig 2 reveals the waveform to be approximately the same shape as the glottal volume velocity but 180° out of phase. The flattened portions of fig 16 represent the open period.

5.3 Discussion of the Laryngograph waveform

A simple model was proposed for predicting the relationship between the volume velocity and the Laryngograph waveform (L_x). This model was built upon Flanagan and Landgraf's vocal cord vibration model (Flanagan J.L. et al (1968)) with the aim of obtaining reasonable impedance predictions.

Their model had one flat-faced mass vibrating at a distance x above a fixed base, as shown in fig 17(a). If one considers a conductance measurement across this, a relationship like that in

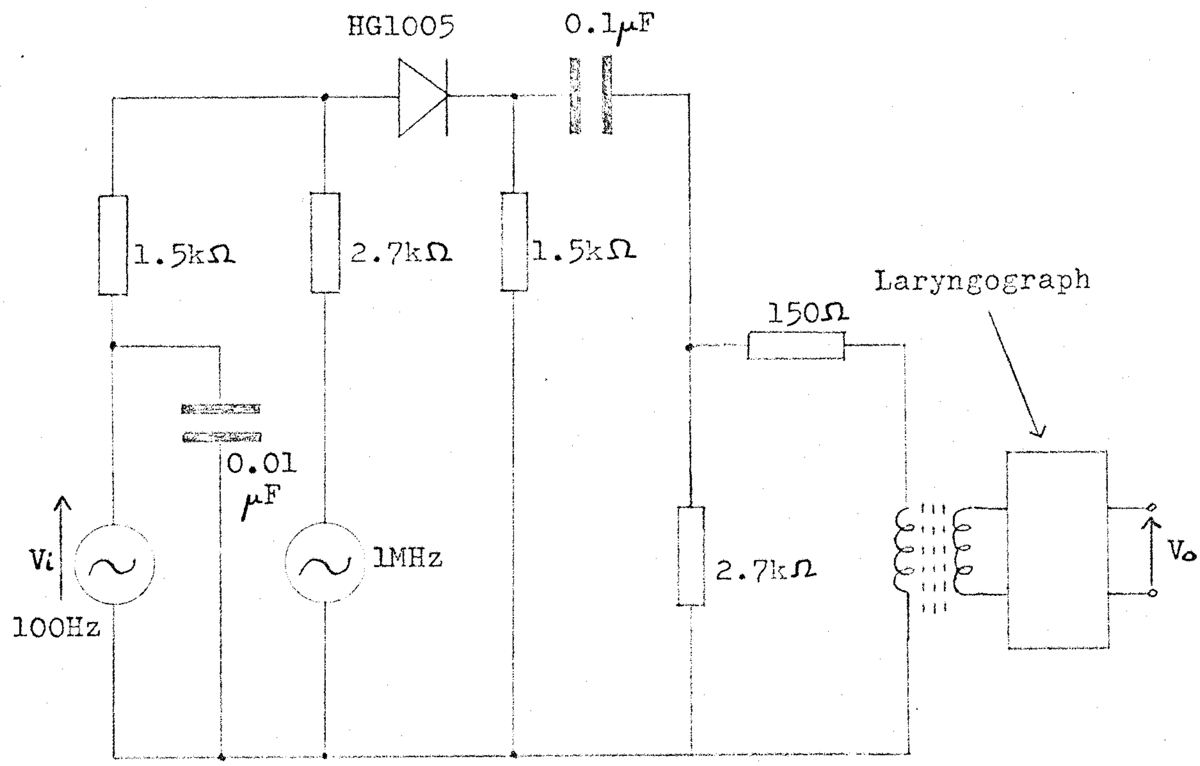


fig. 15(a) Circuit used for testing the linearity of the Laryngograph.

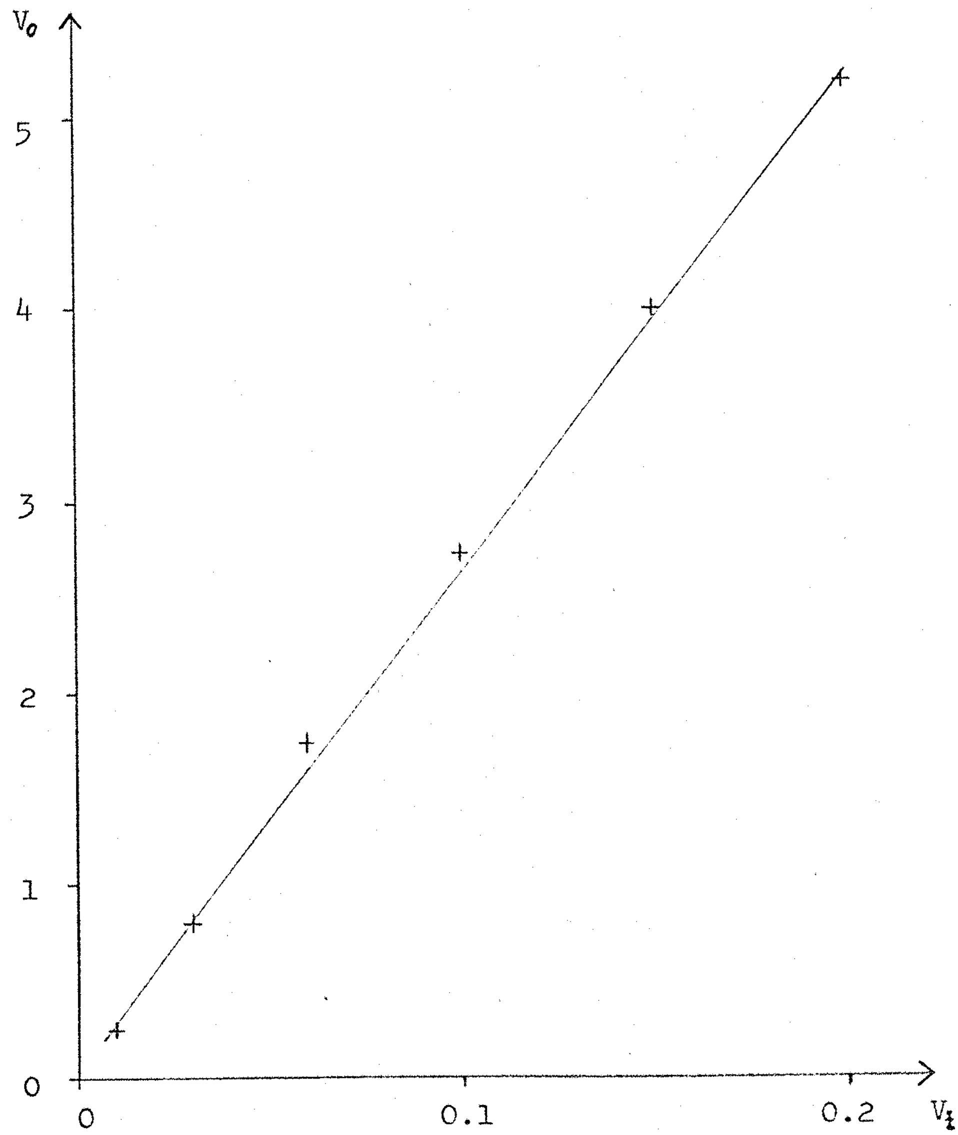


fig. 15(b) Linearity of Laryngograph circuit

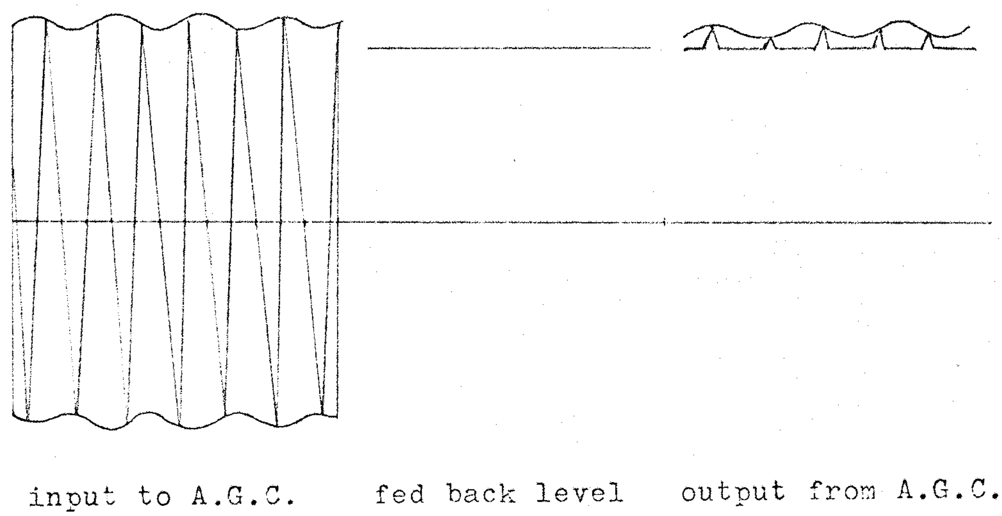


fig. 14 A.G.C. action

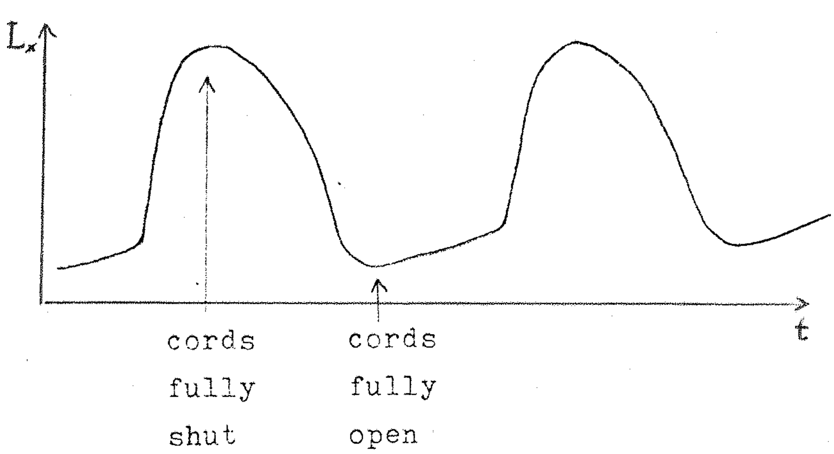


fig. 16 Laryngograph waveform (I_x)

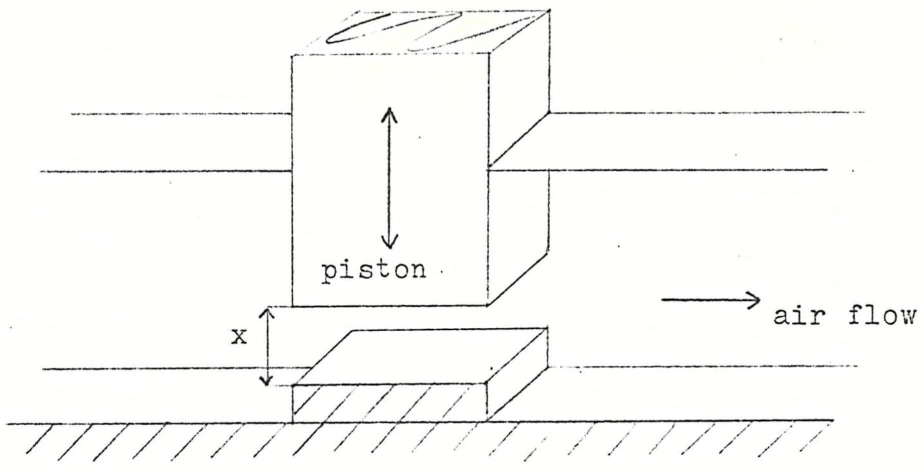


fig. 17(a) Vocal cord model

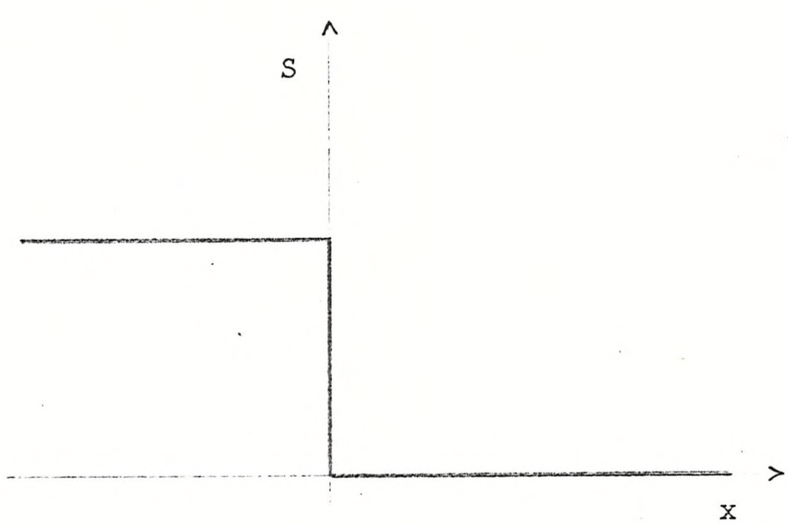


fig. 17(b) Conductance curve
S, against distance x.

fig 17(b) is obtained.

The simplest improvement to this is to add a conductance that is proportional to $\frac{1}{x+C}$ (where C is a constant to prevent the conductance from becoming infinite when x falls to zero) between the vibrating piston and its fixed base. Also one must add a constant series conductance (S_1) to account for the resistance of the non-vibrating outer tissue of the throat, and a conductance (S_3) that is switched in on contact of the cords. This gives a model as in fig 18(a), with a conductance function as in fig 18(b).

This cannot be considered realistic due to the discontinuity caused by the instantaneous cord contact. To overcome this S_3 must also be taken as being x dependent (and so it is placed in parallel with S_2). A reasonable function for S_3 is that shown in fig 19(a) and this gives a total conductance function as that shown in fig 19(b). Proceeding with this as a hypothetical transfer function between the two waveforms, it is necessary to determine first the relationship between x and the volume velocity, and secondly the relationship between the mark/space ratio of the conductance waveform and the closed/total ratio of the volume velocity waveform. Having done this the function drawn in fig 10 can be used to predict a relationship between the mark/space ratio of the Laryngograph waveform and the sub-glottal pressure.

Before doing this it is convenient to define the mark/space ratio of waveforms of this shape as being the ratio of the time spent above a set level to that spent below it. This level, by observation of the waveforms, would be at about 10% of their

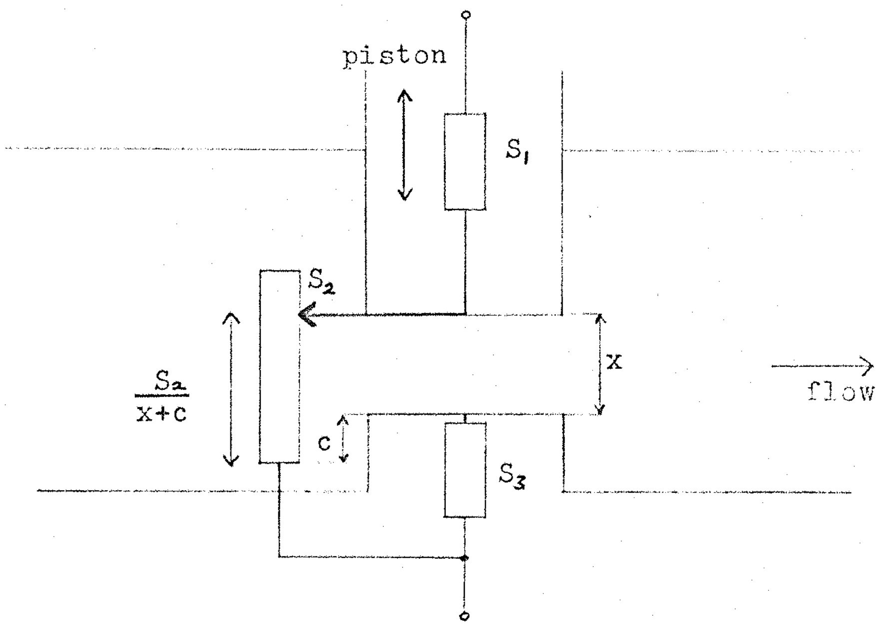


fig. 18(a) Resistance model of the vocal cords

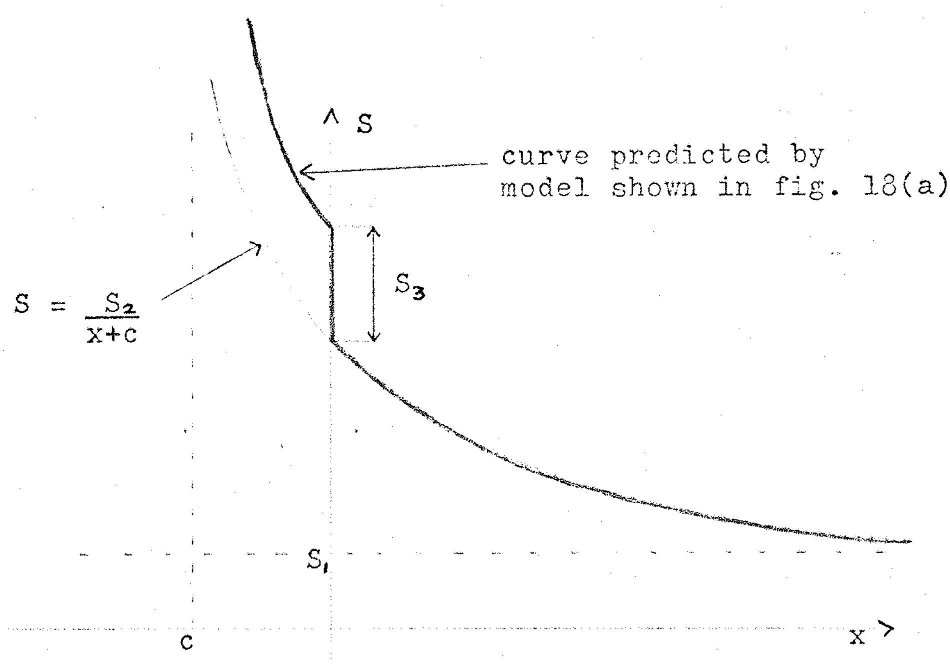


fig. 18(b) Conductance curves

S, against distance x.

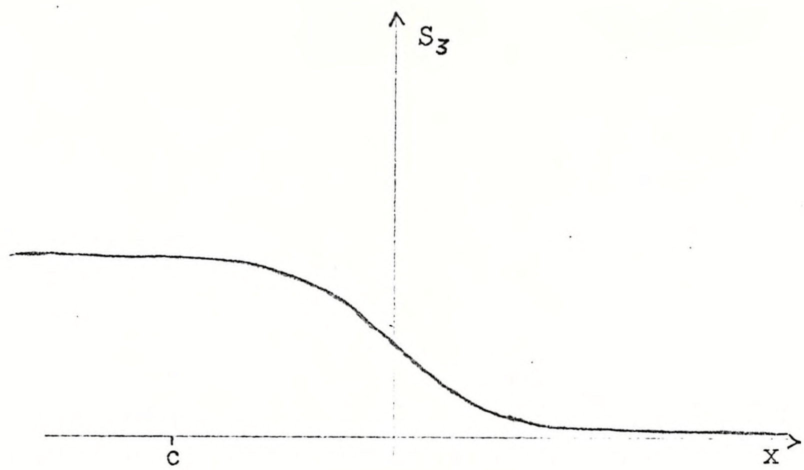


fig. 19(a) Possible function for S_3

S, against distance x.

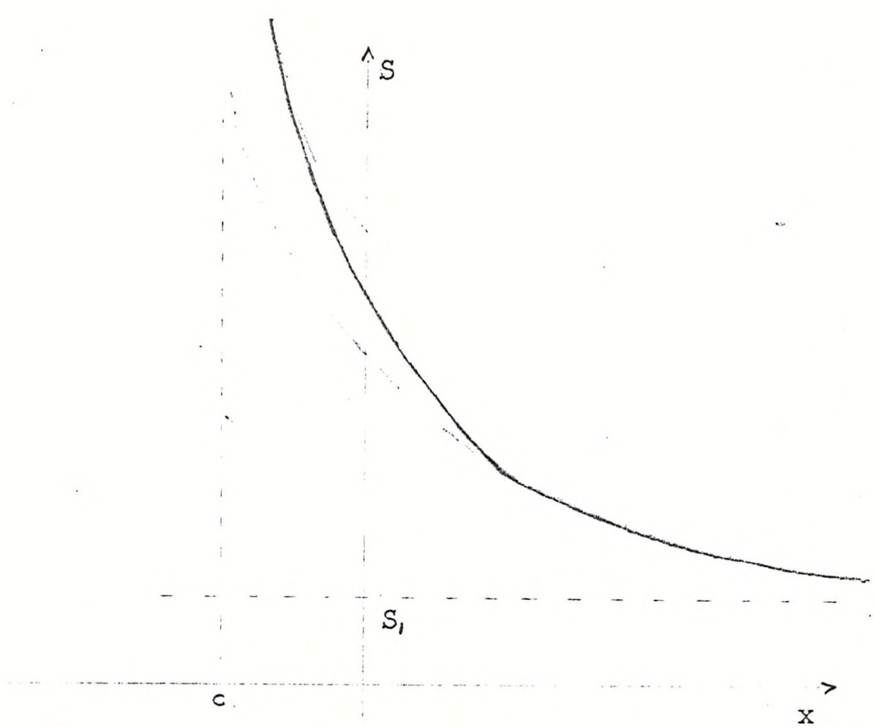


fig. 19(b) Proposed conductance function

S, against distance x.

height. It is shown a little later that this value is non-critical.

The detailed shape of the volume velocity waveform at the glottis is dependent on the loading effect of the upper tract, so it cannot be determined unless all the characteristics of the upper tract are known. In the present study, however, the only parameter of interest is the mark/space ratio. As defined above this is concerned only with the points of opening and closure of the glottis. As flow cannot occur when the open area is zero the mark/space of the volume velocity must be the same as the mark/space of x. This is confirmed by the results of Ishizaka et al (1972) and Lindqvist J. (1970).

Consider the waveforms shown in fig 20. In this a L_x(conductance) waveform and a volume velocity waveform (v) are drawn and various periods of time are indicated.

t_1 is the space period of v

t_2 is the space period of S

Υ is the total period.

From the diagram it is evident that:

mark/space of v $= \frac{\Upsilon - t_1}{t_1}$

mark/space of S $= \frac{\Upsilon - t_2}{t_2}$

and $\Upsilon = t_1 + t_2 + \Delta_1 + \Delta_2$.

Therefore, mark/space of v $= \frac{\Upsilon - t_1}{t_1}$

$$= \frac{\Upsilon - \Upsilon + t_2 + (\Delta_1 + \Delta_2)}{\Upsilon - t_2 - (\Delta_1 + \Delta_2)}$$

$$= \frac{(t_2 + \Delta_1 + \Delta_2)}{\Upsilon - (t_2 + \Delta_1 + \Delta_2)}$$

This is of the same form as the inverse of the mark/space of S.

If $(\Delta_1 + \Delta_2)$ could be assumed to be constant, or in any way predictable, the similarity would be very considerable. However insufficient knowledge about these waveforms means that these assumptions cannot be made, so one can go no further towards predicting the relationship between the conductance waveform and the volume velocity.

Having reached this theoretical dead end some additional empirical data was sought. This appeared in the form of some experiments performed at University College London while the the Laryngograph was being tested. The data consists of four sets of waveforms, each set containing a Laryngograph waveform and an area waveform. (The area waveforms were obtained by high speed cinematography of the vocal cords.) An example set is reproduced in fig 21. From these waveforms graphs of S vs. A were drawn. An example of one of these is shown in fig 22. The other three sets of data produced similar graphs. As can be seen the relationship is not a single line as predicted, see fig 19(b), but a loop instead. This implies a phase difference between the two waveforms.

This unpredicted phase difference between the conductance measure and the open area of the cords is probably due to the 'rolling' vertical component of the motion of the vocal cords as described in Appendix I. The simple mechanical model that was used to predict this relationship considered only horizontal motion of the cords, and so was inadequate in this respect.

The most relevant items of information contained in this data are the values of the Laryngograph waveform at the moments of opening and closure of the glottis, ie. when the area function is just increasing from zero and just returning to zero.

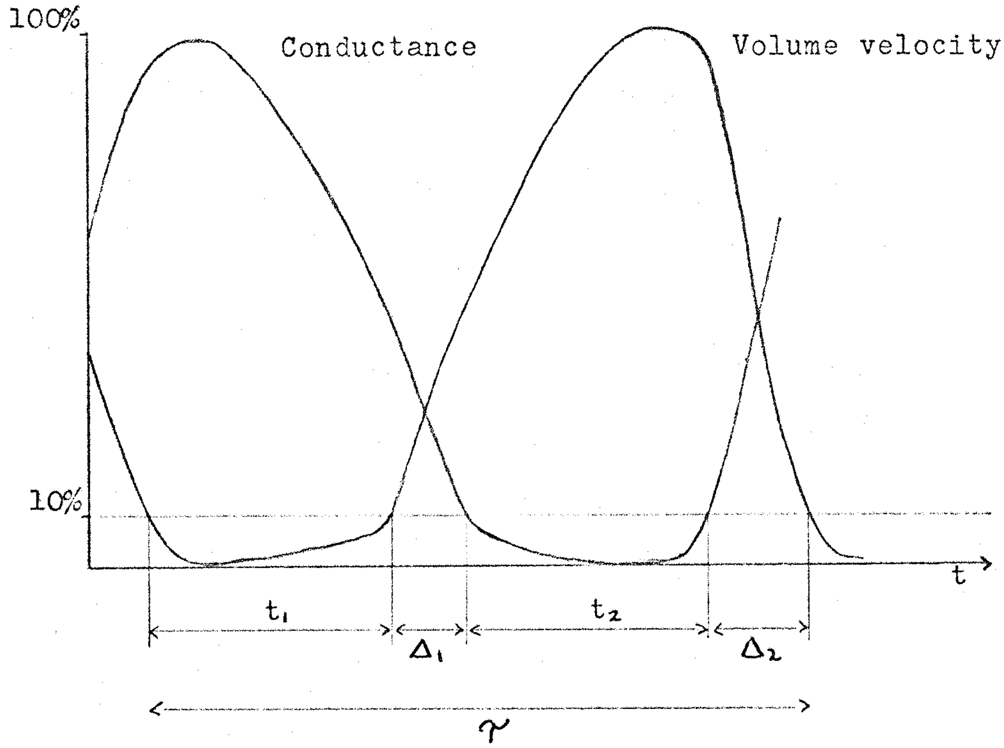


fig. 20 L_x and v waveforms, with various periods marked.

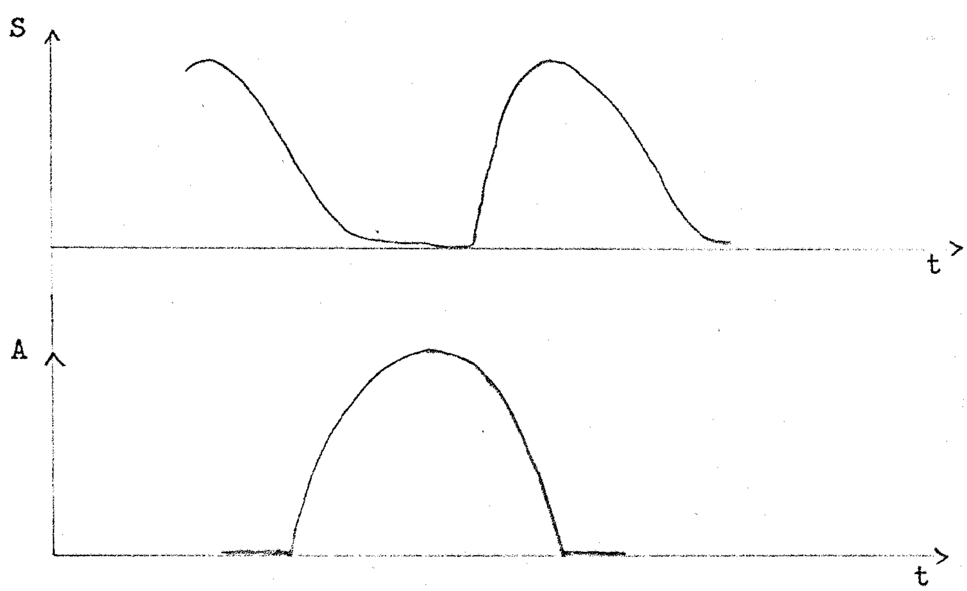


fig. 21 Example of Laryngograph vs. Area data

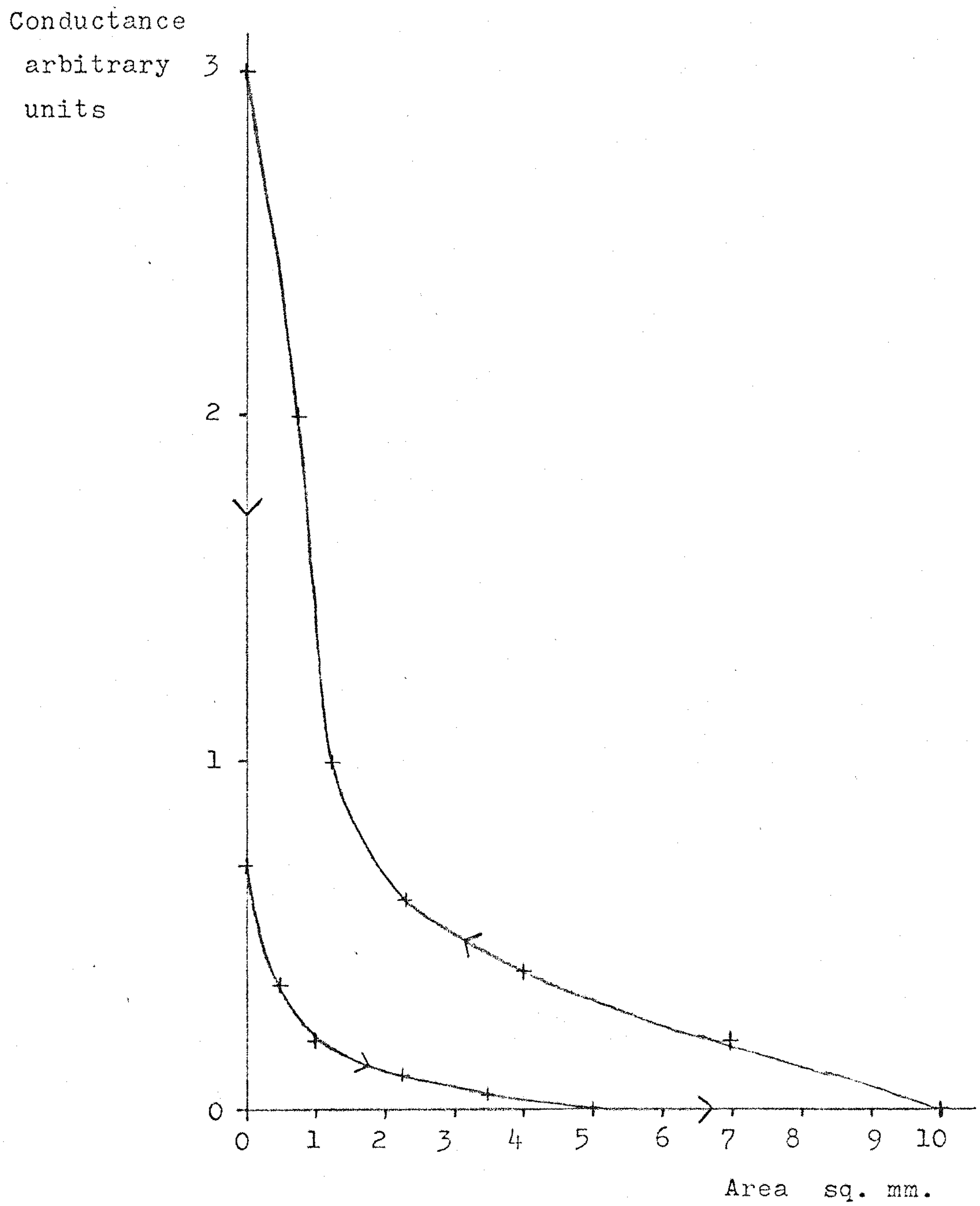


fig. 22 Plot of conductance against area

At the opening point the Laryngograph output is always between 0.6 and 1.0 on this arbitrary scale and the closure occurs at, or very shortly before, the maximum of the Laryngograph waveform.

Thus the mark/space ratio of the area (and so the volume velocity) should be obtained by defining a different mark/space for the Laryngograph based on the above criteria ie. as defined in fig. 23.

5.4 The circuitry: implementation of objectives

5.4.1 The chopping circuit

A circuit was designed and built to square-up the Laryngograph waveform. This is designed to chop the input at two independent levels, so providing the sort of hysteresis described in the previous section. Both levels are pre-settable proportions of the total amplitude of the waveform. The upper threshold level is set very near the maximum amplitude and the lower level is set at about 30% above the minimum. The lower level is not quite as proposed in the previous section as it is relative to the maximum amplitude and not a constant voltage. However the constant voltage technique is not possible in practice, as the Laryngograph will give considerably different ampltudes of signals from different people, and from different electrode positions on the same person. It was decided that a compromise must be made and a relative amplitude was chosen.

An important design objective was to minimise the time constant of the overall proposed teaching aid as a large time constant implies a large delay between a signal change at the

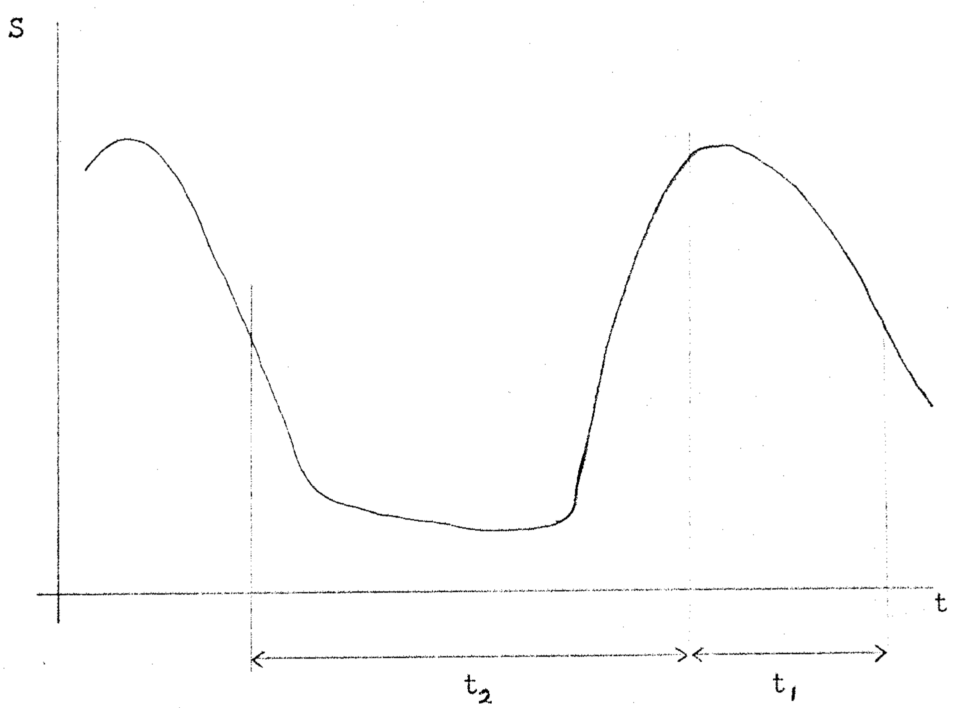


fig. 23 Mark to space ratio of the Laryngograph = $\frac{t_1}{t_2}$

input due to an articulatory event and the corresponding output. The longer the time delay the less effective the machine will be at performing its designed function.

The circuit of the chopper is shown in fig 24 and the associated waveforms are shown in fig 25.

There follows a brief description of its operation. Transistors T_1 and T_2 give amplification - gain varied by P_1 - and filtering. The high pass cut-off is at 10 Hz and the low pass cut-off is at 1 kHz, both are single order. Thus the input to T_3 is a large signal, several volts from peak to peak. The sole function of T_3 is to adjust the quiescent voltage at this point so as to minimise the offsets in the following circuits.

The voltage across C_1 and C_2 follow the positive and negative peak voltages almost exactly; the voltage across D_1 and D_2 being compensated for by ICI with its associated diode feedback circuit.

There are two time constants involved in this peak-picking circuit: rise and fall. The rising time constant is limited mainly by the output impedance of the operational amplifier and is less than 1 ms. The fall time constant was minimised to a value of 40 ms. Any less than this would have made the ripple voltage across the capacitors unacceptably high.

The two comparators IC2 and IC3 produce two different squared waveforms. These are combined by the logic and IC4 returns the signal voltage to a full 30 volts swing.

This circuit has been tested with sine and square waves and it works well over the range of input voltages and frequencies that are likely to be encountered in operation.

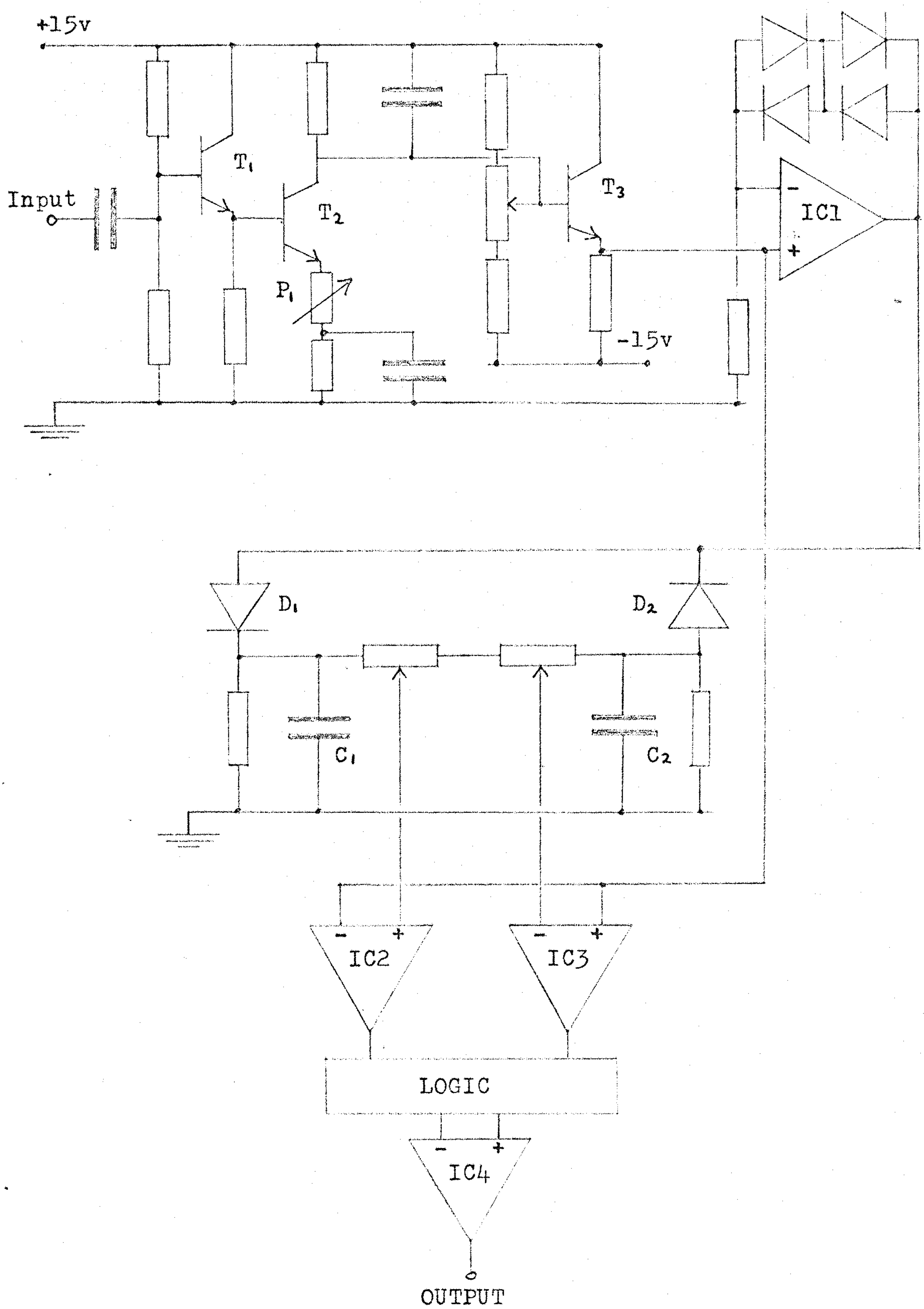


fig. 24 Waveform chopping circuit

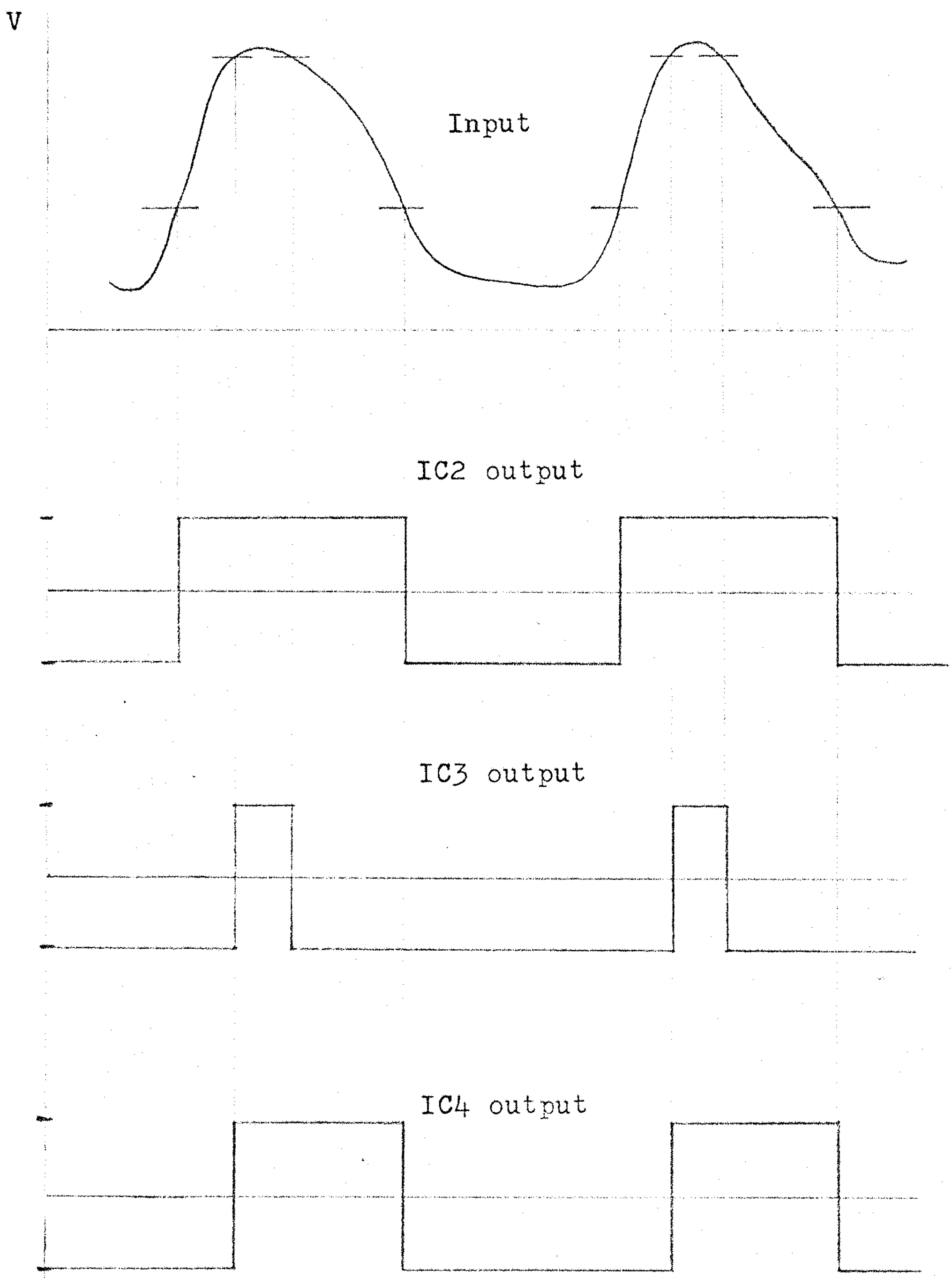


fig.25 Waveforms corresponding to fig.24

5.4.2 Circuit to calculate mark/space ratio

The standard method of obtaining a signal that is proportional to the mark/space ratio of a squared signal is by an R.C. integrating circuit. This simple method was rejected because of the need to keep the time constant of the device as low as possible. Instead a fairly complex circuit was constructed, a block diagram of which is shown in fig 26. This produces the mark/space ratio of the input every two complete cycles of the input, as shown by the waveforms of fig 27.

Integrator no. 1 measures the length of every positive pulse and integrator no. 2 measures the length of every other negative pulse. In this way once every two cycles the positive time is present at the output of integrator 1 simultaneously with the negative time being present at the output of integrator 2. The output of integrator 1 is divided by the output of integrator 2 and the result sampled over the appropriate period and stored on a capacitor.

The circuit was tested initially with sine and square wave inputs. The accuracy was found to be within 5% over the frequency range 80- 400 Hz, and it is linear over the mark/space range of 3:1 to 1:3 within the same accuracy.

Due to the large amount of gain in the chopping circuit the system is very sensitive even to low amplitude signals at the input. During any period of non-voicing the Laryngograph outputs a low amplitude noise signal. The chopping circuit will tend to trigger on this so causing a highly undesirable random output in the absence of a real input signal. To prevent this a threshold circuit is included, see fig 26. This operates by shorting the

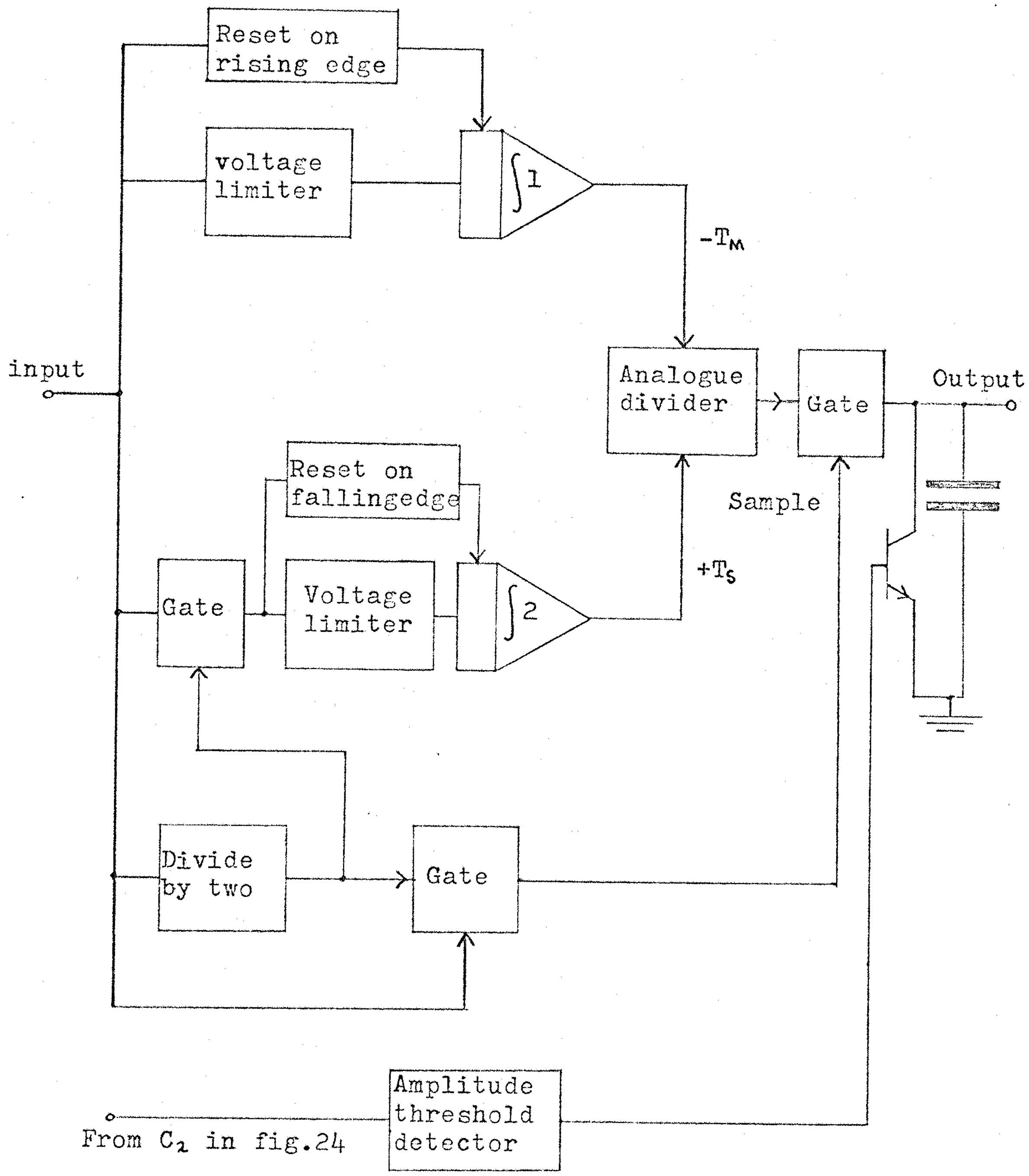
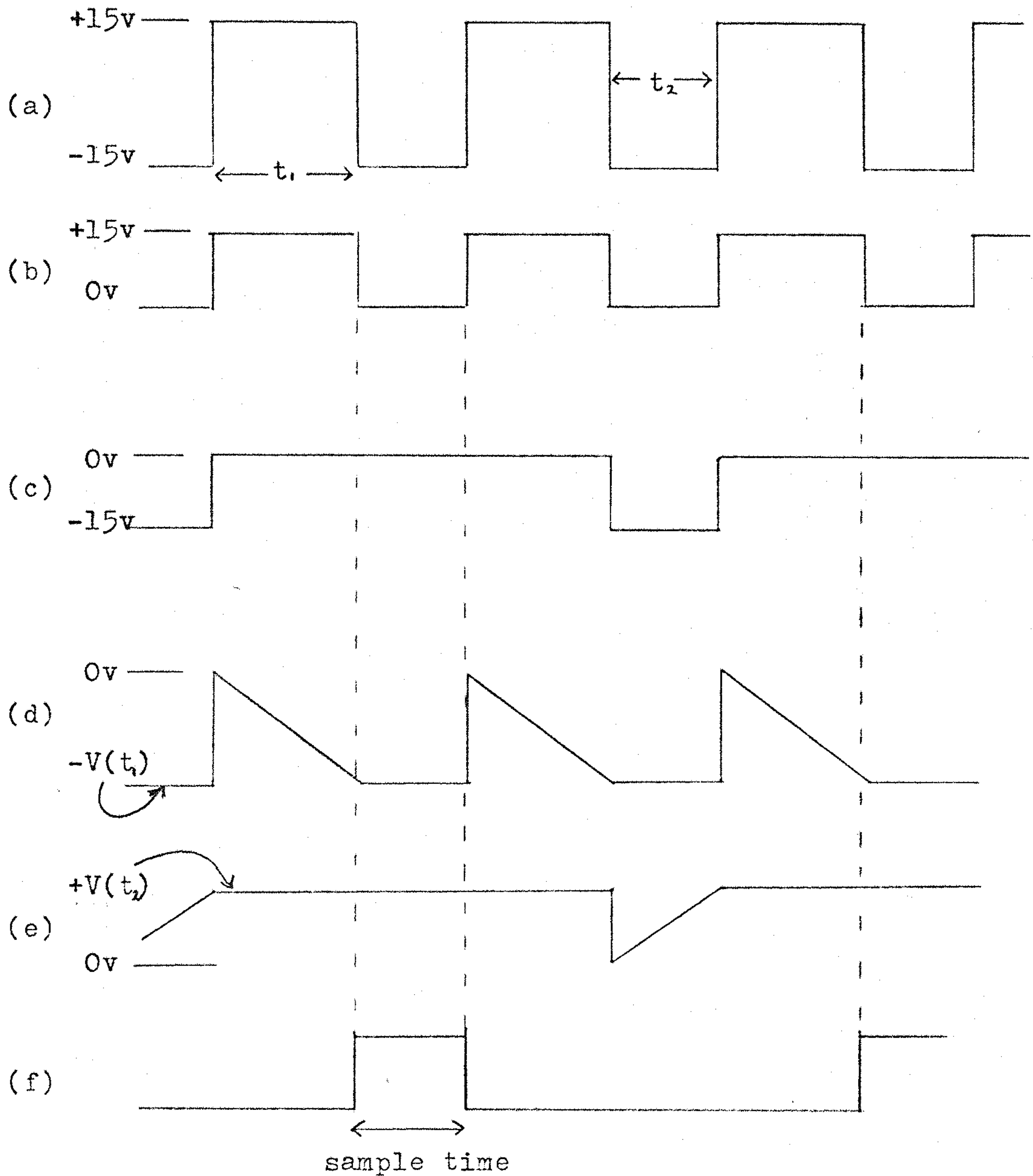


fig.26 Schematic of the Mark/Space measuring circuit



(a) Input
(b) Input to integrator 1
(c) Input to integrator 2
(d) Output of integrator 1
(e) Output of integrator 2
(f) Sampling time for the output gate.

fig. 27 Waveforms produced by the circuit of fig. 26

circuit's output to ground when the height of the positive signal peaks is below a pre-set level. The time constants involved in this are the same as those in the peak-picking circuit and so the time delay on initiation of voicing will be very short but the delay at the cessation of voicing will be considerable. This means that at the end of a period of voicing there is a section of meaningless output.

Examples of some waveforms obtained when testing this circuit (including the chopping circuit) with a switched sine wave input are shown in fig 28: (a) and(b) show how quickly the circuit attains the correct output and (c) and (d) show the effect of the time delay on cessation of the input. The results of two similar tests with the Laryngograph output as input are shown in fig 29. Upon consideration of these waveforms it was decided that the speed of decay at the end of an utterance would have to be increased at some future date before the device would give an adequate response. However the more important aspect at this stage was testing the overall system to see if it fulfills the objectives set out in section 4. This is discussed in section 6.

5.4.3 Pitch determining circuit

Referring to section 4.3.2 and in particular to fig 11, it is evident that to obtain an output that is related to the vocal cord tension there are two requirements: a measure of P_s and a measure of f_o . A measure of f_o was also required during the testing described in section 6.

A circuit was therefore built to calculate f_o. It is shown in fig 30.

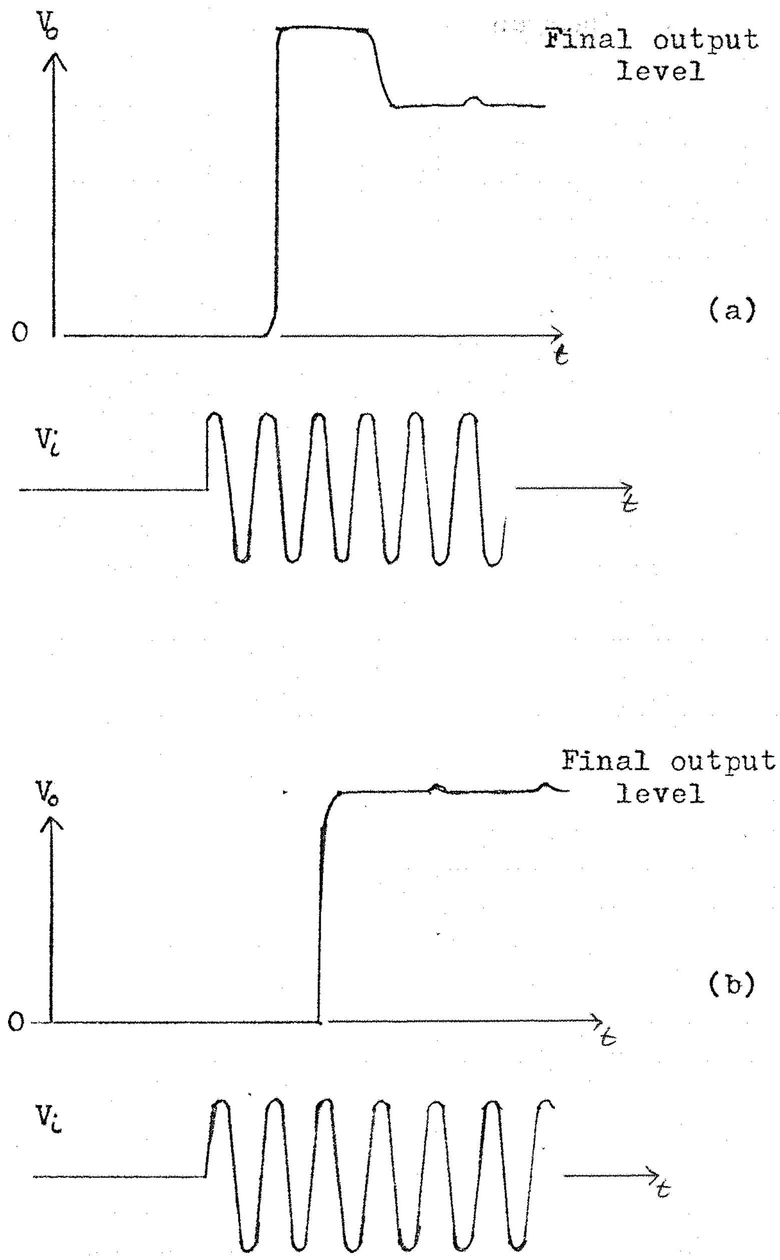


fig. 28 (a) & (b) Test results showing response of the equipment to switched tone input....leading edge. (300Hz signal...5mS/cm)

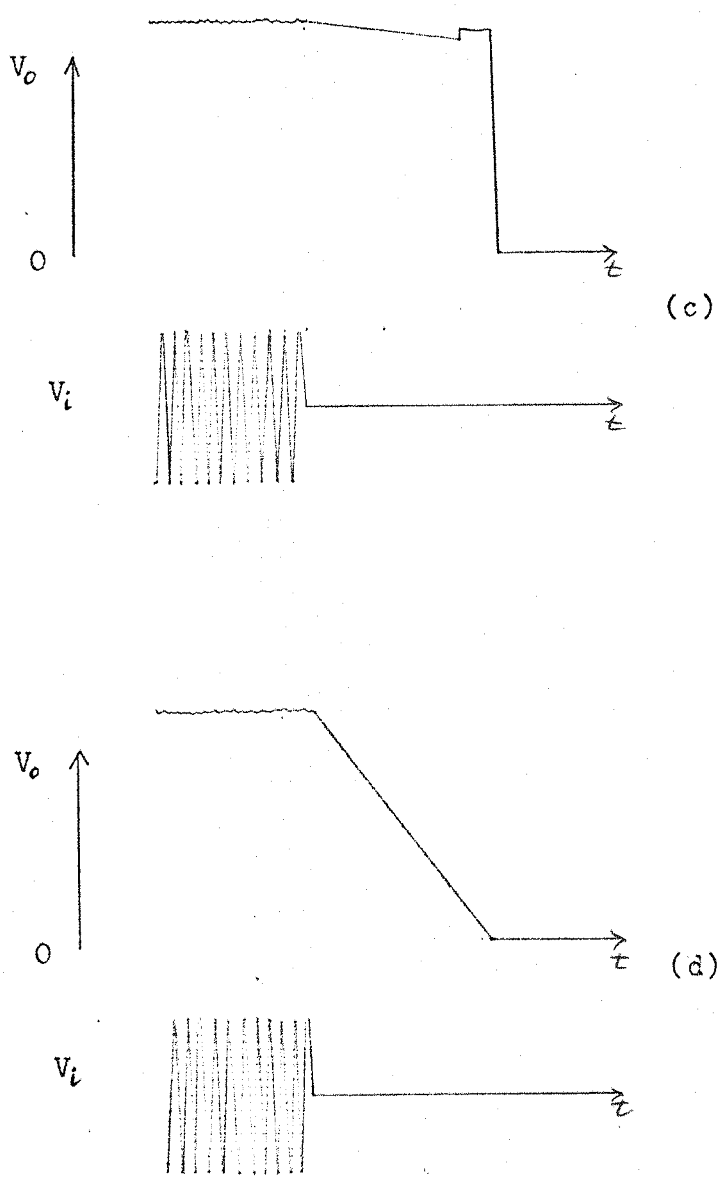


fig. 28 (c) and (d) Test results showing response of the equipment to switched tone input... trailing edge. (300Hz signal...20mS/cm)

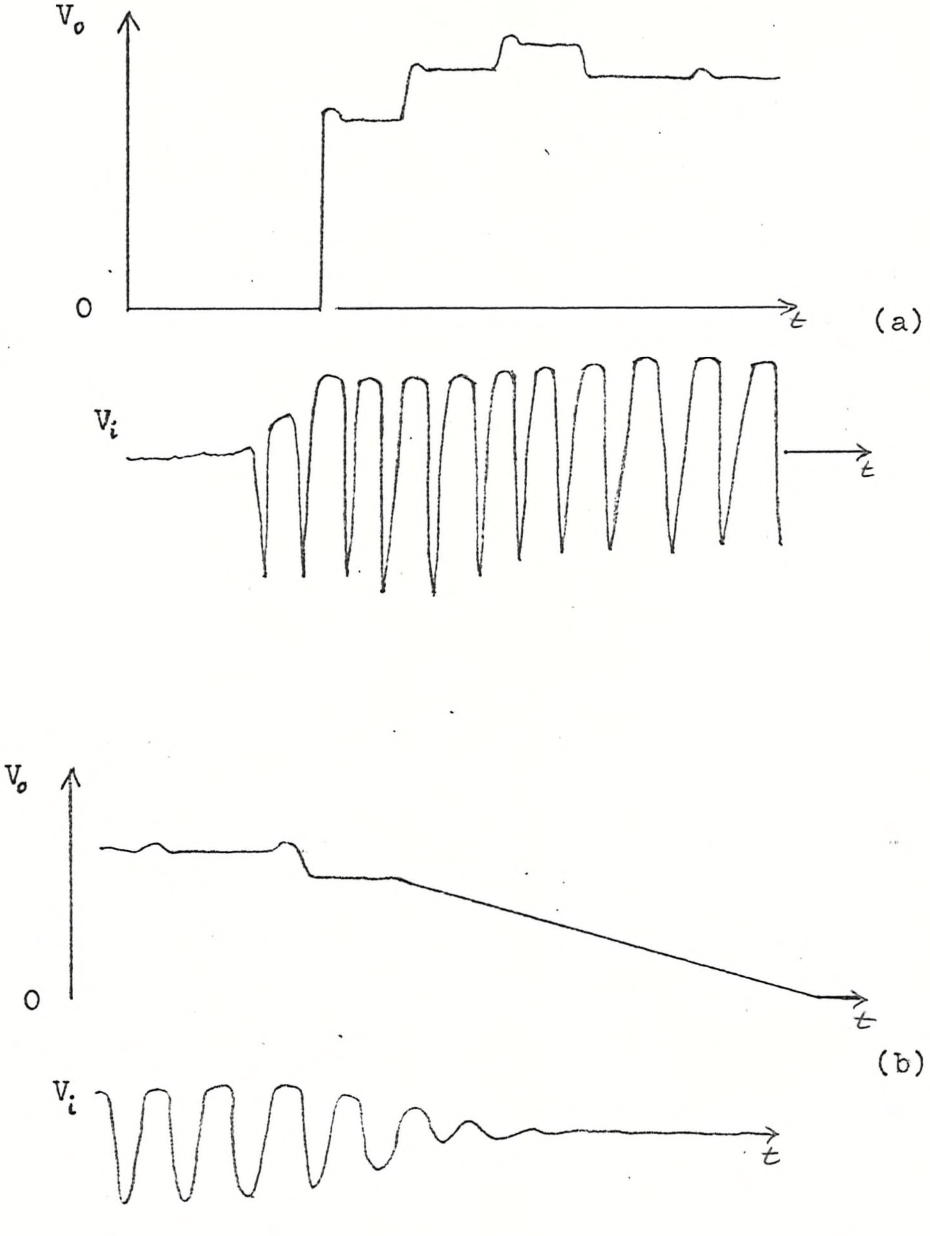


fig. 29 (a) and (b) Test results using L_x waveforms (1mS/cm.)

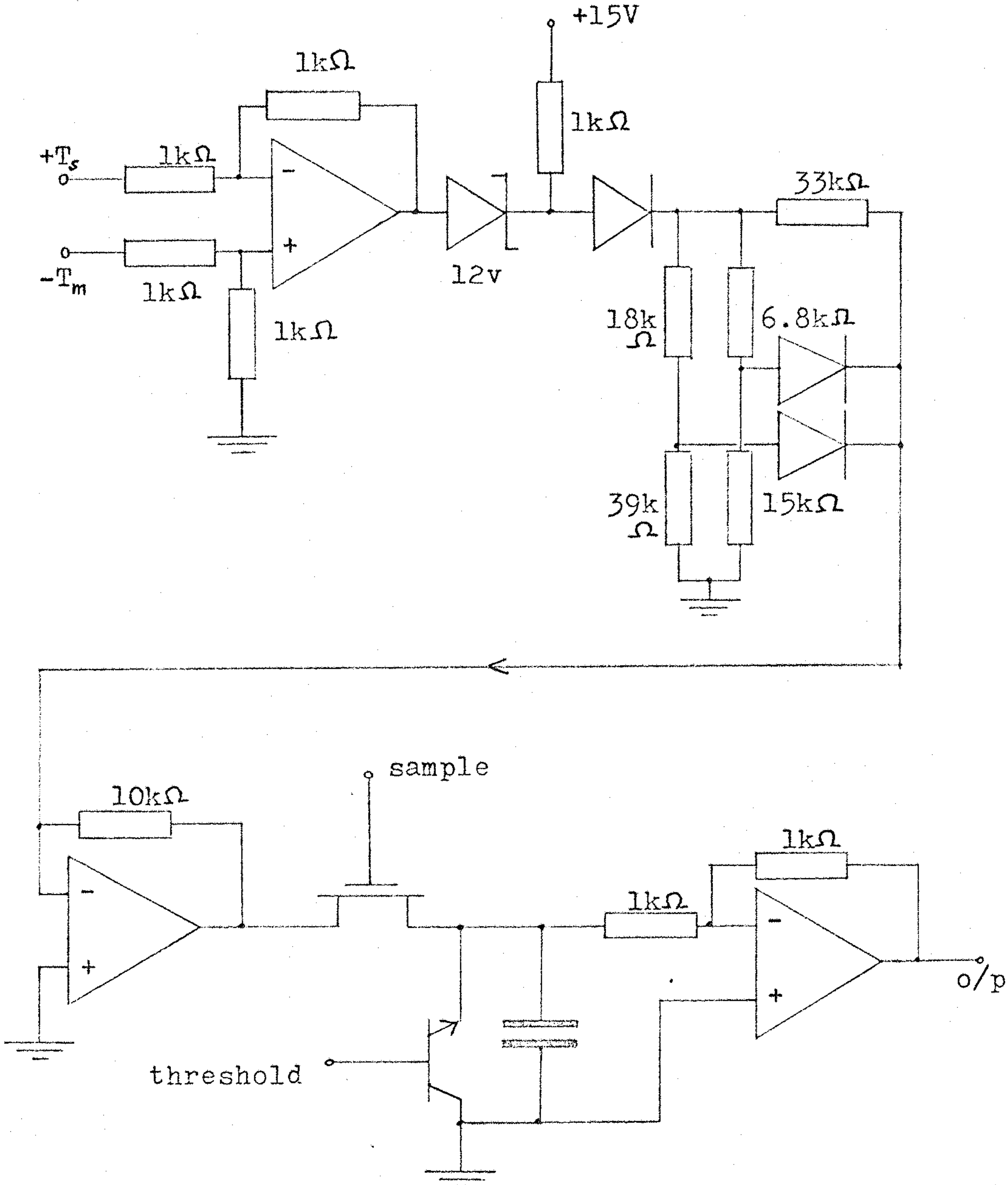


fig. 30 Circuit used to calculate frequency f_o
(all input signals come from the circuit shown in fig. 26)

Both halves of the period are taken from the mark/space calculating circuit and added together so giving the total period. This then passes through a non-linear network and is sampled over the appropriate period (the same sampling period as in the mark/space circuit). The output from this is a measure of f_o, which is plotted in fig 31 and is almost linear over the range 100 - 500 Hz. The transient time for this circuit is identical to that of the mark/space calculating circuit.

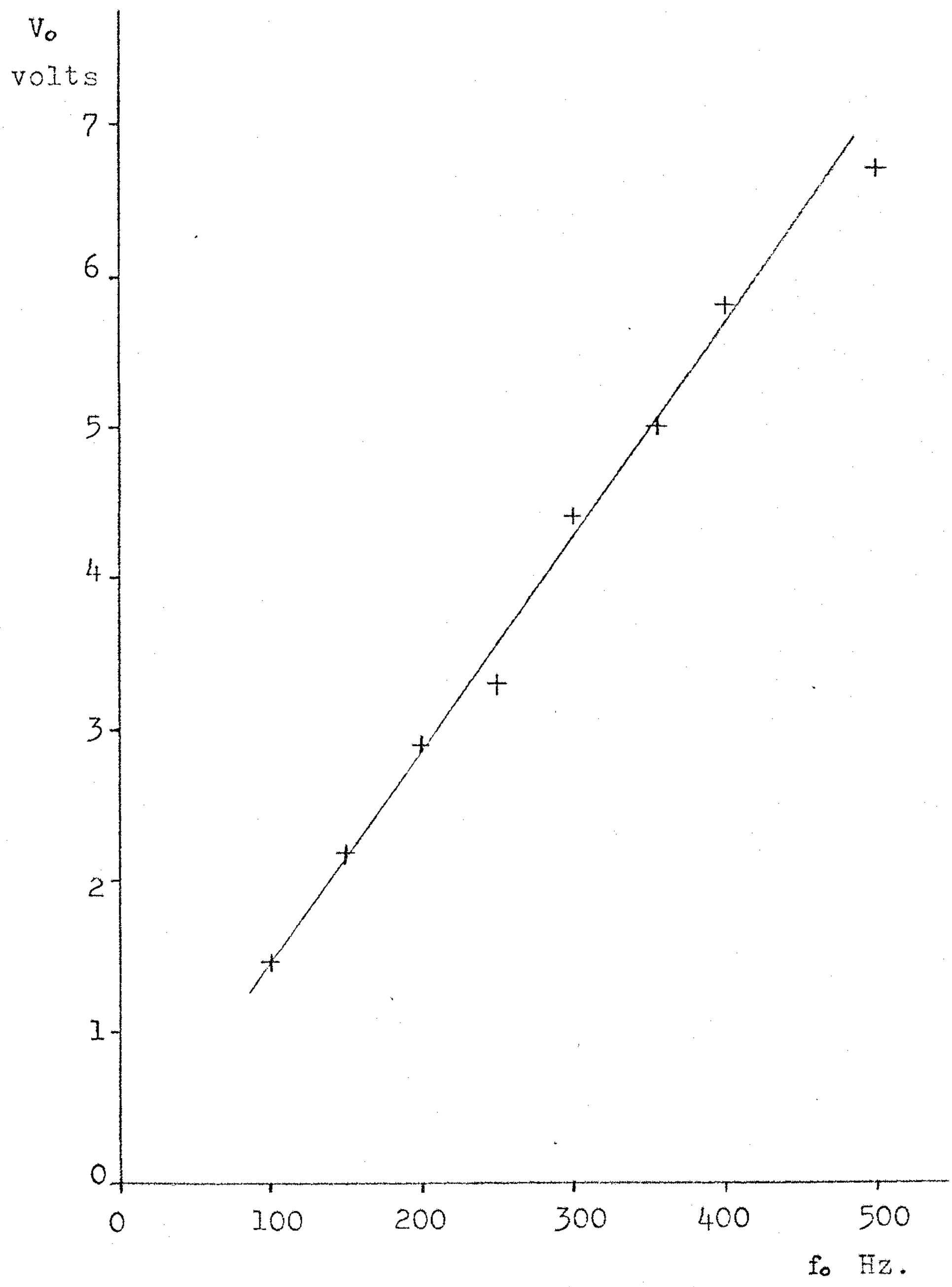


fig. 31 Linearity of f_o -determining circuit

6. Testing the effectiveness of the constructed system

6.1 Initial impressions

The output of the total system consisting of the Laryngograph, chopper and mark/space measuring circuit was displayed on a storage oscilloscope using a slow time base (between 1 and 5 seconds per cm.).

Several points were noted during the preliminary setting up and observation:

1. The output was sensitive to changes of vocal effort
2. The output was independent of upper tract positioning
3. The output was not very sensitive to the setting of the thresholds in the chopper circuit
4. The output was erratic at low vocal effort
5. The output was sensitive to the position of the electrodes.

The first of these is obviously of most interest as it represents the desired response. The second and third points are desirable characteristics of the system , while the last two imply inadequacies in the method of measurement.

These features are discussed below.

6.2 Testing the output for dependence on sub-glottal pressure

6.2.1 Discussion of methods

In order to perform this test one must have another measure of P_s to compare the output with. The features of speech that might cause a variation in output that is unrelated to P_s are:

1. Pitch
2. Vocal tract shape
3. Vocal register ie. chest, mid or falsetto.

1. and 3. are not identical as the same pitch can be produced in more than one register, see Van den Berg J.W.(1956).

According to Ladefoged P. (1967) P_s correlates well with a subjective loudness judgement. He found that the loudness judgement was proportional to the 1.9th power of the sub-glottal pressure (using an arbitrary scale of loudness). Therefore considerable experimentation was performed to determine whether the output of the device was dependent only on the subjective loudness. If it was then the output would be shown to be dependent only on the sub-glottal pressure. The results of the experiments were promising but there did seem to be some dependence on the pitch of the vocal input. The pitch dependence was not quantified as the experimental method was too approximate to allow this.

A more accurate method for obtaining an alternative measure of P_s for comparison was sought. The methods of measurement already discussed in section 4.3 were again considered. While the drawbacks of each method remain unchanged the previous condition that the measurement should be made as non-intrusively as possible can be relaxed. The subject may accept for a short experimental period a technique that would be unsuitable for a teaching aid.

It was noted in section 6.1 above that subjectively the output of the device constructed is independent of the positioning of the articulators in the upper tract. (This is to be expected as L_x itself varies only slightly with articulator position, as stated in section 5.1) Thus it quite valid to test the device under conditions such that the tract position is held constant.

With this condition the intensity of the acoustic speech output may be assumed to be dependent only on P_S. (The justification for this assumption and a prediction of the degree of accuracy that can be expected are to be found in Appendix III)

An experiment was set up to plot the calculated mark/space against pitch at several constant sound intensity levels. Fig 32 shows the equipment which was used.

The frequency response of the level measuring equipment - Truvox microphone amplifier and Avo model 8mkIII on 2.5 V RMS - was measured by plotting a calibration curve, using borrowed 'B & K' sound level measuring equipment,(this had a flat response ($\pm \frac{1}{2}$ dB) over the entire speech frequency range) and a sine wave sound source. The response curve could only be plotted up to 2 kHz due to the frequency response of the sound source, however the effect of the response above 2kHz will not be great due to the small amplitudes of components of the speech wave above this frequency. (The speech spectrum decays at about 9 dB/octave above 500 Hz.) The calibration curve is shown in fig 33: there is a 3 dB/octave decay with a 3 dB attenuation at 125 Hz. The effect of this will be to attenuate the fundamental frequency by about 3 dB but it will have no effect on the harmonics. The fundamental is generally at least 10 dB below the level of the harmonics around the first formant frequency and so this lower cut-off in the response will have little effect on the results.

In order to check this some readings were taken using the 'B & K' equipment instead of the Truvox; the difference in the results was negligible.

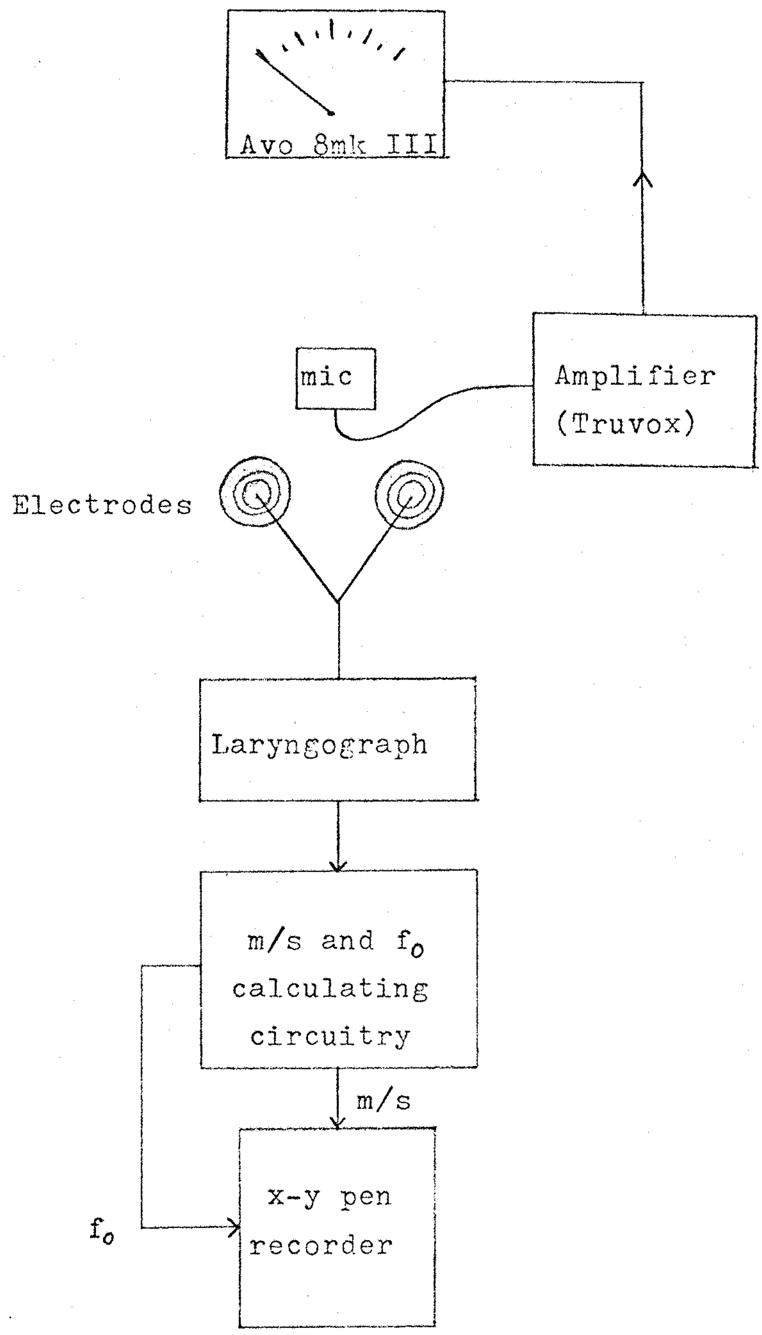


fig. 32 Experimental set-up

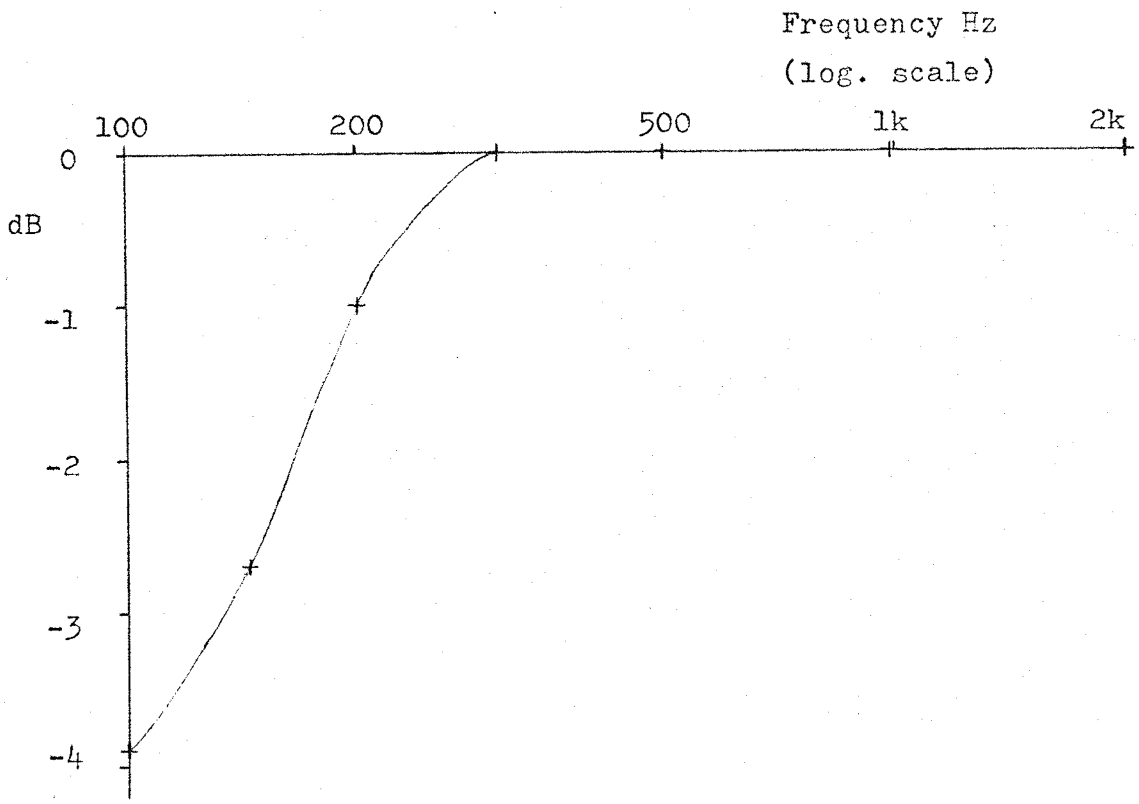


fig. 33 Frequency response of Truvox (measured using B&K equipment which is accurate to within ½dB)

At this stage also the frequency response of the mark/space measuring circuitry was tested using a simulated Laryngograph waveform. Relative to the output at 100 Hz the error at 200 Hz was 5% and at 500 Hz it was 20%. This shows a larger error than was found when testing the circuit with sine and square waves, however it was not considered excessive at this stage. The test circuit and test results are shown in fig 34.

6.2.2 Experimental procedure

The experimental procedure was as follows: the subject sat in front of the microphone with the sound intensity meter at eye-level, and was shown how to hold the electrodes and obtain an output from the Laryngograph. His mouth was positioned over a fixed wire frame that maintains the jaw and lips at constant opening and distance from the microphone. The subject was then asked to practice producing continuous |a| (as in bar) sounds at constant intensities - displayed by the level meter. This practice was very important as the task was found by all of the subjects to be much harder than expected. When the subject became reasonably proficient at performing this task over as wide a range of pitch as possible the experiment itself began.

The subject was asked to produce three pitches at each of three intensity levels, the pitches being subjectively low, medium and high (the actual frequencies vary both from subject to subject - due to physiological inter-personal differences - and from low to high intensity). The intensities to be used were defined as steps of 5 dB. Subjects used either -5, 0, +5 dB or 0, +5, +10 dB depending on their ability. Each subject performed

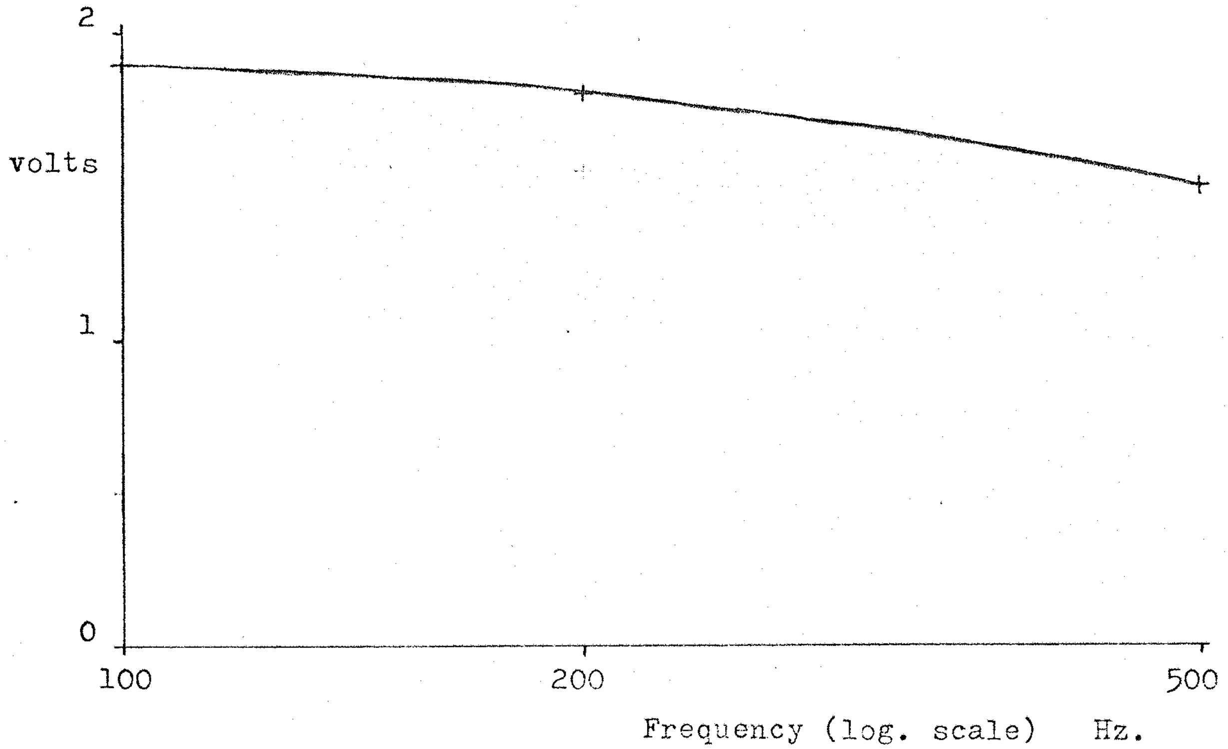


fig. 34(a) frequency response of the mark/space measuring circuit, using simulated Laryngograph waveforms.

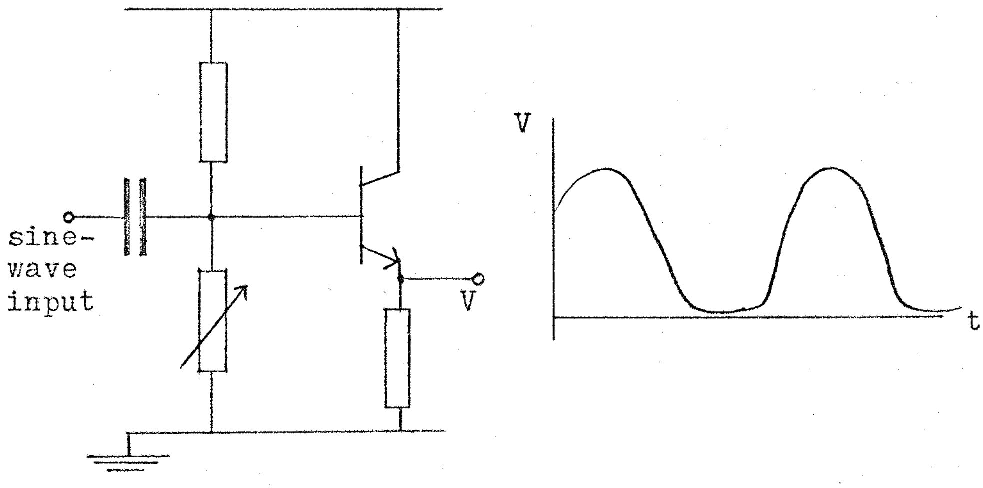


fig. 34(b) circuit used to generate simulated Lx waveform used above the sequence three times.

The tests were planned in two stages : the first group of subjects to be normal adult males, to determine the degree of consistency obtainable within a group with fairly similar vocal apparatus; the second group to include females and children, to determine whether the device was applicable to all, with or without alteration.

6.2.3 Results of the first stage

In this test results were collected from four normal adult males, A, B, C, and D. It was found to be virtually impossible to get accurately repeatable results from each subject, and there was considerable inter-personal variation in the degree of consistency obtained. Two of the subjects (A and B) managed to produce sufficiently consistent results to show a definite trend in the pattern. The other two, however, produced rather more random patterns in which the trend so clear in A and B's results could barely be discerned.

Had the assumptions made earlier in this report been upheld the results would have been vertical lines as shown in fig 35, demonstrating that the mark/space is dependent on the sound intensity, I_x, and not the pitch.

The trend indicated by the results is towards sloping parallel lines as fig. 36, a set of results from subject B, shows. For some reason the slope on these graphs varies considerably between tests, even with the same subject. It has not been possible to determine the cause of this.

This and similar subsequent figures were plotted with f_o as the dependent variable because in terms of the use of the equipment it is the mark/space that will be the indepent variable.

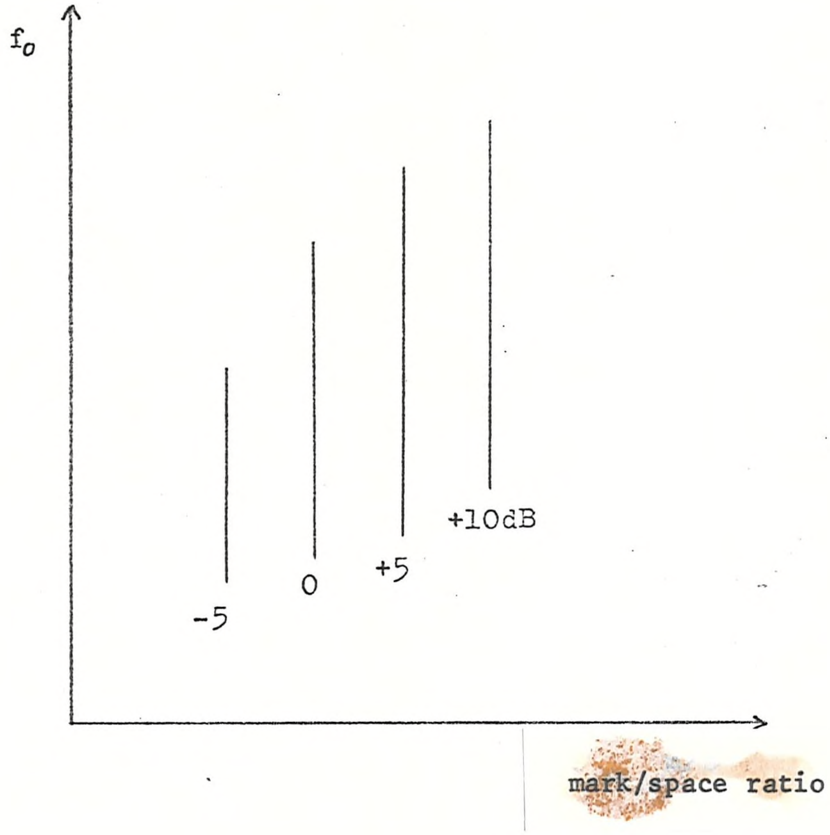


fig. 35 The predicted pattern of results

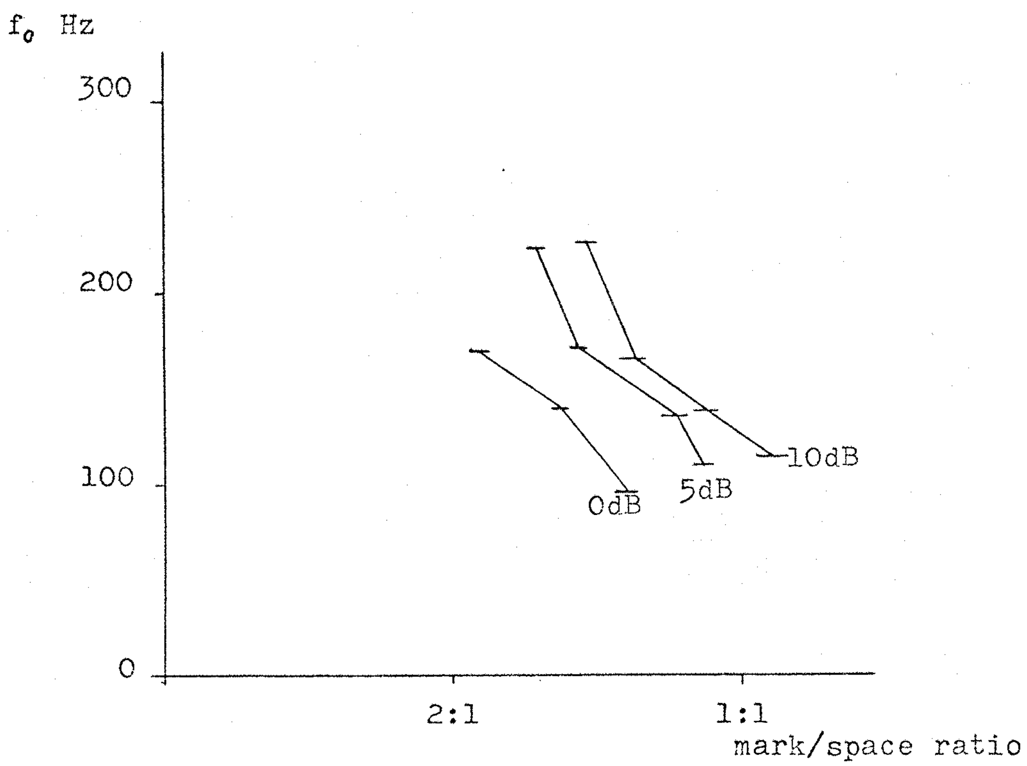


fig. 36 A set of results typical of those produced by 2 of the 4 subjects

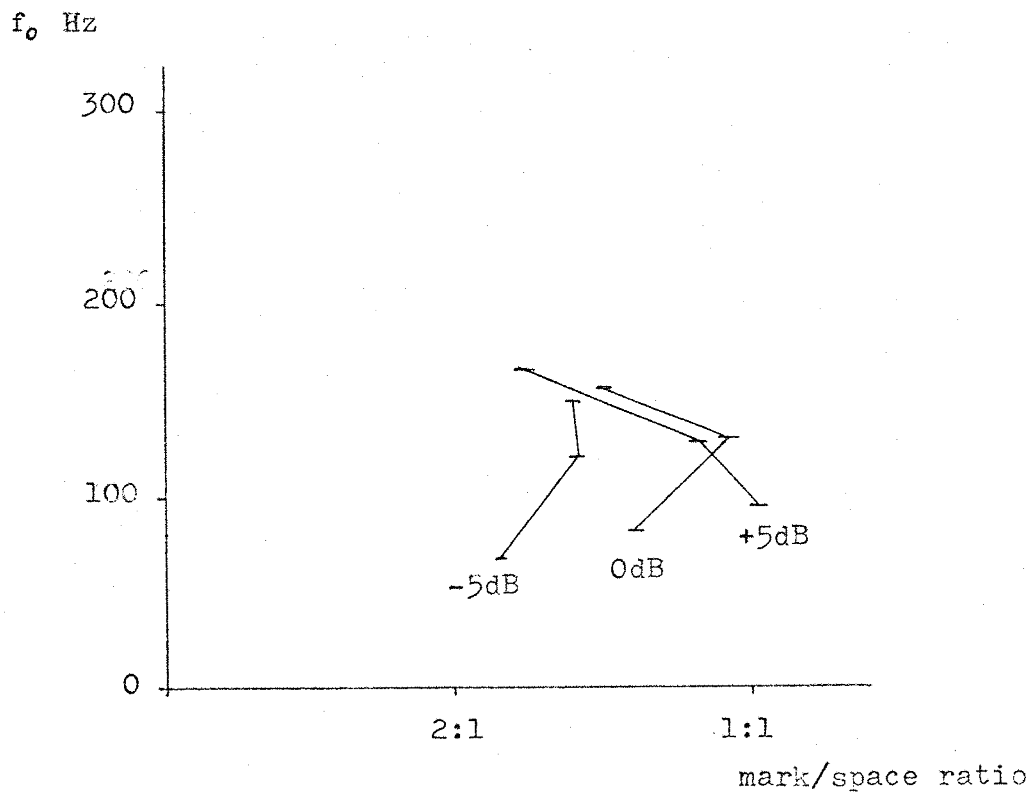


fig. 37 A set of results that shows the erratic results obtained from 2 of the subjects

Fig. 37 shows a set of results from one of the more erratic subjects. These appear significantly different at each repeat.

Due to the variability in the results it was decided to continue just with the four male subjects and to abandon the second stage of tests.

6.2.4 Search for the reasons behind the erratic results

Each of the four subjects noted that during a set of three readings taken at a constant sound intensity (ie. along a line on the graph of results) their subjective vocal effort reduced considerably with increasing pitch. This observation is in contradiction to the conclusion drawn from two relationships noted earlier. Firstly that the subjective voice effort is dependent only on P_s, and secondly that the sound intensity is also dependent only on P_s, both relationships having been obtained from the literature. Thus it seems that one or both of these results is questionable: either subjective voice effort or sound intensity or both is dependent on pitch as well as on P_s.

Subjective voice effort is rather difficult to define. Intuitively one feels that it must relate to the muscular tension forcing air out of the lungs, and so it should not be dependent on pitch. Thus one is drawn to the conclusion that the sound level intensity is dependent on pitch as well as on P_s.

Two more experiments were now performed with the same four subjects to retest the equipment with the above thoughts in mind.

Both tests involved asking the subject to say |a| at a constant subjective effort and to smoothly vary the pitch over their vocal range. Otherwise the experimental set up was the same as before (except that the sound intensity meter was disconnected so that the subject's vocal effort was controlled purely subjectively).

The first test was to determine whether intensity does increase with increasing pitch, at constant subjective effort. In order to do this it was necessary to be able to record loudness levels on the x-y plotter. The signal at the output of the Truvox amplifier was rectified and smoothed to provide an indication of mean amplitude.

The frequency response of the network used was quite flat over the whole speech spectrum, it was tested with both sine waves and speech using the Avo indication for comparison.

In this first test sound intensity was plotted against pitch for a constant subjective effort. In all cases the sound intensity increases drastically as the pitch increases. An example is shown in fig 38 (most of the plots show sharp decreases in sound intensity at one particular frequency which is arrowed in this example). The sudden decrease is thought to be the result of a sudden change from chest to mid voice.

In the second test the subject was asked to perform the same task but the recording equipment was returned to its original state - plotting f_0 vs. mark/space. Again vertical lines were hoped for, and indeed some were obtained, but there was still excessive variability and a great lack of repeatability in the results. A graph for each of the four subjects is shown in fig

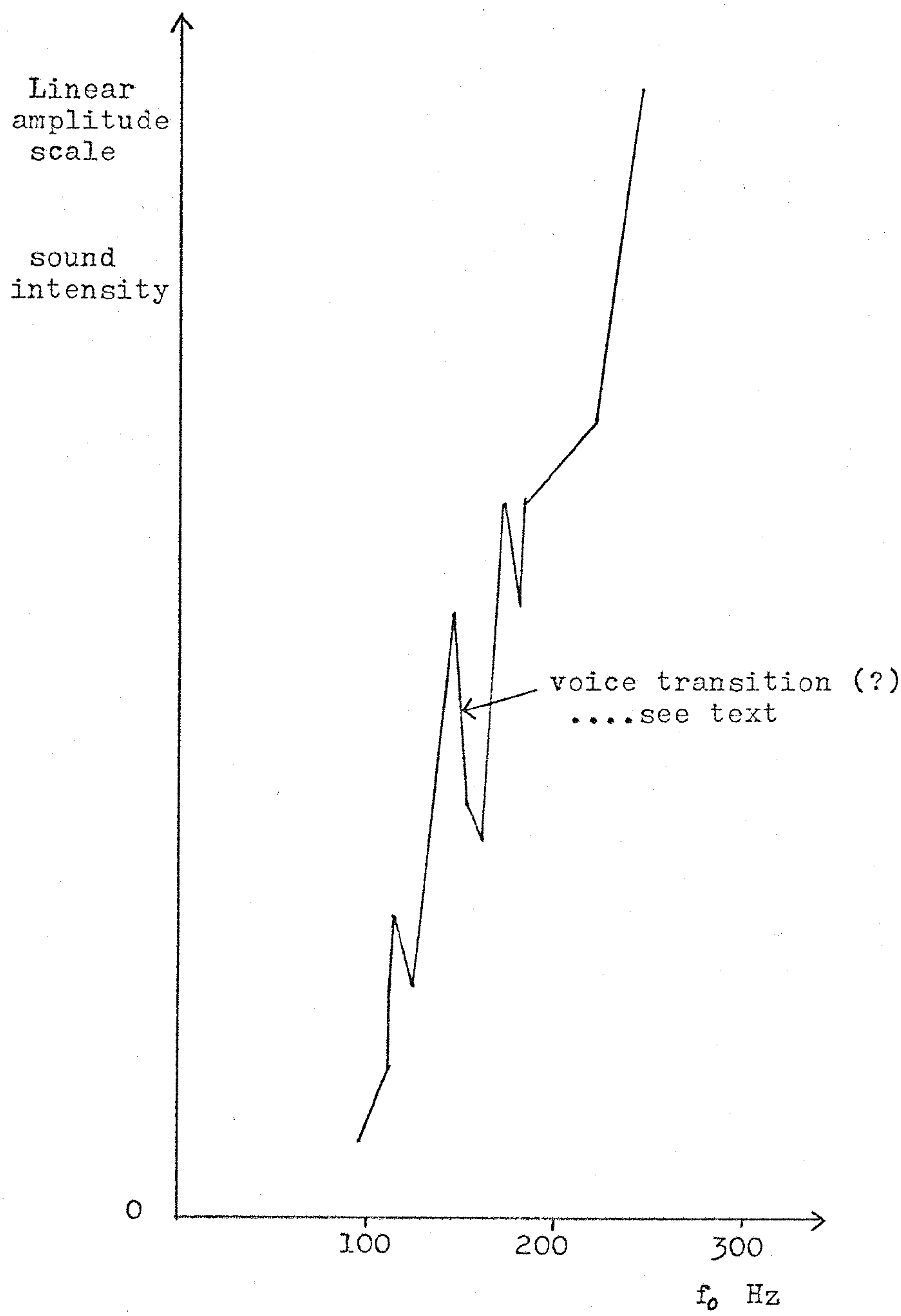


fig. 38 sound intensity vs. f_0 for constant subjective effort

39. These lines are closer to being vertical than were the results shown in figs 36 and 37, however the variability in them makes it impossible to state that the proposals made in the previous sections have been fulfilled.

It is possible that some of the variability between tests has been due to differing positioning of the electrodes. The Laryngograph waveform was displayed continuously during the experiments so that the subject could adjust the electrode positions to those giving the maximum signal before each set of readings was taken. It is possible that tests were performed with slight variations in that position and, although the signal appears subjectively not to change over a variety of electrode positions, this may have caused small variations in the signal's shape.

This does not explain the erratic nature of each set of results. This may be due to the method of using either sound intensity or subjective effort as a measure of sub-glottal pressure. Neither of these methods can be relied upon to any great accuracy. A future objective might be to use a more accurate measure of P_s for comparison. A hypodermic needle through the trachea is the only way to obtain a totally acceptable measure of P_s, and if further work is to be done on the proposed aid an experiment using this method is called for.

The fluctuations in the results may also be caused by actual variations in the vocal effort applied during each test. All subjects found considerable difficulty in maintaining an even vocalisation. The basic task of producing a long continuous note at a constant intensity is found very difficult, especially at the extremes of the subject's intensity and pitch ranges. If any

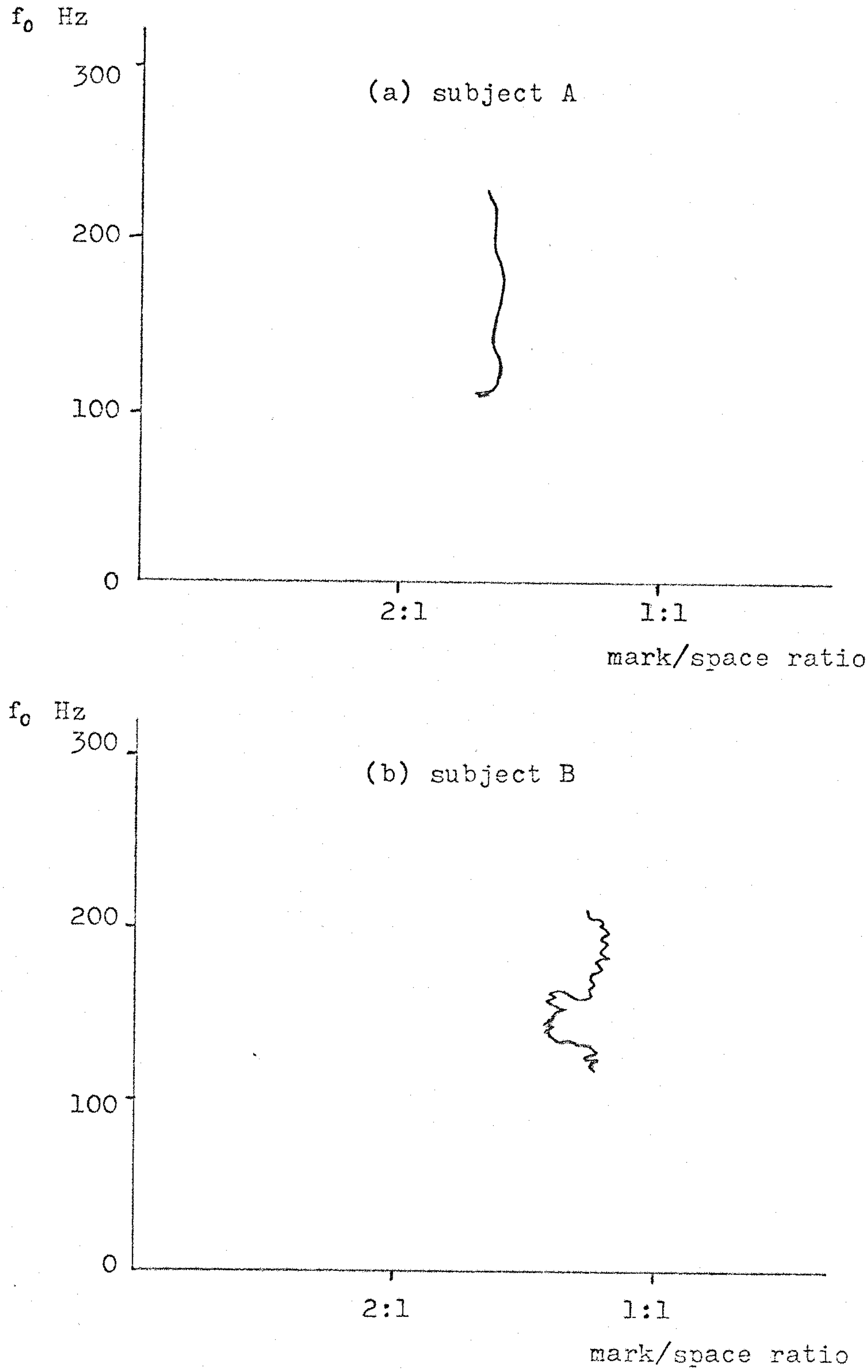


fig. 39 Dependence of mark/space ratio on f_0

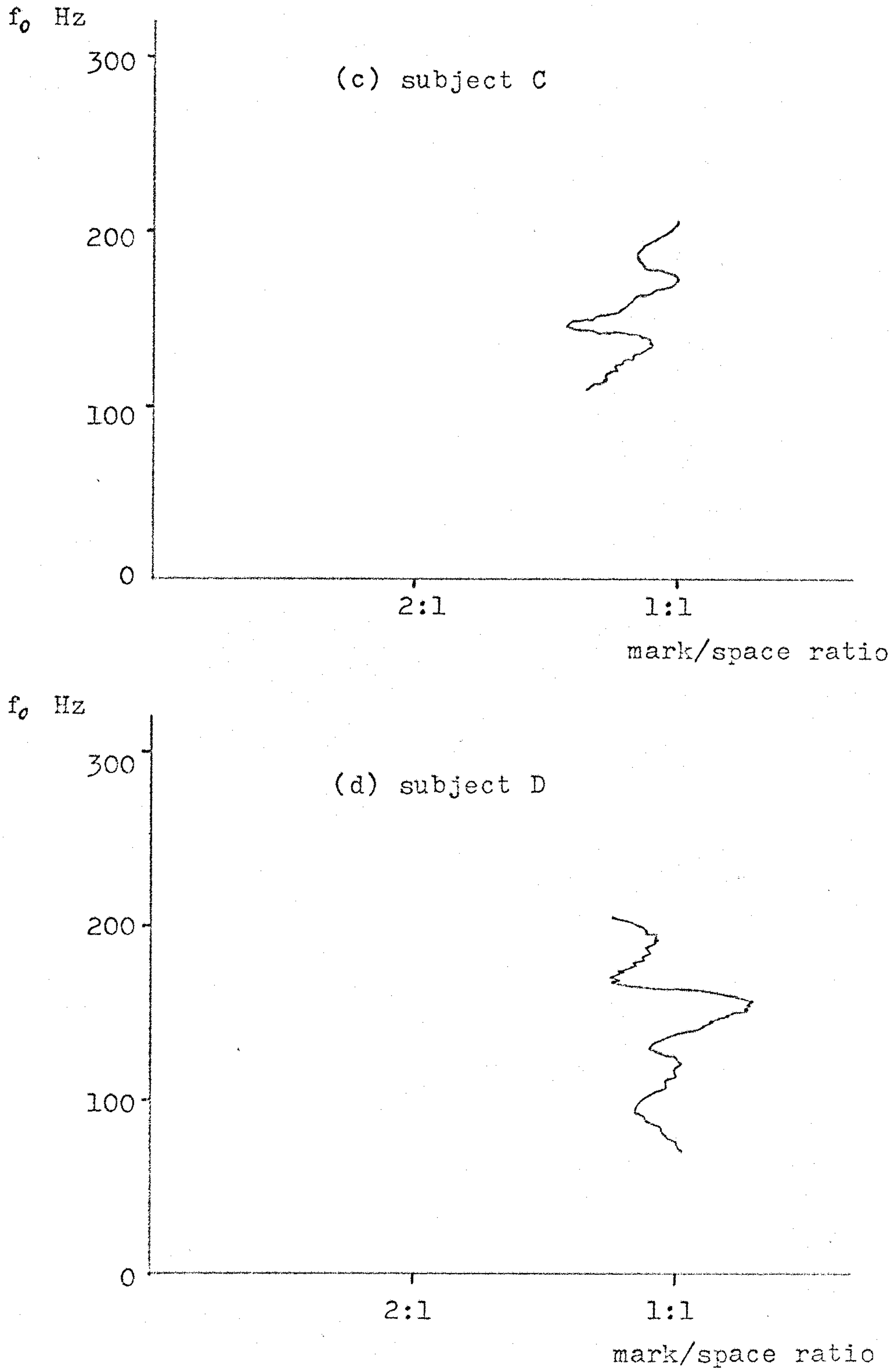


fig. 39 Dependence of mark/space ratio on f_0

experiments of this type are to be performed in future it would be interesting to include some trained singers in the subjects to see if they could produce a more controlled vocal effort.

7. Concluding discussion

In working towards the aim of producing a new type of articulation aid for the deaf the following objectives have been achieved:

1. A novel approach to the solution of the problem was found as a result of constructing simplified flow chart models of speech production. This solution involved the display of physiological parameters as feedback information to the subject.

2. A list of the basic parameters that were thought to be important was drawn up and the two most important were selected for initial study; these were the sub-glottal pressure and vocal cord tension.

3. Methods of measuring these parameters were studied and one possible method was chosen; this was thought to be the method most likely to produce usable results allowing for the condition that the apparatus would have to be acceptable to the user. This method involved the measurement of the mark to space ratio of the Laryngograph waveform.

4. Investigations were carried out to try and determine the relationship between the Laryngograph output, the glottal area waveform and an electrical resistance model of the neck as these could have given a considerable theoretical basis to the proposed method of measuring P_s and Q. However these investigations could not be drawn to a satisfactory conclusion.

5. Equipment was designed and built to implement the first stage of the proposed method, the aim being to obtain a signal that was dependent solely on P_s.

6. Methods of testing this equipment were proposed and

enacted, and the results recorded. No totally satisfactory explanation can be given for the erratic nature of the results obtained, though several possibilities are mentioned and some suggestions are made for other experiments to clarify the matter.

Further investigation in the field of fundamental speech research is obviously necessary as this work has highlighted the lack of knowledge concerning the vocal cord oscillating system. It is felt that the Laryngograph waveform in particular should be fully investigated as this device has the major advantage of providing a simple and easy to use method of observing the glottal vibration during free speech. If suitable research (either analytical or empirical) could produce useful data on the relationships between this waveform and parameters such as P_s, Q, volume velocity, effective vibrating mass, open glottal area, and area of mutual glottal contact then this device could become a very useful tool for research into the mechanisms of speech. In the present state of ignorance it can only be used as an easy method of displaying pitch.

More generally it is felt that mathematical models, such as that of Titze, could produce valuable information on the modes of vibration of the vocal cords and their dependence upon the various physiological parameters.

As a final comment on the work described in this thesis I must point out that the inconclusiveness of these findings in no way invalidates the assertion that in order to teach a deaf person to articulate in a particular manner one should display physiological parameters as feedback information.

APPENDIX I Anatomical Details

The following is an abbreviated description of the anatomy of the larynx, derived from the relevant pages of Ballantyne J. et al (1971), Lindqvist J. (1969) and Hollingshead W.H. (1968).

In the larynx there are three paired muscular structures that can be brought together to block air flow. In descending order these are

1. the aryepiglottic folds
2. the false vocal folds
3. the true vocal cords.

The aryiepiglottic folds are situated at the entrance to the larynx, and are approximately vertical. The major use of these is to prevent material such as food entering the lungs. This protective closure is performed by the aryepiglottic muscles in conjunction with the oblique arytenoid muscles and the thyro-epiglottic muscles.

The false vocal cords are positioned just above the true vocal cords, in the horizontal plane. They do not normally project into the air stream. They are used to prevent air flow out of the lungs when sneezing and coughing.

The true vocal cords consist of the vocalis muscle (pair) stretched between the cricoid and thyroid cartilages. Both of these cartilages can be moved by a combination of various muscles. Fig Ii is taken taken from Katika Y. (1974) and is a schematic diagram showing the basic mechanics of the glottal system, but it is a simplified model, missing out several muscle groups which might have some part to play in tensioning

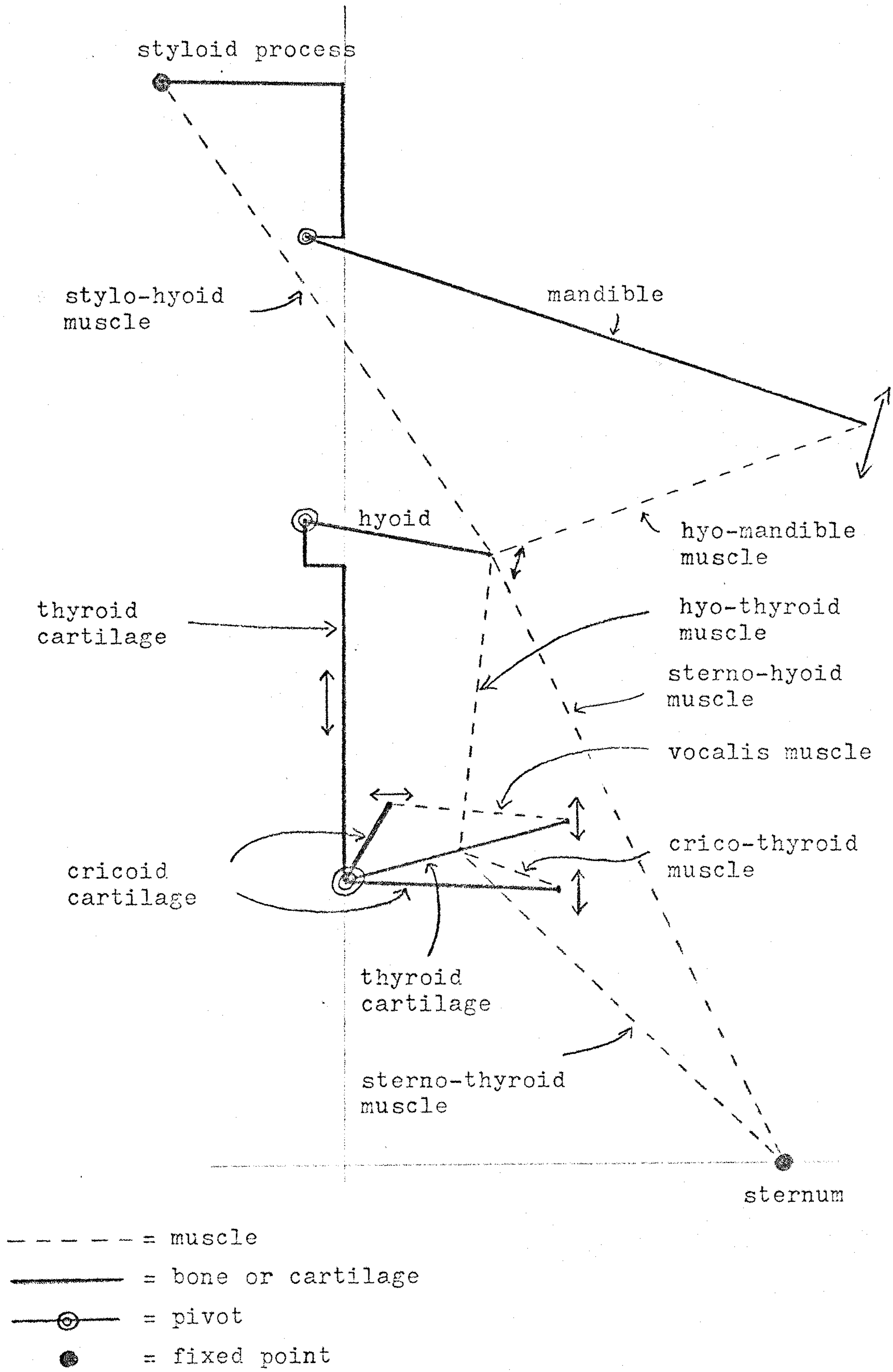


fig. Ii Schematic diagram showing the mechanics of the larynx the cords. It must be stated that it is still not clear exactly which combinations of muscle actions are used during speech, however it seems from the literature that the major contributor to the tensioning of the cords is the cricothyroid muscle, aided by the vocalis and thyro-arytenoid muscles.

There are two theories concerning the vibration of the cords; the neuromuscular and the aerodynamic theories.

The neuromuscular, clonic theory proposes that each cycle of vibration of the cords is due to a separate contraction in the muscle, ie. at a pitch of 100 Hz the glottal muscles are undergoing a contraction every 10 ms. However, electromyographic data has failed to confirm this.

The generally accepted aerodynamic or tonic theory proposes that the tension in the muscles is continuous, not pulsating, and that the vibration is actually caused by the passage of the air through the constriction of the cords.

A cross-section through the vocal cords shows them to be shaped approximately as in fig Iii, gently curving lower surfaces and flat upper surfaces. Thus as air pressure below them (the sub-glottal pressure) is increased they tend to peel apart. (If the shape was reversed then pressure below would tend to force them together.) Having parted, air flows between the cords and thus out of the lungs. This causes a drop in pressure below the cords and also a negative pressure between them due to the Bernoulli effect. These two effects combine to cause the cords to shut again, the sub-glottal pressure then rising to complete the cycle. High speed cinematography has shown that the vocal folds execute a movement in which the cord margins are rolled outwards and upwards. The lower parts of their opposed margins separate before the upper parts. Thus the cord movement has a vertical as well as a horizontal component.

The pitch-frequency of the vibration of the cords-depends on:

1. the tension of the vibrating matter. This is governed by the vocalis muscle and the stretching effect of the cricothyroid and others.

2. the length of the vibrating matter. This again is governed by the stretching effect of the cricothyroid muscle and others.

3. the cross-sectional shape of the vibrating matter (as in fig Iii). This changes from broad and thick to narrow and thin at high frequencies. The alteration of contour of the cord margins is a basic feature of the three different vocal registers (chest, mid and falsetto) and is performed by the thyro-arytenoids. In chest voice a large amount of the tissue around the cords vibrates and the lateral displacement is sufficient to fully shut off the air flow for a portion of the period. Mid voice involves a smaller amount of tissue vibrating a smaller distance, and falsetto is characterised by the air flow being never cut off and just the very edge of the cords vibrating with the cords being brought together to form a small oval window.

4. the sub-glottal pressure. The role of the relevant sets of muscles in the control of sub-glottal pressure is shown in fig Iiii (from Ladefoged P. (1967) p. 12).

In section 4.2 of this report I have made the assumption that the effects of 1, 2 and 3 can be combined into a general

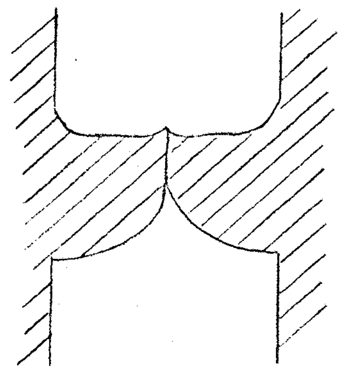


fig. Iii The vocal cords in a closed position

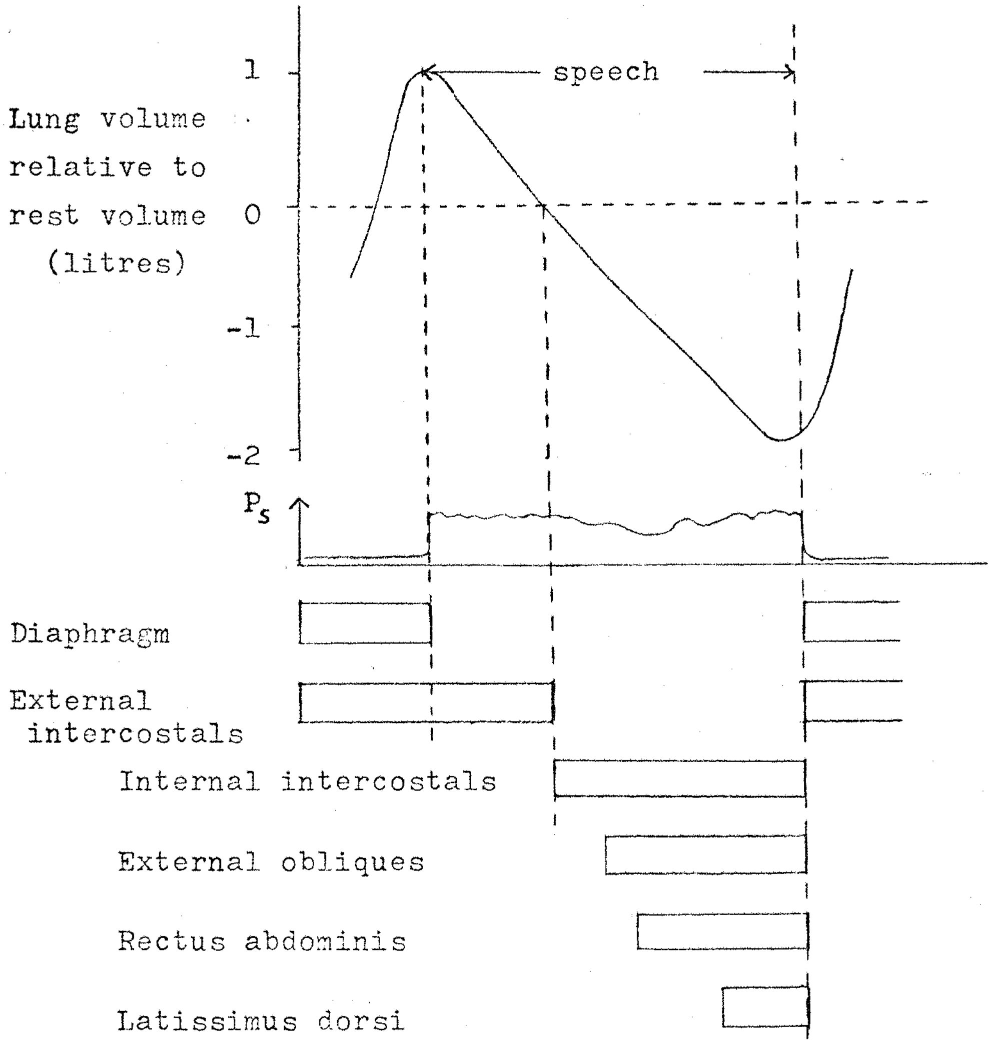


fig. Iiii Muscles used in breath control
during speech

parameter which I have called 'vocal cord tension' (this is also suggested in Ladefoged P. (1967) p. 49) that can be visualised as a descrition of the overall physiological stress in the larynx.

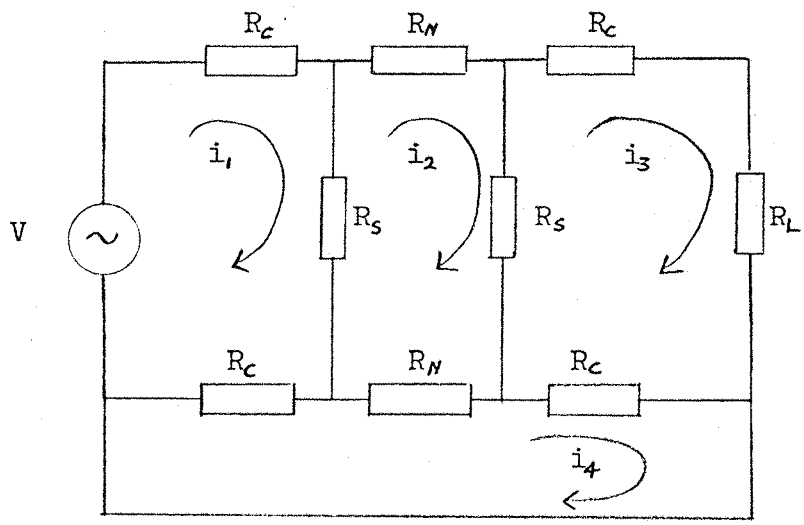
APPENDIX II The resistance model of the neck

As stated in section 5.2 an analysis of fig 13 was desired in order to determine the sensitivity of the Laryngograph to changes in contact, surface and neck resistances. In order to simplify the analysis, all contact resistances, surface resistances and neck resistances were assumed equal. This will inevitably reduce the usefulness of the result, but as can be seen from fig IIi the problem remains quite complicated.

It was desired to plot the relationships between i_3 and each of the resistances. In order that these results should be meaningful some realistic values for the resistances had to be found. The input resistance of the Laryngograph was calculated to be about 10Ω. There are in practice only two measurements that can be made to determine values for the others. One is a surface resistence measurement using one twin electrode, the other is a through resistance between the two electrodes. The measurement circuits are shown in fig IIii.

The measured value of resistance was found to vary considerably in both cases, depending on the force with which the electrodes were pressed against the skin. This introduces an interpretation problem: should all of this pressure-dependent variation be attributed to contact resistance variation, or does the bulk of the neck tissue vary in resistance when compressed? It was decided to consider both possible extremes, ie. first taking the variation as being due to contact resistance, and then taking the whole variation as a bulk resistance change.

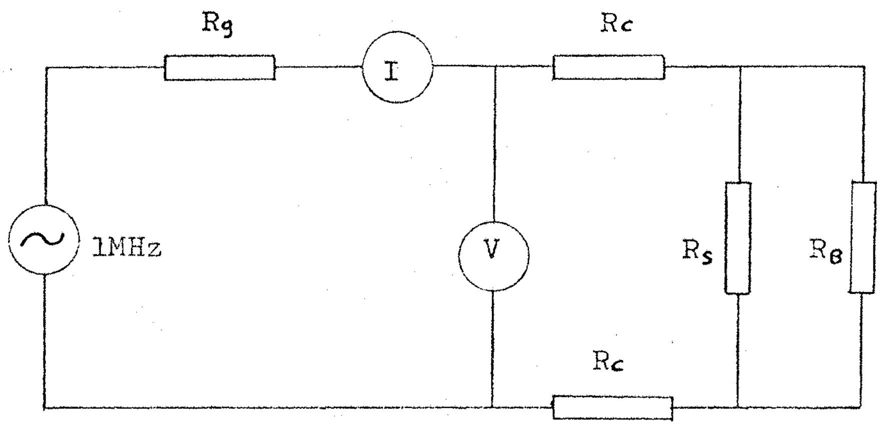
For the resistance measurement the following results were



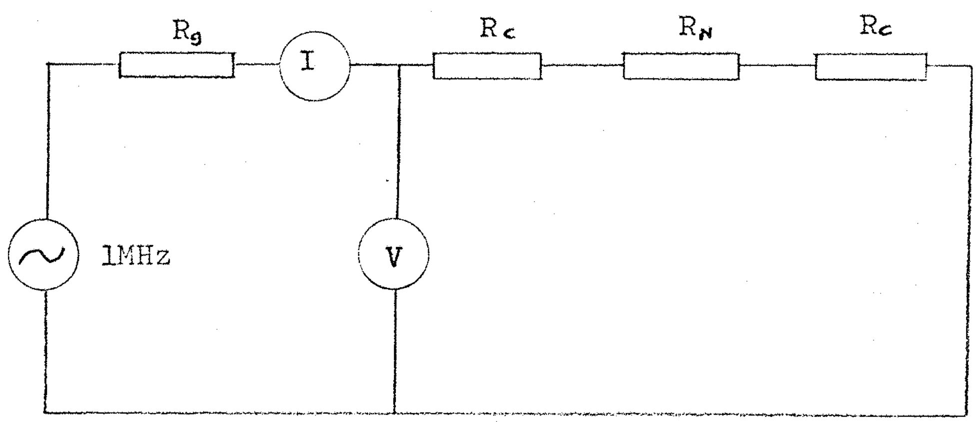
R_L = load presented by Laryngograph detector circuit

$$\begin{bmatrix} V \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} (2R_C + R_S) & -R_S & 0 & -R_C \\ -R_S & (2R_N + 2R_S) & -R_S & -R_N \\ 0 & -R_S & (2R_C + R_S + R_L) & -R_C \\ -R_C & -R_N & -R_C & (2R_C + R_N) \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix}$$

fig. IIi Resistance model of the neck and its matrix equation



(a) circuit for surface measurements



R_g = internal resistance of the generator

R_B = resistance of the bulk of the neck

$$\left(\sim 2R_N + \frac{R_S(2R_c + R_L)}{R_S + 2R_c + R_L}\right)$$

(b) circuit for through measurements

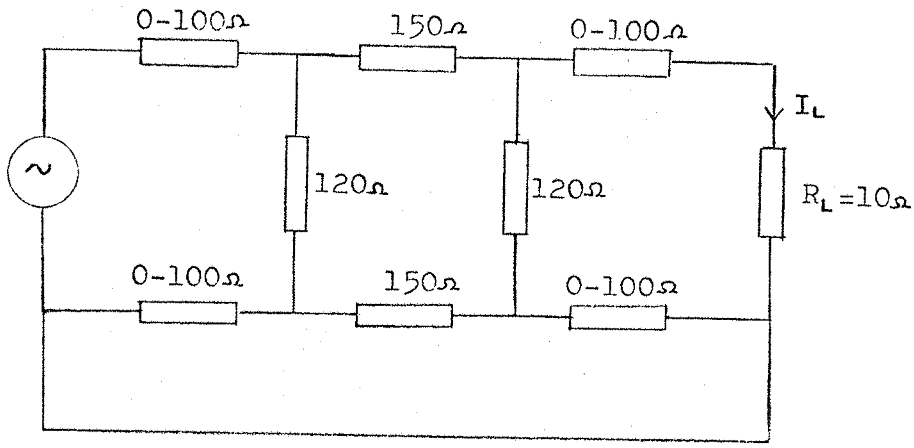
fig. IIii Neck resistance measurement, model circuits obtained:

	surface resistance	through resistance
low pressure	300 Ω	350 Ω
medium pressure	180 Ω	240 Ω
high pressure	120 Ω	160 Ω

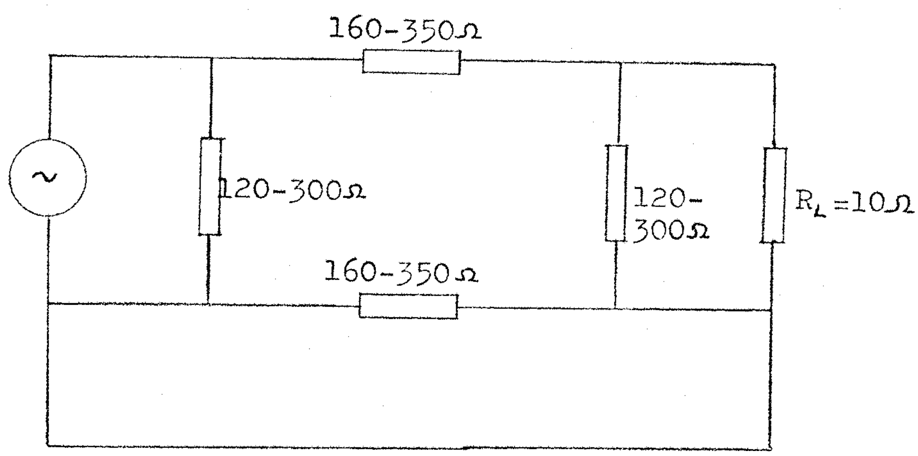
Two models were then constructed using the two principles outlined above. Both could account approximately for these results. They are shown in fig IIiii. The graph, fig IIiv, shows the degree of dependence on the change in the various resistances, taken in pairs for convenience, for the circuit shown in fig IIiii(a). As expected the resistors along the top of the circuit shown in fig 13 (R_{N1}, R_{C1}, R_{C3}) have much greater effect on the output than does the lower resistor chain. These results were not obtained by solving the matrix equation stated in fig IIi but by taking measurements from a real implementation of the circuit. It can be seen from fig IIiv that although the effect of varying R_N is the greatest, the effects of varying the contact and surface resistances are considerable.

The circuit shown in fig IIiii(b) is obviously much simpler and the output current will be almost entirely dependent on R_N.

The considerable difference between the results of these models shows their gross inadequacy. The concept of the neck as a network of only four resistances is thus an over-simplification Also, if this modelling were to be pursued, much more physiological data would have to be taken into consideration before a resistor network model could be found that would give more realistic results. It was considered that further research into this interesting red-herring was not compatible with the aims of this project.

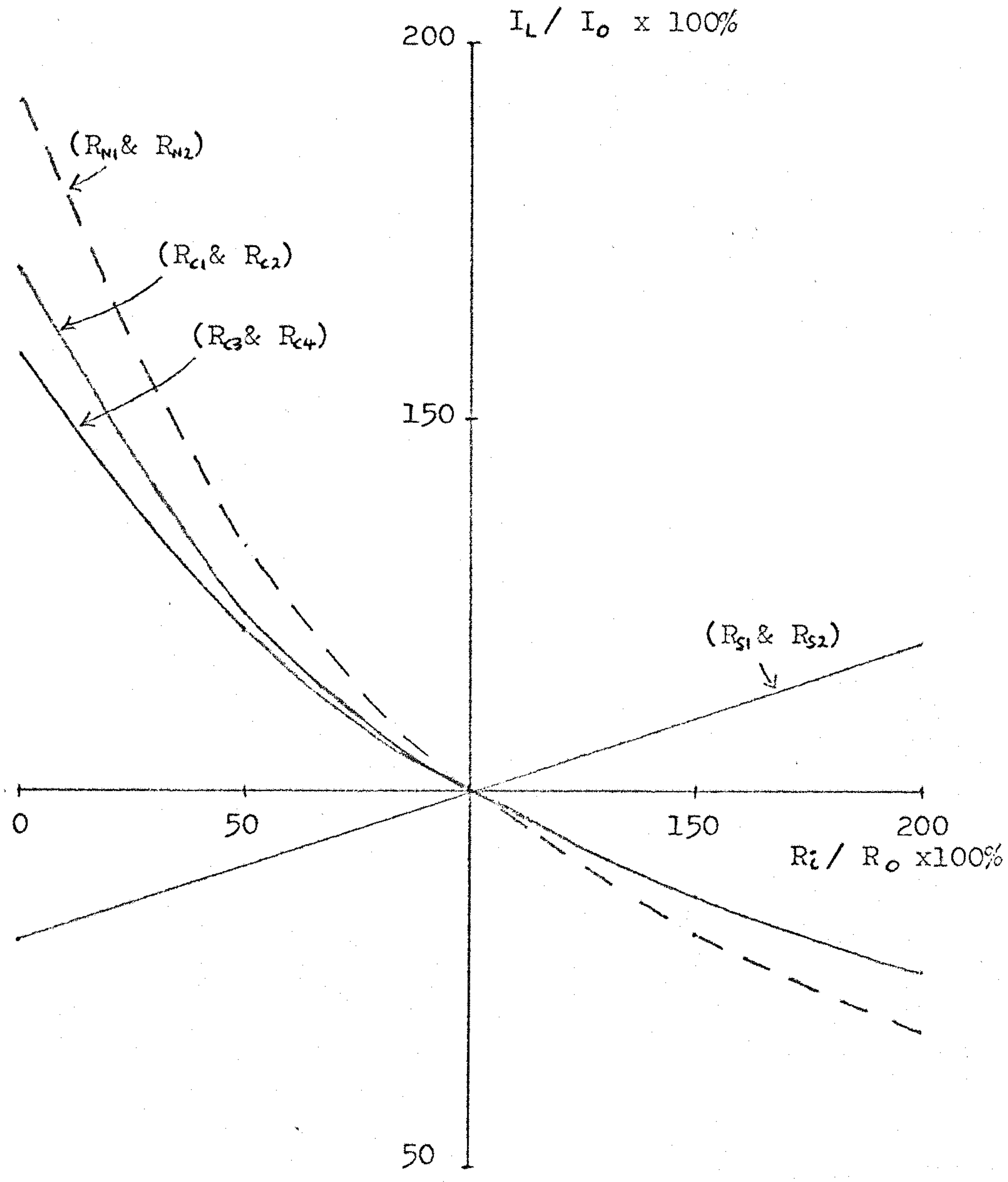


(a) Assuming that pressure-dependent variations are due only to changes in contact resistance



(b) Assuming that change is due only to variation of bulk tissue resistance

fig. IIiii Two resistance models of the neck



I_o = current through R (fig. IIi) when all R's are as in fig. IIiii(a).

R_o = R values as shown in fig. IIiii(a)

R_i = the value of(R_{N1} & R_{N2}); (R_{c1} & R_{c2}); (R_{c3} & R_{c4}) or (R_{S1} & R_{S2}), so giving the 4 curves.

fig. IIiv relationship between I_L and the various resistances in fig. IIiii(a)

APPENDIX III Justification of using sound intensity as a measure of P_s, with one stipulated procedural constraint.

This problem can be approached from two different directions,

1. theoretical propositions/models
2. experimental results obtained from the literature.

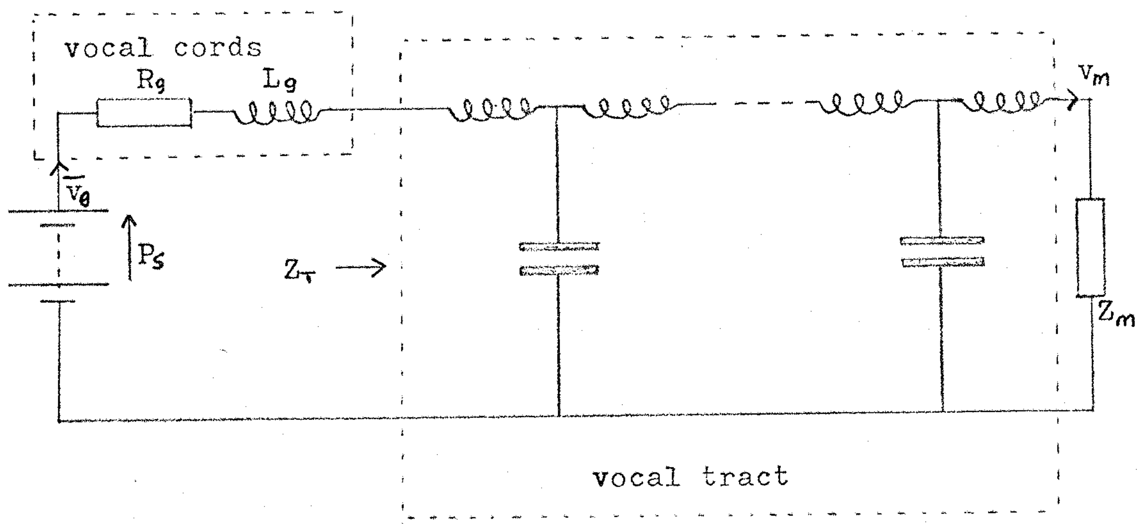
An electrical model of the vocal tract can be drawn, see fig IIIi, this is the model used in Flanagan J.L. (1968). The equations for R_g and L_g come also from the simulation in IEEE ASSP-23 (1975) with support from Van den Berg J.W. (1957) in which measurements were taken on excised human larynxes. (These measurements were taken when the larynxes were not vibrating, however Ishisaka and Flanagan (1972) p. 1239, suggest that these results hold for vibrating cords also, because of the high velocity of glottal flow - compared with the vocal cord velocity - and the small dimensions of the glottis compared with the wavelengths of the sounds of interest.)

If one states the condition that the articulators are stationary and assumes that all the L's and C's in the model can be considered as linear components, then $\bar{v}_m \propto \bar{v}_g$ and thus if the radiation resistance from the mouth, z_m, is constant then

$$\bar{L} \propto \bar{v}_g$$

Unfortunately an equation cannot be derived from this model to relate v_g to P_s, due to a surplus of unknowns. Experimental data must therefore be employed for this, Van den Berg J.W. (1956), and Ladefoged P. (1967) both obtain a linear relationship between v_g and P_s, and so one can predict that

$$\bar{L} \propto P_s .$$



$R_g = c_1 \overline{A}^{-3} + c_2 \overline{A}^{-2} \overline{v_g}$

$L = c_3 \overline{A}^{-1}$

$\overline{A}$ = mean glottal area

c_1, c_2, c_3 = constants

fig. IIIi Electrical analogue of vocal tract

The relationship between $\bar{L}$ and P_s is also plotted experimentally in these two references. The graphs are reproduced in figs IIIii and iii. Both of these indicate that $\bar{L} \propto P_s^n$ where n lies between 1.6 and 2. The reason for the difference between this result and the one predicted seems to be the assumption that $\bar{L} \propto \bar{v}_m$. Van den Berg plots this relationship also and produces the unexpected result that $\bar{L} \propto v_m$. He offers no explaination for this but it does make the predicted relationship agree with the measured one.

The error-deviation from the straight line prediction in Ladefoged's results is quite small. Most points are within 1 dB and 1 cm of water of the straight line. Van der Berg's results show a greater variability, but this is perhaps to be expected as his work was performed at an earlier date when the standard of instrumentation was probably lower.

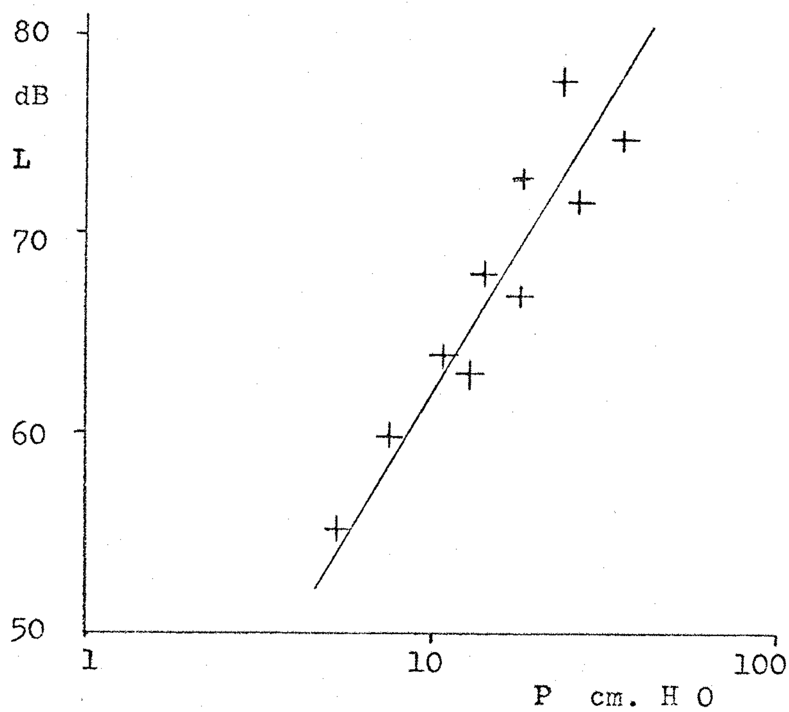


fig. IIIii Loudness vs. subglottal pressure at various pitches (from Flanagan 1968)

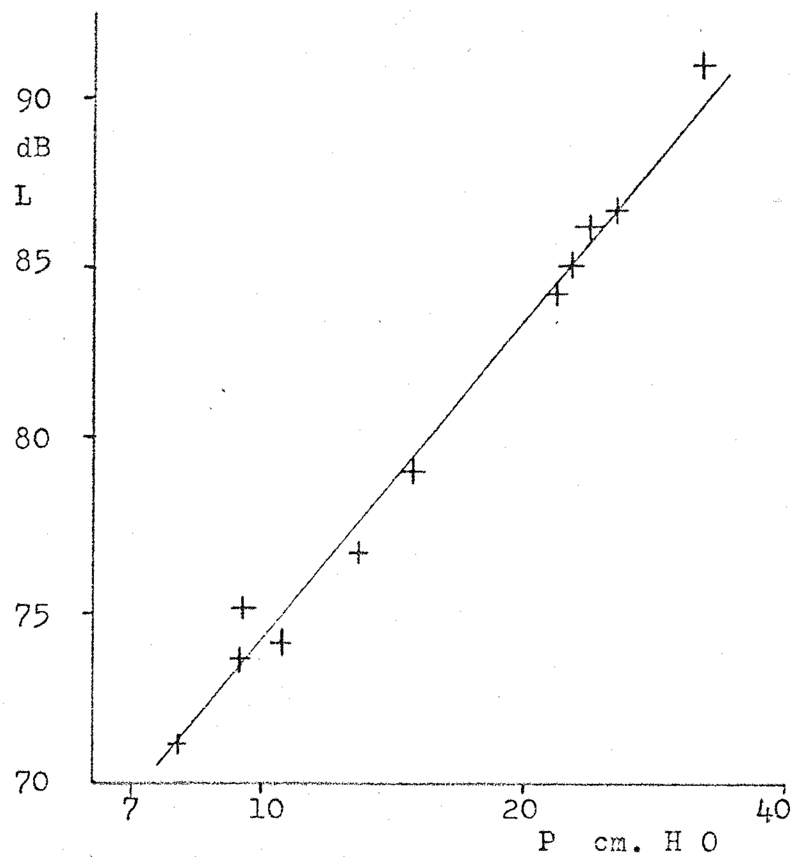


fig. IIIiii Loudness vs. subglottal pressure with no pitch constraint (from Van den Berg 1956)

References

Allinson R.D. et al. Volumetric dynamics of respiration as measured by electrical impedance plethysmography. J. App. Physiol. vol. 19, pp 166-173, 1964.

Anderson A.B. et al. Electrical stimulation of nerves in the skin at audio frequencies. J. Acou. Soc. Am. vol. 23, p155, 1951

Baker L.E. Biomedical applications of electrical impedance measurement. IEEE Medical Electronics Monograph no. 1, 1972.

Ballantyne J. and Groves J. eds. Diseases of the ear nose and throat, vol.1 Basic Sciences. Butterworths, London, 1971.

Biondi E. and Biondi L. The sampling of sounds as a new means of making speech intelligible to the profoundly deaf. Alta Frequenza vol.37, pp728-739, 1967.

Blake P. and Moss T. The developement of socialisation skills in an electively mute child. Behav. Res. and Therapy vol.5, pp349-356, 1967.

Boothroyd A. Some experiments on the control of voice in the profoundly deaf using a pitch extractor and a storage oscilloscope display. IEEE AU21 p274, 1973.

Borrild K. Experience with the design and use of technical aids for the training of deaf and hard of hearing children. Am. Annals of the Deaf vol.113, p168, 1968.

Cernak L.S. Human memory - research and theory. The Ronald Press Co., New York, 1972.

Crichton R.G. and Fallside F. Linear prediction model of speech production, with applications to deaf speech training. Proc. IEE vol.121, pp865-873, 1974.

Dobelle W. et al A prothesis for the deaf based on cortical stimulation. Ann. Otol. Rhinol. Laryngol. vol.82, p445, 1973.

Dolansky L. An intonation display system for the deaf. International Symposium on speech communication ability and profound deafness. Stockholm, Sweden. Paper 30, August 1970.

Erber N.P. Speech envelope cues as an acoustic aid for profoundly deaf children. J. Acou. Soc. Am. vol 51, p1224, 1972.

Fineman K.R. et al. Establishing and increasing verbalisations in a deaf schizophrenic child through the use of contingent visual color reinforcement. Percept. and Mot. Skills vol.29, p647, 1969.

Flanagan J.L. and Landgraf L.L. Self oscillating source for vocal tract synthesiser. IEEE AU16 pp57-64, 1968.

Fourcin A.J. and Abberton E. First applications of a new Laryngograph. Med. and Biol. illust. vol.21, pp172-182, 1969.

Furth H.G. Thinking without language. New York Free Press, 1966.

Georgiou V.J. et al. A multiparameter visual display. Proc. 1973 Carnahan conf. on electronic prosthetics, University of Kentucky p 24.

Gopinath B. and Sondhi M.M. Determination of the shape of the human vocal tract by acoustic measurements. J. Acou. Soc. Am. vol 41, p1002, 1967.

Guttman N. and Nelson J.R. An instrument that creates some artificial speech spectra for the severely hard of hearing. Am. Annals of the Deaf vol.113, pp295-302, 1968.

Hamlet S.L. and Reid J.M. Transmission of the ultrasound through the larynx as a means of determining vocal fold activity. IEEE BME19 p34, 1972.

Hixon T.J. Kinematics of chest wall during speech production: volume displacements of rib-cage, abdomen and lung. J. Speech and Hearing Res. vol.16, pp78-115, 1973.

Hollingshead W.H. Anatomy for surgeons. vol.1 The head and neck. Harper and Row, New York. pp465-486, 1968.

Hyde S.R. Automatic speech recognition: literature survey and discussion. GPO research report no. 45, 1968, PO Research Dept, London.

IEEE ASSP-23 no.1 Special issue on IEEE Symposium on speech recognition, 1975.

Ishizaka and Flanagan J.L. Synthesis of voiced sounds from a two mass model. Bell Syst. Tech. J. vol.51, p1233, 1972.

John J.E.J. Teaching children to talk. Manchester University Press, 1964. Chapter IV, pp82-105.

Kakita Y. A study of laryngeal control for voice pitch based on an anatomical model. SCS Stockholm, Aug. '74 vol.2, pp45-54.

Kalmus H.P. Electronic flowmeter system. Rev. Sci. Inst. vol.25, p201, 1954.

Kirman J.H. Tactile communication of speech - a review and analysis. Pschol. Bull. vol.80, no.1, pp54-74, 1973.

Kung pu li et al. Segment classification in continuous speech. IEEE AU21, p50, 1973.

Kunze L.H. Evaluation of methods of establishing subglottal air pressure. J. Speech and Hearing Res. vol.7, p151, 1964.

Ladefoged P. Three areas of experimental phonetics. Oxford Uni. Press, 1967.

Levitt H. et al. Acoustic, articulation and perceptive characteristics of the speech of deaf children. SCS Stockholm Aug. 1974.

Lieberman P. Direct comparison of subglottal and oesophagal pressure during speech. J. Acou. Soc. Am. vol.43, p1157, 1968.

Lieberman P. Determination of rate of change of fundamental frequency with respect to subglottal air pressure. J. Acou. Soc. Am. vol.45, p1537, 1969.

Lindqvist J. Laryngeal mechanisms in speech. STL Stockholm, Q.P.S.R. no.2-3, pp26-32, 1969.

Lindqvist J. The voice source studied by means of inverse filtering. STL Stockholm, Q.P.S.R. no.1, p3, 1970.

Ling D. and Druz W.S. Transposition of high frequency sounds by partial vocoding of the speech spectrum. Its use by deaf children. J. Audio. Res. vol.7, pp133-144, 1967.

Martony J. On the correction of the voice pitch level for severely hard of hearing subjects. Am. Annals of the Deaf vol.113, pp195-202, 1968.

Mason J.L. et al. Cutaneous stimulation for sound localisation aids. Proc. Carnahan conference on electronic prosthetics, Uni. of Kentucky, 1973, pp1-4.

Mermelstein P. Determination of vocal tract shape from measured formant frequencies. J. Acou. Soc. Am. vol.41, p1283, 1967.

Miller J.D. et al. Preliminary research with a three channel vibrotactile speech reception aid for the deaf. SCS Stockholm, Aug. 1974, vol.4, p97.

Miller R.L. Nature of the vocal cord wave. J. Acou. Soc. Am. vol.31, p667, 1959.

Nickerson R.S. et al. Teaching speech to the deaf - can a computer help. IEEE AU21 p445, 1973.

Obenour J.L. et al. A digital computer simulation of the human speech production mechanism. IEEE BME19, p114, 1972.

Pacela A.F. Impedance pneumography - a survey of instrumentation techniques. Med. Biol. Eng. vol.4, no.1, 1966.

Person S.C. A computer based visual aid for the deaf. Proc. Carnahan conference on electronic prosthetics Uni. of Kentucky, 1973, p15.

Pickett J.M. Some applications of speech analysis to communication aids for the deaf. IEEE AU17 p283, 1969.

Richardson A. Mental Imagery. Routledge and Kegan Paul, London, 1969.

Risberg A. et al. Teaching machine for training experiments in speech perception. STL Stockholm Q.P.S.R. no.2-3, pp72-75, 1967.

Risberg A. Visual aids for speech correction. Am. Annals of the Deaf vol.113, pp178-194, 1968.

Rothenburg M. Inverse filtering to obtain the volume velocity at the glottis. J. Acou. Soc. Am. vol.53, p1632, 1973.

Sokolov A.N. Inner speech and thought. Plenum Press, New York, 1972.

Sondhi M.M. Measyrement of the glottal waveform. J. Acou. Soc. Am. vol.57, pp228-232, 1975.

Stansfield E.V. and Boyner R.E. Determination of the vocal tract area function from the transfer impedance. Proc. IEE vol.120, p153, 1973.

Stark R.E. et al. Preliminary work with the new Bell Telephones V.S.T. Am. Annals of the Deaf vol.113, pp205-214, 1968.

Stark R.E. Teaching of features of speech to profoundly deaf children by means of real-time visual display. Paper 34, Internat. Symp. of Speech Communication ability and profound deafness. Stockholm Aug. 1970.

Terwilliger R.F. Meaning and mind. Oxford Uni. Press, New York, 1968.

Titze I.R. The human vocal cords: a mathematical model. Phonetica vol.28, pp 129-170 and vol.29, pp1-21, 1973-4.

Van den Berg J.W. Direct and indirect determination of the mean subglottic pressure. Folia Phoniat vol.8, pp1-24, 1956.

Van den Berg J.W. et al. On the air resistance and Bernoulli effects of the human larynx. J. Acou. Soc. Am. vol.29, pp626-631, 1957.

Wakita H. Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms. IEEE AU21 p417, 1973.

Willemain T.R. Tactile pitch displays for the deaf. IEEE AU20 p9, 1972.