THE ACOUSTIC CONTRAST BETWEEN THE DUTCH CONSONANTS /T/ AND /D/ IS REDUCED IN TRACHEO-ESOPHAGEAL SPEECH

K. van Sluis[1,2], M. Kapitein[1], R.van Son[1], P. Boersma[2]

[1]Department of Head and Neck Oncology and Surgery, Netherlands Cancer Institute-Antoni van Leeuwenhoek, Amsterdam, The Netherlands

[2]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands

k.v.sluis@nki.nl

ABSTRACT

The purpose of this study is to describe the acoustic changes of the consonants /t/ and /d/ in Dutch speaking individuals before and after total laryngectomy. The speech of seventeen participants was recorded before and after treatment. Eighteen tokens from a read-aloud text were obtained with /t/ or /d/ in initial position. Prevoicing, burst duration and duration of the vowel following the consonant were analyzed. The results show that the acoustic contrast of the /t/ and /d/ in initial position is reduced after treatment. Post-operatively, the presence of prevoicing of /d/ decreases, and burst duration increases. In the post-operative situation, therefore, /d/ becomes more similar to /t/ acoustically.

Keywords: Total laryngectomy – Tracheo-esophageal speech – Voicing contrast – Alveolar plosives – Acoustics.

1. INTRODUCTION

A total laryngectomy implicates surgical removal of the larynx. This procedure is mostly carried out because of advanced laryngeal cancer. With the removal of the larynx the natural voice is lost and individuals have to regain speech with the help of a substitute voice. The preferred substitute voice method in the Netherlands is tracheo-esophageal speech, with help of a voice prosthesis. Voice is generated by vibrations of the pharyngo-esophageal segment. Tracheo-esophageal speech is usually perceived as rough, irregular and reduced in dynamic range [11].

Several studies have described the acoustic characteristics of tracheo-esophageal speech [7, 2, 10]. Weak periodicity and a high noise component are characteristics of tracheo-esophageal speech [8]. Tracheo-esophageal speakers have higher values for jitter and shimmer compared to laryngeal voices [11]. This is a result of using the PE-segment as voicing source, as the tissue is less suited to sustain stable vibrations compared to the vocal folds. Jitter and shimmer will not be discussed in this paper. The focus is on fundamental frequency (F_0), harmonics-to-noise ratio (HNR), percentage voiced (%V), and maximum voicing duration (MVD).

The main difficulties for tracheo-esophageal speakers regarding intelligibility are in the production of word-initial consonants, rather than consonants in word-final position [9]. Tracheo-esophageal speakers also encounter difficulties distinguishing between voiced and voiceless consonants, voicing vowels, maintaining pitch, phrasing, and producing /h/ and /k/, the consonants that are produced in areas most affected by the operation [9].

According to Jongmans et al. [9], listeners often confuse laryngectomees' initial plosives. In Dutch, word-initial voiced plosives are produced with prevoicing (a negative voice onset time), while voiceless plosives are produced without aspiration (zero voice onset time) [1]. Van Alphen [1] suggests that presence of prevoicing is the most reliable cue to voicing distinction for listeners. In the study of Jongmans et al. [9] prevoicing is not taken into account. Jongmans et al. [9] suggest that in tracheo-esophageal speech, the speaker's effort to create a voicing contrast affects the duration of the burst of the plosive as well as the duration of the following vowel.

The aim of this research was to investigate how the acoustic features (*prevoicing*, *burst duration* and *vowel duration*) of initial plosives /t/ and /d/ change in Dutch speaking individuals before and after total laryngectomy.

2. METHODS

2.1. Participants

Seventeen subjects who underwent total laryngectomy were included in the study (2 female, mean age 65, range 47─78). Participants were included in this study when a pre- and post-

treatment recording was present. Subjects in palliative setting were excluded. All participants did receive speech language therapy sessions, until satisfactory speech was reached. Post treatment recordings were made after on average 6 months after total laryngectomy (range 3─35). This study has been approved by the Institutes Research Board (registration number IRBd18005).

2.2. Data collection

Recordings are made during regular appointments at the speech language pathologist. Over time, three different microphones were used during consultations; HS5 Samson Aerobic Headset Mini Jack/XLR STAGE 55/CONCERT 77/AKG Version, Shure SM10A-CN headset with a Blue Icicle USB microphone preamplifier and Samson Qv10e micro-phone.

Participants were instructed to pronounce a sustained /a/ as long as possible and to read aloud a Dutch text of neutral content (*Tachtig dappere fietsers* 'eighty brave cyclists'). In this text, 18 token words occurred (9 starting with /t/, 9 starting with /d/). For two participants, only the first and last two sentences of this text were recorded, since they had difficulty reading the text in its entirety resulting in 11 tokens. .

2.3. Acoustic analysis

Speech recording data were processed in Praat [6]. Voicing characteristics before and after treatment are analyzed on sustained /a/, to check whether voice contrast analysis is possible. All /a/ recordings were cut to a length of 2.0 seconds of the most stable part (as done in van As-Brooks et al. [2]), to ensure optimized and similar data for all participants. The measured acoustic parameters are: F_0 (median/SD), percentage voiced (%V), Maximum Voicing Duration (MVD), and Harmonics-to-Noise Ratio (HNR). Pitch was determined with the standard settings of the *To Pitch* command using the cross-correlation method with the exception of: pitch floor 40 Hz, ceiling 200 Hz; voicing threshold 0.40. The median F_0 was determined as the median value of all the voiced frames. %V was determined by counting the number of voiced and unvoiced frames as determined by the Pitch algorithm. MVD was taken as the longest voiced segment after an automatic annotation with the *To TextGrid (vuv): 0.2, 0.1* command. HNR was calculated adapted from van As-Brooks et al. [2] as the smoothed maximum found after calculating *To Harmonicity (cc)* with a minimum pitch of 40Hz, silence threshold 0, and 4.5 periods per window (1 period per window if no HNR was found).

For each t/d token, the begin and end of the release burst were marked (at nearest zero-crossing). The length of the release burst was measured. The presence of an F_0 value inside a 35ms window before the start of the release was considered evidence of prevoicing. A window of 35ms was chosen to preclude an influence of the preceding word. Vowel duration was calculated from the end of the release burst to the vowel offset.

2.4. Statistical analysis

All data is collected and processed for statistical analysis in R [12]. A Kolmogorov-Smirnov test showed the data was not normally distributed. Therefore, the Wilcoxon Matched Pairs Signed Rank test (WMPSR) is performed to analyse differences between pre- and post-total laryngectomy voice recordings. Statistical significance is corrected for false discovery rate [5].

To estimate the importance of the factors studied for the acoustic realizations of the alveolar plosives, linear mixed effect models were created [3, 4]. The model analyses the relationship between *prevoicing*, *burst duration*, and *vowel duration* on the one hand and the fixed effects *time* (pre- / post-treatment) and *phoneme* identity (/t/ /d/) on the other hand, including the interaction term (see Table 4.). The by-subject and by-word intercepts and random slopes of the effects of *time* and *phoneme* identity were used as random effects. P values were estimated by approximating the Student t-test statistics of the coefficients by Z-test statistics, following Barr et al. [3].

An R Markdown script and the original /t/ /d/ data have been made available on http://www.fon.hum.uva.nl/rob/ICPhS19/.

3. RESULTS

In Table 1 the acoustic characteristics of the sustained vowel /a/ recorded from the participants before and after surgery are presented. In three participants after treatment no median F_0 could be determined. After treatment acoustic features are changed. Median F_0 values of tracheo-esophageal speech have a mean of 61Hz, coming from a pre-operative value of mean 138Hz. The mean percentage voiced frames drops from 93% to 55%. Harmonics-to-noise ratio (HNR) and maximum voicing duration (MVD) decreases. A statistically significant difference between pre- and post-

treatment values was found for F_0-median, %V, HNR and MVD ($p \leq .05$, WMPSR).

Table 1. Range, mean, standard deviation and statistical testing of the acoustic values pre and post-operative of sustained vowel /a/ *n*=17. Abbreviations: F_0: fundamental frequency, %V: percentage voiced, HNR: harmonics-to-noise ratio, MVD: maximum voicing duration. **p*≤.05 tested with WMPSR test ^(*n*=14)

Acoustic Parameter		**Range**	**Mean**	**SD**	**p-value**
F_0-median (Hz)	Pre	79-277	138^	56	
	Post	17-132	61	40	.002*
F_0-SD (Hz)	Pre	1-27	8	10	
	Post	1-42	11	13	.391
%V	Pre	15-100	93	21	
	Post	0-100	55	38	.003*
HNR (dB)	Pre	4-20	15	5	
	Post	0-8	3	4	<.001*
MVD (sec)	Pre	.3-2	1.9	.4	
	Post	0-2	1.2	.8	.013*

Table 2 shows the results with acoustic features of the consonant /t/ for the pre- and post-operative speech conditions. No statistically significant difference was seen for the parameters *prevoicing* and *burst duration* between pre- and post-operative condition. An increase in *vowel duration* following the initial consonant /t/ was seen for the post-operative speech condition (*p*=.017).

Table 3 shows the results with acoustic features of the consonant /d/ and outcomes of statistical testing for the pre- and post-operative speech conditions. The presence of *prevoicing* for the consonant /d/ decreases significantly in the post-operative speech condition (*p*=.003). *Burst duration* of /d/ increases significantly in the post-operative speech condition (*p*<.001). No difference was found for the vowel duration for the vowel following the initial /d/ (*p*=.120).

Table 2. Mean and standard deviation, minimum and maximum value and statistical testing of the acoustic values for the /t/ pre- and post-operative *n*=17. **p*≤.05 tested with WMPSR test

Acoustic Parameter		**Mean**	**SD**	**Min**	**Max**	**p-value**
Prevoicing (%)	Pre	1.3	3.7	0	11.1	
	Post	1.3	5.4	0	22.2	1.00
Burst dur. (ms)	Pre	37.2	10.9	25.2	64.2	
	Post	38	5.5	29.7	47.7	.747
Vowel dur. (ms)	Pre	93.9	16.6	73.6	138.1	
	Post	107.5	32	68.2	174	.017*

Table 3. Mean and standard deviation, minimum and maximum value and statistical testing of the Acoustic Values for the /d/ pre- and post-operative *n=17*. **p*≤.05 tested with WMPSR test

Acoustic Parameter		Mean	SD	Min	Max	p-value
Prevoicing (%)	Pre	57.5	36.7	0	100	
	Post	14.4	23.1	0	77.8	.003*
Burst dur. (ms)	Pre	24.5	5.9	15.9	37.4	
	Post	34.9	6.5	23.6	47.3	<.001*
Vowel dur. (ms)	Pre	101.5	12.6	77.3	129.9	
	Post	109.1	21.2	79.8	150.1	.120

Table 4. Linear mixed effect models analysis *Y~phoneme*time* + (1 + *time*|*speaker*) + (1 + *time*|*word*) *p* determined assuming *t* follows Z-test (normal) statistics. †: $t = 2.43$, $p = 0.015$

Y:	Prevoicing (%)	Burst dur. (ms)	Vowel dur. (ms)
Intercept	1.31	37.21	93.23
Phoneme	56.21	-12.66	8.3
Time	0.77	0.99	12.79†
Phoneme:time	-44.87	9.44	-4.91
t (phoneme:time)	-8.24	3.28	-1.07
p	$1.7 \cdot 10^{-16}$	0.001	0.28

To obtain a rough estimate of the sizes of the effects of *time* (pre/post treatment) and *phoneme* identity (t/d) on the acoustic measurements, an analysis of linear mixed effect models was performed using maximal random factors (see Table 4, Figure 1) [3]. Models using random *phoneme* slopes did not converge and only random slopes for *time* were used. Interactions between *phoneme* identity and *time* were significant for *prevoicing* and *burst duration*.

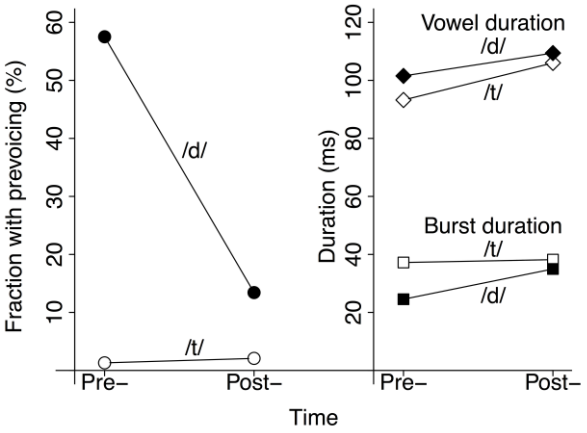

Figure 1. Visualization of the linear mixed effect models based on the estimates of Table 4

4. DISCUSSION

Several studies have compared differences between healthy laryngeal speech and speech after total laryngectomy [11]. After total laryngectomy the pharyngo-esophageal segment functions as the new voice source. The resulting tracheo-esophageal speech is has a lower F_0 and high noise components [2, 11]. In this study, post-treatment values of F_0, %V, HNR, and MVD worsened compared to the pre-operative values. The voices of the participants included in this research are deviant from healthy laryngeal voices. In the pre-operative situation, the voices are distorted due to either tumor presence on the vocal cords, earlier treatment with (chemo) radiotherapy and/or presence of a trachea cannula.

These acoustic changes in the vowel /a/ suggests an effect on acoustic features of consonants as we studied specific in the /t/ and /d/ tokens. This study has shown that the acoustic contrast of the /t/ and /d/ in word initial position reduces after total laryngectomy (Tables 2 and 3). *Prevoicing* and *burst duration* values for /t/ do not change significantly after surgery. For the /d/, presence of *prevoicing* decreases and *Burst duration* increases post-operatively. Values for /d/ become more like /t/ in the post-operative speech condition. *Vowel duration* following the initial consonant /t/ was higher in the post-operative condition compared to pre-operative condition (Tables 2 and 4). The duration of the vowel following the initial consonant /d/ also increased but this increase was not significant. An explanation might be an overall slowing of the speech rate.

With help of linear mixed effect models, the combined effects of *phoneme* identity and treatment were analyzed (Table 4, Figure 1). There was a strong interaction of *phoneme* identity and *time* (pre- / post-treatment), that confirmed the conclusion that the effect of treatment was limited to the initial /d/. And thus, that the difference between /t/ and /d/ became smaller after treatment.

For tracheo-esophageal speakers, the reduced acoustic contrast between /t/ and /d/ in initial position could lead to intelligibility issues. In Dutch /t/ and /d/ in phonology are marked as phonemes. Earlier research on tracheo-esophageal speech has shown that an intended /d/ was more often misheard as /t/ than that an intended /t/ was misheard as /d/ ([9], Table 3 confusion matrix). Our study provides evidence that *prevoicing* and *burst duration* changes for /d/ might explain at least part of these intelligibility issues.

The current study has some strengths and limitations. An advantage of the current research is that our analysis is performed on running speech. Recording running speech leads to most natural speaking conditions. Another strength of the study is the pre- and post-treatment within subject design. To our knowledge, there have not been studies that compared pre- and post-operative voice characteristics within the same individual. In this approach, changes in voice and speech can be spotted at an individual level.

Limitations of the current study include the use of different microphones and that the used text was not phonetically balanced. Only /t/ and /d/ in initial position were frequently present in the text which left out other phonemes with voicing distinction for analysis. Therefore, our data did not contain enough tokens containing other plosives (/b/ and /p/; /g/ and /k/) to investigate the effect of treatment on other places of articulation. The aerodynamics of voicing and the size of the air cavity before the constriction, suggest that these effects might be different for /g/ and /k/, for /b/ and /p/ than for /t/ and /d/.

5. CONCLUSIONS

The acoustic features of initial consonants /t/ and /d/ do move closer together, with /d/ becoming more like /t/ in tracheo-esophageal speech. This could explain results from earlier research that showed asymmetric confusion between /d/ and /t/ from these speakers. Further research on larger sets of running speech is recommended to create better understanding of the intelligibility issues. Total laryngectomy patients are a vulnerable group with communication deficits and need for speech rehabilitation. Special attention towards intelligibility issues is recommended.

6. ACKNOWLEDGEMENTS

The Netherlands Cancer Institute receives a research grant from Atos Medical (Malmö, Sweden), which contributes to the existing infrastructure for quality of life research of the Department of Head and Neck Oncology and Surgery. The authors have no other funding, financial relationships, or conflicts of interest to disclose.

The authors would like to express their gratitude to the Speech Language Pathologists of The Netherlands Cancer Institute for their contribution to the data collection.

7. REFERENCES

[1] Alphen van, P.M., Smits, R. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing, *Journal of Phonetics*, 32, 455-491.

[2] As van-Brooks, C.J., Koopmans-Beinum van, F.J., Pols, L.C., Hilgers, F.J. 2006. Acoustic signal typing for evaluation of voice quality in tracheoesophageal speech, *Journal of Voice*, 20, 355-368.

[3] Barr, D.J., Levy, R., Scheepers, C., Tily, H.J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal, *Journal of Memory and Language*, 68, 255-278.

[4] Bates, D., Mächler, M., Bolker, B., Walker, S. 2014. Fitting linear mixed-effects models using lme4, *arXiv* preprint arXiv:1406.5823.

[5] Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural brain research*, *125*(1-2), 279-284.

[6] Boersma, P., Weenink, D. 2017. PRAAT. Amsterdam: University of Amsterdam.

[7] Drugman, T., Rijckaert, M., Janssens, C., Remacle, M. 2015. Tracheoesophageal speech: A dedicated objective acoustic assessment, *Computer Speech & Language*, 30, 16-31.

[8] Gogh van, C.D., Festen, J.M., Verdonck-Leeuw de, I.M., Parker, A.J., Traissac, L., Cheesman, A.D., Mahieu, H.F. 2005.Acoustical analysis of tracheoesophageal voice, *Speech Communication*, 47, 160-168.

[9] Jongmans, P., Hilgers, F., Pols, L., As van-Brooks, C. 2006. The intelligibility of tracheoesophageal speech, with an emphasis on the voiced-voiceless distinction, *Logopedics Phoniatrics Vocology*, 31, 172-181.

[10] Most, T., Tobin, Y., Mimran, R.C. 2000. Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production, *Journal of Communication Disorders*, 33, 165-181.

[11] Sluis van, K.E., Molen van der, L., Son van, R.J., Hilgers, F.J., Bhairosing, P.A., Brekel van den, M.W. 2018. Objective and subjective voice outcomes after total laryngectomy: a systematic review, *European Archives of Oto-Rhino-Laryngology*, 1-16.

[12] Team R.C. 2015. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna.