Proceedings of the Institute of Acoustics

ANALYSIS OF FUNDAMENTAL FREQUENCY AND TEMPORAL CHARACTERISTICS OF NEOGLOTTAL, ESOPHAGEAL AND NORMAL SPEECH

W.A.Ainsworth (1) and W.Singh (2)

(1) Department of Communication and Neuroscience, University of Keele, Keele, Staffordshire, ST5 5BG
(2) ENT Department, St. John's Hospital at Howden, Howden Road, Livingston, West Lothian, EH54 6PP

1. INTRODUCTION

After laryngectomy it is often possible for speech to be produced, but the production mechanism will obviously be different from that of normal speech. There are a number of mechanisms which might be employed depending upon whether total or partial laryngectomy was performed and whether any protheses are employed to assist speech production. In this investigation some of the acoustic characteristics of two types of alaryngeal speech, esophageal and neoglottal, have been analysed and compared with those of normal speech.

In esophageal speech the speaker uses air from the esophagus to vibrate the pharyngo-esophageal segment at the junction of the esophagus and the lower pharynx [1]. The amount of air per breath is thus quite limited compared with normal speech and precise control of the vibration is difficult. In neoglottal speech the parts of the larynx which are preserved in near-total laryngectomy are used as the voice source [2]. Again control of the vibration is limited, but the lungs are filled via a valve in the trachea, so the airstream is more normal.

It has recently been found that neoglottal speech is rated nearer to normal speech than esophageal speech in terms of intelligibility, normality, fluency and rate of speaking [3]. It is therefore of interest to determine the acoustic characteristics responsible for these perceptual judgements.

2.DATABASE

Recordings were made of 44 speakers reading the Rainbow Passage [4]. Both the acoustic signal and the laryngographic signal [5] were recorded. The passage was read twice by 15 neoglottal speakers (14 male and 1 female), 12 esophageal speakers (8 male and 4 female) and 11 normal speakers (6 male and 5 female). Recordings were also made of 5 esophageal patients (all male) who had not yet

acquired speech and one tracheo-esophageal speaker, but these have not been included in the present analysis. Most of the speakers came from the same geographic region (Edinburgh). An attempt was made to match the age range of the normal speakers to those of the alaryngeal speakers.

3. FUNDAMENTAL FREQUENCY CHARACTERISTICS

The mean, mode, median (and their standard deviations), and the 80% and 90% ranges of the fundamental frequency were computed for each speaker from the laryngographic signals. The averages of these for each class of speaker are shown in Table 1.

Table 1. Means and standard deviations of various first order statistics of fundamental frequency (Hz) for normal, neoglottal and esophageal male and female speakers.

	Normal		Neoglottal		Esophageal	
	Male	Female	Male	Female	Male	Female
Number	6	5	14	1	8	4
Mean	105.2	184.2	141.9	175.6	131.4	122.6
S.D.	15.0	19.1	33.1	-	21.4	20.6
Mode	103.6	184.6	150.7	264.8	122.7	133.6
S.D.	17.4	8.9	49.3	-	41.8	47.0
Median	100.9	189.9	146.0	201.3	125.9	127.0
S.D.	13.7	12.5	40.2	-	29.0	36.1
80% range	73.7	101.4	259.0	267.8	364.3	238.8
90% range	171.1	155.8	428.3	348.3	550.5	439.1

It can be seen that whereas the normal male speakers have an average mean fundamental of 105.2 Hz (88.5 Hz for the lowest normal male speaker to 129.9 Hz for the highest) which is much lower than the 184.2 Hz of normal the female speakers (157.3 Hz to 206.9 Hz), the average mean fundamentals of alaryngeal male and female speakers are much closer. For male neoglottal speakers it was 141.9 Hz (101.5 Hz to 201.3 Hz) compared with 175.6 for the one female neoglottal speaker, and for male esophageal speakers it was 131.4 Hz (101.5 Hz to 161.7 Hz), actually higher than the 122.6 Hz (98.7 to 144.9 Hz) for the female esophageal speakers.

In order to investigate the statistical significance of these differences, t-tests were carried out. They showed that the average mean fundamental frequency of normal the male speakers was significantly different from the normal female speakers and from all the alaryngeal speakers except the female esophageal speakers at the 5% level, and that of normal female speakers was significantly different from all other types of speakers at that level. However, the average mean fundamental frequency of neoglottal male speakers was not significantly different from the

other alaryngeal speakers, and that of the neoglottal female speaker was not significantly different from the neoglottal male speakers. Also the average mean fundamental frequency of the esophageal male speakers was significantly different from that of the other alaryngeal speakers, and that of the esophageal female speakers was not significantly different from the esophageal male speakers.

A histogram showing the distribution of mean fundamental frequency for the male speakers is shown in Figure 1. It can be seen that the fundamental of the alaryngeal speakers is generally higher than that of normal speakers, but there is an overlap of the ranges so that some of the alaryngeal speakers fall into the normal range. There does not seem to be any difference in the ranges of the neoglottal and esophageal speakers.

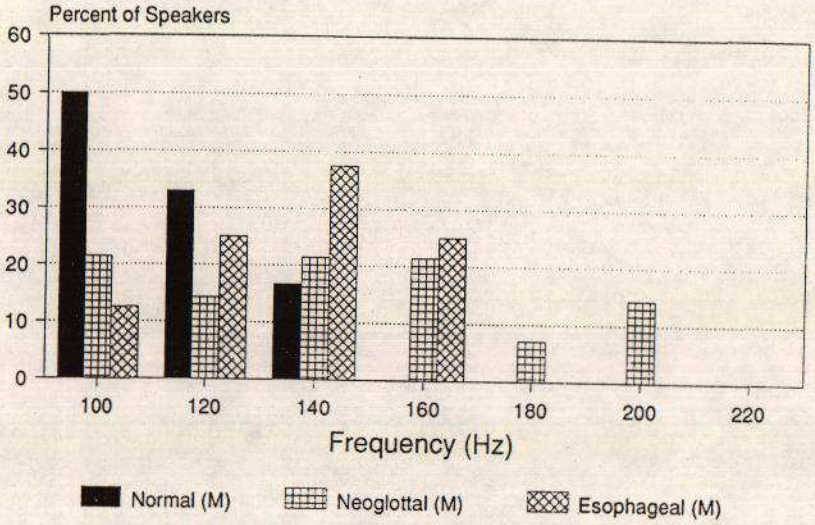

Figure 1. Mean fundamental frequency for normal, neoglottal and esophageal male speakers.

An equivalent histogram for the female speakers is shown in Figure 2. The single neoglottal female speaker falls into the normal range, but the esophageal female speakers mostly have lower

fundamentals than the normal female speakers.

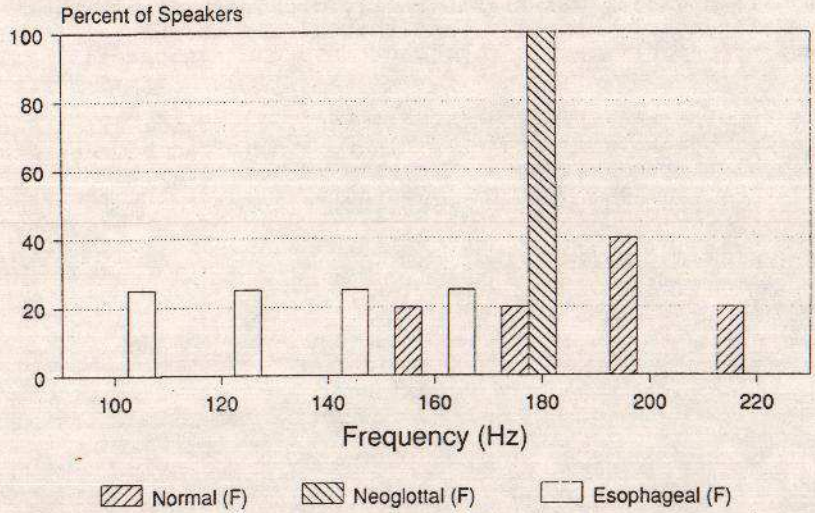

Figure 2. Mean fundamental frequency for normal, neoglottal and esophageal female speakers.

Table 1 also shows the average 80% and 90% ranges of the fundamental frequency (80% and 90% respectively of the glottal periods fall within these ranges). These are about two or three times greater for the alayngeal speakers than for the normal speakers indicating a rather broad distribution.

On the whole the fundamental frequency statistics suggest that alaryngeal speech will sound different from normal speech, but they do not suggest that neoglottal speech should sound more normal than esophageal speech.

4. TEMPORAL CHARACTERISTICS

Besides fundamental frequency the other gross determinants of speech quality are the temporal characteristics such as duration and syllable rate. Consequently a number of temporal characteristics which can easily be measured from the recordings were chosen. These were (a) duration of the entire passage, (b) number of syllables per second, (c) total number of breaths to read the passage, and (d) the average number of syllables uttered per breath.

ANALYSIS OF FREQUENCY AND TEMPORAL CHARACTERISTICS

The duration of the passage was timed for each speaker, the number of syllables in the passage was counted, and the number of breaths each speaker drew to read the passage was estimated from the recording. From these the above measures were calculated. The mean values for each type of speaker are shown in Table 3.

Table 3. The average duration (seconds), number of syllables per second, number of breaths, and the number of syllables per breath (and their standard deviations) for normal, neoglottal and esophageal male and female speakers to read the passage. Each speaker read the passage twice.

	Normal		Neoglottal		Esophageal	
	Male	Female	Male	Female	Male	Female
Number	12	10	28	2	16	8
Duration	37.4	33.8	43.3	45.9	73.2	55.3
S.D.	5.1	2.4	7.6	2.7	27.3	6.7
Syllables/sec	3.46	3.82	3.05	2.80	1.97	2.35
S.D.	0.49	0.28	0.54	0.14	0.64	0.28
No. of breaths	10.4	9.6	11.7	12.5	61.3	33.6
S.D.	3.7	1.5	4.5	0.7	37.1	6.3
Syll./breath	13.9	13.7	12.7	10.2	3.0	3.9
S.D.	5.4	2.3	4.9	0.6	1.9	0.7

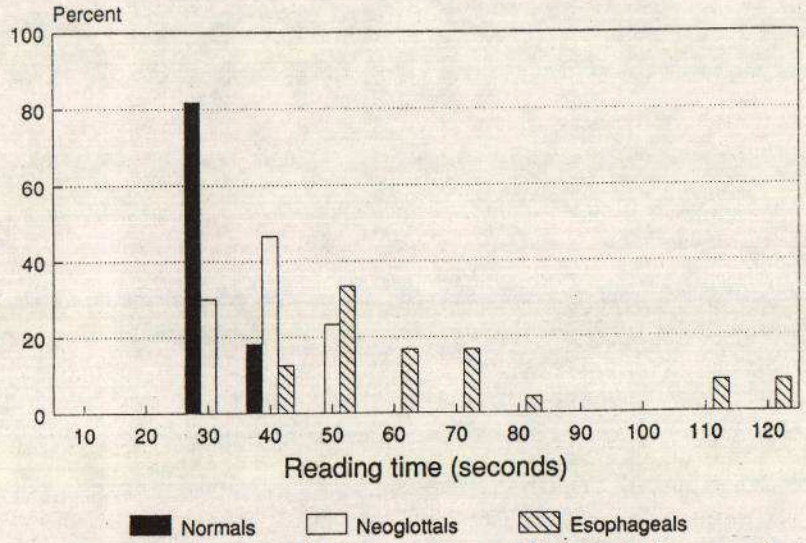

Figure 3. Duration of the passage read by normal, neoglottal and esophageal speakers

ANALYSIS OF FREQUENCY AND TEMPORAL CHARACTERISTICS

It can be seen that there were considerable differences in the time it took the various types of speaker to read the passage. The normal speakers read it fastest, the neoglottals a little slower and the esophageals considerably slower. Histograms of reading time are shown in Figure 3. This shows that there was a good deal of variation amongst the esophageal speakers.

Table 3 shows the same pattern for the number of syllables uttered per second. The normals spoke at an average rate of 3.62 syllables/sec, the neoglottals at 3.03 syllables/sec, and the esophageals at 2.10 syllables/ sec. Histograms of the number of syllables uttered per second are shown in Figure 4.

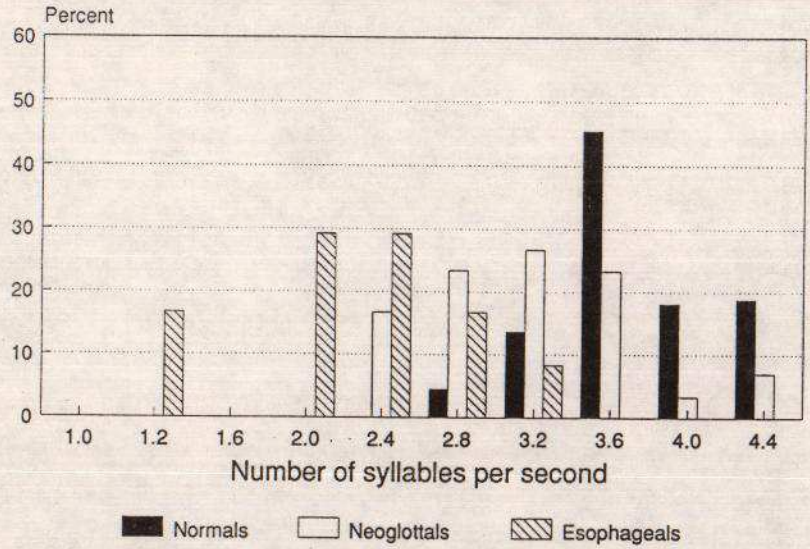

Figure 4. Number of syllables per second uttered by normal, neoglottal and esophageal speakers.

The differences between the different types of alaryngeal speakers are shown to an even greater extent when the number of breaths taken to read the passage are examined. The normal speakers on average took 10 breaths to read the passage, the neoglottals took 11.8 breaths, but the esophageals took 52.1 breaths. The distribution is shown in Figure 5. There was a wide variation in the number of breaths taken by the esophageal speakers.

A large difference between neoglottal and esophageal speakers is also shown in the number of syllables uttered per breath. The esophageals only managed 3.3, whereas the neoglottals produced

12.5, only slightly less than the normals' value of 13.8. A histogram of the number of syllables per breath is shown in Figure 6. This shows a large overlap for the normal and neoglottal speakers, but not much overlap for these speakers and the esophageals.

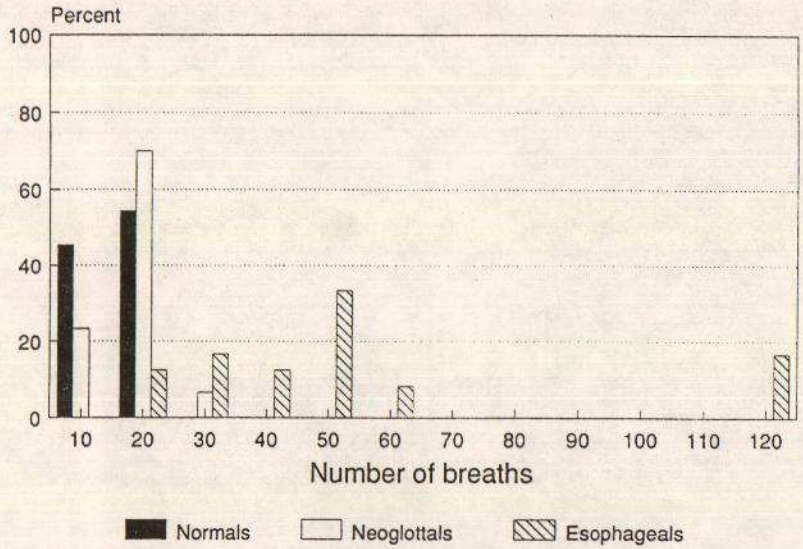

Figure 5. Average number of breaths to read the passage by normal, neoglottal and esophageal speakers.

5. CONCLUSIONS

An analysis of the fundamental frequency and temporal characteristics of normal, neoglottal and esophageal speech has been made. It has been found that the fundamental frequency of alaryngeal speech is generally higher than that of normal male speech but lower than that of normal female speech. There was little difference between the fundamental frequency of neoglottal and esophageal speech.

An analysis of the temporal characteristics, however, revealed differences between neoglottal and esophageal speech. The duration and number of syllables per second of neoglottal speech was nearer to normal than that of esophageal speech. The number of breaths and the number of syllables per breath of neoglottal speech was quite similar to that of normal speech and very different from that of esophageal speech.

This suggests that the reason why neoglottal speech is judged to be nearer to normal than esophageal speech in terms of intelligibility, normality, fluency and speaking rate [3] is that the temporal characteristics rather than the frequency characteristics are more normal.

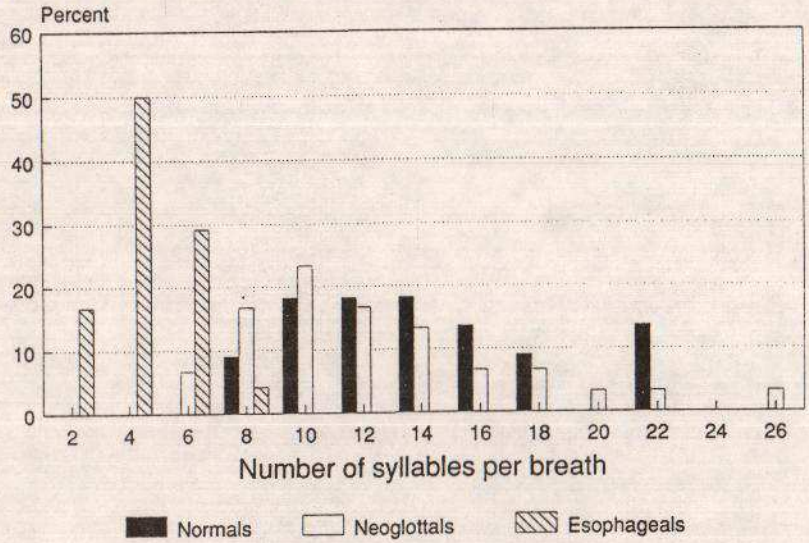

Figure 6. Average number of syllables per breath uttered by normal, neoglottal and esophageal speakers.

6. ACKNOWLEDGEMENTS

We are indebted to Gordon McKenna who made the recordings. The work was supported by the Lothian Health Board.

7. REFERENCES

[1] W M DIEDRICH 'The mechanism of esophageal speech', Ann NY Acad Sci, 155, 303-317 (1968)
[2] W SINGH and P HARDCASTLE 'Near-total laryngectomy with a myo-mucosal valved neoglottis', J Laryng Otol, 99, 581-588 (1985)
[3] W A AINSWORTH, G McKENNA and W SINGH 'Perceptual comparison of neoglottal, esophageal and normal speech' (in press)
[4] G FAIRBANKS 'Voice and Articulation Drillbook' 2nd edition, Harper and Brothers, New York (1960)
[5] A J FOURCIN and E ABBERTON 'First applications of a new laryngograph', Medical and Biological Illustration, 21, 172 (1971)