Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 21 (1997), 143-153

FORMANT FREQUENCIES OF DUTCH VOWELS IN TRACHEOESOPHAGEAL SPEECH*

Corina J. van As[1,2], *Annemieke M.A. van Ravesteijn*[1], *Florien J. Koopmans-van Beinum*[1], *Frans J.M. Hilgers*[2], *Louis C.W. Pols*[1]

[1]Institute of Phonetic Sciences, University of Amsterdam, Amsterdam, The Netherlands
[2]The Netherlands Cancer Institute/Antonie van Leeuwenhoek Hospital, Amsterdam, The Netherlands

Abstract

In the present study Dutch vowel formant characteristics of laryngectomized tracheoesophageal speakers are investigated. Both vowels in CV nonsense syllables (/a/, /i/, and /u/), and vowels of stressed syllables in read-aloud text are studied. It appeared that in the nonsense syllables the first formants were comparable to those of normal speakers, but that the second formants were higher than in normals for the /a/, and /i/, and lower than in normals for the /u/. In read-aloud text however a large portion of the formant frequencies F_1 and F_2 was significantly higher than the formant frequencies in normal speakers. It is thought that the higher formant frequencies can be explained by the fact that the vocal tract is shorter in laryngectomized speakers, since the neoglottis in these speakers has a higher position than the vocal folds in normal speakers.

1. Introduction

In a total laryngectomy the whole larynx, and thus the vocal folds are removed. The most widely used methods of restoring speech following total laryngectomy are tracheoesophageal (TE) and esophageal speech. Since the past two decades, TE speech has become the most preferred method of voice restoration. The disconnection of the upper and lower airways implies that, for esophageal speech, the air supply consists of injected volumes of air from the mouth into the esophagus. For TE-speech, a fistula is created between the trachea and the esophagus; this opening allows the insertion of a prosthesis, which acts as a one way valve through which pulmonary air can be directed into the esophagus. The main difference between esophageal speech and TE speech is therefore the air supply.
In a few earlier studies on vowel formants in alaryngeal speech, esophageal speech is studied. Sisty and Weinberg (1972) studied both male and female esophageal speech. For both groups of speakers the systematic changes in formant frequency were similar, and the mean formant frequencies of esophageal speakers were found

to be consistently higher than those of normal speakers. For men the average increases were 122 Hz for the first formant, and 325 Hz for the second formant. Comparable observations were made by Rollin (1962) who studied English vowels in esophageal speech, and Kytta (1964) who studied Finnish vowels in esophageal speech. Sisty and Weinberg (1972) state that the consistency of this finding across languages, sex, and vowels shows that removal of the larynx does alter vocal-cavity transmission characteristics. They conclude that differences in tongue position (Nichols, 1968), and mouth opening do not fully explain this effect (Stevens and House, 1952; Fant, 1960); but that a reduction in the effective length of the vocal tract may account for these changes in formant frequencies. Kytta (1964) found a higher first formant for all vowels except /u/, /o/, and /e/, and a higher second formant for all vowels. His explanation for these changes is that following removal of the larynx, after which the base of the tongue is directly connected with the esophagus, the vocal tract loses a portion of its most posterior resonance cavity, which becomes apparent as a rise of the mean frequency results for all formants studied. In cineradiographical studies that Kytta (1964) performed in laryngectomees (n = 6), the neoglottis could be defined at the level of the sixth cervical vertebra in four patients, between the fifth and sixth cervical vertebra in one patient, and between the fourth and the fifth cervical vertebra in one patient. He also found that the shape and function of the neoglottis were not affected by the articulation of the three most extremely shaped vowels /a/, /i/, and /u/.

Schilling and Binder (1926), and Beck (1931) studying German vowels in esophageal speech (n =1, and n =2, respectively), and Luchsinger (1952) studying Swiss esophageal vowels (n = 3) demonstrated in the cases investigated only little differences between vowel formant frequencies of normal and esophageal speakers. According to Damsté (1958) this is understandable since the buccopharyngeal cavity has changed little. Later however, Diedrich and Youngstrom (1966) who obtained cinefluorograms of a patient one day prior to and 20 months following surgery, demonstrated that the effective length of the vocal tract of this patient was reduced postoperatively.

In the present study, in contrast to the earlier mentioned studies, tracheoesophageal speech is studied. For this purpose Dutch vowels are studied. Since tracheo-esophageal puncture is nowadays the most widely used method of vocal rehabilitation, it is interesting to get insight in the differences that occur between vowel formant frequencies of tracheoesophageal and normal speech. To the best of our knowledge no comparative data are available yet.

2. Subjects and methods

2.1 Subjects

Subjects were 17 male tracheoesophageal speakers. All of them underwent a standard total laryngectomy, and used a Provox® voice prosthesis (Hilgers and Schouwenburg, 1990). Ages varied from 45 to 81 years, with a mean of 65 years. The time after surgery varied from 9 months to 11 years, mean time after surgery was 6 years. Stoma occlusion was normally performed with thumb or finger in 5 patients, with a Provox®

Stomafilter (Hilgers et al., 1996) in 10 patients, and with a Blom-Singer Adjustable Tracheostoma Valve (Blom et al., 1982) in 2 patients.

2.2 Speech Material

Speech material was taken from an earlier study in this patient group in which the influence of stoma occlusion was investigated (van As et al., in press). For each patient series of nonsense syllables (CV) and a read-aloud text were recorded twice, one recording was made while the stoma was occluded by finger, and one recording was made while the stoma was occluded by a Provox® Stomafilter. In the nonsense syllables the vowel was always an /a/, /i/, or /u/, the consonant differed (/p/, /b/, /t/, /d/, /f/, /v/, /s/, /z/, /k/, and /g/ were used). The text contained all Dutch vowels except /ø/, and was the same as used in a study of Dutch vowel frequencies by Koopmans-van Beinum (1980). Since it is not expected that stoma occlusion may affect vowel formant frequencies (the use of an extratracheal device does not affect the source-filter system), for each patient both texts and both series of nonsense syllables were used in the investigation. The vowel formant frequencies for both occlusion conditions are grouped together for the nonsense syllables as well as for the texts.

2.3 Methods

For the recordings as well as for the formant frequency analysis, the Computerised Speech Lab of Kay Elemetrics Corporation (Lincoln Park, NJ, USA) was used. Via the external module of the Speech Lab, the speech data are digitally stored on a DAT tape. The DAT recorder was a portable Sony TCD8 recorder. The speech data are stored with a sample frequency of 48000 Hz. The microphone was a head-set microphone (AKG-c410), which is standardly used with the Computerised Speech Lab (CSL). The mouth-to-microphone distance was 2,5 centimeters.

All vowels were auditorily and visually selected from the oscillogram and the audio signal. For the nonsense syllables only the vowels /a/, /i/, and /u/ were available from the speech material, each of the vowels was selected 5 times. From read-aloud text all Dutch vowels except the vowel /ø/, which was not used in the text, were available. From the other vowels in the read-aloud text 5 items were selected, except for the vowels /y/ and /Y/, which could only be used twice, and the vowel /u/ which could only be selected three times. The vowels that were used were the same for each subject, and selected on the basis of vowel environment criteria, i.e. stressed syllable, no nasal, and no /r/ or other surrounding consonants that could influence the formant frequencies.

For each vowel the first (F_1) and second (F_2) formant frequency were measured using FFT (Fast Fourier Transform) analysis, performed with CSL. The frame length used was 10 ms, a Hamming window was used for the analysis. Exact positions of the formant frequencies were determined visually at the energy peaks in the frequency spectrum. As a control formant frequencies of 10 /a/ vowels were also measured via the signal processing software package Praat (Boersma and Weenink, 1996).

2.4 Controls

The formant frequencies found in this study were compared to those of normal Dutch speakers from literature.
Control values for the /a/, /i/, and /u/ formant frequencies for our nonsense syllables were obtained from a study by Pols et al. (1973). They studied vowel formants in 50 male Dutch speakers. Their words were of the type /h/-/V/-/t/. Control values for the vowel formant frequency in our read aloud text were compared with the formant frequencies found in a study of Koopmans-van Beinum (1980), who used the same text. The formant frequencies found in stressed syllables in read aloud text of both an untrained speaker and a trained speaker were used. In the present study the ratios between the first and the second formant were determined for comparison, for this comparison control values were also taken from studies mentioned above.

2.5 Statistics

A t-test for one sample was used to investigate possible differences in formant values and ratios of the first and second formant between TE-speakers and the control values of normal speakers.

3. Results

Both first and second formant frequencies could be measured in 81% of the vowels, as a whole 93,1% of the formants could be measured. Failures were spread over all data, and not specifically in one speaker or vowel. In some cases vowels were too short, or were recorded too loud and then clipped, in other cases formants were 'missing' or fell together. In Figure 1(a) a spectrum of a vowel /a/ with clear formants is given, in Figure 1(b) a spectrum of a vowel /a/ is shown in which the formants also are clearly visible, but the second formant is absent, most probably it falls almost together with the first formant. These figures are drawn with the program Praat developed by Boersma and Weenink (1996). For 8 /a/ vowels spoken by tracheoesophageal speakers results of the programs CSL and Praat were compared. The CSL program only draws a spectral envelope, but the formants that are found are comparable with those found in the spectrum drawn by Praat, the averaged absolute difference between formant values for both programs was 16 Hz.
The two spectra in both figures below also give a clear indication of the large amount of noise in these tracheoesophageal voices. Only a few harmonics can be seen, the large amount of noise results in the absence of higher harmonics. These harmonics also show that the voice shown in Figure 1(a) contains more noise (fewer harmonics), than the voice shown in Figure 1(b).

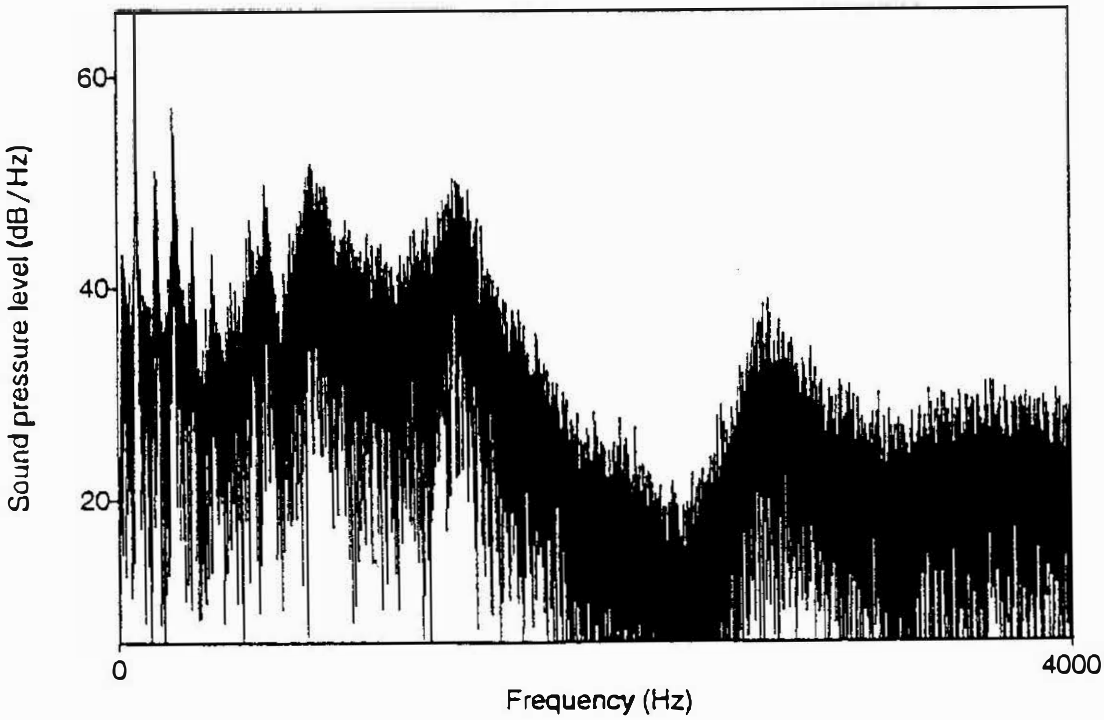

Figure 1(a). Spectrum of the vowel /a/, spoken by a tracheoesophageal speaker. Clear formants can be seen.

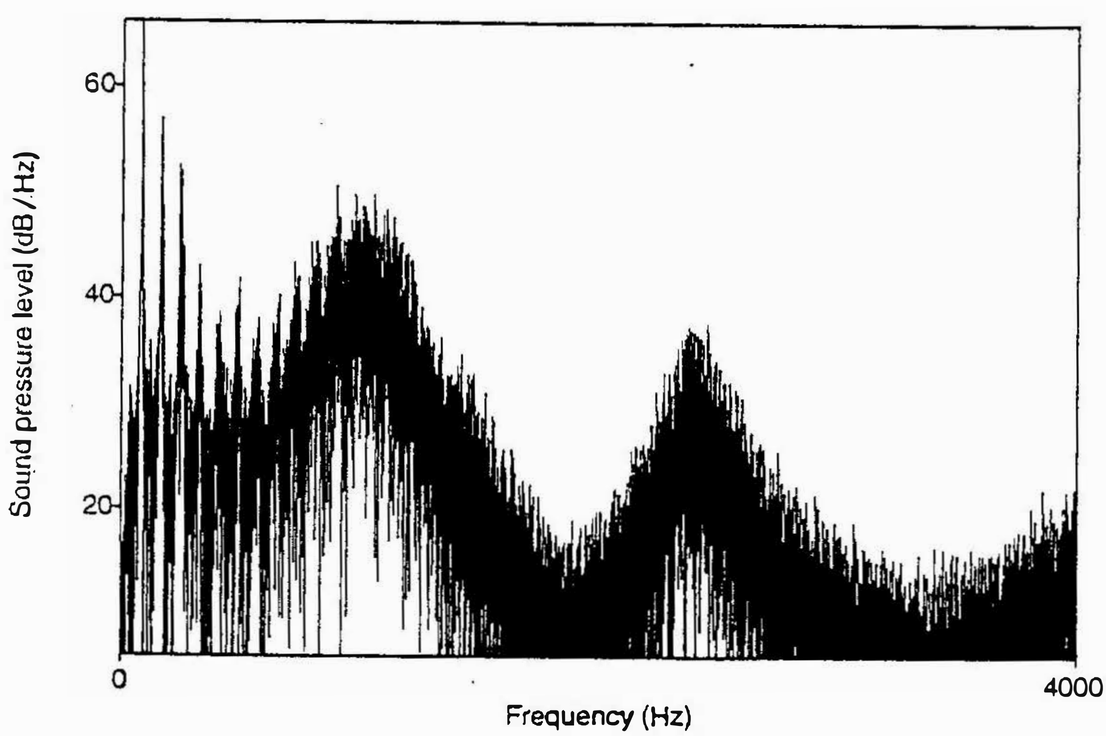

Figure 1(b). Spectrum of the vowel /a/ spoken by a tracheoesophageal speaker, the second format is absent, and most probably falls together with the first formant.

3.1. Tracheoesophageal speech versus normal speech

3.1.1 CV syllables

First the vowels /a/, /i/, and /u/ extracted from the nonsense syllables were studied. These vowels were chosen since they represent the most extreme articulation positions of the vocal tract. Comparison of the vowel formant frequencies found in the vowels /a/, /i/, and /u/ of the nonsense syllables with the vowel formant frequencies that were found by Pols et al. (1973) were performed by means of a Student's t-test in which formant frequencies of each separate tracheoesophageal speaker were compared with the mean values of the normal speakers. The t-test showed certain differences between both speaker groups. The first formant showed no significant differences for either of the three vowels, the second formant showed a significant difference for all three vowels: the second formant of the /u/ was found to be significantly lower than the normal frequency, the second formants of the /a/ and /i/ were significantly higher than the normal frequency.
Mean formant frequency values found for tracheoesophageal speech and normal speech are given in Table 1. In figure 2 a graphic representation of these differences is given.

Table 1. Mean formant frequencies of Dutch vowels in syllables for male tracheoesophageal speakers (n = 17) and in words for male normal speakers (n = 50). Also the t-value and probabilities are given.

Vowel		Normal (Hz)	Tracheoesophageal (Hz)	t-value	Probability
/a/	F1	795	772	-1.01	0.328
	F2	1301	1451	5.09	0.000
/i/	F1	294	306	1.03	0.316
	F2	2208	2675	5.71	0.000
/u/	F1	339	354	0.88	0.391
	F2	810	712	-7.09	0.000

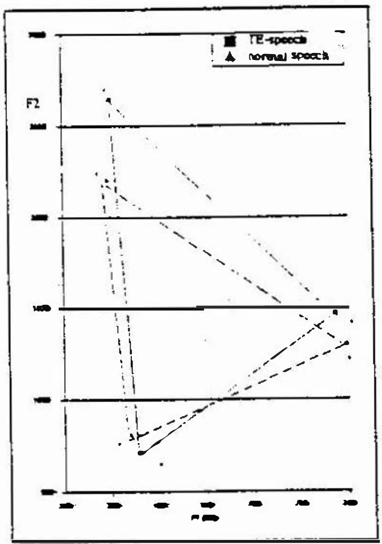

Figure 2. Plot of the vowel formant frequencies in nonsense syllables of normal male Dutch speakers (Pols et al. 1973, average value of 50 speakers) and male tracheoesophageal Dutch speakers (average value of 17 speakers, this study).

3.1.2 Read-aloud text

Regarding the first and second vowel formant frequencies of stressed syllables in read-aloud text, a Student's T-test between the control values of the trained and the untrained speaker, and the values found by the tracheoesophageal speakers showed significant differences between both speaker groups. In Table 2 the vowel formant frequencies are given, with t-value and probability. In Table 3 an overview is given of the differences that were found.

Table 2. Vowel formant frequencies of vowels of stressed syllables in read-aloud text of one trained and one untrained speaker (Koopmans-van Beinum, 1980), and the mean vowel formant frequencies of vowels of stressed syllables in read-aloud text of the tracheoesophageal speakers (n = 17). For both conditions t-value and probability (p) are given.

Vowel (IPA)		TE (n = 17)	Trained speaker	t-value	p	Untrained speaker	t-value	p
/u/	F1	315	342	-3.07	0.004	349	-3.87	0.000
	F2	857	827	1.50	0.143	948	-4.62	0.000
/o/	F1	468	421	4.42	0.000	427	3.68	0.001
	F2	925	877	2.43	0.021	886	1.98	0.056
/ɔ/	F1	494	440	6.67	0.000	446	5.92	0.000
	F2	1022	1035	-0.32	0.755	937	2.01	0.053
/ɑ/	F1	653	612	3.17	0.003	549	8.06	0.000
	F2	1279	1093	7.25	0.000	1096	26.56	0.000
/a/	F1	740	649	6.13	0.000	665	5.06	0.000
	F2	1407	1344	3.34	0.002	1268	7.35	0.000
/e/	F1	432	399	3.75	0.001	409	2.61	0.014
	F2	2161	2143	0.41	0.686	2034	2.84	0.008
/ɛ/	F1	540	590	-4.71	0.000	554	-1.92	0.205
	F2	1848	1780	2.04	0.049	1710	4.14	0.000
/i/	F1	234	322	-16.13	0.000	296	-11.38	0.000
	F2	2550	2450	1.97	0.057	2298	4.98	0.000
/ɪ/	F1	407	387	2.54	0.016	382	3.18	0.003
	F2	2068	2280	-5.49	0.000	2207	-3.59	0.001
/y/	F1	392	306	2.96	0.006	323	2.38	0.024
	F2	1657	1774	-1.85	0.074	1715	-0.92	0.367
/Y/	F1	440	424	1.50	0.144	407	3.06	0.004
	F2	1382	1509	-5.42	0.000	1434	-2.22	0.033

Table 3. Summary of all differences for first and second vowel formant frequencies in read-aloud text, between normal and tracheoesophageal (TE) Dutch speakers. Vowels are represented with IPA (International Phonetic Alphabet) symbols.

	tracheoesophageal speakers					
control speakers	lower F1	higher F1	no difference in F1	lower F2	higher F2	no difference in F2
untrained speaker	i, u	o, ɔ, ɑ, a, y, Y, I, e	ɛ	u, y, Y, I	ɑ, a, i, e, ɛ	o, ɔ
trained speaker	i, u, ɛ	o, ɔ, ɑ, a, y, I, e	Y	Y, I	ɑ, a, ɛ, o	u, ɔ, y, e, i
nonsense syllables (n = 50)			a, u, i	u	a, i	

A graphic representation of the formant frequency values found in text for the normal trained speaker and the normal untrained speaker and tracheoesophageal speakers (n = 17) is given in Figure 3.

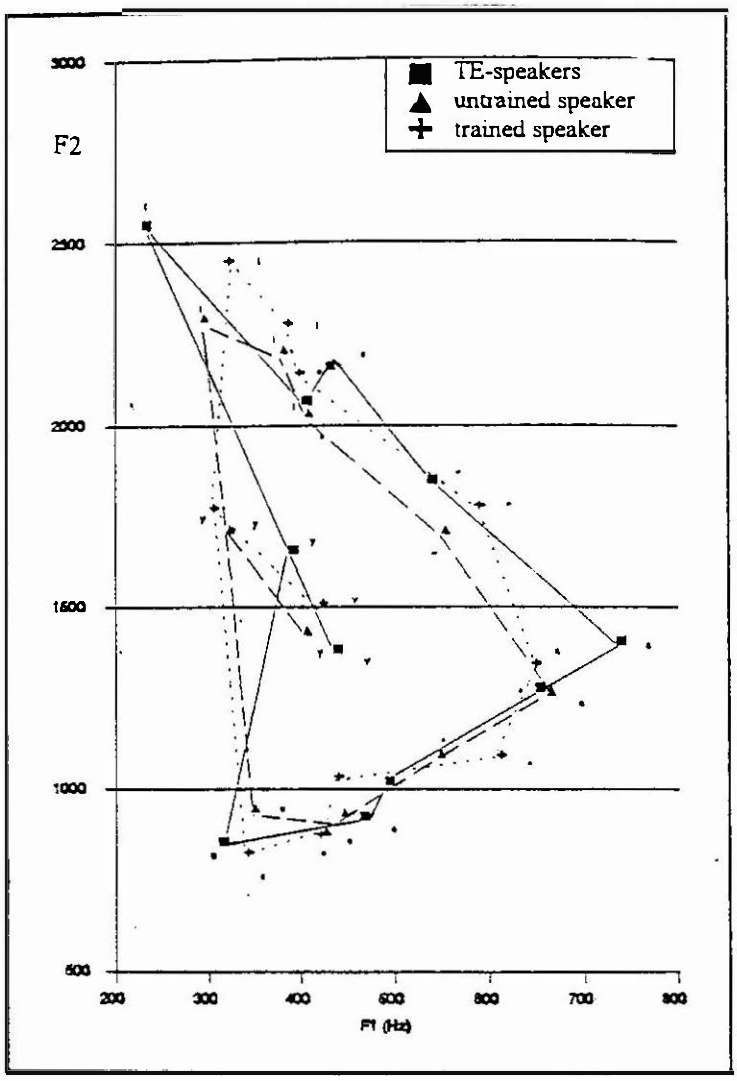

Figure 3. Plot of the vowel formant frequencies in stressed syllables of read aloud text of a male untrained Dutch speaker (Koopmans- van Beinum, 1980), a male trained Dutch speaker (Koopmans-van Beinum, 1980), and tracheoesophageal Dutch speakers (average value of 17 speakers).

3.2. Ratios between F1 and F2.

Each separate mean vowel ratio that was found for each TE-speaker was compared to the control value of the normal speaker group by means of a Students t-test for one sample. In Table 4 the mean ratios of the control groups and the TE speakers are given, both for syllables and text. Also the vowels for which the ratios were significantly different, are indicated.

Table 4. Averaged ratios between first and second formant frequencies, of normal speakers in words, of TE-speakers in nonsense syllables, of an untrained and a trained speaker in stressed syllables of read-aloud text, and of TE-speakers in stressed syllables or read-aloud text.

Vowel (IPA)	Words-normal	Syllables TE	Text Untrained	Text Trained	Text-TE
/u/	2.39	2.10*	2.42	2.72	2.92[1]
/o/			2.08	2.07	2.12
/ɔ/			2.35	2.10	2.27
/ɑ/			1.79	1.99	2.02[1]
/a/	1.64	1.88*	2.07	1.91	1.95[1]
/e/			5.37	4.97	5.13
/ɛ/			3.02	3.09	3.52[1,2]
/i/	6.51	8.81*	7.61	7.76	10.40[1,2]
/ɪ/			5.89	5.78	4.93[1,2]
/y/			5.80	5.31	4.23[1,2]
/Y/			3.56	3.52	3.21[1,2]

* Statistical significant difference between ratios of vowels in syllables ($p < 0.05$).
[1] Statistical significant difference with ratio of untrained speaker ($p < 0.05$).
[2] Statistical significant difference with ratio of trained speaker ($p < 0.05$).

As can be seen in the Table for part of the vowels the ratios were significantly different from those of the control group. For the syllables the ratios of the vowels /a/ and /i/ were higher, and the ratios of the vowel /u/ were lower in tracheoesophageal speech. For the vowels from the text the ratios of the vowels /u/, /a/, /ɑ/, /i/, /ɪ/, /ɛ/, /y/, and /Y/ were significantly different from the control ratio of an untrained speaker. The ratios of the vowels /i/, /ɪ/, /ɛ/, /y/, and /Y/ were also significantly different from the control ratio of the trained speaker.

3.3 Interspeaker differences between the tracheoesophageal speakers

By means of a paired Students t-test it was investigated for each speaker separately whether or not his formant frequencies extracted from read aloud text, differed from those of the trained and the untrained speaker. It appeared that 9 speakers significantly differed from both the trained and the untrained speaker. One tracheoesophageal speaker differed significantly only from the trained speaker. For the remaining seven tracheoesophageal speakers no significant differences with the trained or untrained speaker were found.

4. Discussion and conclusions

The aim of this study was to investigate the first and second vowel formant frequencies in male Dutch tracheoesophageal speakers. Studies reporting on vowel formant frequencies are from Beck (1931), Luchsinger (1952), Schilling and Binder. (1926), and Damsté (1958), who reported only a small difference between esophageal and normal speech, and from Sisty and Weinberg (1972), Rollin (1962) and Kytta (1964), who found higher formant frequencies in esophageal speech. To the best of our knowledge no reports on formant frequencies in tracheoesophageal speech are available yet. It can, however, be expected that the vocal tract is comparable to the vocal tract of esophageal speakers, since the type of surgery is the same. The only difference between esophageal and tracheoesophageal speech lies in the fact that tracheoesophageal speech, like normal speech, is pulmonary driven.

As found in earlier studies on Finnish (Kytta, 1964), and English (Sisty and Weinberg, 1972; Rollin 1962) esophageal speech, also in the present study on Dutch tracheoesophageal speech higher vowel formant frequencies were found compared to normal speech. It was found that the first and second vowel formant frequencies of male Dutch tracheoesophageal speakers do differ significantly of those of normal male Dutch speakers. As a whole the "vowel triangle" is enlarged for the TE speaker group. An explanation for this might be, according to Sisty and Weinberg (1972), that the length of the vocal tract is shorter compared to normal speakers. Also the back of tongue might be a little lowered, due to the removal of the larynx. The formant frequencies found in this group had a large range of variation, which is in concordance with the observations of Rollin (1962) and Sisty and Weinberg (1972). The differences between TE-speakers may be larger than those between normal speakers, since the anatomy of the voice source and the vocal tract depends on the type and extent of the surgical intervention. These differences in vocal tract most probably also explain the intraspeaker differences that were found between the TE-speakers. This leads to significant differences in vowel formant frequencies compared to normal speakers in one part of the TE-speakers, and similar vowel formant frequencies compared to normals in another part of the TE-speakers. Also the ratios between the first and second formants are significantly different from the control ratios for part of the vowels. Performing videofluoroscopy recordings during speech may give more information about the length of the vocal tract and the position of the back of the tongue compared to normals, and can also give insight about the influence of the extent of the surgical intervention on this phenomenon. Also the influence of the different vowel formant frequencies and ratios on the intelligibility of the vowels should be studied.

5. Acknowledgements

The second author's masters thesis was the basis for this report (van Ravesteijn, 1997). The authors wish to acknowledge the Maurits and Anna de Kock Stichting for their financial support of the equipment used for the recordings and formant analysis. They also wish to thank all patients for their participation in the study.

6. References

Blom, E.D., Singer, M. I., and Hamaker, R.C. (**1982**). "Tracheostoma valve for postlaryngectomy voice rehabilitation". *Annals of Otology, Rhinology, and Laryngology*, **91**, 576-578.

Boersma, P. and Weenink D.J.M. (**1996**). "Praat, a system for doing phonetics by computer, version 3.4". *Report of he Institute of Phonetic Sciences, Amsterdam*, **132**, 182 pp.

Beck, J. (**1931**). "Zur Phonetik der Stimme und Sprache Laryngektomierter", *Zeitschrift für Laryngologie usw.*, Bd. **21**, H6, 506-521.

Damsté, P.H. (1958). *Oesophageal speech after laryngectomy,* Ph.D. Thesis, University of Groningen.
Diedrich W.M., and Youngstrom, K.A. (1966). *Alaryngeal speech.* Springfield: Charles C. Thomas.
Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague, The Netherlands: Mouton, 's Gravenhage.
Hilgers, F.J.M., and Schouwenburg, P.F. (1990). "A new low-resistance, self retaining voice prosthesis (Provox®) for voice rehabilitation after total laryngectomy". *Laryngoscope,* 100, 1203-1207.
Hilgers, F.J.M., Ackerstaff, A.H., Balm, A.J.M., and Gregor, R.T. (1996). "A new heat and moisture exchanger with speech valve (Provox® stomafilter". *Clinical Otolaryngology,* 21, 414-418.
Koopmans-van Beinum, F.J. (1980). *Vowel contrast reduction. An acoustical and perceptual study of Dutch vowels in various speech conditions.* Ph.D. Thesis, University of Amsterdam
Kytta, J. (1964). "Finnish oesophageal speech after laryngectomy: Sound spectrographic and cineradiographic studies", *Acta Otolaryngologica (Stockholm),* Suppl. 195.
Luchsinger, R. (1952). "Der Mechanismus der Sprech- und Stimmbildung bei Laryngektomierten und die Ubungsbehandlung", *Practica Oto-Rhino-Laryngologica,* XIV, Fasc. 4/5, 304-323.
Nichols, A.C. (1968). "Loudness and quality in esophageal speech and the artificial larynx. In: J.C. Snidecor (Ed.), *Speech Rehabilitation of the Laryngectomized* (2nd ed). Illinois: Thomas, 108-127.
Pols, L.C.W., Tromp, H.R.C., Plomp, R., (1973). "Frequency analysis of Dutch vowels from 50 male speakers", *The Journal of the Acoustical Society of America,* 53, 1093-1101.
Rollin, W.J. (1962). *A comparative study of vowel formants of esophageal and normal-speaking adults.* Ph. D. Thesis, Wayne State University.
Schilling, R., and Binder, H. (1926). "Experimentalphonetische Untersuchungen über die Stimme ohne Kehlkopf", *Archiv für Ohren-, Nasen- en Kehlkopfheilkunde,* 115, 235-270.
Sisty, N.L., and Weinberg, B. (1972). "Formant frequency characteristics of esophageal speech", *Journal of Speech and Hearing Research,* 15, 439-448.
Stevens, K.N., and House, A.S. (1955). "Development of a quantitative description of vowel articulation", *The Journal of the Acoustical Society of America,* 27, 484-493.
Van As, C.J., Hilgers, F.J.M., Koopmans-van Beinum, F.J., and Ackerstaff A.H. (in press). "The influence of stoma occlusion on aspects of tracheoesophageal voice", *Acta Otolaryngologica.*
Van Ravesteijn, A.M.A, (1997). *Evaluatie van tracheoesophageale spraak.* Master's Thesis, University of Amsterdam.