

PITCH IN ESOPHAGEAL SPEECH

L.W. LANHAM, Ph.D. (WITWATERSRAND)

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

and

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

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

SUMMARY

Most reports on pitch in esophageal speech emphasize that it is low-pitched with a measured fundamental frequency rarely higher than 100 cps. Our investigations show, however, that much esophageal 'phonation' lacks periodicity and, therefore, a fundamental frequency (i.e. pitch in the accepted sense). An auditory impression of pitch modulation can, nevertheless, be created by physical properties other than a varying harmonic structure. Our sample includes a rare case of truly high-pitched esophageal phonation with a fundamental frequency in the upper limit of the voice an octave higher than the highest reported in the literature. High-pitched phonation apparently requires a vibratory source in a 'mode' different from that of low-pitched phonation and should therefore be distinguished from it in discussing pitch in esophageal voice.

OPSOMMING

Meeste verslae oor esofageale spraak beklemtoon dat dit 'n lae toonhoogte het met 'n gemete grondtoon frekwens wat selde hoër is as 100 c.p.s. Ons ondersoek toon egter aan dat in baie gevalle die esofageale fonasie nie periodies varieer nie, en dat daar dus nie sprake kan wees van 'n grondtoon (d.i. toonhoogte in die aanvaarde sin) nie. Die gehoorlike indruk van toonhoogte modulاسie kan nietemin ook veroorsaak word deur fisiese eienskappe ander as harmoniese veranderende struktuur. Die toonhoogte in een van die ondersoekte gevalle was uitsonderlik hoog, met 'n grondtoon in die boonste grensgebied van die stem, 'n oktaaf hoër as die hoogste tot dusver vermeld in literatuur. Hoë toonhoogte fonasie vereis blykbaar 'n vibrerende bron in 'n vibrasietoestand wat verskil van die van 'n lae toonhoogte fonasie, en moet dus onderskei word van 'n lae toonhoogte in die verhandeling van toonhoogte in esofageale spraak.

While the current literature provides a fair consensus as to the limitations of pitch in esophageal speech, certain anomalies and contradictions are nevertheless apparent. Our evidence suggests that this arises from a failure to recognize that a measurable fundamental frequency (the physical correlate of pitch perception in normal voice) is often lacking, or only intermittently present,

in the typical 'low-pitched' esophageal 'voice'. It is nevertheless necessary to account for the weak auditory impression of pitch variation in esophageal voice even without a measurable fundamental frequency. In distinguishing between pitchless 'phonation' and the esophageal voice with sustained, true pitch, we find in the data offered in the literature (in particular, Kytta⁷) and in our own sample, evidence of an upper limit to the pitch range in esophageal voice at least an octave higher than the highest frequency previously reported (see below). Such high pitch is only achievable by a small number of laryngectomees who have both a high-pitch and low-pitch 'mode'. The suggestion that there are two apparently disjoint modes in esophageal phonation leads us to investigate the possibility of two correspondingly different states of the crico-pharyngeus muscle in vibration.

METHODOLOGY

The Kay Sonagraph (Model 6061-B) was used to analyse the physical properties of esophageal voices in the manner described in Kerr and Lanham⁶. As indicated there, narrow-band spectrograms show harmonic structure in narrow lines on the horizontal axis (see, for example, spectrogram C2, narrow, 12-18 in this paper); wide-band spectrograms show each release burst in a sequence of vibratory cycles by vertical lines whose regularity and timing can be measured on the horizontal axis ($\pm 12.33 \text{ cm} = 1 \text{ second}$). Fundamental frequency is measured as the interval between successive lower harmonics using frequency-scale magnification for greater accuracy (possible error is in the region of $\pm 5 \text{ Hz}$). All of our spectrograms are cuts from the flow of speech in normal conversation or, in one case, from an attempt at singing. Three of our cases are presented spectrographically here. One of them (Case C) provided still radiographs drawn from extensive still and cineradiographic investigation of the site and mechanism of this case's two voices – 'low' and 'high-pitched' phonation. These will be found in Kerr and Lanham⁶, (p. 100.)

Case A (female, aged 61) underwent laryngectomy in 1953 when more than the usual segment of trachea was removed. She received speech therapy over a period of 6 months, but no great improvement was recorded. Case B (female, aged 61) underwent laryngectomy in 1950; the infrahyoid muscles were preserved and overlapped to strengthen pharyngeal repair. She received speech therapy shortly after the operation and was greatly helped by it. Case C (female, aged 62) underwent laryngectomy in 1951 with surgery similar to that of Case B. She received speech therapy after the operation, but believes that it was largely due to her own efforts that she 'found her voice'.

A COMPARISON OF FINDINGS IN STUDIES DEALING WITH PITCH IN ESOPHAGEAL SPEECH

1. Fundamental frequencies and pitch ranges.

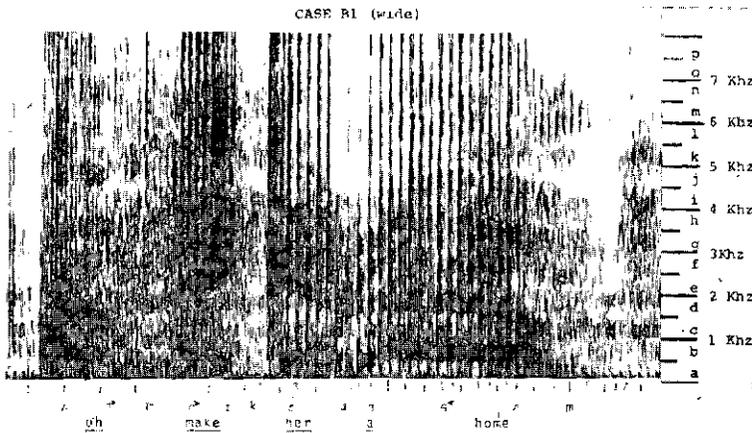
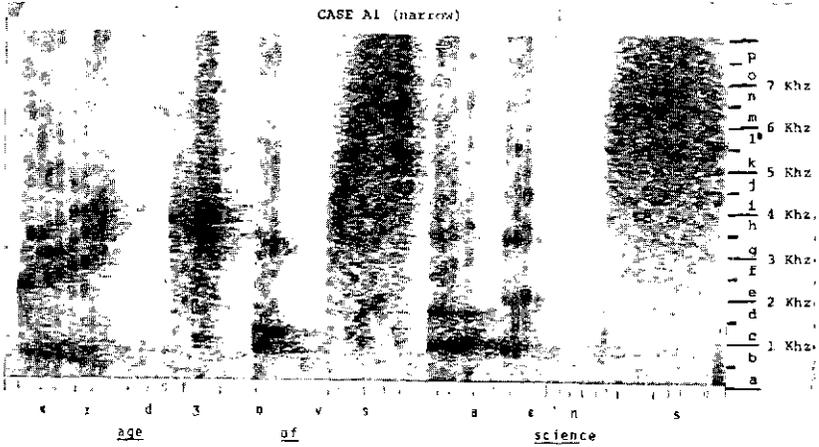
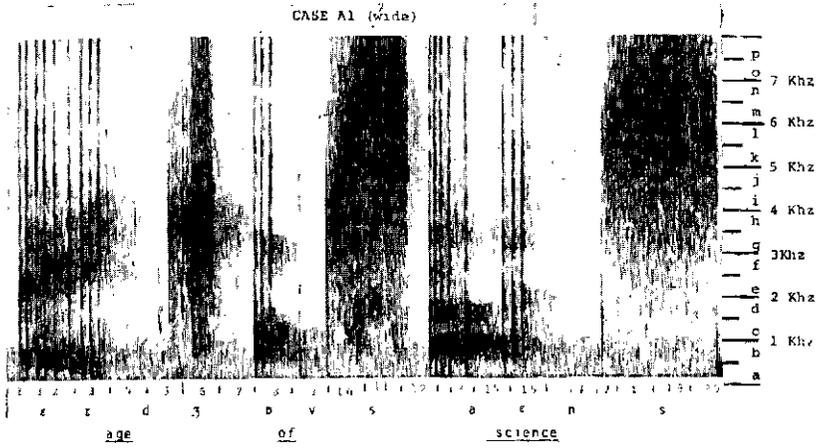
Reports on pitch in esophageal speech give measured fundamental frequencies and pitch ranges which are low in comparison with normal laryngeal speech (a full octave lower according to Snidecor and Curry¹⁰), but the extent of the pitch range may actually be greater than that of normal speakers – a differ-

ence of 13.21 tones against 10.5 tones according to these authors. Obvious anomalies do, however, arise in reports of this kind. After reporting the abnormally wide frequency range of superior esophageal speakers in their sample, Curry and Snidecor¹ continue: '... the esophageal speakers were nonetheless considered to have a restricted pitch range ...' and 'The frequency measured indicated a considerably greater movement than was apparent in pitch to the listener.' In citing a statement by Van den Berg: 'The speech of a clever patient sometimes gives the illusion of agreeable changes of pitch which objectively are not present' these authors do not fully explore the obvious contradiction. Kytta⁷ states: 'The auditory observation group did not in fact notice any appreciable frequency variation ... However, measurement of the fundamental frequency showed an unexpectedly large variation, 3-4 tones ...'. Anomaly and contradiction is, however, partly explained by the conclusions to which our studies have led us:

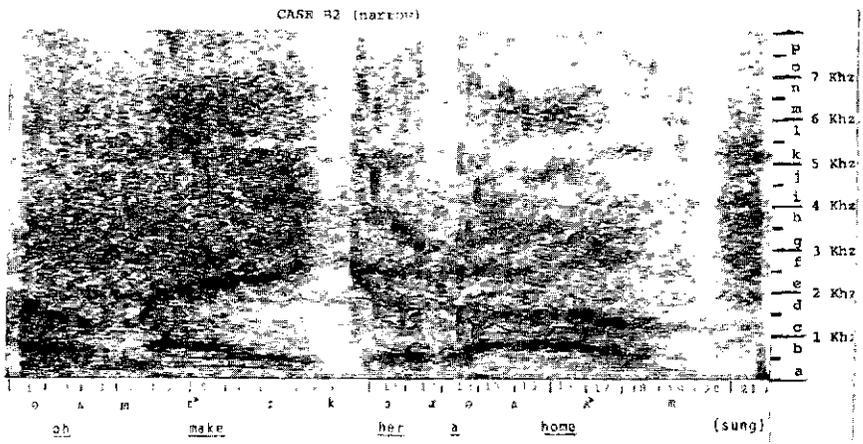
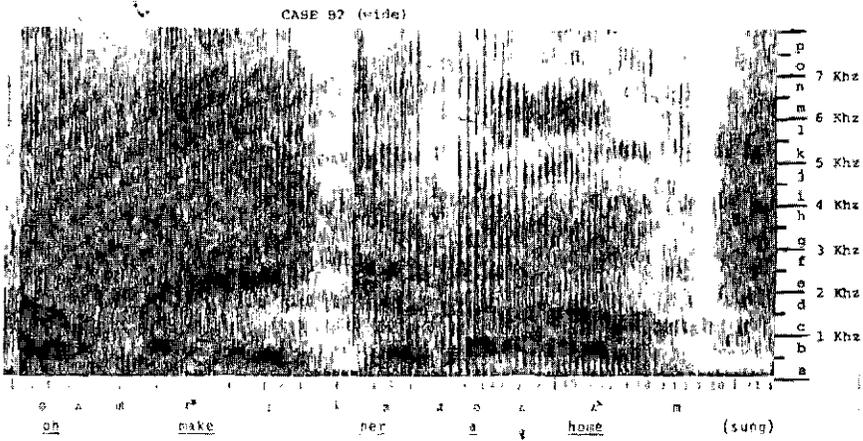
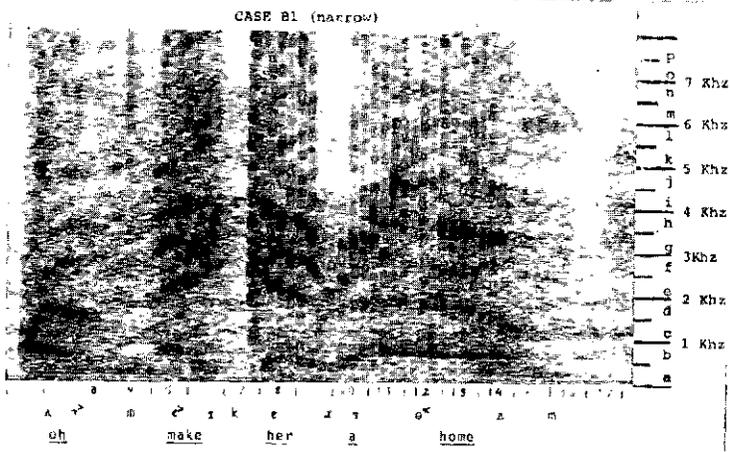
(a) Much 'low-pitched' esophageal phonation is actually pitchless, but does have measurable rates of vibration without periodicity (i.e. repeated cycles) and it is these aperiodic vibrations which are often measured. (b) The voices of many speakers have pitchless stretches interspersed with stretches in which periodicity is achieved and lower harmonics are discernible. (c) An auditory impression of pitch modulation can be achieved by varying prosodic properties other than a harmonic pattern. (d) The most effective pitch modulation is achieved by a small number of laryngectomees who can produce 'high-pitched' phonation with pitch ranges considerably higher than those reported in the literature.

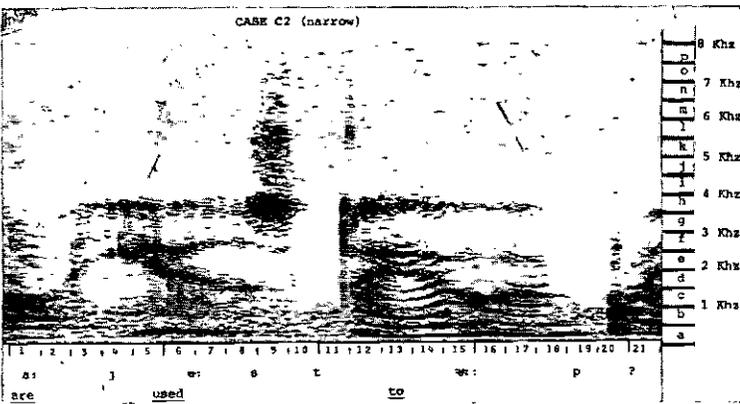
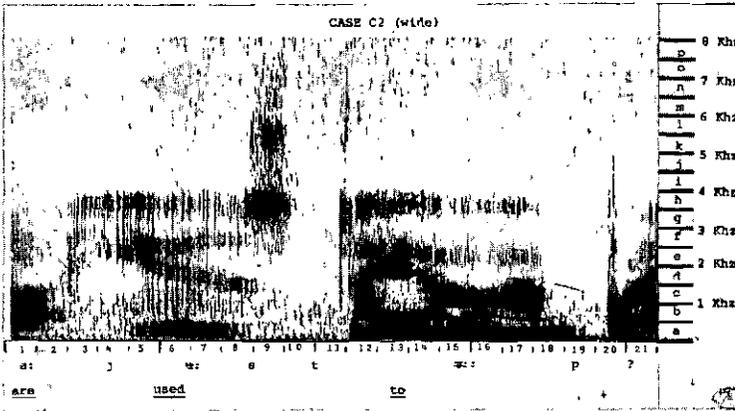
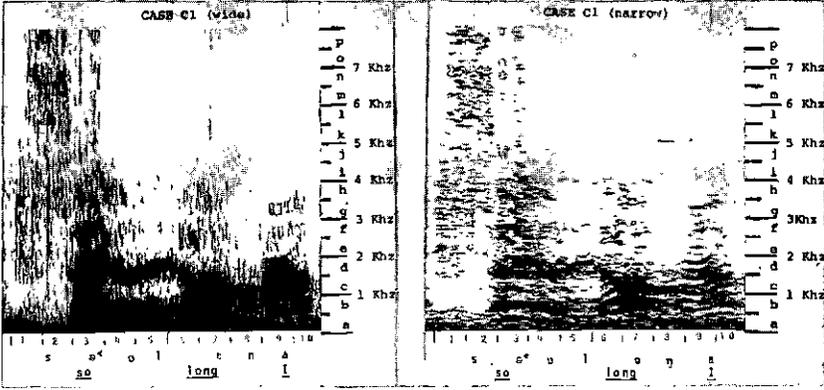
In this article the main focus is on high-pitched phonation which has properties sufficiently distinct to warrant separation from the common low-pitched esophageal phonation. These properties apparently correspond to a different state or mode of the vibratory source. The major differences are found in pitch range and in a sustained harmonic structure with most of the energy concentrated in harmonics rather than in concomitant aperiodic components. Spectrographic evidence of high-pitched phonation is found in the spectrograms below: C2 (narrow) and, less distinctly, C1 (narrow). These demonstrate a clear harmonic structure up to 2 KHz, a measurable fundamental varying between 215 and 317 and, therefore, pitch in the accepted sense of the term. (To our knowledge, the highest reported frequencies in esophageal speech are an estimate of 185 cps by Damste³ and a measured 135.5 cps by Curry and Snidecor¹.)

Our data on low-pitched esophageal voices show that the majority, including some highly intelligible ones, either totally lack a measurable fundamental frequency, or intermittently, even spasmodically, present short stretches of a syllable or less with a few lower harmonics and considerable aperiodic energy between discrete harmonic energies. A minority are capable of sustaining a discernible harmonic structure over several syllables and all these vibrations occur at rates which are within the 'pitch ranges' reported for esophageal speech, e.g.: 32 to 72 in Kytta⁷, 17 to 135 in Curry and Snidecor¹, 50 to 100 in Van den Berg and Moolenaar-Bijl¹³. The output of these slow rates of vibration of the pseudoglottis is, therefore, aperiodic more often than



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





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

periodic, but auditory impressions of pitch variation in esophageal speech are not necessarily dependent on periodicity; our evidence suggests that differences in intensity, duration, etc., can compensate in a limited way in giving an impression of pitch modulation (see our discussion below on Case B's attempt at singing).

Plomp⁸ reports that in normal laryngeal speech, pitch is determined by the lower harmonics; Fry⁵ states that it is the fundamental frequency which determines pitch, but the ear interprets the interval between lower harmonics as the pitch if there is no energy present in the first harmonic. As true pitch is a potential property of esophageal speech (particularly when it is 'high-pitched') we propose that esophageal phonation be recognized as pitchless where there is no evidence of lower harmonics and a fundamental frequency, or these are only spasmodically present over very short stretches.

2. Characteristics of low-pitched esophageal speech

In attempting to characterise low-pitched esophageal phonation we use Kytta's data and our own. Kytta's discussion reveals the basis for discrepancies between his interpretation of pitch properties and ours: In the spectrogram in Fig. 3, p. 31, labelled 'Spectrographic analysis of the fundamental frequency of the fundamental' we see a vibratory pattern producing releases which are bursts of noise irregular in time and amplitude, and varying in complexity. Our spectrogram B1 (narrow) shows a very similar pattern which we recognize as common in low-pitched esophageal speech (all Kytta's spectrograms on pp. 57-63 are of the same type). But Fig. 3, and our spectrogram B1 (narrow) above, both show a vibratory pattern lacking periodicity and devoid of harmonic structure, and a measurable fundamental; they do, however, show measurable rates of vibration and it is apparently this which Kytta measures in stating the limits of the pitch range of his subjects. Vibratory rates in this mode can be considerably faster (varying around 92 at B1 1-2), but do not necessarily acquire periodicity, nor a higher pitch because of this. Fast vibration in this mode (up to roughly 110 in our data) tends to lose release bursts in continuous turbulence (see B2, wide and narrow, 1-3 and Kytta's HIIPI 0-0.25 on p. 63). In some, possibly a minority of esophageal speakers phonating in this mode, discrete lower harmonics (usually below 1 KHz) do indeed emerge intermittently over short stretches — often shorter than a syllable, but a good deal of energy in random aperiodic components partly obscures them. (In our data these intermittent harmonics occur in the range of 60 to 130 cps.) In respect of the quantity of periodicity in low-pitched esophageal phonation, yielding a measurable first harmonic, we find it difficult to accept Curry and Snidecor's² finding that 59.5% out of 61.4% total phonation is periodic.

3. Characteristics of high-pitched esophageal phonation.

High-pitched esophageal phonation is characterized by sustained harmonic structure over more than one syllable; high concentration of energy in the lower harmonics as distinct from accompanying noise; and the pitch range, the lowest limit of which is probably in the region of 150 cps and the highest

close to 400 cps (the latter frequency is seen in spectrogram DVI in Kerr and Lanham⁶, p. 92). In addition to our Case C, Kytta provides spectrographic evidence of high-pitched esophageal phonation similar to Case C (but lacking the vibrato in the voice of the latter) on p. 50 (the utterance VAARA) and p. 54 (TLOIKO). Fundamental frequency in the former is roughly 170 cps at 0.25 and 162 cps at the peak of the first syllable in the latter. Kytta, however, fails to interpret his data correctly and gives the highest recorded fundamental frequency in his sample as 72 cps.

Case C has a pitch range between 210 cps and nearly 400 cps; the fundamental frequency in spectrogram C2 is 270 cps at 1.3 and 317 cps at 16.4. (Notice, however, the break into low-pitched phonation at C2 3-8). Another possible feature of high-pitched phonation is the significantly longer 'duration of air charge'. Snidecor and Curry¹⁰, reporting on mean duration of air charge in their sample, give a maximum of 2.25 secs. Case C can repeat 'all my arms are sore'¹ for approximately 4.5 secs. In tape-recorded conversation her longest stretch without obviously recharging her esophagus is 3.7 secs.

AN ANALYSIS OF THE THREE VOICES

In briefly reviewing our three cases presented here we note that radiography locates the vibratory source at the cricopharyngeus for all. (The postulation of an upper pharyngeal vibrator for Case C's high-pitched phonation made in Kerr and Lanham⁶, has been refuted by recent cineradiography.) The voices of Cases A and B are confined to low-pitched phonation with no evidence of discernible harmonics in the voice of the former and virtually none in the latter.

Case A's voice is croaky, low in power and intelligibility and gives no auditory impression of pitch modulation. There is believed to be some loss of muscular function in the inferior pharyngeal constrictor. Spectrogram A1 shows very slow, highly irregular rates of vibration; at A1 1-3 there are 9 vibrations over some 17 csecs at rates ranging between 30 and 61. Case B on the other hand has a highly efficient voice of considerable power and high intelligibility although unpleasantly harsh. Vibrations are faster and more regular; a variation of ± 11 around a mean rate of 92 at B1 1-2 and ± 2 around a mean rate of 45 at B1 11-12.

Case B's voice is capable of weak impressions of pitch modulation and the sung and spoken versions of the same utterance are shown in spectrograms B1 and B2 respectively as evidence of the physical properties of 'pitch' in aperiodic low-pitched esophageal speech. The sung version differs in the following respects: (a) Where the highest pitch is attempted (*make* at B2 5-8) the syllable duration is more than doubled. (b) The rate of vibration increases by nearly 40% at B2 5-8 and approximately 30% at B2 11 (the vowel nucleus of *her*). (c) The intensity of the prominent syllables is greater by 3.3 db for the vowel nucleus of *make* and 1.3 for the nucleus of *home*. (d) More energy

¹This utterance was chosen because it lacks stop (plosive) consonants. The articulation of these consonants has been shown by Van den Berg, et al.¹² to present opportunities for filling the esophagus.

is located in higher frequencies (above 3 KHz); an impression of this difference can be gained at B2 1-3. Apart from (b), which does not create discernible harmonics to any extent, the other differences, probably collectively, contribute to the auditory impression of higher pitch.

Case C has two 'voices' (low and high-pitched esophageal phonation) and spectrograms C2 are produced in order to show how they alternate in the continuum of speech and reveal the basis of the strong impression of pitch modulation which, over the telephone, can deceive even the most experienced ear as to the true nature of the voice. As seen at C2 3-8, where the second, high-pitched voice breaks and the first voice takes over, Case C's first voice gives no evidence of a harmonic structure. The first voice is heard as low-pitched and it is the alternation between the two voices which contributes to the strong impression of wide pitch variation. Case C reports that her second voice is the product of strenuous, sustained effort in the months after laryngectomy and obviously requires a much higher degree of neuromuscular control; the vibrato effect best seen in C1 (narrow), is evidence of this. With declining morale in recent years Case C tends to lapse into her first voice for longer and longer stretches.

The general impression of Case C's voice is that it is a high-pitched woman's voice (spectrographically at C2 13-17 it is strikingly similar to female laryngeal phonation), but extreme, often abrupt, variation in power is a feature. Pitch change is often sharp, sudden and somewhat unpredictable. The abrupt changes usually correspond to a change from one voice to another; spectrogram C2 shows this with C2 at 1 and 12-21 being the second voice with a one-syllable break into the first voice at 3-8.

STATE AND FUNCTION OF THE ESOPHAGEAL SPHINCTER IN VIBRATION

The site and mechanism of the two modes of phonation in Case C's voice may be examined by reference to radiographs on p. 100 in Kerr and Lanham⁶. In that article they are numbered D1 - D4. D2 and D3 represent phonation in the first (low-pitched) and second (high-pitched) modes respectively, D4 outlines the cricopharyngeus immediately after phonation has ceased. In D3 the shadow over the lower three-fourths of a club-shaped head of the cricopharyngeus outlines a channel of air which rises and tapers abruptly to a point of occlusion in the upper quarter of the clubhead. The constriction ring within the esophageal sphincter is apparently located here. In D2 the constriction ring is less easily identifiable, but an air channel is seen above the protruding fold (marked a) and forward of the lower third of the inferior anterior edge of the clubhead. Blurring at the upper half of this anterior edge suggests a comparatively large mass of the inner margin of the sphincter involved in vibration, which is not the case in D3. This latter state is consistent with high-frequency vibration in which only a short stretch of a tightly constricted inner margin is involved.

In an attempt at simulating the second mode of vibration in the form of a labial trill, L.W.L. produced, in the manner described below, a spectrogram

fairly similar to C1 (narrow) 6-10, which showed a fundamental around a mean of 400 cps with undulating harmonics. It was necessary to make a very tight labial closure with strong intra-oral pressure bursting through a very short stretch (about .5 cm) which produced a high-frequency labial trill. Essential requirements are that each lip present a stiff, tense, edge which is relatively thin. With more flaccid, thicker edges labial trilling becomes slower and less regular, a longer stretch is involved in vibration and periodicity is easily lost. For high-frequency trilling the level of effort required to achieve the balance between air pressure, muscle tension and the configuration of the vibrating edge appears to match that required for Case C's second voice and is as difficult to maintain at an even level.

In comparison with radiograph D4 (Kerr and Lanham⁶) in which there is no vibration, phonation in D2 and D3 is seen to involve a shortening of the 'neck' behind the clubhead of the cricopharyngeus and a lifting and flattening of the head (contrast the drooping head in D4). The head is flattened in the vertical plane more in D3 than in D2 suggesting tighter constriction.

Taking account of the differences shown in the radiographs referred to above, we suggest that first-mode and second-mode vibrations differ in mechanism in the following ways: In the former a relatively thick, fairly relaxed inner margin of the sphincter is drawn into closure or close approximation in a muscular movement not much different from that required for esophageal burping in a normal speaker; consequent vibrations involve a comparatively large mass of the cricopharyngeus. Second-mode vibration, however, requires a much more highly controlled movement which produces greater tension along the inner margin and presents a stiffer, firmer edge at the point where vibration takes place. Much less of the margin is involved in high-frequency, low amplitude vibration with only small quantities of esophageal air released in the process. This could account for the significantly longer 'duration of air charge' in this type of phonation. The distended upper end of the esophagus suggests high intra-esophageal pressure.

CONCLUSION

In the discussion above it has not been our intention to present a case for two mutually exclusive categories of esophageal phonation, but to highlight a type of esophageal phonation of which only a limited number of laryngectomees is capable. This 'second mode' of pseudoglottal vibration is the product of a relatively intact musculature and considerable effort both in acquiring and maintaining it. We suggest, however, that in one or more of the parameters of esophageal phonation the two modes are discontinuous states and not merely different points on the same scales. There is, for example, no evidence that over any 'voiced' stretch in esophageal speech a gradational move from one to the other mode is possible (by progressively increasing muscle tension along the margin of the esophageal sphincter). Spectrogram C2 (wide) shows how Case C moves from one mode to another by a discrete abrupt change; in fact,

²Van den Berg's 'high pitch' is not equatable to our second mode in esophageal phonation. His upper limit of the fundamental frequency is 85 cps and is said to coincide with high intensity and stronger air flow.

we have no evidence at all of harmonic energies in Case C's first, low-pitched voice and, therefore, no evidence of an ability to shift gradually from low to high pitch. Other investigators have suggested that one 'setting' of the pseudo-glottis in phonation cannot be phased gradually into another. Van den Berg et al.¹², state: '... he always separated the section with a high pitch² from that with a low pitch by a new breath and a new injection of air . . .'. There is evidence, therefore, that for stretches of, possibly, the duration of an air charge, the esophageal speaker is locked into a particular mode of phonation and, if a harmonic structure is present, into maintaining more or less the same pitch.

REFERENCES

1. Curry, E.T. and Snidecor, J.C. (1961): Physical measurement and pitch perception in esophageal speech. *The Laryngoscope*. LXXI: 415-424.
2. Curry, E.T. and Snidecor, J.C. (1973): Fundamental frequency characteristics of Japanese Asai speakers. *The Laryngoscope*. LXXIII: 1759-1763.
3. Damste, P.H. (1958): *Oesophageal Speech after Laryngectomy*. Thesis, print. Gebr. Hoitsema. Groningen.
4. Damste, P.H. (1959): The glotto-pharyngeal press. *Speech Pathology and Therapy*. 2:70-76.
5. Fry, D.B. (1970): Prosodic phenomena. In Malmberg, B. ed., *Manual of Phonetics*. Amsterdam.
6. Kerr, W.A. and Lanham, L.W. (1973): Anatomical and spectrographic analysis of the voice in disease: a report on five cases. *J. SA Speech & Hear. Association*. 20:81-107.
7. Kytta, J. (1964): Finnish oesophageal speech after laryngectomy. *Acta Oto-Laryngologica*, Supplementum 195. Vammala. Stockholm.
8. Plomp, R. (1966): *Experiments on tone perception*. Van Gorcum and Comp. Netherlands.
9. Robe, E.Y., Moore, G.P., Andrews, A.H., Holinger, P.H. (1956): A study of the role of certain factors in the development of speech after laryngectomy. *The Laryngoscope*. LXVI: 382-401.
10. Snidecor, J.C. and Curry, E.T. (1959): Temporal and pitch aspects of superior esophageal speech. *The Annals of Otology, Rhinology and Laryngology*. LXVIII: 623-636.
11. Tato, J.M. (1954): Study of the sonospectrographic characteristics of the voice in laryngectomized patients. *Acta Oto-Laryngologica*. 44: fasc. 5-6.
12. Van den Berg, J.W., Moolenaar-Bijl, A.J., Damste, P.H. (1958): Oesophageal speech. *Folia Phoniatrica*. 10:65-81.
13. Van den Berg, J.W. and Moolenaar-Bijl, A.J. (1959): Cricopharyngeal sphincter, pitch, intensity and fluency in oesophageal speech. *Prac. Oto-Rhino-Laryngol*. 21:4.

MODERN HEARING AIDS (Pty.) Limited

305 Rand Central 305 22-4855
165 Jeppe Street/Straat 165
Johannesburg 2001
Box/Bus 83, Johannesburg 2000

THE LATEST FROM TELEX

The TW2/TDR2 FM Auditory Training System:

Featuring:

Teachers microphone and transmitter all in the size of a hearing aid.

Student direct receiver for the following modes of operation: —

- 1) FM and Hearing Aid — with a control for adjusting FM relative to hearing aid sound pick-up.
- 2) Telephone or Loop pick-up alone.
- 3) Hearing Aid alone.

Peak output, compression and tone controls are also provided.

Uses a dry battery or rechargeable cell.

Automatic built in squelch provided.

Operating frequency changed by plug in module allowing free movement from one classroom to another.

No extra wires required — no cross talk between classrooms — 32 channels available.

Our services also include:

- 1) The supplying of the latest *DIRECTIONAL DISCRIMINATING* hearing aids for mid-medium-severe hearing losses.
- 2) Making the most comfortable instant earmoulds for our patients — while they wait.
- 3) First service facilities covered by a 3 month guarantee.

Other products include:

- 1) Screening and Diagnostic Audiometers.
- 2) Sound measuring instruments.