Automatic tracheoesophageal voice typing using acoustic parameters

Renee P. Clapham[3,1]*, Corina J. Van As-Brooks*[1,2]*, Michiel W.M. Van den Brekel*[1,3]
Frans J.M. Hilgers[1,3]*, Rob J.J.H. Van Son*[1,3]

[1]Netherlands Cancer Institute; [2]Atos Medical, Horby, Sweden
[3]ACLC, University of Amsterdam, The Netherlands

R.P.Clapham@uva.nl, R.v.Son@nki.nl

Abstract

The acoustics of isolated vowels, e.g. of /a/, have in many studies been linked to pathological voice types, such as tracheoesophageal (TE) voice. To study the possibilities of objective and automatic classification of pathological TE voice types, the acoustic features of /a/ were quantified and subsequently classified using a suit of machine learning technologies. Best classification was achieved by using a voiced-voiceless measurement and the harmonics-to-noise ratio. Other common acoustic features were correlated to pathological type as well, but were less distinctive in classification. We conclude that for objective and automatic classification of TE voice pathology, voicing distinction and harmonics-to-noise ratio are most relevant.

Index Terms: Tracheoesophageal speech, pathological speech, machine learning

1. Introduction

Cancer of the larynx as well as most treatment modalities have a negative impact on a person's voice and speech quality. In the case of advanced laryngeal cancer, a total laryngectomy is often unavoidable. Although many patients develop functional alaryngeal speech by means of a prosthetic device to direct air towards the neo-glottis, voice quality is variable [1–3].

Presently, the prospects for the development of an adequate substitute voice due to the use of prosthetic devices are good [1, 3, 4]. Subsequent speech therapy will then aim at further improving voice quality and speech intelligibility. Studies have shown that improvements of speech quality and intelligibility can indeed improve the quality-of-life (QoL) of patients [3]. To support and evaluate voice quality after total laryngectomy and subsequent speech therapy, efforts have recently been made to introduce objective methods and automatic evaluations of the intelligibility and quality of alaryngeal speech [2, 5].

A three-type classification of voice quality on sustained vowels by Titze [6, 7] was adapted by Van As-Brooks [1, 8] to a four-type classification for tracheoesophageal speech (TES) on sustained /a/, i.e., acoustic signal typing (AST). Both classifications were based on spectrographic information of a sustained vowel. Both classification systems have consistent links to perceptual evaluation of voice quality of these speakers by speech and language therapists (SLTs).

The link between an objective classification system of voice pathology and the auditory perception of voice quality offers an opportunity to link objective and automatic acoustic measurements to the perception of pathology. Many studies have investigated the correlation between individual acoustic measures and TES voice pathology, see Table 1 for a short list.

It is clear that many acoustic variables are related to the severity of TES voice pathology. However, it is not clear how these should be weighted and combined to get a better understanding of TES voicing pathology. Ideally, one would like to be able to "predict" the AST class from acoustic parameter measurements alone. Such automatic classifications are the subject of machine learning [14–16].

The current study is part of an ongoing effort to understand the evaluation of TE speech and the development of diagnostic aids. The question we want to answer here is: To what extent can acoustic features of sustained /a/ contribute to predicting and understanding the severity of voice pathology in TES?

2. Materials and Methods

2.1. Speech recordings

We used a corpus containing sustained vowel /a/ of 87 TE speakers. Recordings were made between 1995 and 2009 as part of several unrelated studies. At the time of the recordings

Table 1: *Overview of acoustic parameters in studies investigating TES voice quality - limited to analysis of vowels. AST: Acoustic Signal Typing [1, 8], Perc.: Perceptual evaluation. ns: Not significant, +: Significant (versus normal), ⋆: Not included in this study. See section 2.3*

Acoustics	Reference	Type	Sign.
% Voiced	Kazi et al. (2009) [9]	Perc.	+
	Moerman et al. (2004) [10]	Perc.	ns
max. voice dur.	Moerman et al. (2004) [10]	Perc.	ns
F_0	van As-Brooks et al. (2006) [8]	AST	ns
	van Gogh et al. (2005) [11]	AST	ns
	Kazi et al. (2009) [9]	Perc.	ns
F_0 variability	van As-Brooks et al. (2006) [8]	AST	+
	van Gogh et al. (2005) [11]	AST	+
	Kazi et al. (2009) [9]	Perc.	+
Shimmer	Kazi et al. (2009) [9]	Perc.	+
Jitter	van As-Brooks et al. (2006) [8]	AST	ns
HNR	Maryn et al. (2009) [12]	Perc.	ns
	Moerman et al. (2004) [10]	Perc.	ns
	van As-Brooks et al. (2006) [8]	AST	+
HNR$<$700 Hz	van Gogh et al. (2005) [11]	AST	+
HNR$\geq$700 Hz	van Gogh et al. (2005) [11]	AST	ns
High freq noise	van Gogh et al. (2005) [11]	AST	ns
GNE	van As-Brooks et al. (2006) [8]	AST	ns
Rahmonic	Maryn et al. (2009) [12]	Perc.	ns
Intensity	van Gogh et al. (2005) [11]	AST	ns
BED	van As-Brooks et al. (2006) [8]	AST	+
$D_2^\star$, SampEn*	Yan et al. (2013) [13]	Perc.	+

all speakers provided informed consent allowing the recordings to be used for research purposes within the institute. In total there were 74 male and 13 female speakers. Age at treatment was 38-85 (median age 57). All speakers produced sustained /a/ vowels as part of a larger assessment battery. As some speakers had provided multiple recordings for various research projects over the 14 years, we selected the /a/ recording with the earliest recording date. At the time of recording, 83 speakers had a Provox1 or Provox2 prosthesis and the remaining four speakers had a Provox Vega prosthesis (three speakers a 22.5 Fr and one speaker a 17 Fr) [17–19].

Due to the fact that recordings were made over more than a decade as part of unrelated studies, a range of equipment and media were used for recording and storage, but this is not expected to alter acoustic measures below 5 kHz [20]. For this study, all recordings were first digitized and converted to 44.1 kHz sampling rate and 16-bit Signed Integer PCM encoding. No audio compression had been used on the recordings.

2.2. TEVA and Acoustic Signal Typing

The NKI developed a computer program (Tracheoesophageal Voice Analysis tool, TEVA [21]) to assist researchers and SLPs to identify acoustic signal types. TEVA runs as a Praat extension [22, 23] and both programs are available under an Open Source License (GPL). Acoustic signal type classification for TE speakers requires an observer to classify a segment of a spectrogram into one of four signal types: stable and harmonic (*1*), stable with at least one harmonic (*2*), unstable or partly harmonic (*3*) and barely harmonic (*4*) which corresponds roughly to a severity scale from *good* to *bad* [8]. As observers may differ in how they arrive at a classification, a consensus procedure was used for segment selection and classification into signal type.

Using the TEVA program, two experienced SLPs (authors Clapham and Van As-Brooks), classified all 87 recordings into signal type based on visual inspection of the spectrogram. They were blind to speaker characteristics (e.g. prosthesis type or gender) and were unable to listen to the recordings.

The spectrograms were classified according to AST over two steps. During step one, each rater independently classified the segment of the spectrogram that she considered most stable (1.75 seconds) and in step two, a consensus model was used whereby the raters first agreed on the segment of the spectrogram that was the most stable and then agreed on the AST of this stable segment. This interval of 1.75 seconds is shorter than the 2 seconds advised in [1,8] because several of the recordings had been segmented (i.e., the original unedited recordings were no longer available) meaning that the margins of the spectrogram would be invisible for stimuli with a length of 2 seconds. Inter-rater agreement was 58% before consensus with a correlation coefficient of R=0.75 between the AST values ($p<0.001$). See Table 3 for the distribution of the speakers over AST classes.

2.3. Acoustic measurements

The consensus intervals were used to measure the acoustic features. Table 2 lists the acoustic features which were selected for this study, based on the studies presented in Table 1. These features were automatically measured with Praat with a pitch floor of 40 Hz and a window size to 25 ms (see *AcousticMeasureScripts.praat* [24]). Where possible, we used published settings for measurements [1, 11]. MVD was determined on the whole /a/ realization. For practical reasons, the HNR_{low}, HNR_{high}, and cepstral rahmonic intensity as used by Van Gogh et al. [11] were substituted with the HNR of low-passed and band-passed speech, and the cepstral peak prominence (CPP), respectively. Formant quality factors (QF_1-QF_3) were added as non-voice measures. D_2 and Sample Entropy as proposed in [13] could not yet be implemented in Praat.

Table 2: *Overview of acoustic parameters used. With the exception of BED and QF_1-QF_3, all measures depend on the detection of voicing and pitch.*

Feature	Description
VF	Fraction of frames that are voiced
MVD	Maximum voicing duration
F_0	Standard deviation of F_0
Shimmer	
Jitter	
HNR	Harmonics-to-noise ratio (dB)
HNR_{low}	HNR low pass filtered speech (<700Hz)
HNR_{high}	HNR band pass (700Hz - 2300Hz)
GNE	Glottal noise energy
CPP	Cepstral peak prominence
BED	Band energy difference
QF_1-QF_3	F_1-F_3 quality factor (F_i/B_i)

2.4. Acoustic features and machine learning

Automatically evaluating AST based on acoustic information has aspects of both classification (identity) and regression (size): each signal type is distinct and derived from features in the spectrogram (classification), yet the signal types are also ordinal whereby prediction between classes can be seen as an intermediate value (regression). Model performance can be evaluated based on classification error when using a classification algorithm, on the root mean square (RMSE) when using a regression algorithm, or on the explained variance (e.g., correlation coefficient) between observed and predicted signal types.

Although it is not possible to find *the* best classification function in an efficient way, it is still possible to find *an* efficient classification function from examples. Using a variety of machine learning techniques and feature selection, it is also possible to estimate the robustness of the solution under different sets of examples [14–16]. These technologies can also be used to determine the importance or redundancy of individual and combinations of acoustic features for classification. Acoustic features were selected and ordered on explanatory importance using machine learning (ML) techniques as described in [15, 16]. All ML experiments were done using implementations in R [25] (see *model_AST.R* [24]). Seven ML algorithms were tested: Linear model (*LM*), Linear and Quadratic discriminant analysis (*LDA*, *QDA*), Support Vector Machines (*SVM*), Random Forest (*RF*), CaRT (*RPart*), and Neural nets (*NNet*).

Methods were used with their default settings in R [25]. The number of possible settings is too large to allow meaningful optimization for our data set. The results presented here should be interpreted as lower bounds on performance. All ML methods were tested in classification and regression mode. Where necessary, regression results were converted to classification, class *1-4*, by rounding (*LM*, *NNet*). Classification probabilities were converted to regression values by calculating the expected value (*LDA*, *QDA*).

A wrapper methodology with forward selection and backward elimination was used for feature selection [15, 16]. This means that each ML method was used as a black box that outputs a figure of merit given a training and feature set. Stratified bootstrap sampling validation, with 40-fold resampling, was used to check robustness of feature selection. Leave-one-

Table 3: *Effect of signal type (AST) on each acoustic variable (Kruskal-Wallis test), explained variance (R^2) using a linear model, median variable value, and post-hoc comparisons (Mann-Whitney tests, if Effect significant). P-values *: p<.0083, shaded: p<.0035 (correction for multiple comparisons). Exclamation mark highlights comparisons where exact significance cannot be computed due to ties within a category. AST class frequencies (c:N) - 1:14, 2:43, 3:13, 4:17. See Table 2 for abbreviations.*

	Effect	R^2	Median				AST comparisons (p values Mann-Whitney test)					
	p<	(*LM*)	*1*	*2*	*3*	*4*	*1-2*	*1-3*	*1-4*	*2-3*	*2-4*	*3-4*
VF	.000	0.595	1.00	0.95	0.34	0.06	.000 !	.000 !	.000 !	.001 !	.000 !	.002 !
MVD	.000	0.457	5.06	3.57	1.36	0.36	.007*	.000	.000 !	.000	.000 !	.000 !
F_0	.623	0.010	4.44	9.41	6.59	4.29						
Shimmer	.002	0.105	0.11	0.21	0.21	0.21	.001 !	.002	.003 !	.835 !	.269 !	.381 !
Jitter	.018	0.018	0.01	0.02	0.01	0.01	.008* !	.033	.359 !	.721 !	.048 !	.134 !
HNR	.000	0.355	10.42	3.74	2.18	1.38	.000	.000	.000	.107	.000	.118
HNR_{low}	.000	0.282	28.11	17.43	13.78	10.68	.001	.003	.000	.136	.000	.316
HNR_{high}	.000	0.042	15.63	7.46	8.09	9.25	.000	.012	.001	.732	.501	.892
GNE	.003	0.017	0.923	0.850	0.890	0.879	.013 !	.152 !	.152	.613 !	.306 !	.963 !
CPP	.007*	0.084	21.40	17.90	17.41	16.26	.002	.054	.002	.853	.204	.294
BED	.091	0.097	26.94	23.88	17.61	17.73						
QF1	.027	0.190	5.91	3.54	5.23	5.02	.039 !	.560 !	.450 !	.040 !	.027 !	.630 !
QF2	.798	0.010	6.90	6.20	6.70	6.5						
QF3	.307	0.015	5.85	10.00	9.70	7.00						

out cross-validation (LOOCV), where each sample is predicted using all but this sample as training set, was used to estimate the real predictive power of the models. Three recordings had no measurable voicing, and thus, no pitch related features. These were assigned predicted type *4*.

3. Results and Discussion

3.1. Single feature analysis

A summary of the relationship between acoustic features and observed AST is listed in Table 3. Nine of the acoustic features show a main effect for classification type and many of these can differentiate between signal type pairs (see post-hoc results in Table 3). A simple linear regression model using VF alone can explain almost 60% of the variance in classification. There are strong correlations between the acoustic features (not shown), the strongest between VF and MVD (R^2=0.71), HNR and HNR_{low} (R^2=0.58), and between VF and HNR (R^2=0.64). Purely random classification, using permutations, results in a correct classification of 0.33 (sd=0.04) and R^2=0.01 (sd=0.016).

ML methods were trained on the link between AST classification and single acoustical features. The best and median performances are presented in Figure 1. For both R^2 and correct classification, VF, MVD, HNR and HNR_{low} outperform the other features (in this order). Where median values are higher than chance performance plus two standard deviations (see Figure 1), the classification is likely robust. Otherwise, the performance is expected to be erratic. For the leftmost four features (VF, MVD, HNR, HNR_{low}), the classifications seem to be robust. For neither classification nor regression do QF3, QF2 or HNR_{high} reach this level of significance.

3.2. Feature combinations

Bootstrap validation versus LOOCV and forward selection versus backward elimination all resulted in comparable feature selection and performance (not shown). Classification outperformed regression slightly, but was otherwise comparable. Only results for classification with LOOCV and forward selection will be reported unless indicated otherwise.

Classification performance is plotted versus the number of acoustic features in Figure 2 for all ML algorithms used. The ML methods split into two groups: *LDA*, *QDA*, *SVM*, and *RPart*

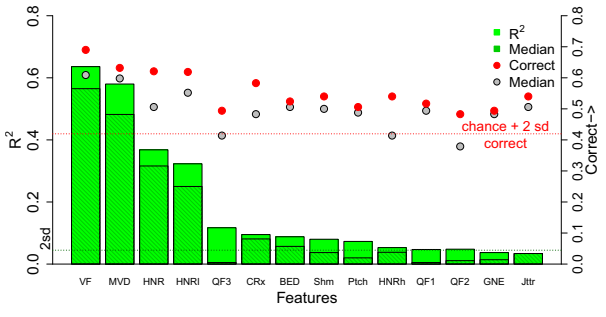

Figure 1: *Maximum R^2 (green bars) and correct classification (red circles) for single acoustical features over all ML methods (ordered on decreasing R^2, LOOCV, see text). Median R^2 and correct classification are indicated with cross-hatched bars and grey circles, respectively. Added are chance level + 2·sd lines for R^2 (0.045, green) and correct classification (0.42, red).*

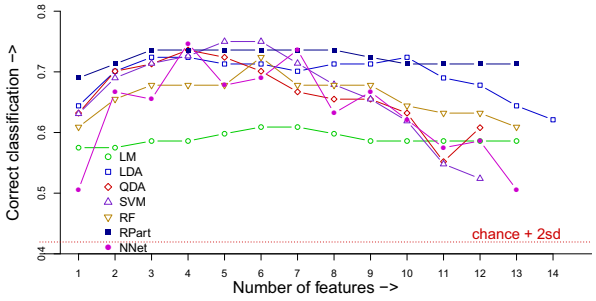

Figure 2: *Correct classification as a function of the number of acoustic features for all ML algorithms (see section 2.4). Features were included using forward selection and leave-one-out cross validation. Added is a chance level + 2·sd line (0.42, red). R^2 values follow comparable curves (not shown, see text).*

all reach high correct classification rates with only three acoustic features. The remaining methods, *LM*, *RF*, and *NNet*, perform worse. The inherently stochastic nature of *RF*, and *NNet* might at least partly explain their erratic results on small data sets. "Good" runs were selected for these two methods.

Classification performance as a function of the number of features varies widely between methods. Complex ML methods such as *SVM*, *QDA*, and *RF* are sensitive to the "curse of dimensionality": Including more features leads to marked decreases in performance due to overtraining [15, 16]. However, *RPart* drops features that do not increase performance. This might be an explanation for the stable performance of *RPart*.

The order of selection of features was investigated with bootstrap validation (see section 2.4). All methods select either VF or MVD as their first feature. The second feature is then one of the HNR features (HNR, HNR_{high}, or HNR_{low}). The third feature selected is more varied, either another from MVD, VF, or the HNR group, but also QF2 and BED were selected (*LM*, *LDA*). The LOOCV results were equivalent, but varied somewhat in the third selected feature (Figure 2).

From this we conclude that VF and MVD alone supply enough information to get well over 60% correct classification. Including the HNR group of features then allows performance to rise to over 70% correct classification (see Figure 2). Members of these groups often appear again as third selected feature, indicating they are not completely interchangeable (redundant). With three features, the high-performance methods get over 70% correct. Increasing the number of features can sometimes improve performance even to 75% correct classification with five features, e.g., for *QDA* and *SVM*. However, differences become rather small and unreliable for our data set. For all ML methods it was found that an analysis which excluded VF, where MVD would substitute for it, resulted in slightly lower performance, still reaching ∼70% correct classification and R^2 up to 0.6 (note, *RPart* performed *better* without VF).

AST classification is also an ordinal scale. Therefore, not all classification errors should be weighted equal. An AST class *1-4* confusion is worse than a class *3-4* confusion. The squared correlation coefficient (R^2) between predicted and consensus classification is a figure of merit that measures such discrepancies. For all ML methods, except *LM*, the R^2 peaked between 0.6 and 0.7 at best classification performance in Figure 2. That is, the ML methods were able to explain more than 60% (close to 70% for *SVM*) of the variance in the consensus classification.

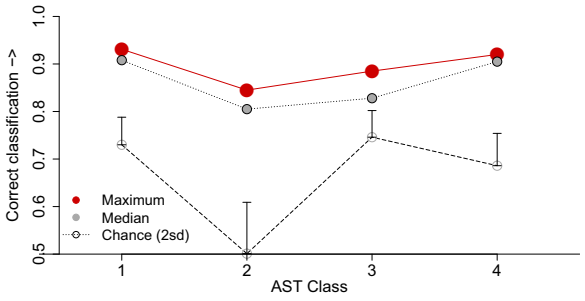

Figure 3: *Best classification for individual types versus all others (see text). Presented are best and median ML classifier performance. Added is chance level for each class distinction and 2·sd error bars for random classification.*

Table 3 presents several other features beside VF, MVD, and HNR that show statistically significant differences between AST classes, e.g., Shimmer, GNE, CPP. However, it seems the ML algorithms applied here are unable to use this information to improve classification. The analysis presented in Figure 2 was repeated with the exclusion of VF and MVD. With this exclusion, classification was regularly over 0.6 but R^2 came only slightly over 0.4 (≤43% explained variance). The first feature selected was always either HNR or HNR_{low}.

When excluding all of VF, MVD, and the HNR group of features, correct classification peaked at 0.62 (for *QDA* with BED + Shimmer), but was well below 0.6 for all other methods (not shown). This might seem high considering the chance level was 0.33. However, R^2 was rarely above 0.2, and generally lower (≤20% explained variance). This indicates that classification errors became much more random. Information in these acoustic features might mainly identify individual classes (c.f., Table 3). The first feature selected under these conditions was four times CPP, and BED, Shimmer, and QF3 each once.

3.3. Individual class type identification

Best performance of AST classification might not be attained using a single model for all types. The above analysis was repeated, but now as four two-type classification tasks. All ML methods (see section 2.4) were trained and tested on a single type with all other types merged into a single class, e.g., type *2* against types *1*, *3*, and *4* combined. Chance classification performance was recalculated for each combination. The results are presented in Figure 3.

As expected, the end-point types *1* and *4* were easier to identify than the inner types *2* and *3*. Behavior of the classifiers was more erratic than with the original four type task. *SVM* could not even classify type *3* versus the others. Number and selection of features varied much more than the patterns seen in Figure 2. This is likely caused by the unbalance between positive and negative samples in this task.

4. Conclusions

Many acoustic measurements correlate, often strongly, with the AST classification (see Tables 1, 3). However, our study shows that only two groups of features can perform a classification to any reasonable extent: voice detection (VF and MVD) and the harmonics-to-noise ratio (HNR, HNR_{low}, and HNR_{high}). Other factors improve classification performance only marginally. This indicates that the presence and duration of voicing and the harmonic-to-noise ratio are the most salient acoustic features that can be used to classify a TE signal into its acoustic signal type. In our study, *QDA* and *SVM* performed best, but *RPART* would perform almost as good and can easily be assessed automatically. A practical tool incorporating these methods will be made available online at [21].

The fact that classical measures of glottal voices, like jitter and shimmer, are less salient in TE speech can possibly be attributed to the inherent instability of neo-glottis vibrations [13].

5. Acknowledgements

This research was made possible by an unrestricted research grant from Atos Medical, Sweden and a grant from the Verwelius Foundation, Naarden. We thank Louis Pols for his helpful comments and suggestions.

6. References

[1] van As, C.J., "Tracheoesophageal speech. a multidimensional assessment of voice quality," Ph.D. dissertation, University of Amsterdam, Sep. 2001.

[2] Haderlein, T., Nöth, E., Toy, H., Batliner, A., Schuster, M., Eysholdt, U., Hornegger, J., and Rosanowski, F., "Automatic evaluation of prosodic features of tracheoesophageal substitute voice," *European Archives of Oto-Rhino-Laryngology*, vol. 264, no. 11, pp. 1315–1321, 2007.

[3] Jongmans, P., "The intelligibility of tracheoesophageal speech: An analytic and rehabilitation study," Ph.D. dissertation, University of Amsterdam, 2008.

[4] Op de Coul, B.M.R., Ackerstaff, A., van As-Brooks, C.J., Van Den Hoogen, F.J.A., Meeuwis, C.A., Manni, J.J., and Hilgers, F.J.M., "Compliance, quality of life and quantitative voice quality aspects of hands-free speech," *Acta oto-laryngologica*, vol. 125, no. 6, pp. 629–637, 2005.

[5] Moerman, M., Martens, J.P., Van der Borgt, M., Peleman, M., Gillis, M., and Dejonckere, P., "Perceptual evaluation of substitution voices: development and evaluation of the (i)infvo rating scale," *European Archives of Oto-Rhino-Laryngology*, vol. 263, no. 2, pp. 183–187, 2 2006.

[6] Titze, I., "Workshop on Acoustic Voice Analysis: Summary Statement." Denver, CO: National Center for Voice and Speech, 1995, pp. 1–36.

[7] Zhang, Y., Jiang, J. *et al.*, "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies." *Journal of voice: official journal of the Voice Foundation*, vol. 22, no. 1, p. 1, 2008.

[8] van As-Brooks, C.J., Koopmans-van Beinum, F., Pols, L., and Hilgers, F., "Acoustic signal typing for evaluation of voice quality in tracheoesophageal speech," *Journal of Voice*, vol. 20, no. 3, pp. 355–368, 2006.

[9] Kazi, R., Kanagalingam, J., Venkitaraman, R., Prasad, V., Clarke, P., Nutting, C., Rhys-Evans, P., and Harrington, K., "Electroglottographic and perceptual evaluation of tracheoesophageal speech," *Journal of Voice*, vol. 23, no. 2, pp. 247–254, 2009.

[10] Moerman, M., Pieters, G., Martens, J.P., Van der Borgt, M.J., and Dejonckere, P., "Objective evaluation of the quality of substitution voices," *European Archives of Oto-Rhino-Laryngology*, vol. 261, no. 10, pp. 541–547, 11 2004.

[11] Van Gogh, C., Festen, J., Verdonck-de Leeuw, I., Parker, A., Traissac, L., Cheesman, A., and Mahieu, H., "Acoustical analysis of tracheoesophageal voice," *Speech Communication*, vol. 47, no. 1, pp. 160–168, 2005.

[12] Maryn, Y., Dick, C., Vandenbruaene, C., Vauterin, T., and Jacobs, T., "Spectral, cepstral, and multivariate exploration of tracheoesophageal voice quality in continuous speech and sustained vowels," *The Laryngoscope*, vol. 119, no. 12, pp. 2384–2394, 2009.

[13] Yan, N., Ng, M.L., Wang, D., Zhang, L., Chan, V., and Ho, R.S., "Nonlinear dynamical analysis of laryngeal, esophageal, and tracheoesophageal speech of cantonese," *Journal of Voice*, vol. 27, no. 1, pp. 101–110, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0892199712001014

[14] Maindonald, J. and Braun, J., *Data analysis and graphics using R: an example-based approach*. Cambridge University Press, 2006, vol. 10.

[15] Guyon, I. and Elisseeff, A., "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[16] Ladha, L. and Deepa, T., "Feature selection methods and algorithms," *International Journal on Computer Science and Engineering*, vol. 3, no. 5, pp. 1787–1797, 2011.

[17] Hilgers, F.J., Ackerstaff, A.H., Balm, A.J., Tan, I.B., Aaronson, N.K., and Persson, J.O., "Development and clinical evaluation of a second-generation voice prosthesis (provox® 2), designed for anterograde and retrograde insertion," *Acta oto-laryngologica*, vol. 117, no. 6, pp. 889–896, 1997.

[18] Hilgers, F.J. and Balm, A.J., "Long-term results of vocal rehabilitation after total laryngectomy with the low-resistance, indwelling provoxtm voice prosthesis system," *Clinical Otolaryngology & Allied Sciences*, vol. 18, no. 6, pp. 517–523, 2007.

[19] Hilgers, F.J., Ackerstaff, A.H., van Rossum, M., Jacobi, I., Balm, A.J., Tan, I.B., and van den Brekel, M.W., "Clinical phase i/feasibility study of the next generation indwelling provox voice prosthesis (provox vega)," *Acta oto-laryngologica*, vol. 130, no. 4, pp. 511–519, 2010.

[20] van Son, R.J.J.H., "A study of pitch, formant, and spectral estimation errors introduced by three lossy speech compression algorithms," *Acta acustica united with acustica*, vol. 91, no. 4, pp. 771–778, 2005.

[21] van Son, R.J.J.H., "NKI TE-VOICE Analysis tool (TEVA)," Computer program: http://www.fon.hum.uva.nl/IFA-SpokenLanguageCorpora/NKIcorpora/NKI_TEVA/, 2012.

[22] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer," Computer program: http://www.Praat.org/, 2009.

[23] Boersma, P., "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, pp. 341–345, 2001. [Online]. Available: http://www.Praat.org/

[24] van Son, R.J.J.H., "Additional Files to Interspeech 13 proceedings," Link: http://www.fon.hum.uva.nl/IFA-SpokenLanguageCorpora/NKIcorpora/NKI_TEVA/, 2013.

[25] R Core Team, "The R project for statistical computing, version 2.15.2 (2012-10-26)," Computer program : http://www.r-project.org/, 1998–2012.