

ANATOMICAL AND SPECTROGRAPHIC ANALYSIS OF THE VOICE IN
DISEASE: A REPORT OF FIVE CASES

W.A. KERR, M.R.C.S., D.L.O.,

*Department of Otorhinolaryngology, Johannesburg
Hospital, (Head: D.R. Haynes)*

and

L.W. LANHAM, PH.D.

*Head, Department of Phonetics & General Linguistics,
University of the Witwatersrand, Johannesburg*

SUMMARY

Five cases are presented. One is a case of ventricular phonation of iatrogenic origin and the remaining four had undergone laryngectomy for carcinoma of the larynx. Points of interest are discussed, particularly the constant ventricular fold phonation of the first case and the clear harmonic structure present in the voice of one of the laryngectomy cases who has both esophageal speech and pharyngeal phonation.

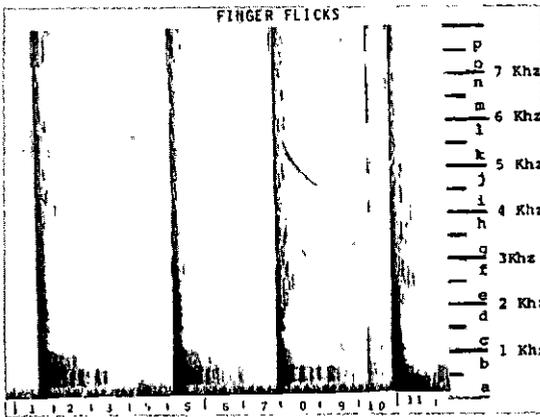
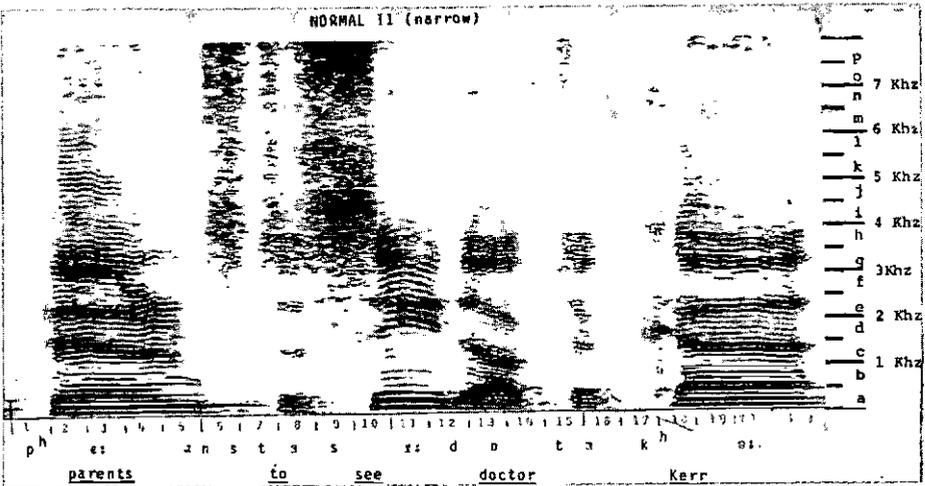
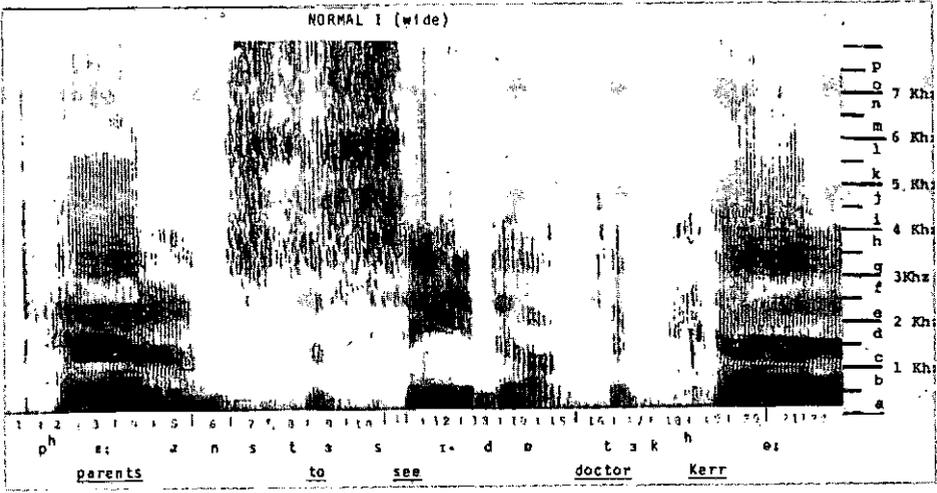
OPSOMMING

Vyf gevalle word voorgelê. Die een is 'n geval van ventrikulêre fonasie van iatrogeniese oorsprong, terwyl die oorblywende vier gevalle laringektomie ondergaan het vir karsinoma van die larinks. Aspekte van belang word bespreek, veral die teenwoordigheid van harmoniese struktuur in die stem van die eerste geval met ventrikulêre fonasie, asook in die stem van een van die laringektomie-gevalle wat beide esofageale en faringeale fonasie gebruik.

The authors do know that much has been written on this subject and that a comprehensive reference to the many authors would be impossible in this article. But they have recorded some publications in the list of references and acknowledge the work of such authorities as: Negus, C. and C.L. Jackson, Kallen, Bateman, Kirchner, Arnold, Huizinga, van den Berg, Moore, Norris, Conley, Tapia, Tato, Moolenaar Bijl and the Bell Telephone Company.

This paper is comprised of two sections -- first (Part I), the discussion of the spectrographic evidence, second (Part II), the anatomical analysis and case histories.

Editor's note: We gratefully acknowledge a grant-in-aid for the publication of this article from The National Cancer Association of South Africa.



It is hoped that the association of the disciplines of Acoustic Phonetics and Otorhinolaryngology, as well as the method of presentation, may be of interest in South Africa; and, if so, this report will be a fitting tribute to Professor P. de V. Pienaar who has worked for many years in this country in Speech Therapy on the foundations of Phonetics.

PART I

THE PHYSICAL ANALYSIS OF THE DIFFERENT "VOICES".

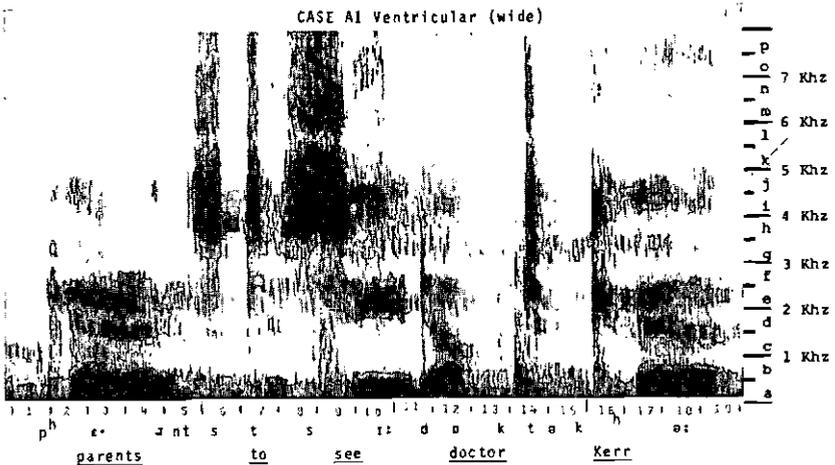
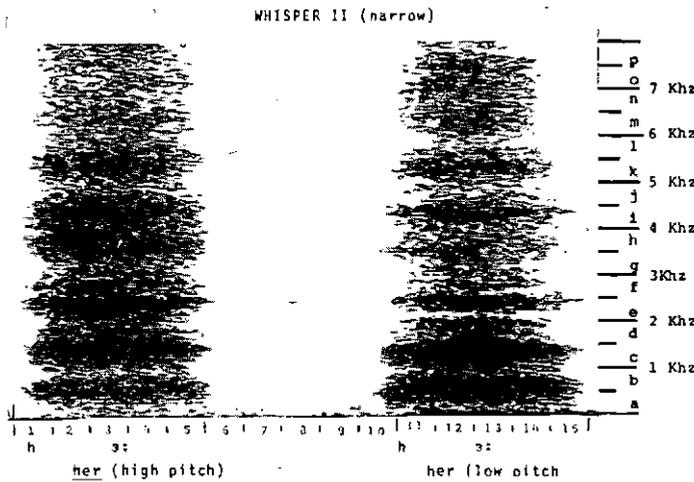
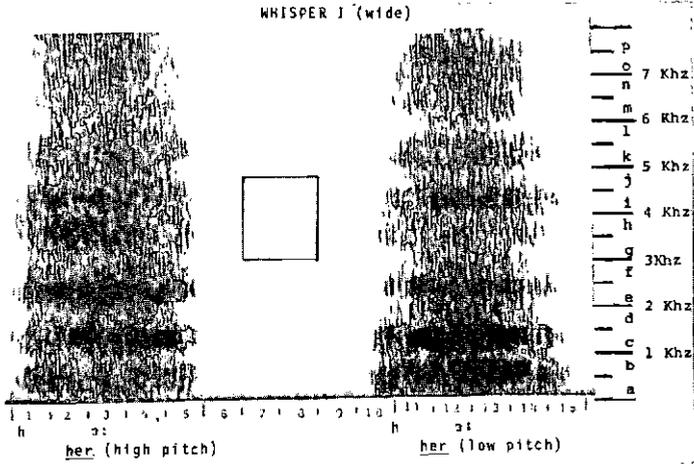
"Voice" refers to the manner in which the upper vocal tract is made to function as a resonating system bringing out the distinctive qualities of vowels and vowel-like speech sounds (i.e. oral and nasal resonants such as l and r, m and n), whose qualities depend on the resonances of the vocal tract. In this section of the paper the acoustic properties of the different voices are analysed in an attempt to correlate this analysis with states of the esophageal sphincter, the pharynx and the organs of normal speech. The acoustic properties are extracted by means of a spectrographic analysis. The measurement of pharyngeal pressure is resorted to in one case in order to confirm pharyngeal constriction.

In physical properties each of the voices of the five cases are different in their own way and are examined in the order of, first, the case of ventricular phonation followed by the four cases of esophageal speech including Case D, an interesting case who has two distinct "voices". The aim of this analysis is to identify features which may contribute to a classificatory framework of voice without vocal folds, and to an understanding of some of the compensatory mechanisms involved in producing intelligible speech.

SPECTROGRAPHIC ANALYSIS OF SPEECH SOUNDS

The Kay Sonagraph Model 6061-B was used to provide spectrograms showing the acoustic properties of the speech sounds made by our five cases. The spectrograms show frequency on the vertical axis from 0-8 KHz., amplitude in varying degrees of darkness in lines and smudges made by the stylus of the Sonagraph, and frequency/amplitude changes in time on the horizontal axis (12,33 cm = 1 sec.). The Sonagraph provides fine-grained analysis on the vertical axis in a *narrow-band* display in which the instrument registers conflated intensity in bandwidths of 45 cps; variations in frequency over very brief intervals are therefore registered separately. The *wide-band* display conflates intensity over bandwidths of 300 cps. Wide-band analysis provides fine-grained analysis in the time dimension. The analysis of variations in time in narrow-band spectrograms is coarse, as are variations in frequency in broad-band displays.

In the analysis of speech sounds which depend on the glottal note, a narrow-band spectrogram highlights harmonic structure. The spectrogram marked Normal II shows a harmonic structure with a fundamental of approximately 120 cps in the region of 19 to 20 on the horizontal cm scale. The resonances of the supra-glottal vocal tract amplify, selectively, harmonics in the glottal note and these amplified harmonics are seen to be darker and thicker at a/b



Reproduced by Sabinet Gateway under licence granted by the Publisher (dated 2012)

and c in the vertical scale. Resonance bars appear much more clearly at equivalent points on the wide-band Normal I spectrogram. Wide-band spectrograms highlight resonance properties on which the distinctive quality of vowels and resonant consonants depend and such resonance properties are the products of the cavities of the supra-glottal vocal tract. Wide-band displays obliterate harmonic structure. Resonance bars are normally termed "formants" and the formant structure for the vowel [3:] of *Kerr* emerges as F1 = 550, F2 = 1340, F3 = 2250 (approximate centre frequencies of lowest three formants). Fine-structure analysis revealing the pulses of the glottal note shows in wide-band spectrograms in the thin parallel vertical lines rising vertically through most of the visible frequency scale where vowels are articulated. Each of the highly regular (in time and amplitude) pressure pulses from the vibrating vocal folds is shown by a vertical line which also thickens and darkens as it enters the bandwidth of a formant. Glottal pulses up to 500 per second are discernible in wide-band spectrograms.

Aperiodic noise, at whatever point in the vocal tract it is created, shows as striations in wide-band displays: i.e. as irregular vertical lines of varying lengths as seen at Normal I 10. Here the consonant [s] is seen as randomly distributed energy concentrated mainly above 3.5 KHz which, in time, has a relatively gradual onset and continues for approximately 16 csecs. Noise in the form of a burst, i.e. instantaneous onset and rapid decay is seen at Case AI 7 corresponding to the release of the [t] of *to*. Aperiodic high-intensity noise is also clearly shown by Whisper I where random, irregular frequency/amplitude components are seen to be amplified by resonances; compare the ill-defined, but discernible, Formant 2 of [3:] in *her* in Whisper I and [3:] of *Kerr* in Normal I. An aperiodic noise component in narrow-band spectrograms appears as a horizontally elongated smudge rather than a thin vertical line. The representation of noise in the two different spectrographic samplings are clearly shown by Whisper I (wide band) and Whisper II (narrow band).

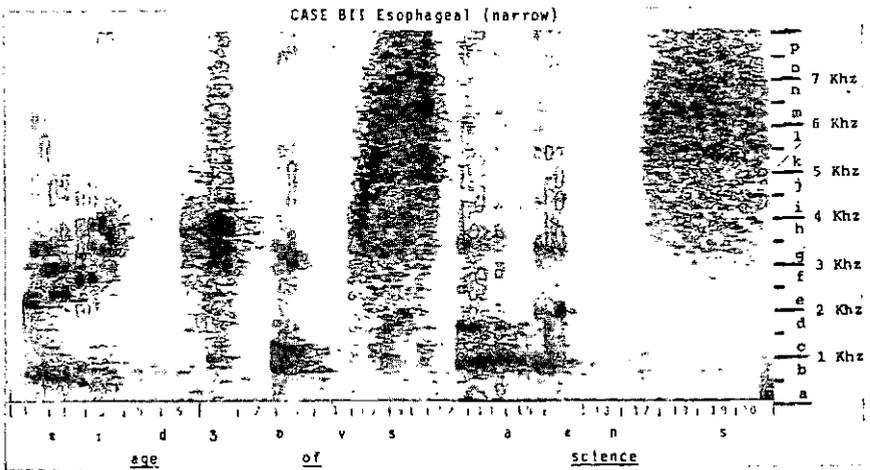
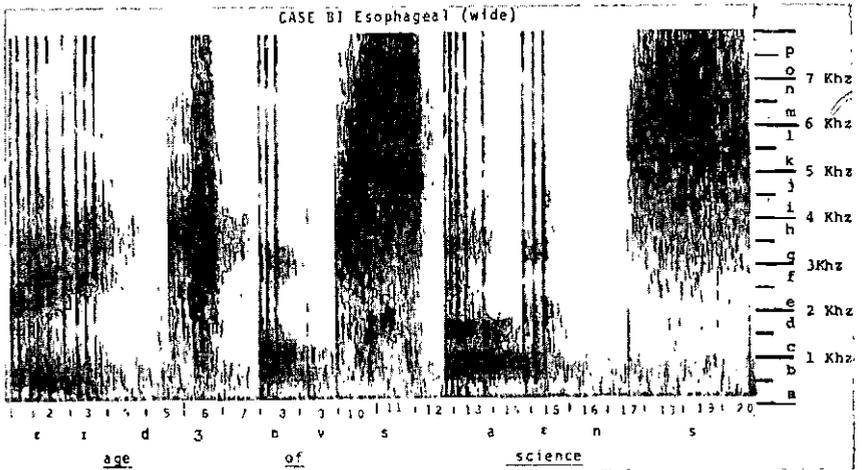
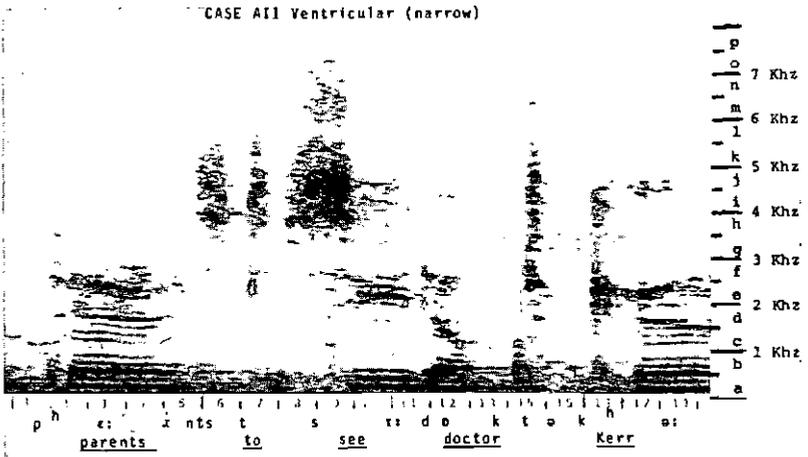
The means of identifying areas on the spectrograms presented here is in terms of coordinates on the horizontal and vertical scales. In illustration, the square box in the centre of spectrogram Whisper I is 7 to 8 - g to j. The title over each spectrogram should be interpreted thus: CASE DV Esophageal (wide) is a wide-band spectrogram identified by the number V and representing the voice of Case D who has esophageal voice. Note that V and VI are wide and narrow-band spectrograms of the same utterance.

The means of interpreting the pharyngeal pressure graphs are discussed below.

VOICE TYPES IN THIS STUDY

The five cases range, in their speech, from high to relatively low levels of intelligibility (with some fluctuation in individual cases) and the correlation between these differences with the different means of exciting the resonators is attempted below.

Resonators function in response to an input in which a driving force is involved. In the human voice the driving force is always an air flow which passes a point of constriction and the resonators of the upper vocal tract can be excited in three different ways. The air flow may be interrupted at the point of



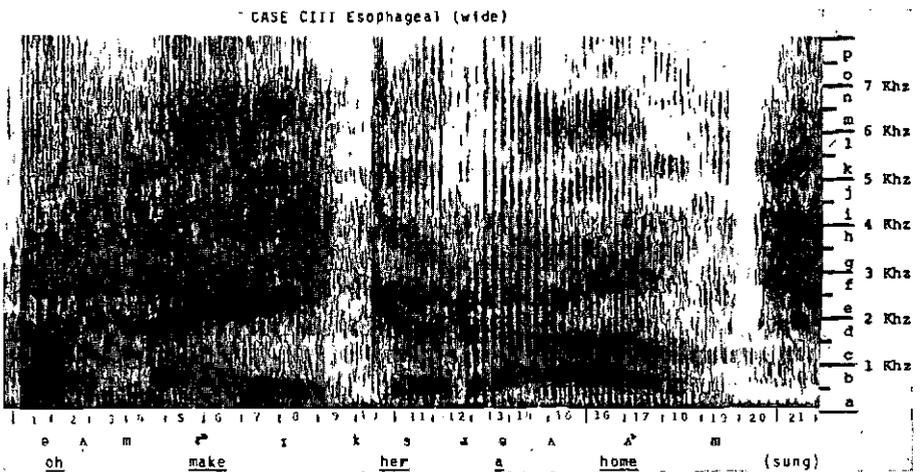
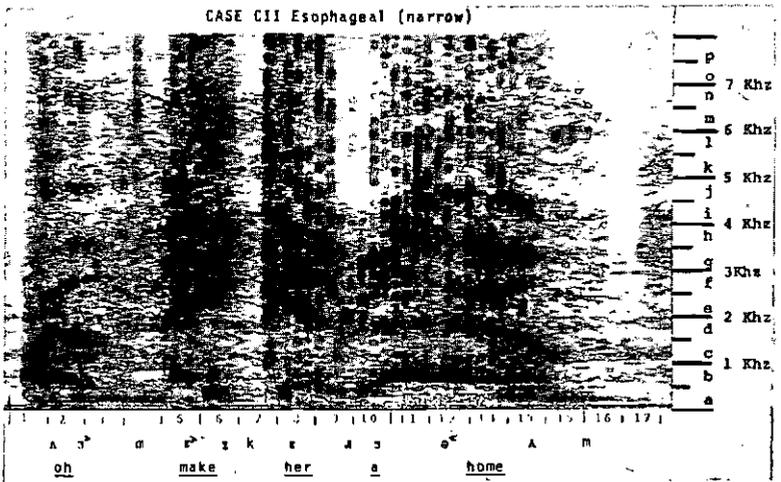
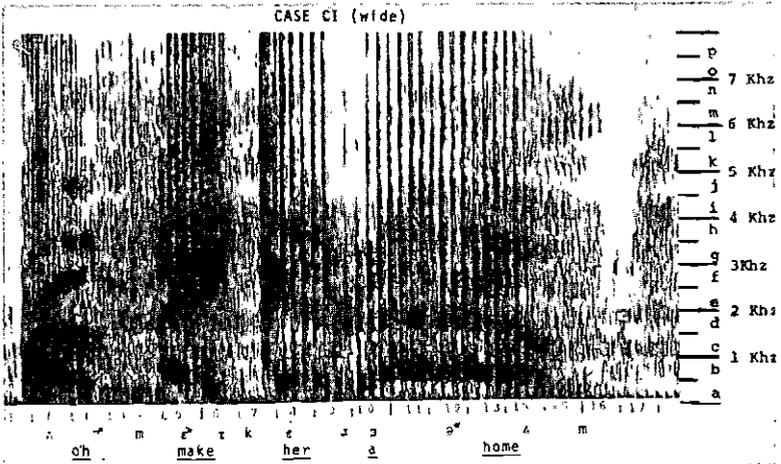
constriction by rapidly alternating closed and open phases. In normal voice lung air is interrupted in highly regular and rhythmical vibratory cycles by the true vocal folds which are drawn medially and, in varying degrees, laterally, into a state of vibration. These vibrations are a consequence of the Bernoulli Effect in which an accelerating air flow through a narrowing channel sucks the elastic edges of the constrictor together. Closure brings a change in pressure at the point of constriction and is followed by rapid opening. The vibrations of normal phonation set up pressure pulses which are highly regular in time and amplitude and only vibrators having the properties of the vocal folds, violin strings, etc. can produce a stream of pressure pulses of this kind. These pulses create a periodic sound wave with an inherent harmonic structure, i.e. the energy in the vibrations is largely concentrated at points in the frequency scale which are multiples of the fundamental (the lowest frequency component). Very little turbulence occurs in this type of interrupted air flow even as the closed phase of the vibratory cycle is approached and there is, therefore, little concomitant aperiodic noise. A vowel as a relatively "pure" note or tone is produced by resonators linked to a vibratory source of this kind. The resonators amplify harmonics set up by the glottal note which fall within their bandwidth and four clear resonance bars (formants) can be seen in wide-band spectrograms in the normal voice (*wide* Normal I 3 to 4 in the articulation of the first vowel of *parents*). To the extent that irregularities develop in time (frequency) and amplitude in the vibratory pattern of the vocal folds, the ear identifies harshness or roughness. A variation of frequency of as little as 1 cps can give rise to perceived roughness.²⁴ Changes in the harmonic structure give rise to perceived differences in pitch (intonation). In the narrow-band spectrogram Normal II 11 - a to i, a higher pitch corresponding to roughly 160 cps falls to a lower pitch of 123 cps at 21 - a to p.

A resonating system can, however, also function by receiving impulses in the form of sharp raps or taps. The forefinger flicking the throat just above the superior edge of the right lamina of the thyroid produces a spectrogram such as that labelled Finger-flicks. Each rap sets up a noise burst of very brief duration with aperiodic noise properties in which energy is distributed over the 8 KHz. visible on the spectrogram. There is no inherent harmonic structure in the noise burst and even a rapid succession of such raps at rates of over 100 per second does not set up a harmonic structure. There is, therefore, no possibility of frequency modulation which would be perceived as pitch variation or intonation, even if the rate of rapping is varied significantly.*

The vocal tract functions as a resonating system in response to the "rapping" input by amplifying frequencies in the noise bursts falling within the bandwidth of the resonators. The energy at such amplified frequencies decays relatively slowly and the emergent formants can be discerned at Finger-flicks 1 and 2 - a and b as horizontally extended smudges.

A third driving force for exciting resonators is continuous turbulence causing

* Our discussion implicitly rejects Scripture's¹⁹ (1906) theory that no over-tones emanate from glottal vibrations, only air puffs which cause the resonators to sound with their own frequencies.

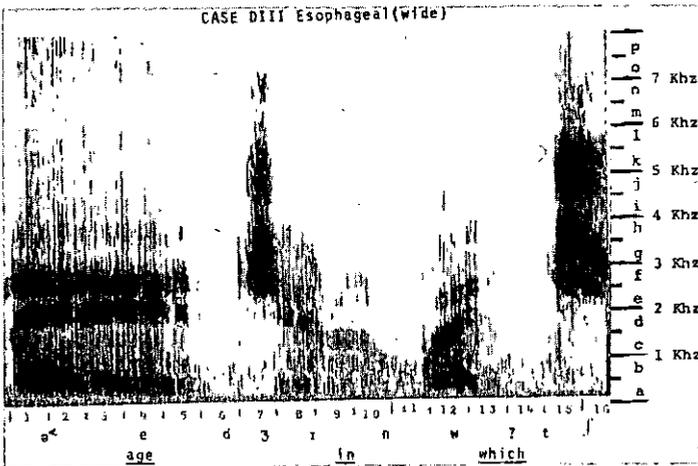
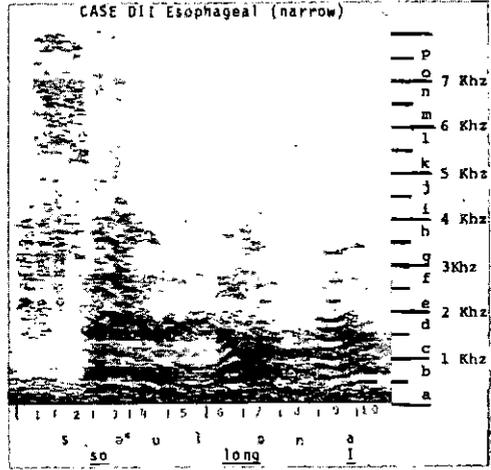
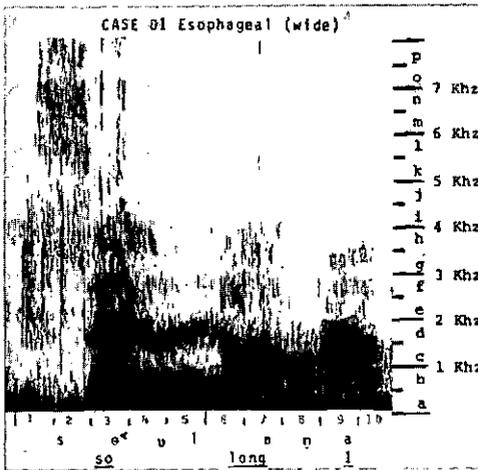
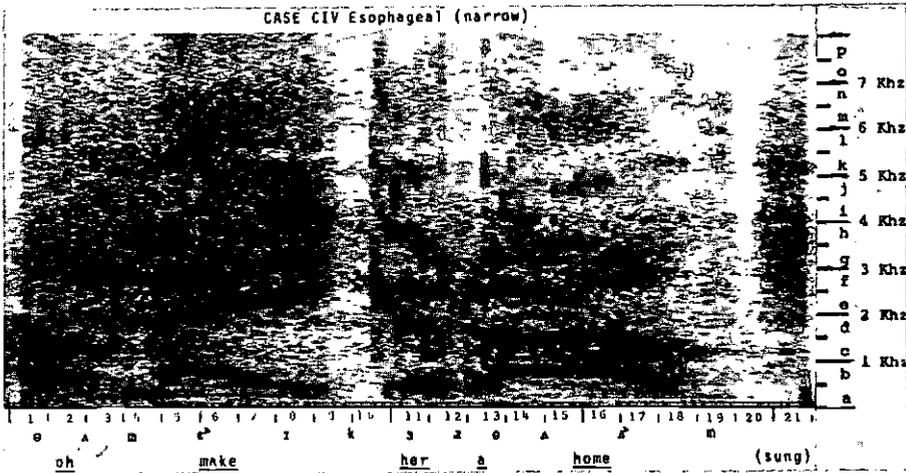


friction of relatively high amplitude emanating from a point of narrow constriction at which air escapes without any interruption of the air flow. Whispering is an excellent example of resonators functioning with this type of input. The point of constriction is a small V-shaped opening at the posterior end of the glottis and the vocal folds do not vibrate. Whisper I (wide-band, and II (narrow-band) show strong, transient, aperiodic noise components over the whole of the visible frequency scale and strong amplification of those frequencies falling in the bandwidths of resonators, in particular; see Formant 2 in Whisper II 11 to 14 - c and the corresponding area in the narrow-band display of Whisper II. This is spectrographic evidence of the discernible quality differences of whispered vowels. Continuous turbulence is, for the human body as a sound producer, highly uneconomical in rapidly draining the air reservoir. For esophageal speech with its small air reservoir it is impracticable. One reason for discussing whisper here is that esophageal "burping" (see below) with a high level of concomitant friction, is seemingly far more effective than esophageal "rapping" with little friction.

A point of interest connected with whisper is that, although totally lacking in harmonic structure (see Whisper II), an auditory impression of pitch variation can be created by so altering the constriction at source that the energy is differently distributed in the frequency scale. The difference between "low-pitch whisper" and "high-pitch whisper" in Whisper I and II (representing a conscious effort by the normal voice recorded on these spectrograms) is that Formant I (at level a/b) and Formant 2 (at level c) have, respectively, less and more energy concentrated at these relatively low frequencies. Formant 3 (at e/f, i.e. 2.5 KHz) shows the approximate point where the low-high energy distribution becomes inverted.

In esophageal "voice" air flow from a reservoir in the upper esophagus is interrupted by crude vibrations of the esophageal sphincter which alternatively opens and closes the esophageal lumen. In esophageal sphincter vibration the closure of the valve-like exit to the air reservoir is probably brought about by muscle tension after the opening caused by air pressure. If this is the case then this vibration is not a consequence of the Bernoulli Effect. Each opening releases a burst of noise into the resonating system at irregular, variable rates. Muscle tension and air pressure would regulate the rate of vibration. Case B illustrates how close this type of vibration can come to "rapping" or "tapping"; compare Case BI 1 to 3 with Finger-flicks. Here rapping is seen to be irregular in time with an average rate of roughly 42 per second. Case CI 8 to 14 shows a somewhat more rapid, regular rate of rapping at approximately 44 per second. The corresponding sections on the narrow-band spectrograms BII, CII show a total absence of harmonic structure and the formants appear as smudges darkened in the vertical bars of the raps. Spectrographic evidence attests to the absence of frequency modulation in esophageal voice.

A significant variation in esophageal sphincter vibration begins to show in comparing BI and II with CI and II (particularly 1 to 4) and is clearest when EI enters the comparison. The actual noise bursts recede in prominence and continuing aperiodic noise components over wide bandwidths are the main input to the resonators. The raps apparently smooth out into a succession of air puffs somewhat more rapid and regular, and lower in amplitude (hereafter



referred to as “burping” in contradistinction to “rapping”). The distinction between rapping and burping is significant because the voice of Case B is very “croaky” and of low intelligibility; that of Case C is unpleasantly croaky, but of high intelligibility; that of E not at all croaky, breathy, but pleasant, and highly intelligible.

Although our postulation as to the site of constriction which produces random aperiodic noise components is unconfirmed experimentally, we suggest that friction noise emanates from the esophageal sphincter where the manner in which the air flow is interrupted ranges over vibratory cycles with tight closure, high esophageal pressure and “clean” break on opening (rapping), to burping with weaker, probably incomplete, closure and considerable concomitant friction. A third possible state is a still weaker interruption with a more or less open lumen allowing a continuous air flow. This would seem to be the nature of the esophageal air mechanism in oral fricatives such as [s].

If turbulence is not a consequence of esophageal vibration then pharyngeal constriction is the next most likely source. Pharyngeal pressure measurements show that Case D uses this mechanism and provides evidence of a vibratory mechanism involving the pharynx. We are doubtful, however, whether Case C has pharyngeal constriction. Case C burps without embarrassment and makes no attempt to cover the esophageal voice by pharyngeal constriction. Compare spoken and sung *make* in CI, II at 5 to 6 and CIII, IV at 5 to 8. The prominence of the raps recedes as the vibrations speed up* and the quantity of random, aperiodic noise is greater. In Case E the significant absence of strong noise bursts in EI is worth investigating.

Two significant dimensions of esophageal voice emerge from this discussion: (a) the manner in which the esophageal sphincter controls the release of air; (b) the presence and nature of pharyngeal constriction.

In classifying cases of the types dealt with here it is useful to distinguish the three types of source input to the resonating system:

Type X Pressure pulses with a harmonic structure and true frequency modulation (friction noise may be present, but is a minor contributor).

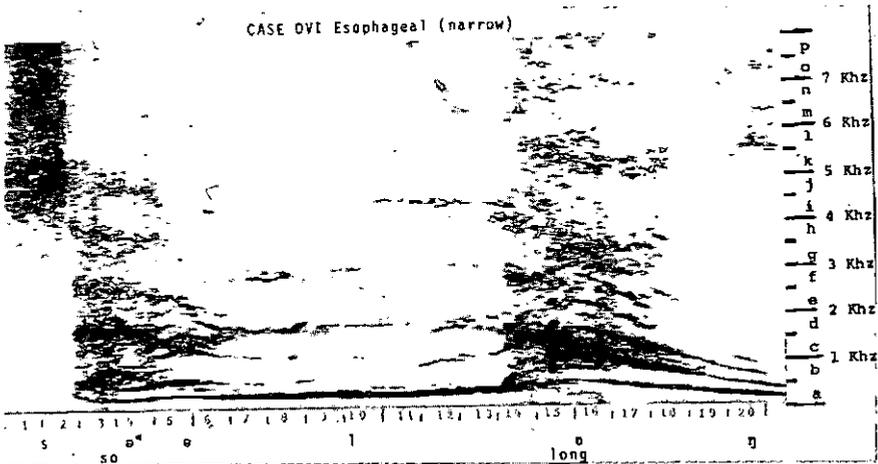
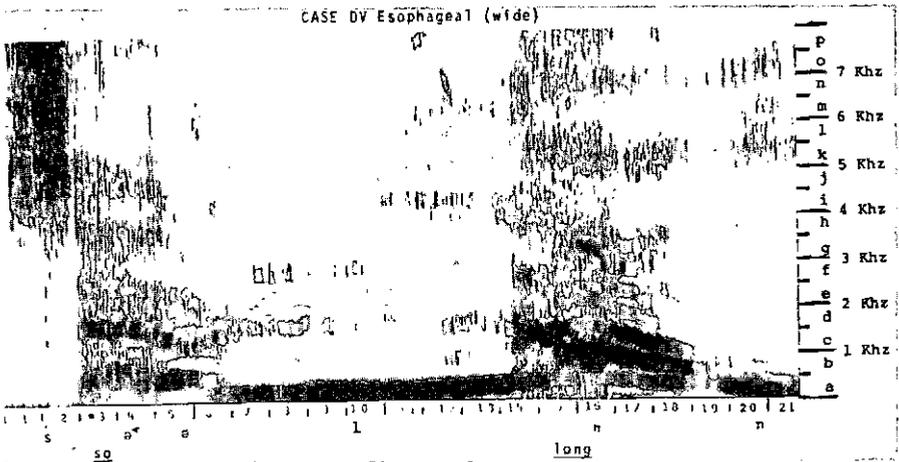
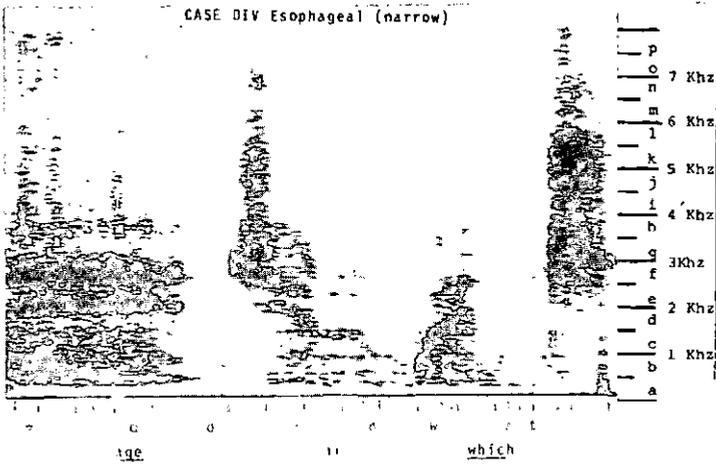
Type Y Vibrations of low frequency in the form of noise bursts lacking a harmonic structure (rapping and burping).

Type Z Continuous, relatively uninterrupted friction noise.

None of our cases is classified as Z, but experiments with a normal vocal tract show that pharyngeal constriction could set up a “voice” of this kind. Obviously there is overlap within this categorisation. Z can clearly overlap with X and Y; X and Y are, as a rule, mutually exclusive.

The phonation of a normal voice is only Type X; a succession of glottal stops or catches cannot be produced at a rate fast enough to provide any real semblance of Type Y. Ventricular phonation in the one case discussed below, is

* The superior-inferior branches of the laryngeal nerve seem to convey the same instruction to the cricopharyngeus muscle as it did to muscles of the larynx.



Type X. Esophageal voice is Type Y, but Case D is a particularly interesting case of both X and Y. A summary of the cases discussed below:

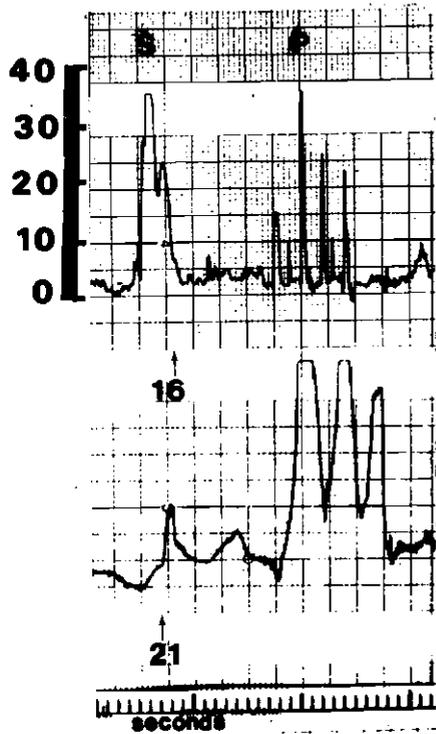
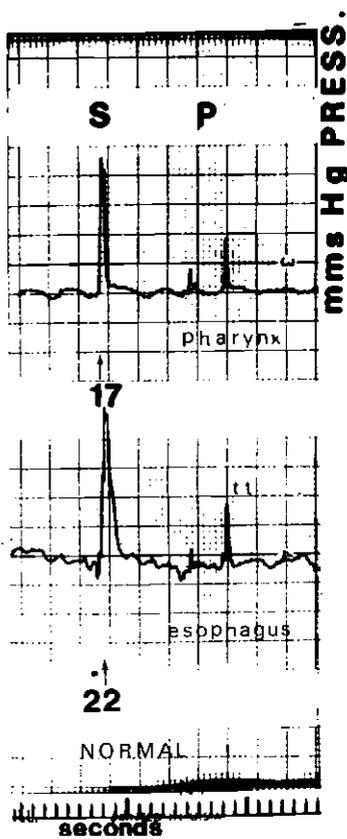
Normal	Case A	Case B	Case C	Case D	Case E
	Ventricular folds intact; no functioning vocal folds	Laryngectomy	Laryngectomy	Laryngectomy	Laryngectomy
Type X	Type X	Type Y	Type Y	Types Y and X	Type Y

DISCUSSION OF CASES

VENTRICULAR PHONATION : CASE A. This voice is highly intelligible with a fair degree of hoarseness or roughness. Case A is capable of a sustained high level of power. He has limited pitch variation in intonation although the stretch of speech represented in spectrogram AII shows little frequency modulation. (The fundamental at AII 2/3 is roughly 168 cps.) With effort, Case A can raise his pitch to a fundamental frequency of about 300 cps. The lowest harmonics in AII are seen to be very strong, but harmonic structure weakens rapidly as the frequency rises and above 2.5 KHz, formants are brought out mainly by high frequency aperiodic noise amplified in the resonance bars of the highest formants. Contrast this with Normal II where harmonic structure shows gradual decrease in amplitude, but continues throughout the visible frequency scale and brings out even the highest formants.

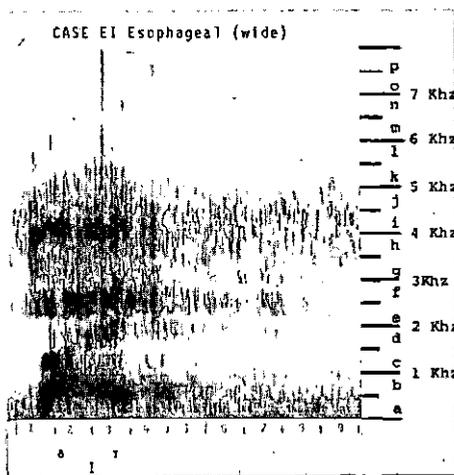
In the region AI 17 to 18, frequency variation is a maximum of 16 cycles around a mean of approximately 178 cps and this variation is a major contributor to the hoarseness of the voice. Ventricular phonation in Case A is therefore periodic to a high degree and these findings apparently differ from those of Moore¹²: *When ventricular folds vibrate, they are quite aperiodic*.... The laryngological report shows, however, that Case A is an exceptional one of ventricular phonation. This case is important in showing how ventricular phonation can approach normal phonation quite closely.

ESOPHAGEAL VOICE.: CASE B. This voice is very "croaky" and of the four esophageal cases has the lowest level of power and intelligibility. The nature of the "voice" is a major contributor to this lack of clarity. All recordings of this case produce an extreme Type Y "rapping" at slow, highly irregular, rates. At BI 1 to 3, (the diphthong [ei] of *age*) there are 9 raps at frequencies ranging from 30 to 61 per second. The irregularity at BI 13 to 15 is even more striking. The poor definition of formants seen at BI 2 to 3 — b, and e to h, is a consequence of the manner in which the resonators are excited. A striking difference between this voice and other more successful esophageal voices is the absence of random aperiodic noise outside the noise bursts of the raps and, in consequence, the weak, shadowy, formant structure. A compari-



CASE D

GRAPH D VII



son of BI with CI and DIII illustrates this. The almost blank vertical strip at BI 16 to 17.5 (the nasal consonant [n]) shows no impulses being transmitted to the resonators and contrasts in this respect with an equivalent nasal consonant of CI 4. Apparently, random aperiodic components produced by turbulence play a major role in filling out resonances to produce quality differences. The "rapping" of Case C (see CI) is much more successful because the rate is faster and more regular and a good deal of energy is located in aperiodic components outside the noise bursts. The raps themselves do not differ significantly in amplitude between Cases B and C. Another contributory factor to the poor quality of Case B speech is the poor timing of the rapping mechanism which switches off for nasals and resonants.

CASE C. This is a voice of considerable power and very high intelligibility although sometimes unpleasantly harsh. It is closest to the "rapping" version of Type Y but, interestingly, the "burping" mode is evident when this case attempts to sing. The spectrographic definition of formants, particularly the higher formants, tends to be clearest with "speaking" rather than "singing" (i.e. rapping rather than burping), but both are highly intelligible. In neither version is there any sustained harmonic structure and the weak auditory impression of a pitch change between sung and spoken *make* is due not to the different rates of rapping or burping, but partly to the lengthening of the vowel in sung *make* and partly to greater energy at the higher frequencies (although this is not as clear as in high-pitch vs low-pitch whisper in Whisper II). Case C burps at a rate of approximately 72 per second at CIII 8 and raps at 47 per second at CI 12, but with a high degree of irregularity.

CASE D. The two "voices" of this case can be seen by comparing DI, II 3 to 10 and DIII, IV 1 to 5. The harmonic structure (with vibrato) emerges clearly in the former and the fundamental frequency is approximately 215 cps at DI 9.3. Harmonics are strong below 3 KHz., but hardly discernible above that. In DI there is a clear definition of formants and the spectrogram is not too unlike that of a normal woman's voice although there is a high degree of accompanying aperiodic noise. DIII, IV 1 to 5 shows burping with an absence of harmonic structure although the burps, with energy mainly below 4 KHz., bring out a fairly clear formant structure. Vibrations at DIII 3 are fast, approximately 110 per second. (In our data Case E gave the fastest esophageal sphincter vibration of 120 per second.)

Spectrograms DI, II were made from a recording in 1966 when the morale of this case was said to be high. DIII, IV were made early in 1973 when the case was somewhat depressed. DV, VI also date from the latter period and represent the response of this case to a request to emulate her 1966 voice making the same utterance *so long* with high emotion expressing surprise. The harmonic structure is more tenuous, intermittent and without vibrato; most of the energy is in the first harmonic which moves from approximately 240 cps at DVI 4 to about 400 cps at DVI 18. Like DI, II there is a clear auditory impression of frequency modulation.

It was seen quite early that Case D was stifling her esophageal voice by pharyn-



Fig. A1
Respiration
a = Vocal Fold

Fig. A2
Phonation
b = Ventricular Fold

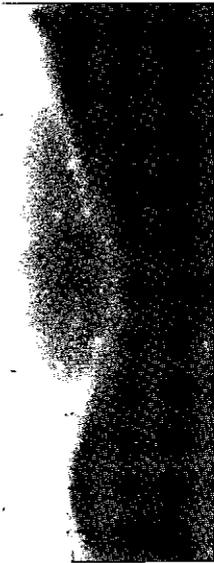


Fig. A3

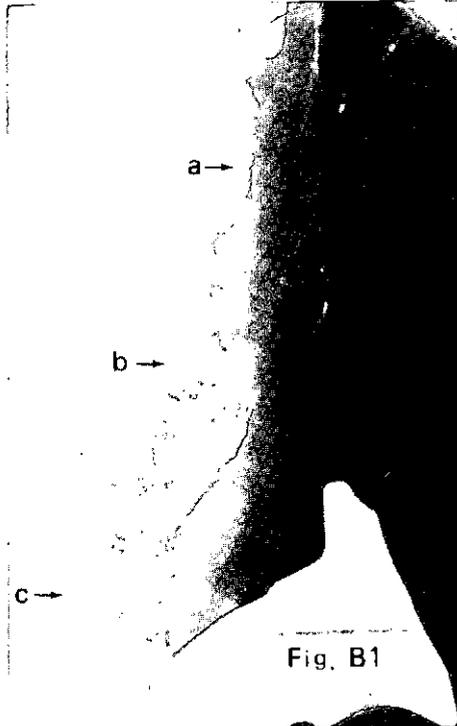


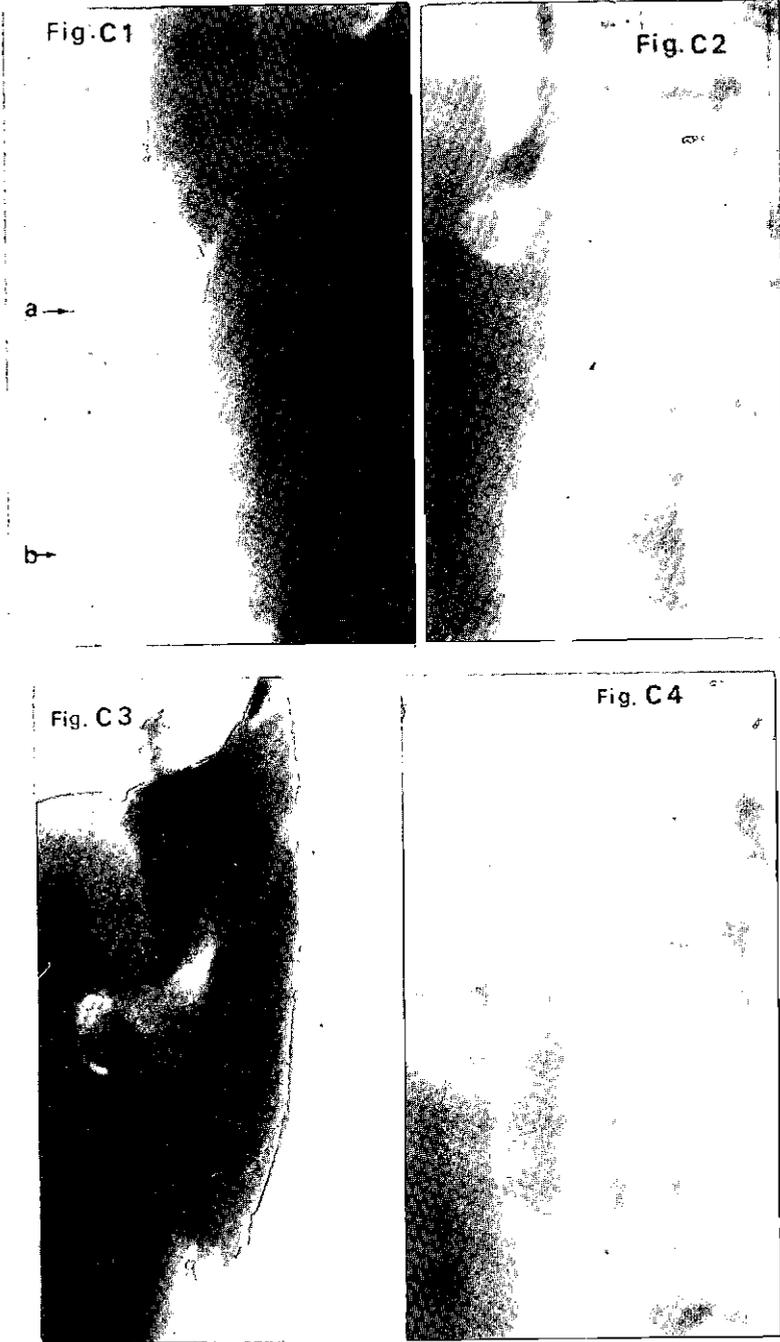
Fig. B1

geal constriction apparently involving the pharyngeal constrictors and the back of the tongue. Pharyngeal constriction is a common feature of Case D's voice except for rare breaks, sometimes occurring in speech and sometimes in pharyngoscopic investigations, when the esophageal burping is clearly audible – to the embarrassment of Case D whose pharyngeal constriction is apparently an attempt to cover up the true esophageal voice.

The site of Case D's Type X voice which reaches such high fundamental frequencies, is a question of some interest. Obviously, there is an extremely delicate balance in muscle function, hence the vibrato. The pharyngeal pressure graphs DVII represent our attempt to obtain firm evidence of the action of the pharyngeal mechanism.^{2,3,25} They support our belief that Type X voice is produced by pharyngeal constriction because of the clear indication of phases of total closure.

Pharyngeal pressure was measured with a Beckman RM Dynograph Recorder using 4-bore polyvinyl tubes perfused at .055 ml/minute connected to P23AA Statham transducers. The pressures for Case D and for Normal (right graph and left graph respectively) were taken in the pharynx about 2 cms and 1 cm respectively above the esophageal sphincter and show a swallow (marked S) and then speech (marked P) for the utterance *I can't*. During the utterance the Normal voice shows two spikes of slight pressure in the pharynx probably corresponding to the "hold" (closure of the vocal tract) of the consonant [k] of *can't*, but there is no significant increase in pressure elsewhere. Case D, on the other hand, shows very high pharyngeal pressure (in comparison with norms for phonation) throughout the utterance which could only be achieved by phases of total adduction. (The utterance was quite abnormally prolonged by Case D.) The high pressure is, however, not sustained and can be seen to fall drastically and quickly rise again. This, we believe, could be due to the inability to maintain the high degree of pharyngeal constriction in achieving Type X voice and the need to release pressure before allowing it to build up again. The whole mechanism of esophageal air release linked with the pharyngeal pseudoglottis is not fully understood by us, but we suggest that there is spectrographic evidence that true Type X voice can only be sustained for limited periods followed by a break with some degree of constrictor abduction. In DIII, IV 12 there appears to be about 4 csecs of Type X voice (note a semblance of harmonic structure at DIV 12 – b to c) preceded by weak Type Y voice from 8 to 11. An interesting question is whether the vibrations in the latter stretch – and 1 to 4 in the same spectrogram – are vibrations of the esophageal sphincter as well as of the pharyngeal constriction. Their rate of 110 per second is within the range of esophageal sphincter vibrations.

In auditory impression, Case D's 1973 voice is squeaky, "squeezed" and low-powered, particularly when in the true Type X mode, with an intelligibility level between that of Cases B and C. It is not at all harsh or croaky. In 1966 the voice was stronger and clearer and the ready ability to achieve frequency modulation is known to have deceived at least one experienced laryngologist in telephone conversation with Case D.



Reproduced by Sabinet Gateway under licence granted by the Publisher (dated 2012)

CASE E. This voice is husky, not unduly harsh and maintains a high, very even, level of power. With Case C it is the most intelligible of the four esophageal voices. It contrasts with C in being typically "burping" rather than "rapping" and demonstrates how a clear formant structure can be brought out with a non-harmonic source. (See EI 2 to 3.) The clear formant structure and, consequently, the high intelligibility, are due to turbulence which provides components from very low in the frequency scale to at least 5.5 KHz. The rate of burping in the region of EI 2 to 3 varies around a mean of approximately 85.

SIGNIFICANT ASPECTS OF THE FIVE CASES

Points arising from our study (all of which call for deeper investigation) which have significance in speech therapy and surgical procedures in laryngectomy are:

- a) The nature of compensatory mechanisms in "speech without a larynx".
- b) The tendency to "cover" the esophageal croak with pharyngeal constriction and the effects thereof.
- c) The superiority of the "burping" mode of esophageal sphincter action, i.e. relatively fast and regular vibration with a good deal of concomitant friction noise, as the most effective means of exciting upper vocal tract resonators.
- d) The presence of a pharyngeal vibrator producing true pitch modulation.
- e) The ability to create an impression of pitch modulation without harmonic structure by a shift of energy in aperiodic noise components from a relatively low to a higher bandwidth.
- f) The effectiveness with which ventricular phonation can, over a period of time, substitute for the function of the true vocal folds.

PART II

ANATOMICAL ANALYSIS AND CASE HISTORIES.

CASE A. MR. W.P.

Mr. W.P. is now 33 years old. In 1944, at 4 years of age, he was examined by a laryngologist who found Laryngeal Papillomata. These were removed and, at the same time, a diathermy needle was used on the vocal cords. This operation was followed by laryngeal obstruction and tracheostomy became necessary three weeks later. Within three months the cords had become "adherent" and a course of dilatation with bougies was started and continued until August 1945; but no improvement had been achieved and the patient was referred to Dr. C.L. Jackson in Philadelphia where further treatment, including laryngostomy and plastic closure, was carried out by Dr. Jackson and Dr. Charles M. Norris.

Fig. D1

Pharynx →

Cricopharyngeus →

Oesophagus →

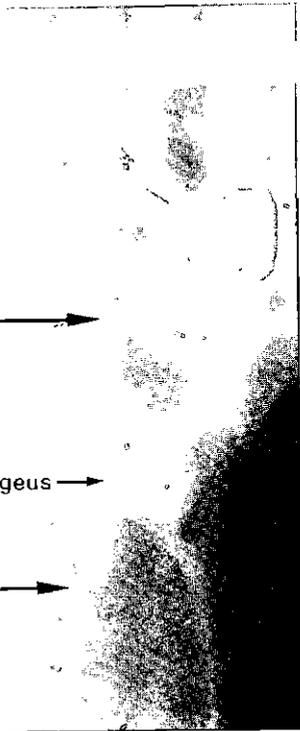


Fig. D2

a →

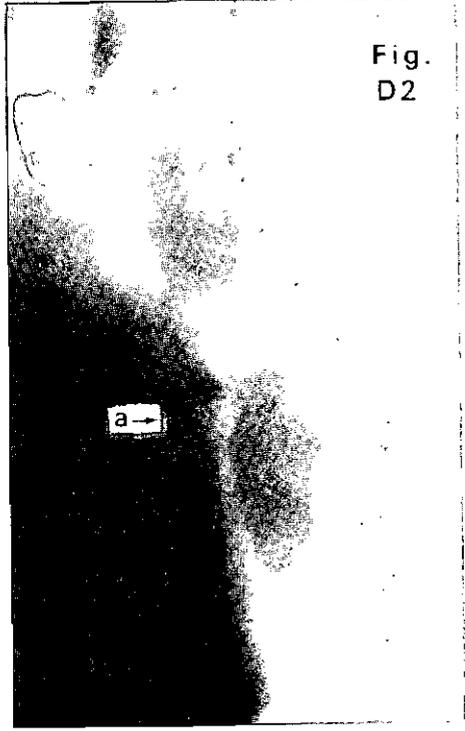
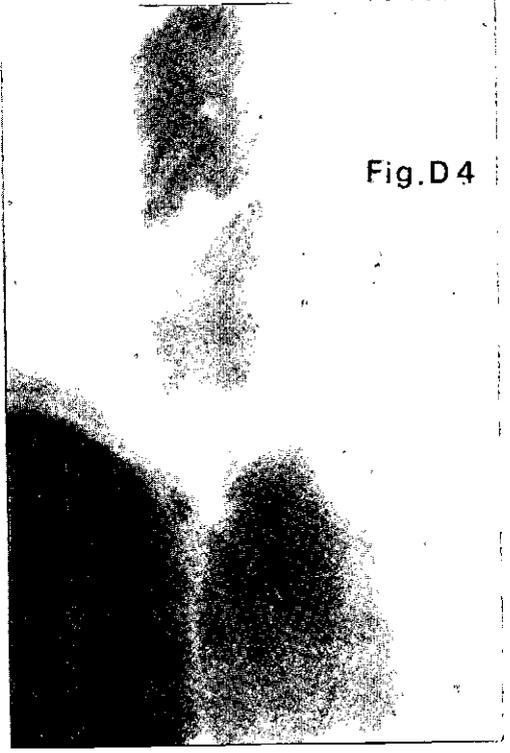


Fig. D3



Fig. D4



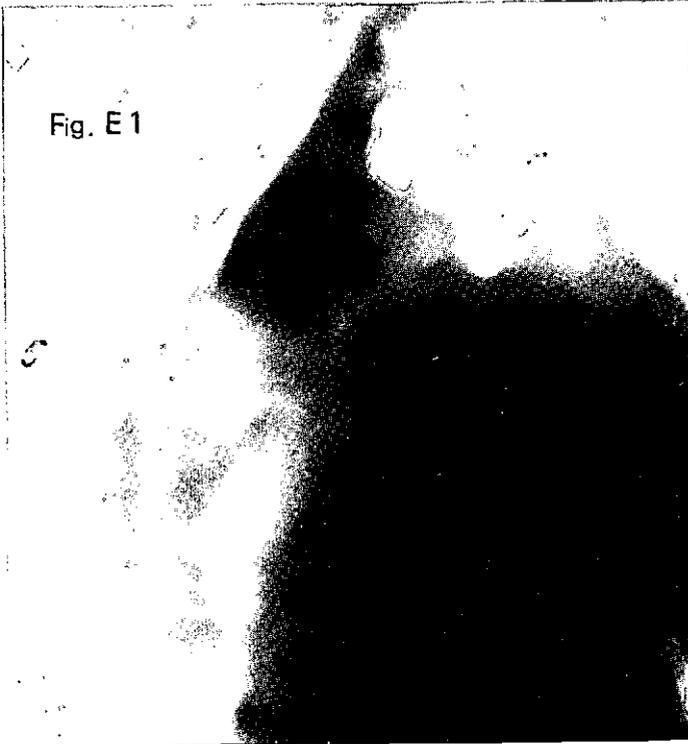
The patient was first examined by W.A.K. in 1948 because of a recurrence of papillomata and stenosis; and direct laryngoscopy confirmed the presence of both, as well as the fact that the cords were not adherent but were hard, fibrous and immobile. He was not seen again until 1963 and, in the interim, he had developed a weak, high-pitched voice which had not been present in 1948 when it was just a whisper. Preliminary mirror examination revealed pale, immobile, rounded vocal cords which were lying in partial adduction but not in contact on phonation. The ventricular bands, however, were mobile and came into contact on phonation; and they showed, also, apposing edges which were more than normally smooth. Radiological examination, using a contrast medium, confirmed the clinical findings of ventricular phonation and of rounded vocal cords. See Figs. A 1 and A 2. Apart from this, he had developed a tracheocele at the site of previous surgery, from the increased intratracheal pressure resultant on the laryngeal stenosis. See Fig. A 3.

A laryngofissure operation was performed by W.A.K. in September 1963 when the hard, fibrous, immobile vocal cords were excised down to cartilage and a skin graft was applied, care being taken not to traumatise the healthy, smooth-surfaced ventricular bands.

At the same time the tracheocele was excised. This operation has been only partially successful and an ellipsoidal stenosis has recurred at the site of the vocal cords; but Mr. W.P.'s ventricular phonation has continued to improve up to the present time and is not accompanied by general contracture of the upper sphincter muscles of the larynx; this was recorded by Saunders.¹⁷ He usually keeps his abdominal muscles in contraction when he speaks; a dry atmosphere improves his voice and he is quite certain that it is also improved by smoking (15 cigarettes a day). Iced drinks cause immediate, temporary deterioration as well as does a humid atmosphere. He can sing in tune and was a member of a boys' choir up to puberty; then the pitch of his voice lowered but it never 'broke'. His almost lifelong courage and his determination to speak, coupled with the numerous surgical procedures, have given rise to an increasing voluntary muscular control of his ventricular bands, as mentioned by Pressman,¹⁴ and to mucosal changes towards the thin and adherent type which is present on true vocal cords. His condition is of the type which has been described as vicarious by Jackson⁷ and by Arnold and Pinto.¹

CASE B. MRS. J.S.

Mrs. J.S. was 40 years of age when she underwent the operation of laryngectomy by Mr. C.P. Wilson in London in 1953; and he found it necessary to remove more than the usual segment of trachea. The hyoid bone was preserved and the epiglottis removed. The patient was first seen by W.A.K. in 1954 as she had not been able to develop anything more than a just-audible voice. Clinical examination revealed a recessed laryngectomy stoma lying at the level of the suprasternal notch, thus confirming that a considerable amount of trachea had been removed. However, in spite of speech therapy and encouragement from other laryngectomees, this lady has never developed satisfactory speech.



Radiographic assessment in 1973 has shown a functioning cricopharyngeal sphincter and the patient swallows food without difficulty; but it has not been possible to obtain an air column in the oesophagus up to this sphincter; see Fig. B 1, in contrast to cases D and E, Figs. D1 and E1 where the sphincter is sharply defined between well marked pharyngeal and oesophageal air columns. The area a. to b. in Fig. B 1 measures about 2.5 - 3.0 cm. which is longer than the normal length of the cricopharyngeus, of 1 to 2 cm., as given by Batson,⁴ and by Dey and Kirchner.⁵ W.A.K. does feel that, in Case B, some loss of normal muscular function has taken place in the inferior pharyngeal constrictor and, particularly, in the upper end of the esophagus, Ref Fig. B 1, b to c; and that her defective voice is partly due to these changes.

CASE C. MRS. G.A.

Mrs. G.A. was 37 years of age when she underwent laryngectomy by W.A.K. in May 1950. At operation the cricopharyngeus on the left was found to be adherent to an enlarged thyroid gland. Care was taken to minimise trauma to the former, the hyoid bone was preserved and the epiglottis removed. The infrahyoid muscles were preserved and overlapped to strengthen the pharyngeal repair.

Recovery was normal and within one month of the operation the patient had started esophageal speech. After six months the voice was satisfactory, though harsh, and her air intake was noisy. Radiographs by Dr. E. Samuel¹⁶ in 1951 and 1953 demonstrated satisfactory air columns in the pharynx and esophagus but no sharply defined cricopharyngeus.

Further studies by Dr. Stanley Lazar¹⁰ in June 1973 confirm the satisfactory air columns, Figs. C1 and C2 and show that the area a. to b. in Case C is almost 3.0 cm. in length. This is much longer than in Cases D and E and resembles, somewhat, the corresponding area in Case B; but the upper esophageal air column in Case C is wide and normal right up to b.; whereas, in Case B, it has always been narrow and tapering. This difference in the two cases may explain, at least in part, why Case C has a powerful and intelligible voice but Case B's voice is weak and lacks clarity.

In 1953 it was first noted by Samuel¹⁶ that Case C's pharyngeal air column was larger immediately after speech than before it, Fig. C3; and this interesting observation has been confirmed by Lazar¹⁰ in 1973, in Cases C and D. Figs. C4 and D4.

CASE D. MRS. B. McL.

Mrs. B. McL. was 39 years of age when she underwent laryngectomy by W.A.K. in February 1951. The hyoid bone was preserved, the epiglottis removed and trauma to the pharyngeal constrictors was minimised. The infrahyoid strap muscles were preserved and overlapped to strengthen the pharyngeal repair. Many authors have stressed the importance of preserving these structures, when possible, so as to give the patient every opportunity of developing satisfactory speech.^{7,8,18,22,23}

The patient was given a postoperative course of orthodox high-voltage radiation for over one month. By October 1951 she had developed good speech and, even from that time, she has always insisted that her voice is pharyngeal and she points to the region just below the hyoid bone on each side. Repeated clinical and radiological examinations have confirmed a functioning cricopharyngeus muscle and excellent esophageal and pharyngeal columns of air. Ref. Figs. D 1, D 2 and D 3. Fig. D 4 demonstrates the enlarged pharyngeal air column immediately after the sudden cessation of speech and this was referred to in the discussion on Case C. Fig. D 2 is a radiograph taken during relaxed esophageal speech and the 'lip' 'a'⁹ can be seen in close contact with the lower anterior margin of the cricopharyngeus, in which is a small area of calcification. This part of the cricopharyngeus can be seen to be in vibration and W.A.K. is of the opinion that this is the site of Case D's pseudoglottis for esophageal speech.^{11,15,21}

Fig. D 3 demonstrates the pharyngeal air column during pharyngeal phonation with emotion. Note that the tongue is in approximation to the posterior pharyngeal wall and reference to the graph D VII shows high hypo-pharyngeal pressure during this type of phonation.²⁰ Furthermore, light palpation with one finger just above the cricopharyngeus, during this phase, reveals vibration which is maximal at this point;¹¹ and mirror examination of the hypopharynx shows two antero-posterior folds in this region which are in apposition during pharyngeal phonation, the fold on the left side having a sharp, fibrous edge.^{6,8,13,20} Apart from the ability to 'blow her nose', provided a 'reasonable-sized' mass of secretion be in it, this lady can phonate immediately after a bolus of food has passed the cricopharyngeus; and can also phonate with her tongue partially protruded.

CASE E. MR. D.G.C.

Mr. C. was 58 years of age when laryngectomy was performed by W.A.K. in 1959. The epiglottis was removed and the hyoid bone preserved. The usual care was taken to minimise trauma to the cricopharyngeus, and the infra-hyoid muscles were preserved and overlapped to strengthen the pharyngeal repair.

Within 3 weeks of the operation this patient had started esophageal speech, within 3 months it was fair, and within a year it was good. He has always maintained that his voice comes from the region of the cricopharyngeus.

In March 1965 X-rays were taken to demonstrate his good esophageal air reservoir; and also the hypertrophy of the cricopharyngeus. Ref. Fig. E 1. Through the years his voice has, if anything, improved. His air intake is silent and his voice has never been loud nor raucous. He can say normal sentences without effort and, as a business executive, he has been capable of conducting interviews throughout the usual working day.

ACKNOWLEDGEMENTS

Pharyngeal pressure measurements were made by Mr. C.G. Bremner F.R.C.S. of the Department of Surgery in the University of the Witwatersrand; and his assistance is gratefully acknowledged, as well as that of Dr. Michael Denny, radiologist, his partners and radiographers for their kindness and generosity in carrying out numerous radiographical examinations on all the patients. In particular, we thank Dr. Monty Said and Dr. Stanley Lazar, as well as Miss J. Robb and her radiographer colleagues. The anatomical photographs were taken by A.R.J. Kerr, A.M. Shevitz and Mark Hudson whom we thank; also, Nursing Sister Pauline Green S.R.N. for carrying out tape recordings and typing, as well as routine nursing duties.

REFERENCES

1. Arnold, G.E. and Pinto, S. (1960): Ventricular Dysphonia. *Laryngoscope*, LXX, No. 12, p. 1612.
2. Bateman, G.H., Dornhurst, A.C. and Leathart, G.L. (1952): Oesophageal Speech. *B.M.J.*, 2: pp. 1177-78.
3. Bateman, G.H. (1953): Oesophageal Speech after Laryngectomy. *Acta Otolaryngologica*, XLIII, Fasc. 11-111, pp. 133-139.
4. Batson, O.V. (1955): The Cricopharyngeus Muscle. *Annals of O.R.L.*, 64, No. 1, p. 52.
5. Dey, F.L. and Kirchner, J.A. (1961): The Upper Esophageal Sphincter after Laryngectomy. *Laryngoscope*, LXXI, No. 2, p.99 and pp.112-114.
6. Equen, Murdock. (1956): The Rehabilitation of the Laryngectomee. *Arch. of Otolaryngol.*, 64, No. 1, pp. 1-2.
7. Jackson, C. and C.L. (1942): *Diseases and Injuries of the Larynx*. The Macmillan Co. New York, 2nd Ed. p.76 and p.456.
8. Kallen, L.A. (1934): Vicarious Vocal Mechanisms. *Arch. Otolaryngol.*, 20, pp. 460-503.
9. Kytta, Jyrki. (1964): Finnish Oesophageal Speech after Laryngectomy. *Acta Otolaryngol.*, Supplementum 195, p. 86.
10. Lazar, Stanley. (1973): Personal Report.
11. Moolenaar-Bijl, Annie. (1953): The Importance of Certain Consonants in Esophageal Voice after Laryngectomy. *Annals of O.R.L.*, LXII, No. 4, p. 980, Fig. 2.
12. Moore, G.P. (1971): *Organic Voice Disorders*. Prentice-Hall Inc., Englewood Cliffs, N.Y., p. 80.
13. Negus, V.E. (1929): *The Mechanism of the Larynx*. Heinemann, p. 350.
14. Pressman, J.J. (1954): Sphincters of the Larynx. *Arch. of Otolaryngol.*, 59, No. 2, p. 236.
15. Robe, Evelyn Y.; Moore, G.P.; Andrews, A.H.; Holinger, P.H. (1956): Speech after Laryngectomy. *Laryngoscope*, LXVI, No. 4, p. 399.
16. Samuel, Eric. (1951 & 1953): Personal Reports.
17. Saunders, W.H. (1950): Dysphonia Plica Ventricularis. *Annals of O.R.L.*, LXV, No. 3, p. 668.

Ban All Sticky Fingers

This is one of our slogans to help you remember the name of the world's finest cassette tapes. We think our tapes are better than most because of the smooth way things go around inside. Never any flutter problems.

Use the letters BASF, create your own slogan, send it to us at P.O. Box 6547, Johannesburg, and win a spy glass or an elephant or a matchstick or whatever. We're sure you can do better than us. The better your entry, the better your prize.



IF IT'S WORTH RECORDING IT'S
WORTH RECORDING ON BASF.
THE WORLD'S FINEST CASSETTE
TAPES.

WELLS, ROSTRON & ROWE ● 2581/5

18. Simpson, I.C.; Smith, J.C.S. and Gordon, Margaret T. (1972): Laryngectomy: The Influence of Muscle Reconstruction on the Mechanism of Oesophageal Voice Production. *J. Laryngology and Otology*, 86, No. 10, pp. 961-990.
19. Scripture, E.W. (1906): *Elements of Experimental Phonetics*. Chas. Scribner's Sons, New York.
20. Scripture, E.W. (1916): Speech without using Larynx. *J. Physiology*, 50, pp. 397-403.
21. Svane-Knudsen, V. (1960): The Substitute Voice of the Laryngectomized Patient. *Acta Otolaryngol.* 52, Fasc. 1, pp. 85-93.
22. Vrticka, K. and Svoboda, M. (1961): A Clinical and X-ray Study of 100 Laryngectomized Speakers. *Folia Phoniatic.*, 13, pp. 174-186.
23. Warner, Jennifer. (1971): Vocal Rehabilitation following Total Laryngectomy. *J. Laryngology and Otology*, LXXXV. No. 6, p. 578.
24. Wendahl, R.W. (1963): Laryngeal Analog Synthesis of Harsh Voice Quality. *Folia Phoniatica*, XV.
25. Zinner, E.M. and Fleshler, B. (1972): Intraesophageal Pressures during Phonation in Laryngectomized Patients. *J. Laryngology and Otology*, LXXXVI, No. 2, p. 138.