AN INTRODUCTION TO THE ASSESSMENT OF INTELLIGIBILITY OF TRACHEOESOPHAGEAL SPEECH

Petra Jongmans, Corina van As, Louis Pols, Frans Hilgers

Abstract

In cases of laryngeal cancer, it is sometimes necessary to perform a total laryngectomy. This procedure changes the anatomy and physiology of the vocal tract, with the most noticeable effect on speech. By applying a voice prosthesis, enabling the patient to use tracheoesophageal speech, speech is of better quality than with esophageal or electrolarynx speech, but still very deviant from laryngeal speech. In different countries, studies investigating the intelligibility of speech have been performed. In the Netherlands, only voice quality has been studied in detail. For the understanding of the physiology of the neoglottis and for improving the results of speech therapy it is important to study intelligibility as well. This paper will review relevant literature of research that has been carried out on this subject and will discuss some first acoustical observations in relation to intelligibility.

1 Introduction

In the Netherlands, annually around 700 individuals are diagnosed with laryngeal cancer (Visser et al., 1998). When diagnosed at an early stage, these tumors can be treated well with radiotherapy or laser surgery. However, when the tumor is in an advanced stage, as well as in cases of recurrence of the tumor, a total laryngectomy is often necessary. This is an operation whereby the entire larynx is removed and the upper and lower airways are disconnected. The digestive tract is re-established and the patient has to breathe through a permanent tracheostoma at the base of the neck.

One of the most obvious consequences of total laryngectomy is the loss of natural voice. Over the years, different methods of voice rehabilitation have been developed, the two earliest being the use of esophageal speech or an electrolarynx. From 1980 onwards a voice prosthesis is often applied, enabling the patients to use tracheoesophageal (TE) speech. The obvious advantage of TE speech is that, like normal laryngeal voicing, it is pulmonary driven, i.e. air from the lungs is used to set the tissues of the pharyngo-esophageal segment (PE segment; also called neoglottis or pseudo glottis) into vibration. This results in longer phonation time and a higher intelligibility rate than is achievable with esophageal or electrolarynx speech (Williams & Watson, 1987; Pindzola & Cain, 1988; Nieboer et al., 1988; Debruyne et al., 1994; Bertino et al. 1996; Max et al., 1996).

Most studies on TE speech have focused on voice quality (Van As et al., 2003), but only few studies have investigated TE speech intelligibility (Hammarberg et al., 1990; Doyle et al., 1988 and Miralles & Cervera, 1995), the topic this paper will focus on.

In our institutes, three pilot studies (MA theses) on intelligibility after total laryngectomy have been performed, two of which have investigated consonant intelligibility (Roeleven & Polak, 1999; Boon-Kamma, 2001) and one of which has investigated vowel intelligibility (Oubrie, 1999). These studies rendered some interesting results that deserve more attention. Therefore a project has been initiated that concerns investigating intelligibility of TE speech in more detail. The project will focus on both consonants and vowels. Intelligibility will be investigated by means of perceptual evaluation and acoustic analyses. In addition, an attempt will be made to investigate intelligibility in relationship to the anatomical and physiological changes after total laryngectomy. Potential methods to investigate those changes are videofluoroscopy, digital high-speed imaging and aerodynamic measures. The possible surgical and therapeutic consequences of the improved insight in the mechanisms which influence intelligibility of TE speech, may be incorporated in this project at a later stage.

In this introductory paper, an overview is given of how intelligibility could be investigated. Also the recordings of Roeleven & Polak (1999) were used for preliminary acoustic analyses in order to explore possible relationships between the perceptual data and acoustic characteristics. Some preliminary results will be discussed.

2 The operation

For a better understanding of TE speech, a brief description of the operation and the voice prosthesis will be given.

As is indicated in the introduction, with a total laryngectomy the complete larynx is removed. After removal the digestive tract is re-established and the trachea is bent forward and sutured into the skin at the base of the neck forming the tracheostoma. To make TE speech possible, a fistula is created between the trachea and the esophagus. The voice prosthesis (in this study the Provox® or Provox 2® prosthesis: Hilgers & Schouwenburg, 1990; Hilgers et al., 1997) is then placed in the fistula and acts as a one-way valve, which allows air to pass from the trachea into the esophagus, but prevents food from entering the trachea and the lungs.

The pharynx with its mucosa overlying the pharyngeal constrictor muscles and/or the cricopharyngeus muscle serves as a new voice source called the neoglottis. The air from the lungs sets this neoglottis into vibration.

It is important to gain insight into the anatomy and physiology of the voice source and the vocal tract after a total laryngectomy to fully understand why certain articulation difficulties exist. This is the reason why not only perceptual and acoustic analyses will be performed, but also methods that study the anatomical and physiological changes of the anatomy and voice production.

3 Perceptual analyses

In the actual study, specific speech samples will be collected from TE speakers and a control group of normal speakers. In the first part of the project, naïve listeners (representing the population laryngectomized patients encounter in daily life) and a small group of speech language pathologists will listen to the speech samples and give

their perceptual evaluations. For each word/syllable the listener will write down what (s)he perceived. These evaluations will be analyzed (confusion matrices, SPSS) with the aim of gaining insight in the sounds that are difficult to produce and what kinds of confusions are made. We will particularly focus on the voiced-voiceless distinction as confusions between these two are found frequently in every study on TE speech intelligibility that has been performed so far. Besides this the (dis)ability to produce this contrast may learn us more about the functioning of the neoglottis and whether TE speech is myoelastic-aerodynamic or only aerodynamic.

In the literature, different opinions exist on the use of two groups of listeners. Most studies found that experienced listeners rate TE speech higher than naïve listeners (Doyle et al., 1989; Finizia et al., 1998). In this study, the results from the experienced listeners will be used diagnostically, as they are also the individuals that provide speech therapy for the patient and decide what needs attention (Van As et al., 2003). The results of the naïve listeners are indicative for TE speech intelligibility in the general population.

As mentioned in the introduction, some studies have already investigated TE speech intelligibility. This has been done for Swedish, Spanish, American English and Dutch.

For Swedish (Hammarberg et al., 1990), it was found that initial consonants showed three times as many identification errors as final consonants. Just as in Spanish and American English (Miralles & Cervera, 1995 and Doyle et al., 1988) problems with voicing distinctions were found and /h/ was often omitted completely.

In American English, besides the voiced/voiceless distinction, fricatives and affricates were often not perceived correctly. In Spanish, fricatives also caused problems, as did velars and stops.

In Dutch, two studies have been performed on TE consonant intelligibility. Boon-Kamma (2001) investigated consonant intelligibility. She had 38 laryngectomees pronounce 48 meaningful CVC words. Perceptual analyses were performed by 4 naïve and 4 experienced listeners. She found that also for Dutch the voiced/voiceless distinction is difficult. In initial position voiceless plosives and fricatives were more difficult to perceive correctly, whereas in final position the voiceless consonants were often perceived correctly. Fricatives caused problems in initial position, whereas nasals did so in final position. Roeleven & Polak (1999) also looked at consonants, using CV, VCV and VC syllables. Nine laryngectomees participated and 15 naïve and 4 experienced listeners performed the listening task. Among other confusions they mostly found confusions of voiced and voiceless consonants and /h/ was often deleted.

Acoustic analyses are necessary as back up for these findings to see if we can find a reason for the above-mentioned confusions. In the next paragraph, literature on acoustic analyses of TE speech will be described.

4 Acoustic analysis

Perceptual analyses are often not precise enough for intelligibility research as they only describe what goes wrong but not where and why it goes wrong. Therefore it is important to perform acoustic analyses of TE speech as well. Some authors have already looked at acoustic speech signals, both for vowels and consonants (Robbins et al., 1986; Most et al., 2000; Saito et al., 2000; Cervera et al., 2001; Searl & Carpenter, 2002). They have mostly used voice onset time (VOT) as a measure to investigate the difference between voiced and voiceless sounds. Also the duration of sounds, voicing

lead, intra-oral pressure, consonant sound pressure level and the fundamental frequency and formant frequencies have been used as a measure.

Most et al. (2000) used speech material consisting of sentences of which they analyzed extracted words or syllables. They found significant correlations in Hebrew TE speech between F0, duration measures and acceptability, and between F0 and intelligibility. Cervera et al. (2001) analyzed the first vowel (stressed vowel) in real CVCV words and found overall that Spanish laryngectomized patients produce vowels with higher formant frequencies and longer durations than laryngeal speakers. Robbins et al. (1986) used real CVC words and state that American English TE speakers produce significantly shorter VOT's and longer vowel durations than laryngeal speakers. However, the shorter VOT is not the only reason why vowels sound longer; the delayed offset time also contributes to longer vowel durations. Saito et al. (2000) have looked at the production of /p/ and /b/ for Japanese TE speakers and laryngeal speakers, using the construction aVa preceded by the word 'desu', meaning 'that is'. They studied VOT, voicing lead and intra-oral pressure. They found that the onset of vibration during the closure period (the voicing lead) appears to be highly significant for voicing consonants. They also state that greater intra-oral pressure is needed for TE speakers to set the mucosa into vibration when producing consonants and that this intra-oral pressure together with the voluminous neoglottis is responsible for the extended VOT in TE speakers. Searl et al. (2002) also studied the production of voiced and voiceless consonants of American English TE speakers using the nonsense word m℘C↔ in a carrier phrase whereby the C was varied. They took four acoustic aspects into account: VOT, consonant sound pressure level, consonant duration and vowel duration. They assume that more factors than just VOT play a role in the correct production of voiced/voiceless consonants. It appeared that three acoustic aspects show the difference between TE and normal speakers: consonant sound pressure level, consonant duration and vowel duration. VOT did not differ significantly between the two groups of speakers. They conclude by saying that TE speakers could conceivably be trained to use secondary cues to enhance the voiced/voiceless consonants.

It needs to be kept in mind that VOT differs across languages. Therefore, VOT measurements in Dutch may show different results than the results presented in here.

In the intended pilot study, an attempt will be made to find acoustic explanations for the confusions found in the perceptual analyses, originally performed by Roeleven & Polak (1999). The program Praat will be used for the acoustic analysis (Weenink & Boersma, www.praat.org).

5 Videofluoroscopy

Videofluoroscopy is a commonly used clinical tool used for visualization of the neoglottis, mostly in lateral view. The anatomy and morphology of the neoglottis as observed with this method appears to predict tracheoesophageal voice quality in a rather consistent way (Van As et al., 2001, see also Op de Coul et al., 2003).

With videofluoroscopy, x-ray images of the neoglottis allow to visualize the anatomy/morphology of the neoglottis and the vocal tract. The patients swallow barium sulphate, a contrast fluid that enables visualization of the tissues. A standard coin is stuck on the cheek as a reference for quantitative measurements. For investigation of tracheoesophageal voice production a sustained vowel is most often used. For intelligibility research different vowels could be used to observe the changes in the neoglottis and vocal tract and how they influence intelligibility.

Observations and quantitative measure of the neoglottis and the vocal tract for different vowels may help finding out why certain vowels cause difficulties. In Van As et al. (2001) great individual differences were found between patients. This might mean that we will also find individual differences in vowel intelligibility.

6 Digital high speed imaging

In laryngeal voicing, stroboscopy is often used to visualize vocal fold vibration. With stroboscopy a virtual slow-motion of the rapid vibration is acquired by subsequent light flashes that are slightly out of phase with the fundamental frequency of the vocal fold vibration. However, as the neoglottis works differently from the vocal folds and often does not show a stable fundamental frequency, stroboscopy is not a very useful instrument to study vibrations in the neoglottis. Other instruments are available that are frequency-independent, such as digital high-speed imaging (Wittenberg et al., 1995) and videokymography (Sveč & Schutte, 1996). These methods have been used to study pathological laryngeal voices and also have proven to be suitable to visualize vibrations of the neoglottis (Van As et al., 1999). They were the first to use digital high-speed imaging to investigate whether this instrument would be useful in TE speech research. These authors showed that useful images, giving realistic visual information about the vibration of the neoglottis, can be collected.

Lundström & Hammarberg (1999) later used digital high speed imaging to study the voiced-voiceless distinction in an esophageal and a tracheoesophageal speaker and conclude that high speed recordings used together with voice signals can answer important questions about the functions of the pharyngoesophageal segment.

7 Aerodynamics of TE speech

According to Van den Berg (1958), natural phonation is a myoelastic-aerodynamic process. This theory implies that Bernoulli forces (negative pressure) cause the vocal folds to be sucked together, thereby creating a closed airspace below the glottis. The air pressure from the lungs continues to build up until it is high enough to press the folds outward, thus opening the glottis and producing sound. The lateral movement of the vocal folds continues until the natural elasticity of the tissue takes over and the vocal folds move back to their original, closed position. The fundamental frequency of the vibration depends on the effective mass and stiffness of the vocal folds that are regulated by the sustained innervation of internal and external laryngeal muscles and lung pressure.

This theory cannot completely be applied to TE speech. As air from the lungs is used for phonation, aerodynamics most probably does play an important role. The question is if there is also a myoelastic component. Some cases are known whereby patients are capable of influencing their voice source in a myoelastic way so that they could change their F0. However, they were not able to do this consistently (Moon & Weinberg, 1987). Also the fact that some patients are able to produce a clear voiced/voiceless distinction suggests muscle activity (Nord et al., 1992).

For therapeutic purposes it is important to understand the processes involved in TE phonation. If it is possible for patients to actively influence the neoglottis, this may well improve their speech intelligibility. Therefore an overview of research in this area will be given here.

As stated earlier, Moon & Weinberg (1987) looked at aerodynamic and myoelastic contribution to tracheoesophageal voice production. Their goal was to assess the

influence of aerodynamic factors on the regulation of F0 and to determine whether tracheoesophageal voice production is an aerodynamic-myoelastic event. They found that there was a significant overall change in F0 as well as in trans-source airflow rate and tracheal pressure as a function of the effort/flow rate level for each subject. This outcome supports the accepted view that TE voice production is, at least in part, an aerodynamically mediated event. However, it should be noted that variation was found and that not all subjects showed the same results. Even though it is safe to say that aerodynamics play an important role, the high variability in F0 values suggests that more factors should be considered and that probably, in some cases, myoelastic contributions can be found.

Desschler et al. (1997) investigated the effect of sound pressure level on fundamental frequency in TE speakers. They found that TE speakers maintain a functional range of pitch variation, but that the cause of this pitch control is not clearly defined. Their results show that SPL was a significant factor in the prediction of F0, but that it accounted only for 9 % of the variability. They also found that subjects were able to increase F0 consistently at stable effort levels. This means that other factors must play a role as well which could include possible contractile changes within the pharyngo-esophageal segment. Their results suggest that TE speech is both an aerodynamic and myoelastic event.

Searl (2002) looked at oral pressure (OP) in TE speakers to investigate if TE speakers maintain phonemic class differences. Measures of peak OP can also be used to assess the stability of TE speech production. He found that TE speakers have mean OP values about 2-3 times greater than those of laryngeal speakers. The speakers also showed significantly greater OP occurring on the voiceless compared to the voiced phonemes, which is the same for laryngeal speakers. The difference in mean OP for voiced and voiceless stops was significantly greater than the difference between voiced and voiceless fricatives. This could explain the higher confusion rates between fricative cognates found in for example Roeleven en Polak (1999). Even though differences are found, OP in TE speakers shows high variability in comparison with laryngeal speakers. Between studies that looked at oral pressure, also great differences exist in OP values (Saito et al., 2000; Motta et al., 2001). This might be caused by different set ups, but also might be due to different types of fistula construction, as all the studies used different techniques. OP differences might also be attributed to significant inter-speaker variability, which has been found before. If this is the case, it could explain the fact that some speakers are better in maintaining the voiced-voiceless contrast than others. Searl (2002) only included consonants that were perceived correctly. It is important to investigate if laryngectomized individuals who cannot produce a proper voiced/voiceless contrast show deviant OP values from patients who can. If so, speech therapy might focus on teaching the patients to 'push harder' when producing voiceless consonants in order to increase OP. Studies have reported (Christensen & Dwyer, 1990; Connor et. al., 1985) that this approach has helped to increase the intelligibility of consonants in esophageal and electrolarynx speakers.

The studies described above show that it is important to study aerodynamics.

8 Pilot study on consonants intelligibility by Roeleven and Polak

8.1 Patients

In 1999, Roeleven & Polak performed an experiment to investigate the intelligibility of consonants in TE speech. They had nine speakers, who all had undergone a standard total laryngectomy. The speakers' age varied between 48 and 75 years with a mean age of 66. They were all male and had Dutch as their mother tongue. They were laryngectomized at least a year prior to the study. They all used an indwelling Provox voice prosthesis and digitally occluded their stoma through a Provox HME (heat and moisture exchanger) in a peristomal adhesive (Hilgers et al., 1991). None of the patients had any anatomical or neurological problems that might influence their speech.

8.2 Methods

The speech material used in the experiment consisted of syllable lists testing consonants in initial, medial and final position, using the three vowels /a/, /i/ and /u/ e.g. paa, aapaa, aap. They also used meaningful CVC words, short sentences and a reading passage. Only the syllables for initial and medial positions were taken into account as most mistakes are made in these two positions.

All speakers pronounced 54 CV syllables and 63 V1CV2 (V1=V2) syllables in a sound treated room, using a DAT recorder for recording, the Computerized Speech Lab (Kay Elemetrics Corp., Lincoln Park, USA) and a headset microphone (AKG-c410).

Four experienced and fifteen naïve listeners listened to the recordings and for each syllable they wrote down what they had heard. These identification results have been statistically investigated by Roeleven & Polak (1999) for average consonant intelligibility and its standard deviation and t-tests were used to investigate differences in intelligibility between TE speakers and a reference group of normal speakers, who participated in a somewhat comparable study by Pols (1981) where CVCVC-type nonsense words were used. Per subject, confusion matrices were composed and studied. In our reanalysis, only the cumulative perceptual data from the 15 naïve listeners were used.

8.3 Perceptual results

The perceptual data were reanalyzed to ensure their consistency and to optimize them for acoustic analysis. No critical differences were found between the original and the re-analysis data except for some insignificant differences in the number of confusions for some of the phonemes. The confusion matrices show that most mistakes were made within the feature voice. The voiceless consonants were more often replaced by voiced ones (usually cognate pairs) than the other way round (see Table 1 for initial position and Table 2 for medial position). This is consistent with results from other studies. We see a higher amount of confusions for the fricatives than for the plosives. These confusions are voice-source related and were found both in initial and in medial position.

Table 1. Confusion matrix of consonants in initial position

	p	b	t	d	k	g	f	v	s	z	sj	h	m	n	w	l	r	j	tj	zj	add	del	
p	300	89	2										1	2							8		402
b	9	363										1	8	19							5		405
t	9	1	327	52	15		1																405
d		2	13	375	6	7											2						405
k	4	1	19	5	357	13											1		1		1	3	405
g					2	357	2	1		2	1	22					8	1			3	6	405
f							228	168	1						5						3		405
v				1			135	237			2				29						1		405
s							1	1	209	184	7						1		1		1		405
z									25	364	6							7		3			405
sj	1		2		16		2		33	54	215							16	4	58	4		405
h	3	1	1	1							2	311	1		12	7	8	14			33	11	405
m	1	1											372	15	4	10					2		405
n				3								1	19	366		14	1	1					405
w		2					7	23				8	20	4	332	6	2				1		405
l			1											3	3	388	2	5			2	1	405
r												6	1	1	1	6	383				2	5	405
j				1		11	1	2		3	3	12	1			16		344	1		6	4	405

Table 2. Confusion matrix for consonants in medial position

	p	b	t	d	k	g	f	v	s	z	sj	h	m	n	ng	w	l	r	j	tj	nj	zj	dj	add	del	
p	359	40	4	1	1																					405
b	21	344	1	4			3	3				1				22		5							1	405
t	7	4	354	28	7		1	1									1						1		1	405
d		8	36	345	3	1			1				3				1	1	4				1		1	405
k	5			9	1	378	8									1	1	1							1	405
g					5	350		4	1	1		5		1	1			33	1		1				2	405
f	4	5					192	189								14								1		405
v	6	3					113	250					1			30									2	405
s							1		154	234	8											3		3	2	405
z			4						81	305	7						1			1	2	3			1	405
sj			3			3			42	32	188	1					1		6	11		101	12	3	2	405
h				1	2	6					1	331		1		11	4	13	20	1				4	10	405
m		11										8	325	35	4	6	6	8					1	1		405
n				1	2				1	1			11	366	1		14	2			4			1	1	405
ng						12						2	40	136	174				2		12		1	16	10	405
w	1	45		1		3	8	29					5	3	2	297		9			1				1	405
l							1						1	6		11	375	10					1			405
r				2	1				1	1				1			5	394								405
j		1	1	1		4			2			25		1	1	4	8	1	353				1	2		405
tj	1		18	1	6	1	1		1	4	5	1							1	287			65	8	5	405
nj											4		3	28	12			1	4		336	2	4	7	4	405

When we look more closely at the kind of confusions made, we see that fricatives and plosives are mostly confused with phonemes that have the same manner of articulation. This is not the case with nasals and approximants, which show confusions with phonemes with a different manner of articulation. Approximants are often confused with fricatives and nasals. Nasals show all sorts of confusions, but a prominent confusion is /n/ with /l/. This means that with plosives and fricatives it seems that the manner of articulation is perceived correctly, but not the place. Apparently with nasals and approximants it is the manner that is difficult to perceive. An articulation-based confusion is found between /b/ and /w/ in both positions with /w/ becoming a /b/ only in medial position.

8.4 Acoustic analysis

Even though perceptual analyses are still considered as the gold standard in investigating voice and speech, it is also interesting to back up findings with acoustical data. The question is whether consonants with poor intelligibility also show deviant acoustic patterns and how much these patterns differ from those for normal speakers. Some studies have already looked at the acoustics of TE speech (Robbins et al., 1986; Most et al., 2000; Saito et al., 2000; Cervera et al., 2001; Searl & Carpenter, 2002), but for Dutch, acoustic analyses have only been made in connection with voice quality (Van As, 2001). Therefore we performed acoustic analyses of some of the speech stimuli of Roeleven & Polak, with special emphasis on the confusions of /p/→/b/, /b/→/p/, /t/→/d/ and /d/→/t/ in both initial and medial position. It is important to keep in mind that the percentage correct for these particular phonemes is still rather high, which means that the confusions were not consistently made by each listener. However, TE speakers do show lower scores than the laryngeal speakers of Pols (1983).

Several preliminary observations were made for the confusion /p/→/b/: in some cases vibration of the neoglottis mucosa was present, which made /p/ sound like /b/. Sometimes however, the signal for /p/ showed a clear burst, but it was still perceived as /b/. One patient consistently inserted /↔/ in front of the initial consonant. It could be the case that it is difficult for TE speakers to rapidly change from voiced to voiceless, which would explain why the voiceless consonants of this patient were perceived as voiced. One patient whose /p/ was perceived correctly in most of the cases, showed aspiration. As aspiration typically occurs with voiceless plosives, this might be an important cue for listeners.

For /t/ in general it was noticed that the noise in the signal was longer for the lower scoring subjects than the higher scoring ones. For the investigated phonemes it was found that at the end of each period in the signal a little noise was present, something that was not found in the speech signal of a laryngeal speaker. Most of all there was an enormous diversity between subjects and more analyses are necessary to draw any meaningful conclusions about causes for misperception.

9 Summary/discussion

In this paper, different ways of investigating the intelligibility of TE speech have been discussed and results of earlier studies have been presented. Perceptual analyses are still considered the gold standard in intelligibility research, and even though it is a good start to get an overview of the problems, acoustic analyses really are necessary to back up findings. These two methods only provide information about the speech

signal: how it is perceived and why it might be perceived like that. They do not tell us anything about the cause of the mistakes TE speakers make in producing certain sounds. Therefore it is important to look at the anatomy and physiology of the vocal tract as well. Both videofluoroscopy and digital high-speed imaging are tools to study the anatomy and function of the neoglottis. Videofluoroscopy is expected to be suitable to study vowel production, and digital high-speed imaging may be suitable for studying the vibratory behavior of the neoglottis. Aerodynamic insight in the anatomy and physiology of TE speech is of great significance, first, because it clarifies the working mechanism of the neoglottis and second, because some manipulations at a different 'level', like, for example, oral pressure, seems to improve the intelligibility of certain consonants, at least in esophageal speakers.

The results of the pilot study so far show that acoustic analyses of TE speech signals are complicated due to the diversity between subjects and because many perceptual cues are used to distinguish between voiced and voiceless sounds. Obviously, further analysis is necessary to clarify the different issues addressed.

All the described methods together will give a more detailed description of TE speech and might help improve the results of speech rehabilitation through adapted surgical and/or clinical intervention approaches. The practical implications of this study do not only make this study valuable for scientists, but also for the TE speakers themselves as they might benefit from it at the end.

Acknowledgements

References

As, C.J. van, Koopmans-van Beinum, F.J., Pols, L.C.W. & Hilgers, F.J.M. (2003): "Perceptual evaluation of tracheoesophageal speech by naïve and experienced judges through the use of semantic differential scales", *Journal of Speech, Language and Hearing research* **46(4)**, 947-959.

As, C.J. van (2001). *Tracheoesophageal speech. A multidimensional assessment of voice quality*, PhD. Thesis, University of Amsterdam.

As, C.J. van, Op de Coul, B.M.R., van den Hoogen, F.J.A., Koopmans-van Beinum, F.J. & Hilgers, F.J.M. (2001): "Quantitative videofluoroscopy: a new evaluation tool for tracheoesophageal voice production", *Archives of Otolaryngology and Head and Neck Surgery* **127**, 161-169.

As, C.J. van, Tigges, M., Wittenberg, T., Op de Coul, B.M.R., Eysholdt, U. & Hilgers, F.J.M. (1999): "High-speed digital imaging of the neoglottic vibration after total laryngectomy", *Archives of Otolaryngology and Head and Neck Surgery* **125,** 891-897.

Berg, J.W. van den (1958): "Myoelastic-aerodynamic theory of voice production", *Journal of Speech and Hearing Research* **1**: 227-244.

Bertino, G., Bellomo, A., Miani, C., Ferrero, F., & Staffieri, A. (1996): "Spectrographic differences between tracheo-esophageal and esophageal voice", *Folia Phoniatrica et Logopaedica* **48**, 255-261.

Boon-Kamma, B. (2001): *Verstaanbaarheid na totale laryngectomie. De verstaanbaarheid van sprekers met een Provox® 2 stemprothese*. Master's thesis, University of Amsterdam. 70 pp.

Cervera, T., Miralles, J.L. & González-Alvarez, J. (2001): "Acoustic analysis of Spanish vowels produced by laryngectomized subjects", *Journal of Speech and Hearing Research* **44**, 988-996.

Christensen, J.M. & Dwyer, P.E. (1990): "Improving alaryngeal speech intelligibility", *Journal of Communication Disorders* **23**, 445-451.

Connor, N.P., Hamlet, S.L. & Joyce, J.C. (1985): "Acoustic and physiologic correlates of the voicing distinction in esophageal speech", *Journal of Speech and Hearing Disorders* **50**, 378-384.

Debruyne, F., Delaere, P., Wouters, J. & Uwents, P. (1994): "Acoustic analysis of tracheo-esophageal versus esophageal speech", *Journal of Laryngology and Otology* **108**, 325-328.

Desschler, D.G., Doherty, E.D., Reed, C.G., & Singer, M.I. (1999): "Effects of sound pressure levels on fundamental frequency in tracheesophageal speakers", *Otolaryngology-Head and Neck Surgery* **121**, 23-26.

Doyle, P.C., Danhauer, J.L. & Reed, C.G. (1988): "Listeners' perceptions of consonants produced by esophageal and tracheoesophageal talkers", *Journal of Speech and Hearing Disorders* **53**, 400-407.

Finizia, C., Lindström, J. & Dotevall, H. (1998): "Intelligibility and perceptual ratings after treatment for laryngeal cancer laryngectomy versus radiotherapy", *Laryngoscope* **108**, 138-143.

Hammarberg, B., Lundström, E. & Nord, L. (1990): "Consonant intelligibility in esophageal and tracheoesophageal speech. A progress report", *Phoniatric and Logopedic Progress Report. Department of Logopedics and Phoniatrics, Huddinge University Hospital & Karolinska Institute*, Stockholm, Sweden. Report no. 7.

Hilgers, F.J.M. & Schouwenburg, P.F. (1990): "A new low-resistance, self retaining prosthesis (Provox®) for voice rehabilitation after total laryngectomy", *Laryngoscope* **100**, 1202-1207.

Hilgers, F.J.M., Aaronson, N.K., Ackerstaff, A.H., Schouwenburg, P.F., van Zandwikj, N. (1991): 'The influence of a heat and moisture exchanger (HME) on the respiratory symptoms after total laryngectomy', *Clinical Otolaryngology* **16(2)**, 152-156.

Hilgers, F.J.M., Ackerstaff, A.H., Balm, A.J.M., Tan, I.B., Aaronson, N.K. & Persson, J.O. (1997): "Development and clinical evaluation of a second-generation voice prosthesis (Provox® 2), designed for anterograde and retrograde insertion", *Acta Otolaryngologica (Stockholm)* **117**, 889-896.

Lundström, E. & Hammarberg, B. (1999): "The pharyngo-esophageal voice source in laryngectomees as studied by means of high-speed imaging technique with focus on the voiced-voiceless distinction", *Phoniatric and Logopedic Progress Report, Department of Logopedics and Phoniatrics, Huddinge University Hospital & Karolinska Institute*, Stockholm, Sweden, Report no. 11.

Max, L., Steurs, W. & De Bruyn, W. (1996): "Vocal capacities in esophageal and tracheoesophageal speakers", *Laryngoscope* **106**, 93-96.

Miralles, J.L. & Cervera, T. (1995): "Voice intelligibility in patients who have undergone laryngectomies", *Journal of Speech and Hearing Research* **38**, 564-571.

Moon, J.B. & Weinberg, B. (1987): "Aerodynamic and myoelastic contributions to tracheoesophageal voice production", *Journal of Speech and Hearing Research* **30**, 387-395.

Most, T., Tobin, Y. & Mimran, R.C. (2000): "Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production", *Journal of Communication Disorders* **33(2)**, 165-180.

Motta, S., Galli, I. & DiRienzo, L. (2001): "Aerodynamic findings in esophageal voice", *Archives of Otolaryngology and Head and Neck Surgery* **127**, 700-704.

Nieboer, G.L.J., De Graaf, T. & Schutte, H.K. (1988): "Esophageal voice quality judgments by means of the semantic differential", *Journal of Phonetics* **16**, 417-436.

Nord, L., Hammarberg, B. & Lundström, E. (1992): "Voiced-voiceless distinction in alaryngeal speech-acoustic and articulatory observations", *Phoniatric and Logopedic Progress Report. Department of Logopedics and Phoniatrics, Huddinge University Hospital & Karolinska Institute*, Stockholm, Sweden, report 2-3, 19-22.

Op de Coul, B.M.R., van den Hoogen, F.J.A., van As, C.J., Marres, H.A.M., Manni, J.J. & Hilgers, F.J.M. (2003): "Evaluation of the effects of primary myotomy in total laryngectomy on the neoglottis using quantitative videofluoroscopy", *Archives of Otolaryngology and Head and Neck Surgery* **129**, 1000-1005.

Oubrie, A. (1999): *Formantwaarden en verstaanbaarheid van Nederlandse klinkers bij tracheo-esophageale sprekers*, Master's thesis, University of Nijmegen, 50 pp.

Pindzola, R.H. & Cain, B.H. (1988): "Acceptability ratings of tracheoesophageal speech", *Laryngoscope* **98**, 394-397.

Pols, L.C.W. (1983): "Three-mode principal components analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation", *Speech Communication* **2**, 275-293.

Robbins, J., Christensen, J.M. & Kempster, G. (1986): "Characteristics of speech production after tracheoesophageal puncture: voice onset time and vowel duration", *Journal of Speech and Hearing Research* **20**, 499-504.

Roeleven, S. & Polak, R. (1999): *De Verstaanbaarheid van consonanten in de Nederlandse trachea-esophageale spraak.* Bachelor's thesis, Rotterdam College of Logopedics.

Saito, M., Minoru, K. & Amatsu, M. (2000): "Acoustic analysis clarify voiced-voiceless distinctions in tracheoesophageal speech", *Acta Otolaryngologica* **120**, 771-777.

Searl, J.P. (2002): "Magnitude and variability of oral pressure in tracheoesophageal speech", *Folia Phoniatrica et Logopaedica* **54**, 312-328.

Searl, J.P. & Carpenter, M.A. (2002): "Acoustic cues to the voicing feature in tracheoesophageal speech", *Journal of Speech, Language, and Hearing Research* **45**, 282-294.

Sveč, J. G. & Schutte, H.K. (1996): "Videokymography: High-speed line scanning of vocal fold vibration", *Journal of Voice* **10**, 201-205.

Visser, O., Coebergh, J.W.W., Otter, R. & Schouten, L.J. (Eds.) (1998): *Head and neck tumors in the Netherlands 1989-1995*, Utrecht: The Netherlands Cancer Registry.

Williams, S.E. & Watson, J.B. (1987): "Speaking proficiency variations according to method of alaryngeal voicing", *Laryngoscope* **97**, 737-739.

Wittenberg, T., Moser, M., Tigges, M. & Eysholdt, U. (1995): "Recording, processing and analysis of digital high speed sequences in glottography", *Machine, Vision and Applications* **8**, 399-404.