SCIENTIFIC REPORTS
nature research

OPEN

Machine learning based identification of relevant parameters for functional voice disorders derived from endoscopic high-speed recordings

Patrick Schlegel✉, Stefan Kniesburges, Stephan Dürr, Anne Schützenberger & Michael Döllinger

In voice research and clinical assessment, many objective parameters are in use. However, there is no commonly used set of parameters that reflect certain voice disorders, such as functional dysphonia (FD); i.e. disorders with no visible anatomical changes. Hence, 358 high-speed videoendoscopy (HSV) recordings (159 normal females (N_F), 101 FD females (FD_F), 66 normal males (N_M), 32 FD males (FD_M)) were analyzed. We investigated 91 quantitative HSV parameters towards their significance. First, 25 highly correlated parameters were discarded. Second, further 54 parameters were discarded by using a LogitBoost decision stumps approach. This yielded a subset of 12 parameters sufficient to reflect functional dysphonia. These parameters separated groups N_F vs. FD_F and N_M vs. FD_M with fair accuracy of 0.745 or 0.768, respectively. Parameters solely computed from the changing glottal area waveform (1D-function called GAW) between the vocal folds were less important than parameters describing the oscillation characteristics along the vocal folds (2D-function called Phonovibrogram). Regularity of GAW phases and peak shape, harmonic structure and Phonovibrogram-based vocal fold open and closing angles were mainly important. This study showed the high degree of redundancy of HSV-voice-parameters but also affirms the need of multidimensional based assessment of clinical data.

In the field of laryngology, high-speed videoendoscopy (HSV) is an assessment technique about to be established in clinics[1,2]. This technique is already commonly used in research settings and large clinics to investigate the oscillations of the vocal folds in the larynx, forming the basis signal for our voice[3–5].

During voice production an airstream rises from the lungs through the trachea and sets the vocal folds in motion. Vibrating at oscillation frequencies between 150 and 400 Hz[6], the vocal folds divide the continuous airstream in a series of flow pulses producing the fundamental frequency and basis signal of the voice. The flow pulses are then further modulated by the vocal tract, tongue and lips producing audible voice and speech[7,8]. However, during singing the vocal folds can vibrate much faster. Oscillation frequencies of up to 1568 Hz with complete glottal closure are reported[9].

In general, periodic and symmetric vocal fold oscillations with complete glottal closure indicate a healthy voice[10–12]. Respectively asymmetric, aperiodic oscillations or a large continuously open part of the glottis indicate a voice disorder[13–15]. Different systems exist to classify voice disorders, such as subdivisions in central and peripheral dysphonias; neurogenic, psychogenic and myogenic dysphonias or mucosal and neuromuscular disorders[16]. In this work, the European classification in organic and functional voice disorders will be used since only healthy subjects and subjects suffering from functional dysphonia (FD) were investigated.

FD is a diagnosis of exclusion meaning that the subject has no organic voice disorders i.e. visible changes in the vocal tract or injuries of the vocal folds[17]. Symptoms of FD include hoarseness, changes in pitch or other changes in voice quality[16]. Also, purely psychologic causes are in the range of possibilities[18].

Department of Otorhinolaryngology, Division of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany. ✉e-mail: patrickschlegel93@yahoo.de

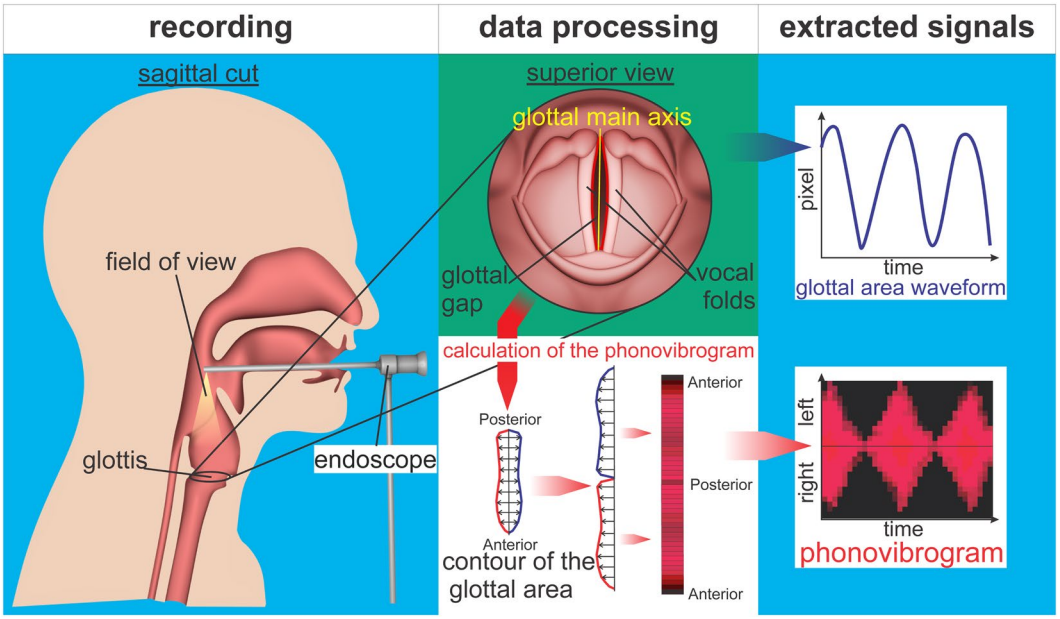

Figure 1. HSV recording using a rigid endoscope yielding 1D-GAW and 2D-PVG signals.

The current "gold standard" in clinics to investigate vocal fold oscillations and respectively voice disorders is stroboscopy[19–21]. However, stroboscopy only produces an artificial slow motion perspective of the vocal fold vibration and therefore data that cannot be interpreted in the case of irregular vibrations; HSV does not have this disadvantage[22].

As depicted in Fig. 1, during HSV recording, a rigid endoscope is inserted in the throat of the subject recording the vocal folds from above. The fast oscillations of the vocal folds are recorded with sampling rates of about 4000 Hz severely exceeding the oscillation frequencies of the vocal folds[22]. Based on the recorded video data the 1D Glottal Area Waveform (GAW) that represents the area between the vocal folds (the "glottal area") over time can be computed; i.e. the GAW is the function of glottis-pixels over time (see Fig. 1, top right). Another (2D) signal determinable from HSV-recordings is the Phonovibrogram (PVG) introduced by Lohscheller *et al.*[23]. The PVG depicts the whole oscillation pattern of the contour of the glottal area over time in one image, as shown in Fig. 1, bottom right.

From these signals different kinds of GAW- and PVG-based parameters are calculable[24].

In recent years machine learning based approaches have grown in popularity in voice research[25–27]. Machine learning was also used in combination with parameters to separate healthy from disordered voices[28–30]. Callan *et al.* trained a self-organizing map using acoustic parameters to differentiate normal from disordered voices and achieved an overall accuracy of 0.76[28]. Awan and Roy achieved 0.75 accuracy for separation of normal, breathy, rough and hoarse voices also using acoustic parameters[29]. PVG based parameters were used by Voigt *et al.* to differentiate normal and FD voices with 0.81 accuracy[30]. Also a few more recent studies were published reporting accuracies of up to 1.00 using only acoustic measures[31,32]. However, because of the perfect accuracies stated, the reliability of these findings may be questionable.

It is known that many features associated with FD (for instance incomplete glottis closure) also frequently occur in healthy subjects[33]. This indicates that multidimensional approaches applying different parameters are needed to separate healthy and disordered subjects. Furthermore, many parameters describing laryngeal features are redundant[5,34]. However, the redundancies of parameters are not yet fully explored, and it is not known which parameters best characterize FD or other voice disorders. For this reason, this study uses a multidimensional approach investigating GAW- and PVG-based parameters in regard of their linear dependencies and expressiveness in differentiating healthy and disordered voices. The aims of this work are:

1. Determine linear relations between a large set of parameters using clinical data and discard redundant parameters.
2. Investigate GAW- and PVG-based parameters and which combination of them is best suited for separating healthy from FD subjects.
3. Discuss why the final parameter set is able to differentiate groups and which features of the vocal fold oscillation process are described by these parameters.

Methods

358 HSV recordings of 260 female and 98 male subjects were investigated. The recordings were taken using a 70° rigid endoscope attached to a clinically used Photron Fastcam MC2 camera (frame rate: 4000 fps, resolution: 512×256 pixels). All subjects phonated the vowel /i/ at a comfortable (i.e. habitual) pitch and loudness level (sustained phonation) and all recordings had a length of at least 250 ms. The study was approved by the ethic

	Healthy	Disordered
Females	159 (N_F)	101 (FD_F)
Males	66 (N_M)	32 (FD_M)

Table 1. Number of HSV-videos in healthy and disordered groups.

committee of the Medical School at Friedrich-Alexander-University Erlangen-Nürnberg (no. 290_13B) and all methods were carried out in accordance with relevant guidelines and regulations. Written consent was obtained by the subjects. Recordings of females and males were each subdivided into one healthy group and one disordered group:

- Recordings of healthy subjects with normal sounding voices (Females: N_F, Males: N_M).
- Recordings of disordered subjects before treatment (Females: FD_F, Males: FD_M).

Disordered patients were diagnosed by our clinicians. All disordered patients have only FD with no concurrent organic disorders. Table 1 contains the numbers of recordings from female and male subjects in healthy (N_F, N_M) and disordered (FD_F, FD_M) groups. For each subject one HSV-recording was performed.

Segmentation of the glottal area. The glottal area of the collected videos was segmented using an in house developed version of Glottis Analysis Tools (GAT-2018). At the moment, GAT is used by 27 voice groups in 7 countries. A screen shot of GAT featuring glottis segmentation is depicted in Fig. 2.

The segmentation process is illustrated in Fig. 3. A summary of the process is given here:

1. A section of the video containing the entire glottis region was selected.
2. A segment of 1000 frames (250 ms) of the video was selected during which the subject holds sustained phonation.
3. Seed points were chosen and the brightness thresholds were adjusted to segment the dark glottal area between the vocal folds.
4. The contour of the glottal area was calculated as described in[35] and a midline was selected dividing the total glottal area in two sides. Left (GAW_L) and right (GAW_R) partial GAWs were computed for each side by numerical integration over the distances between the midline and the left and right contour lines.
5. The total GAW of the entire area (GAW_T), GAW_L and GAW_R as well as the Phonovibrogram (PVG) for all 1000 segmented frames were extracted (for a detailed explanation of the PVG see[23]).

Parameter computation. For each of the 358 recordings one GAW_T, GAW_L, GAW_R and PVG signal were calculated. Extremum based cycles were determined for the GAW_T signals and conferred to the PVG, GAW_L and GAW_R. Then for each GAW_T, 41 parameters were computed. 18 symmetry parameters were calculated using GAW_L and GAW_R and further 32 parameters based on PVG. In the supplementary information in Table S1, names, abbreviations, sources and descriptions of all 91 parameters (starting parameter set: $\mathbf{HSV_0}$) are given. Parameters were calculated for maximum based cycles (i.e. each cycle starts at a sufficiently distinct local maximum and ends before the next distinct local maximum) with exception of PhA[Mean] PhA[StD], PhAI[Mean] and PhAI[Std] which required minimum based cycles (analogously to maximum based cycles but using local minima instead) by their definition. The following investigations were performed using MATLAB (version 9.3.0.713579, R2017b).

Linear dependencies. In a first step the parameters were investigated for linear dependencies by calculating Pearson Correlation Coefficients (PCC) between all parameters over all healthy and disordered groups. Parameters being correlated "very high" (corr. ≥ 0.9 following the suggestions of Mukaka[36]) were removed. Furthermore, based on previous studies[5], four additional parameters were removed that were only correlated "high" ($0.9 >$ corr. ≥ 0.7).

By calculating PCCs over all healthy and disordered groups, regardless of health status or gender, only correlations were found that were consistent for all cases i.e. correlated parameters behave strongly similar for all data. This implies that the parameters are redundant. For this reason based on the found PCCs, the parameter set $\mathbf{HSV_0}$ was reduced yielding parameter set $\mathbf{HSV_1}$.

Influence of subject age. A large difference in age between healthy and disordered groups exists. This is a common problem in clinical studies[37,38]. For this reason, it was investigated if subject age had a substantial influence on parameters for females and males. In Fig. 4, the age distribution of the healthy and disordered subjects is shown for females and males.

PCCs between each parameter and the age of the subjects for the groups FD_F and FD_M were calculated. The influence of subject age was investigated only for disordered subjects since all healthy subjects had a similar age (see Fig. 4). Furthermore, the p-value and a confidence interval of each PCC were calculated using the Matlab function "corrcoef"[39]. The p-value states if the correlation is statistically significantly different from zero (alpha = 0.05). The confidence interval calculated with Matlab is an estimator of the 95% confidence interval of the calculated PCC (see[39]). In this way a statement regarding the degree of linear dependencies between parameters and age can be made.

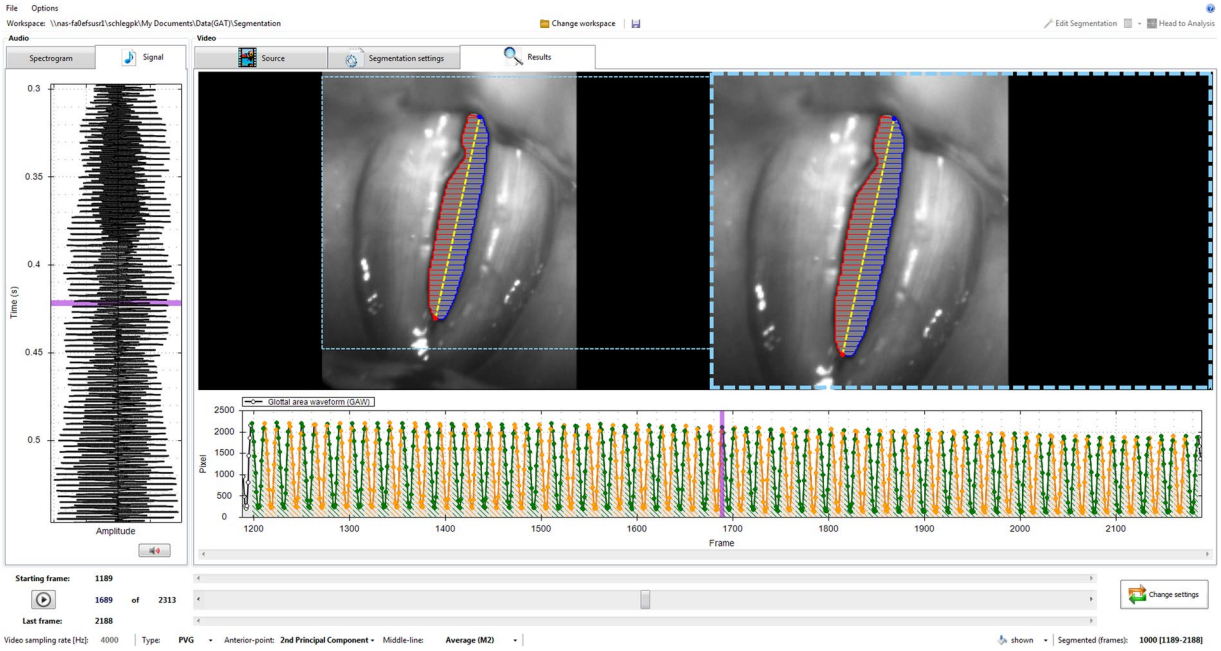

Figure 2. Glottis segmentation using Glottis Analysis Tools.

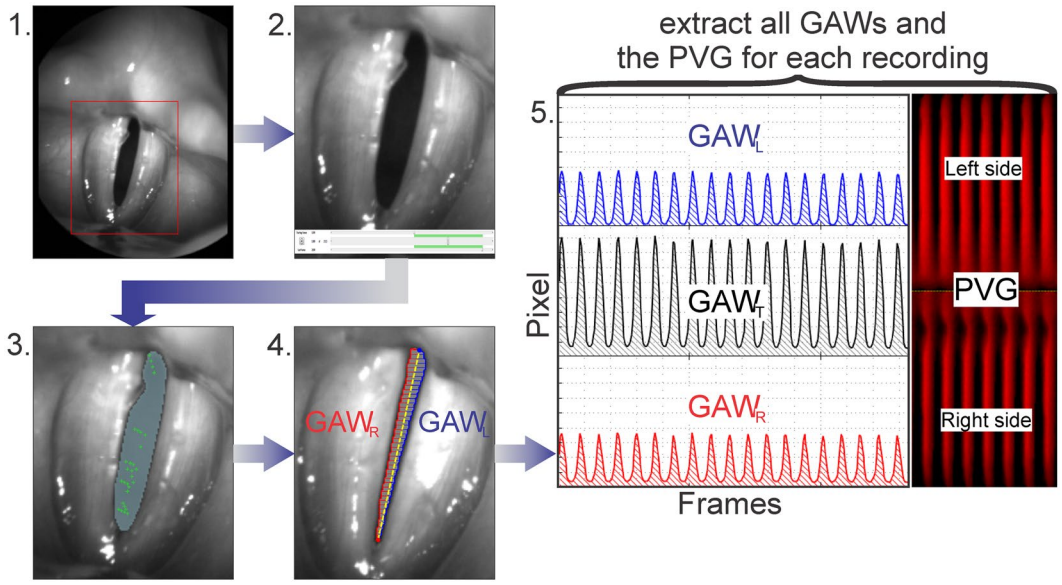

Figure 3. Segmentation Process. 1. Selection of the glottis region; 2. Selection of a 1000 frames section; 3. Segmentation of the glottal area; 4. Calculation of the partial GAWs (GAW_L and GAW_R) and 5. Extraction of all GAWs (GAW_T, GAW_L and GAW_R) and the PVG.

Following the suggestions of Mukaka a correlation was seen as negligible if it was 0.3 or lower[36]. Only a little number of "low" correlations (between 0.3 and 0.5[36]) were detected and no PCC was higher than 0.5. For this reason, the influence of subject age on this data was seen as negligible. Also, non-linear dependencies were investigated by reviewing scatter-plots of the parameter values against subject age but no obvious relations were found.

Model selection and optimization. Exclusion of redundant parameters yielded parameter set HSV_1. Now, two group comparisons were used for classification:

1 N_F vs. FD_F
2 N_M vs. FD_M

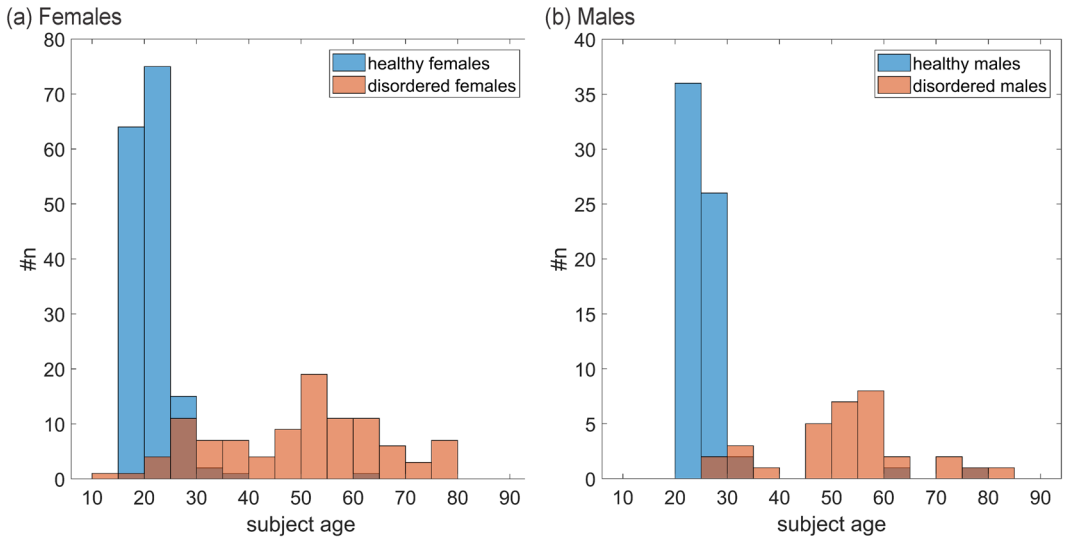

Figure 4. Distribution of subject age; for (**a**) females and (**b**) males for healthy and disordered groups with #n being the number of subjects.

For each comparison, models applying the supervised learning classification approach of single level boosted trees (also: boosted stumps) were generated. This approach uses trees consisting of one node and two leaves each for data separation. After each added tree stump, data weights are recalculated allowing the separation of otherwise hardly separable subjects (this process is called "boosting")[40,41]. We decided to use boosted stumps, since they performed comparatively well for the separation of a range of different data sets using various classification performance measures in comparison to other classification algorithms[42]. However, to avoid overfitting we decided to use boosted stumps instead of boosted trees, which achieved best overall performance in class separation[41,42]. For all models the "name value pair arguments" of the MATLAB function "fitcensemble"[43] that was used for model generation were set as follows:

1 'prior' was set to 'uniform' because of imbalanced class sizes,
2 'surrogate' was set to 'off' since no data was missing,
3 'MaxNumSplits' was set to 1 to avoid overfitting (i.e. trees consisted of only one node),
4 'LearningRate' was set to 0.1 for training with shrinkage to find a better optimum.

For performance measure calculation, ten-fold cross validation was used. To prevent influences by random partitioning, each model was recalculated ten times. All performance measures were averaged over testing partitions and recalculated models. In the following **five steps**, it will be shown how the models were generated:

Step 1 - Determine the boosting algorithm best suited for this problem: Three boosted decision stumps algorithms "AdaBoost", "LogitBoost" and "RUSBoost" were investigated. "AdaBoost" was included since it is one of the most widely applied boosting algorithms and hence a common choice[41]. "LogitBoost" is an algorithm designed for hardly separable classes and "RUSBoost" is designed for unbalanced class sizes[43]. Both are the case for our data.

Algorithm performance was rated using the performance measures, area under curve (AUC) and accuracy (ACC) (the higher the better). These measures complement each other to some degree. ACC can be misleadingly high for unbalanced class sizes but AUC is not influenced by class sizes. On the other hand, AUC can be misleadingly low for extremely sharply separated classes. However, for the final models, also sensitivity and specificity are given to show that no class is overly preferred[44].

Further, it was investigated how much these algorithms weighted two added random parameters (a normal and an equally distributed variable) by measuring feature importance (FI). FI is a measure that states how important each feature (i.e. parameter) is for group separation (for more details see[45]). Therefore it is expected that the two added random parameters only achieve low importance. If random parameters would achieve high FI the algorithm would be unsuitable for this investigation.

Step 2 - Determine the number of decision stumps to include in the model: Applying the best algorithm determined in **Step 1**, models consisting of one up to 500 consecutive tree stumps were generated (without random parameters). AUC and ACC were plotted over the number of included stumps, i.e. model complexity. Based on these plots, an optimal number of stumps was chosen for the following models.

Step 3 - Find the parameters that achieve the best result in separating N_F vs. FD_F: The FI for HSV_1 parameters was determined for the group comparison N_F vs. FD_F. Afterwards, models (as many as remaining parameters) were generated: The first of these models included only the parameter that was rated most important by FI, the second model included the parameter that was rated most important and the parameter rated second most important by FI and the last model included all parameters. From these models, one model was selected that

Correlated parameter values	Kept value	Reasoning
corr $\geq$ 0.9		
AP [Mean], AP [Std]*, APQ3, APQ5, APQ11, MShim, APF	MShim	Widely applied, straightforward
TP [Mean], Jit(%), PPQ3, PPQ5, PPQ11, PPF, RAP_K	Jit(%)	Widely applied, straightforward
EPQ3, EPQ5 EPQ11, EPF	EPF	Unexpected behavior found for EPQ-based parameters in[5]
PhAI[Mean], WaSI[Mean]	PhAI[Mean]	Faster to calculate
AmSI[Std], AmS[Std]*, DyRSI[Std], DyRS[Std]*	AmSI[Std]	Consistent with[5]
PhA[Std], PhAI[Std]	PhAI[Std]	No risk of cancellation of inverse phase shifts
SpA[Std], SpAI[Std]	SpAI[Std]	Consistent with PhAI[Std]
CAS^{CA}[Std], $CASI^{CA}$[Std]	$CASI^{CA}$[Std]	Otherwise possible under- estimation of asymmetry because of cancellation effects
0.9 > corr $\geq$ 0.7		
TP[Std], F0[Std]	F0[Std]	TP [Mean] already removed
DyRS[Mean], AmS[Mean]	AmS[Mean]	Consistent with[5]
DyRSI[Mean], AmSI[Mean]	AmSI[Mean]	Consistent with[5]

Table 2. Groups of redundant parameters (corr: $\geq$ 0:9) It is stated which of multiple parameters are kept and why. 25 out of 36 parameters were discarded. The *-symbol indicates that some of the correlations of this parameter in this group are marginally below 0.9 for some cases.

achieved high AUC and ACC with only a small number of parameters. The parameters included in this model were rated as best set of parameters for this model comparison.

Step 4 - Find the parameters that achieve the best result in separating N_M vs. FD_M: Analogous to Step 3 but for the group comparison N_M vs. FD_M

Step 5 - Find the combined parameter set that best separates female and male group comparisons: Models including different combinations of the parameters found in **Step 3** and **Step 4** were generated. Investigation of all possible combinations was not feasible. Therefore, only certain combinations (e.g. only PVG or GAW based parameters) were investigated. A final parameter set ($\mathbf{HSV_2}$) that achieved the best compromise between high performance measures and a low number of parameters for both comparisons N_F vs. FD_F and N_M vs. FD_M was determined.

Results and Discussion

Parameters were reduced in two main steps yielding parameter sets $\mathbf{HSV_1}$ and $\mathbf{HSV_2}$. In the following the steps leading to these parameter sets and their possible applicability are discussed.

Linear dependencies. Table 2 shows the parameters that were correlated "very high" (corr $\geq$ 0.9[36]). It is stated which of the parameters were kept and why. After discarding 25 of 91 parameters, the parameter set HSV_1 consisting of 66 parameters remains. The 25 discarded parameters are marked in Table S1.

It is stated which of multiple parameters are kept and why. 25 out of 36 parameters were discarded. The *-symbol indicates that some of the correlations of this parameter in this group are marginally below 0.9 for some cases.

By only excluding parameters that were correlated very high across all subjects, a conservative approach on parameter reduction was taken. Since the correlation was consistently high, it is reasonable to assume that it is due to the mathematical similarity of the underlying parameters. Parameters contained in HSV_1 may not be completely independent but all obviously superfluous parameters were removed.

Influence of subject age. Influence of age was investigated for all HSV_1 parameters. The calculated PCCs and estimated confidence intervals for groups FD_F and FD_M are listed in the supplementary information in Tables S2 and S3. Table S2 contains GAW-based parameters i.e. based on GAW_T (or GAW_L and GAW_R in case of symmetry measures). Table S3 contains PVG-based parameters. The highest absolute correlation values considering all parameters were -0.335 for $CA^{R,OP}$ [Mean] in FD_F and -0.497 for $CASI^{CA}$[Mean] in FD_M. The scatter plots of these parameters for the respective groups against subject age are shown in Fig. 5. In addition, each plot contains a fitted line. Investigating the scatter plots in Fig. 5, no clear linear or nonlinear coherence between age and the depicted parameters is evident. Scatterplots of the remaining parameters were similar. Therefore, it was concluded that correlations of parameters with age are negligible for this study.

Model selection and optimization. **Step 1:** The Algorithm judged as best was LogitBoost since it provided the highest AUC and ACC on average for both group comparisons (for models with and without added random parameters) and still did not rate the random parameters as important. This is also illustrated in Fig. 6 depicting (a1/b1) the normalized FI of the ten parameters rated most important and (a2/b2) average AUC and ACC for all three tested algorithms.

Step 2: A number of 300 stumps was chosen for the following models, since neither for females nor for males AUC and ACC increase after approximately 300 stumps are reached (See Fig. 7).

Step 3 and 4: In Fig. 8, the results for group comparison N_F vs. FD_F and N_M vs. FD_M are depicted. Normalized FI of the parameters rated as most important for comparisons (a1) N_F vs. FD_F and (b1) N_M vs. FD_M are shown.

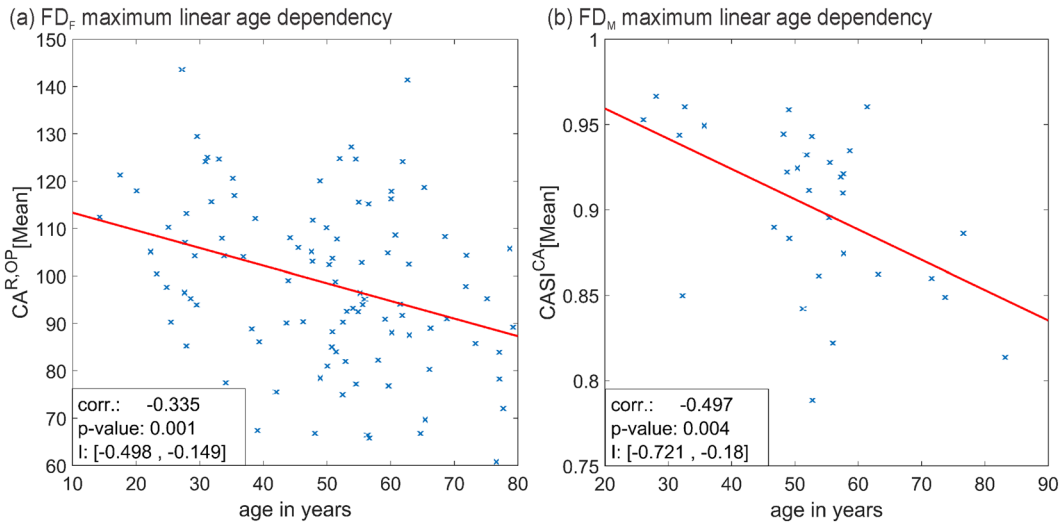

Figure 5. Parameters correlated highest with age; for (**a**) FD_F (**b**) FD_M.

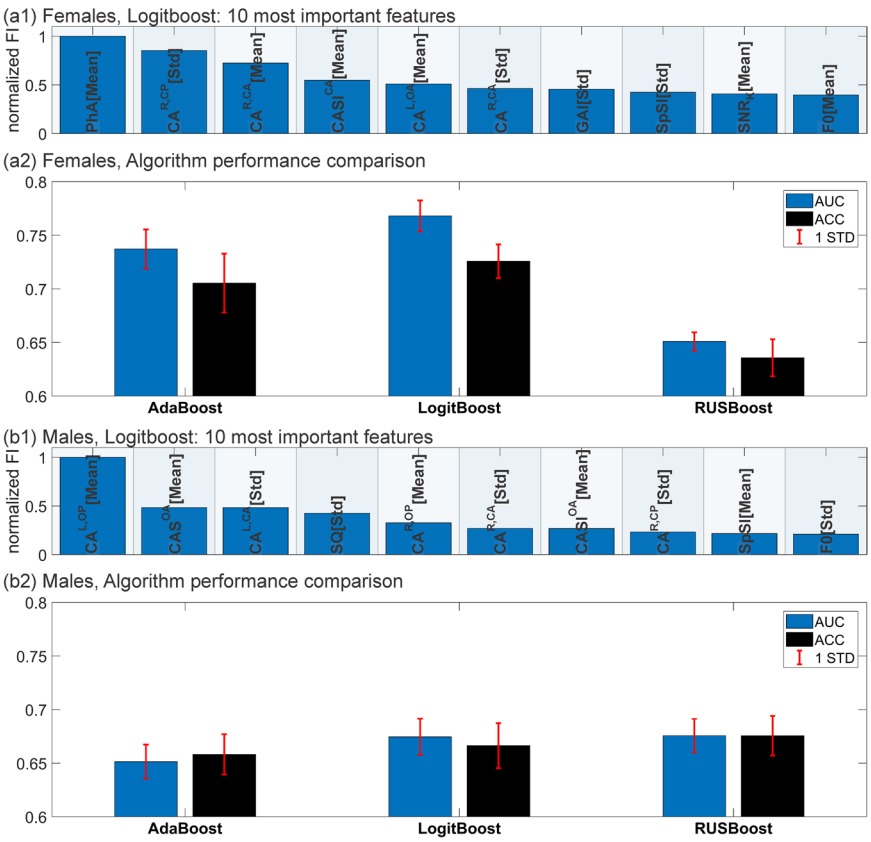

Figure 6. Comparison of boosting algorithms for (**a**) females and (**b**) males. (a1/b1) normalized feature importance of the 10 highest ranked parameters for Logitboost using a 300 stumps model. (a2/b2) comparison of AUC and ACC of all tested algorithms.

For comparison (a2) N_F vs. FD_F 13 parameters need to be included until AUC and ACC do not improve substantially anymore. Analogously, (b2) 11 parameters are included for N_M vs. FD_M. Afterwards the model performance decreases. These parameters are respectively the 13 parameters in Fig. 8 (a1) and 11 parameters in Fig. 8 (b1). Since two parameters are included in both comparisons (marked in red), altogether 22 parameters were found to be relevant.

Step 5: In Table 3, AUC and ACC values of models for relevant parameter combinations are given for both group comparisons. The table also shows the number of included parameters reasoning which types of parameters

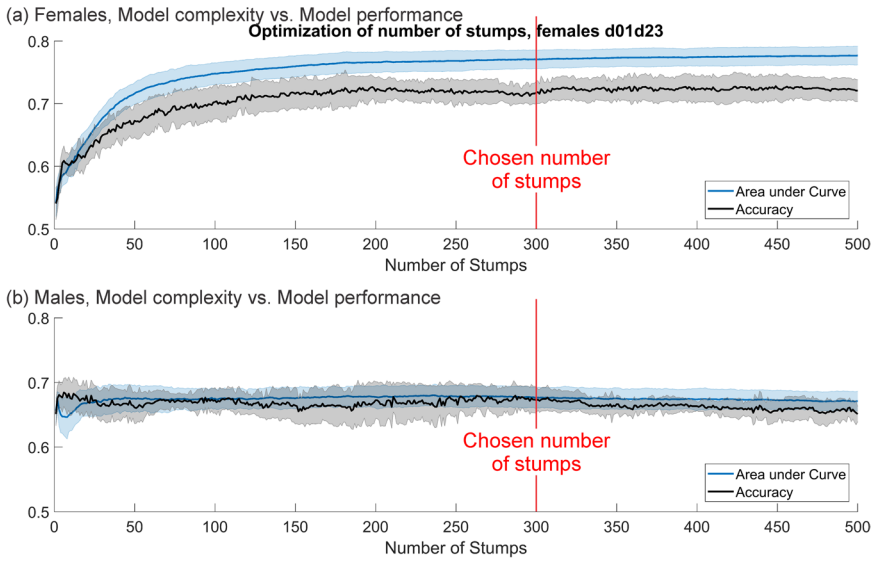

Figure 7. Choosing the optimal number of stumps for (**a**) females and (**b**) males. Number of stumps (model complexity) included in the model versus performance in measured in AUC and ACC.

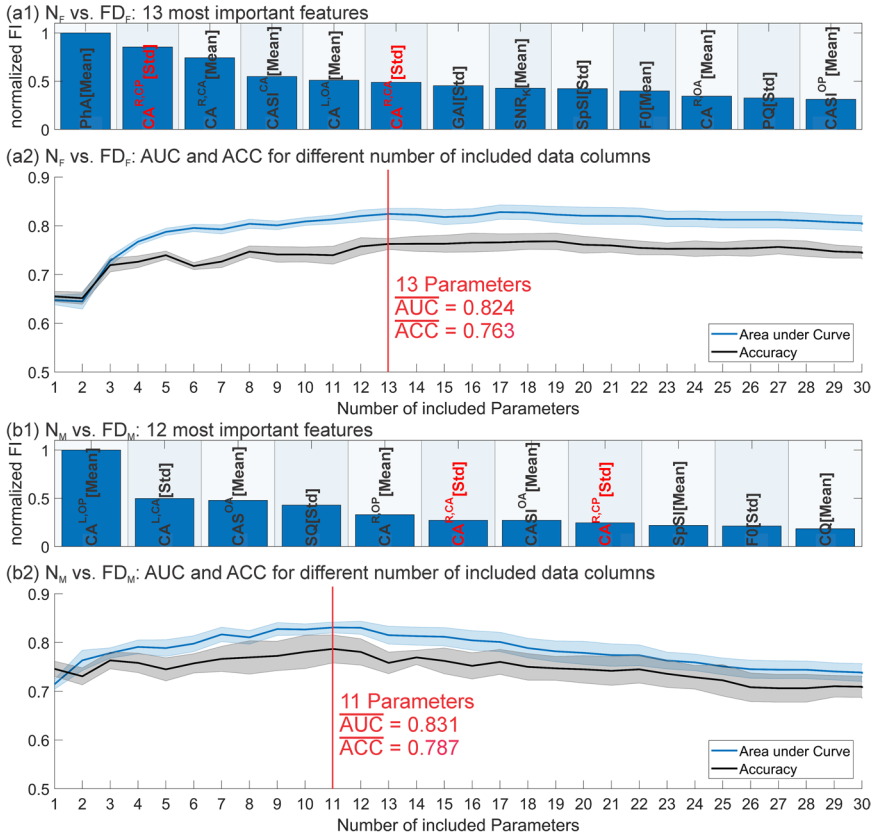

Figure 8. Determination of best parameter subset for group comparisons (**a**) N_F vs. FD_F and (**b**) N_M vs. FD_M. (a1/b1) normalized FI of the 13/11 parameters ranked as most important. (a2/b2) AUC and ACC of models including only the best rated parameter, the best and the second best rated parameter,…. Parameters that were included in the best set for both group comparisons are marked in red.

were contained. The final parameter set HSV_2, given in Table 4, was determined as best compromise between a still comparatively high AUC and ACC and a small number of included parameters. Average specificity (healthy subjects correctly identified as healthy) and sensitivity (disordered subjects correctly identified as disordered) for this set were 0.766 and 0.712 for group comparison N_F vs. FD_F and 0.767 and 0.772 for N_M vs. FD_M. Average

Number	Type of parameters	AUC females/males	ACC females/males
23	both parameter sets	0.812/0.771	0.752/0.722
13	only parameters found relevant in N_F vs. FD_F	0.824/0.558	0.763/0.547
12	only parameters found relevant in N_M vs. FD_M	0.694/0.831	0.647/0.787
13	only PVG-based parameters	0.770/0.763	0.713/0.756
10	only GAW-based parameters	0.716/0.599	0.686/0.597
12	Best parameter subset (HSV_2)	0.788/0.804	0.745/0.768

Table 3. Relevant combinations of parameters and resulting AUC and ACC.

	GAW-based	PVG-based
parameter [Mean]	SNR_K, PhA	$CA^{L,OP}$, $CA^{R,CA}$, CAS^{OA}, $CASI^{OA}$, $CASI^{CA}$
parameter [Std]	SQ, PQ	$CA^{L,CA}$, $CA^{R,CA}$, $CA^{R,CP}$

Table 4. Final parameter set HSV_2.

Group comparison	AUC	ACC
Females: HSV_2 (12 parameters) vs. HSV_1: (66 parameters)		
N_F vs. FD_F	0.788/0.012 vs. 0.771/0.015	0.745/0.012 vs. 0.718/0.019
Males: HSV_2 (12 parameters) vs. HSV_1: (66 parameters)		
N_M vs. FD_M	0.804/0.014 vs. 0.676/0.017	0.768/0.029 vs. 0.673/0.020
Average over all group comparisons		
Averaged	0.796/0.013 vs. 0.724/0.016	0.757/0.021 vs. 0.696/0.020

Table 5. Comparison of mean/standard deviation for AUC and ACC between parameter sets HSV_1 and HSV_2.

difference between specificity and sensitivity was 0.061 (N_F vs. FD_F) and 0.088 (N_M vs. FD_M). Therefore, no group was overly preferred.

Table 5 contains mean and standard deviation of AUC and ACC for models using HSV_2 and HSV_1. The AUC of the model using HSV_2 was on average 0.072 better than the model using the larger parameter set HSV_1. The ACC was 0.061 better on average. In Table 6, mean values and standard deviations of all parameters assembled in HSV_2 are given separated by groups (N_F, FD_F, N_M and FD_M).

HSV_2 was able to clearly outperform the larger parameter set HSV_1 even though all parameters included in HSV_2 are also assembled in HSV_1. This means that most parameters in HSV_1 do not provide valuable information for group separation and only complicate the distinction. However, even the best achieved accuracies never exceeded 0.8. This implies that not all information that is needed for a definite distinction between healthy and disordered subjects is represented by the investigated parameters.

In the final parameter set HSV_2, the GAW based parameters are underrepresented. This is especially noticeable since in HSV_1, GAW based parameters were in the majority (GAW: 36 to PVG: 30). The indication that GAW based parameters may be less important than PVG based ones can also be concluded from Table 3. Including only GAW based parameters from the combined set yielded distinctly less AUC and ACC than including only PVG based parameters, especially for males. Since disordered voices are generally associated with aperiodic oscillations[13–15], the GAW, as a measure exclusively of the glottal area, may not be sufficient to describe all features of such irregular vocal fold oscillations. Furthermore, by compressing the entire actual 3D-information of the vocal fold motion[46–48] into a 1D-GAW-signal, much information is lost. In the PVG, the information is only compressed in 2D-space meaning less information loss in comparison to the GAW.

The initial parameter set HSV_1 found for the group comparison N_F vs. FD_F did not perform well for the group comparison N_M vs. FD_M and vice versa (see Table 3). This illustrates the considerable difference in vocal fold dynamical characteristics between females and males.

The final subset HSV_2 performed as well as the gender combined subset of 22 parameters and in some cases even better (see Table 3). The parameter set HSV_2 consists of four types of parameters:

Type 1: Phonovibrogram (**PVG)** contour angle measures and contour angle symmetry measures. Different contour angles describe if the glottis opens or closes from anterior to posterior direction or vice versa and how fast this process is (see Table S1). For instance, a contour angle $CA^{L,OA}$ [Mean] of 90° means that the left vocal fold (L) on its anterior half during opening phase (OA) opens simultaneously from the anterior part until its middle part. All CAS and CASI measures describe the symmetry of left and right pairs of contour angles. The different contour angles are illustrated in Fig. 9. Contour angle measures and contour angle symmetry measures describe roughly the oscillation pattern of the vocal folds. Therefore, it seems natural that they play the most important role in differentiating between normal and FD groups.

	N_F	FD_F	N_M	FD_M
Mean/standard deviation				
PVG-based				
$CA^{L,OP}$ [Mean] (°)	100.8/12.9	99.9/15.8	96.9/17.2	80.9/16.3
$CA^{R,CA}$ [Mean] (°)	87.6/5.9	87.7/9.4	83.1/9.1	78.5/8.6
CAS^{OA} [Mean] (a.u.)	0.976/0.141	1.001/0.154	0.995/0.084	1.020/0.103
$CASI^{OA}$ [Mean] (a.u.)	0.883/0.070	0.880/0.073	0.933/0.045	0.923/0.049
$CASI^{CA}$ [Mean] (a.u.)	0.934/0.035	0.904/0.061	0.921/0.044	0.902/0.048
$CA^{L,CA}$ [Std] (°)	3.4/1.7	4.3/3.0	2.9/1.6	4.0/1.7
$CA^{R,CA}$ [Std] (°)	3.4/1.8	4.4/2.1	3.1/1.5	3.3/1.0
$CA^{R,CP}$ [Std] (°)	6.6/4.3	6.5/5.1	5.3/6.5	2.9/2.1
GAW-based				
SNR_K [Mean] (dB)	11.2/1.4	10.5/1.6	11.1/1.3	11.0/1.4
PhA [Mean] (a.u.)	−0.031/0.080	0.001/0.113	−0.001/0.078	−0.011/0.092
SQ [Std] (a.u.)	0.151/0.065	0.174/0.085	0.155/0.057	0.165/0.100
PQ [Std] (a.u.)	0.047/0.011	0.052/0.014	0.043/0.013	0.051/0.018

Table 6. Mean and standard deviation of groups N_F, FD_F, N_M and FD_M.

Type 2: SNR_K [Mean] is the only **noise measure** included in HSV_2. It describes the relative energy of the harmonics in relation to the total energy of the signal in the Fourier spectrum[49]. A higher value implies a greater proportion of harmonics in the total spectrum and, as can be seen in Table 6, the GAWs of healthy subjects seem to be slightly more "harmonic" on average.

Type 3: The **symmetry measure** PhA [Mean] describes if the oscillations of the left and right vocal folds are in phase or time shifted. In the healthy case, this measure is expected to be close to zero. PhA [Mean] is a mean value. This means that positive and negative phase shifts in different cycles will cancel each other out. However, there is also a parameter that measures the absolute phase shift (PhAI [Mean]) which would not cancel out during averaging. This parameter was in no case selected as relevant by the boosting algorithm. This could be a hint that time-shifted vocal fold oscillations are only associated with FD if the time-shift is consistent.

Type 4: Standard deviations of two glottal dynamic characteristic parameters (SQ [Std] and PQ [Std]) were selected. These parameters describe the ratio between closing and opening phase and the "peakiness" of the GAW_T[50,51]. The fact, that the standard deviations and not the average values of these parameters were selected, indicates that the actual shape of a GAW_T seems to be not as important as that this shape is consistent over time (i.e. cycles). Also in Table 6, the mean values of these parameters are slightly higher for the disordered cases. This means that SQ and PQ change more strongly on average between cycles for disordered subjects.

Summary. From 91 investigated HSV-parameters (HSV_0) only 12 parameters (HSV_2, 13%) were required to separate healthy and FD subjects with fair accuracy of 74.5% respectively 76.8%. This final parameter set HSV_2 also outperformed parameter set HSV_1 (consisting out of 66 parameters). This indicates a large number of unneeded parameters for this separation task. However, no accuracies exceeding 0.8 could be achieved, hinting that not the entire information needed is contained in these parameters. Accuracies found in this work are mostly on a par with literature values of 0.76[28] and 0.75[29] for similar tasks. One study achieved a slightly better performance of 0.81 accuracy using only PVG-based measures[30]. Since in this study not the same PVG-based features were investigated as in our study, this may explain the difference. However, performance measures also varied considerable between recalculated models with different partitioning, so the observed difference may also be explainable purely by chance. The main gains from this investigation are the following:

1. 25 of the investigated 91 parameters are highly redundant (see section Linear dependencies in Results and Discussion and Table 2).
2. GAW-based parameters are less suited in differentiation healthy and FD subjects than PVG-based parameters. However, they provide valuable additional information.
3. Average values and standard deviations of parameters are both relevant. Regularity of GAW phases (SQ) and peak shape (PQ), harmonic structure (SNR_K) and regularity and average values of different contour angles are mainly important.

Shortcomings. Only parameters based on HSV-recordings were investigated. Other recording techniques, like stroboscopy or videokymography, were not applied. It is possible that better performance in separating healthy and FD subjects could have been achieved if more parameters from more signal sources, e.g. simultaneously recorded audio, would have been investigated in this work.

Due to the different age ranges of the healthy and the disordered group, results could have been influenced by subject age. An influence of subject age for different signal types and voice parameters is well documented in the literature[52–54]. However, in this study this influence should be low or even negligible as the variations in the data caused by FD seemed to outclass the influence of subject age by far.

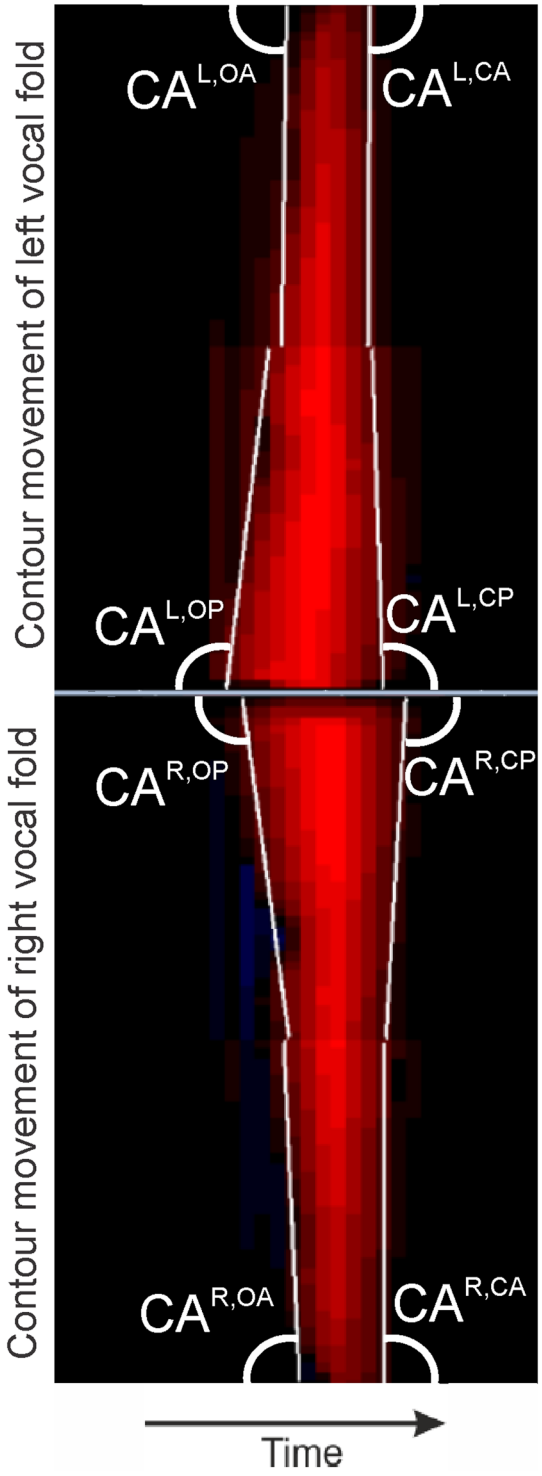

Figure 9. PVG oscillation cycle of healthy vocal folds with indicated contour angles.

Finally, more parameters, alternating parameter definitions and signal types exist that were not investigated in this study. However, with the investigation of 91 different parameters, we covered a large partition of the HSV parameters in use in voice research[24].

Conclusion

In this study we derived the subset HSV_2 of 12 relevant HSV-parameters (mean of SNR_K, PhA, $CA^{L,OP}$, $CA^{R,CA}$, CAS^{OA}, $CASI^{OA}$, $CASI^{CA}$ and standard deviation of SQ, PQ, $CA^{L,CA}$, $CA^{R,CA}$, $CA^{R,CP}$) from a set of 91 parameters (HSV_0). Parameters in HSV_2 reflected FD induced impairments and were sufficient to separate healthy and FD subjects with fair accuracy. The high degree of redundancy within parameters is shown by (1) exclusion of 25

parameters from HSV_0 due to very high correlations yielding HSV_1 and (2) 12 parameters in HSV_2 even outperforming 66 parameters in HSV_1 during group separation. Sources for investigated parameters can be found here:[55–70].

Furthermore, this work shows that PVG-based parameters may be more relevant for differentiation between healthy and FD subjects than GAW-based parameters. However, best results were achieved by a combination of both. Also, the combination of boosted stumps and the FI measure were confirmed as a reliable approach to find relevant parameters and it was shown that the influence of subject age on our results is negligible.

This study affirms the need of multidimensional approaches for assessment of clinical data. Single parameters based on single signal sources are not sufficient to identify disorders. However, a too large amount of parameters also negatively affects results. By finding the best set of parameters, clinically applicable tools could be created assisting in assessment and therapy judgement of voice disorders. This could significantly objectify and improve current clinical routine.

Received: 28 January 2020; Accepted: 20 May 2020;
Published online: 29 June 2020

References

1. Döllinger, M. The next step in voice assessment: High-speed digital endoscopy and objective evaluation. *Curr. Bioinform.* **4**, 101–111 (2009).
2. Zacharias, S. R. C., Deliyski, D. D. & Gerlach, T. T. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *J. Voice.* **32**, 216–220 (2018).
3. Birk, V. *et al.* Automated setup for *ex vivo* larynx experiments. *J. Acoust. Soc. Am.* **141**, 1349, https://doi.org/10.1121/1.4976085 (2017).
4. Deliyski, D. & Hillman, R. State of the art laryngeal imaging: research and clinical implications. *Curr. Opin. Otolaryngol. Head Neck Surg.* **18**, 147–152 (2010).
5. Schlegel, P. *et al.* Influence of spatial camera resolution in high-speed videoendoscopy on laryngeal parameters. *PLoS ONE.* **14**, e0215168, https://doi.org/10.1371/journal.pone.0215168 (2019).
6. Wendler, J., Seidner, W. & Eysholdt, U. *Lehrbuch der Phoniatrie und Pädaudiologie* (4th ed.) 113 (Thieme, 2005).
7. Titze, I. R. Principles of voice production (2nd ed.) (National Center for Voice and Speech, 2000).
8. Stevens, K. N. Acoustic Phonetics (MIT Press, 1999).
9. Echternach, M., Döllinger, M., Sundberg, J., Traser, L. & Richter, B. Vocal fold vibrations at high soprano fundamental frequencies. *J. Acoust. Soc. Am.* **133**, 82–87 (2013).
10. Inwald, E., Döllinger, M., Schuster, M., Eysholdt, U. & Bohr, C. Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging. *J. Voice.* **25**, 576–590 (2011).
11. Unger, J., Schuster, M., Hecker, D. J., Schick, B. & Lohscheller, J. A generalized procedure for analyzing sustained and dynamic vocal fold vibrations from laryngeal high-speed videos using phonovibrograms. *Artif. Intell. Med.* **66**, 15–28 (2016).
12. Uloza, V., Vegienė, A., Pribuišienė, R. & Šaferis, V. Quantitative evaluation of video laryngostroboscopy: reliability of the basic parameters. *J. Voice.* **27**, 361–368 (2013).
13. Roy, N. Functional dysphonia. *Curr. Opin. Otolaryngol. Head Neck Surg.* **11**, 144–148 (2003).
14. Eysholdt, U., Rosanowski, F. & Hoppe, U. Vocal fold vibration irregularities caused by different types of laryngeal asymmetry. *Eur. Arch. Otorhinolaryngol.* **260**, 412–417 (2003).
15. Bonilha, H. S., Deliyski, D. D., Whiteside, J. P. & Gerlach, T. T. Vocal fold phase asymmetries in patients with voice disorders: a study across visualization techniques. *Am. J. Speech Lang. Pathol.* **21**, 3–15 (2012).
16. Wendler, J., Seidner, W. & Eysholdt, U. *Lehrbuch der Phoniatrie und Pädaudiologie* (4th ed.) 139-189 (Georg Thieme, 2005).
17. Wilson, J. A., Deary, I. J., Scott, S. & MacKenzie, K. Functional dysphonia. *BMJ* **311**, 1039, https://doi.org/10.1136/bmj.311.7012.1039 (1995).
18. Aronson, A. E. Importance of the psychosocial interview in the diagnosis and treatment of "functional" voice disorders. *J. Voice.* **4**, 287–289 (1990).
19. Hartnick, C. J. & Zeitels, S. M. Pediatric video laryngo-stroboscopy. *Int. J. Pediatr. Otorhinolaryngol.* **69**, 215–219 (2005).
20. Stemple, J. C. & Fry, L. B. *Laryngeal Evaluation.* 110–119 (Georg Thieme, 2010).
21. Vaca, M., Cobeta, I., Mora, E. & Reyes, P. Clinical assessment of glottal insufficiency in age-related dysphonia. *J. Voice.* **31**, 128.e1–128.e5, https://doi.org/10.1016/j.jvoice.2015.12.010 (2017).
22. Deliyski, D. Laryngeal evaluation., 245–270 (Georg Thieme, 2010).
23. Lohscheller, J., Eysholdt, U., Toy, H. & Döllinger, M. Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2D-diagrams for visualizing and analyzing the underlying laryngeal dynamics. *IEEE Trans. Med. Imaging* **27**, 300–309 (2008).
24. Pedersen, M., Jønsson, A., Mahmood, S. & Agersted, A. Which mathematical and physiological formulas are describing voice pathology: an overview. *J. Gen. Pract.* **4**, 253, https://doi.org/10.4172/2329-9126.1000253 (2016).
25. Laves, M.-H., Bicker, J., Kahrs, L. A. & Ortmaier, T. A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **14**, 483–492 (2019).
26. Cordeiro, H., Fonseca, J., Guimarães, I. & Meneses, C. Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies. *J. Voice.* **31**, 384.e9–384.e14, https://doi.org/10.1016/j.jvoice.2016.09.003 (2017).
27. Moccia, S. *et al.* Learning-based classification of informative laryngoscopic frames. *Comput. Methods Programs Biomed.* **158**, 21–30 (2018).
28. Callan, D. E., Kent, R. D., Roy, N. & Tasko, S. M. Self-organizing map for the classification of normal and disordered female voices. *J. Speech Lang. Hear. R.* **42**, 355–366 (1999).
29. Awan, S. N. & Roy, N. Acoustic rediction of voice type in women with functional dysphonia. *J. Voice.* **19**, 268–282 (2005).
30. Voigt, D. *et al.* Classification of functional voice disorders based on phonovibrograms. *Artif. Intell. Med.* **49**, 51–59 (2010).
31. Panek, D., Skalski, A., Gajda, J. & Tadeusiewicz, R. Acoustic analysis assessment in speech pathology detection. *Int. J. Appl. Math. Comput. Sci.* **25**, 631–643 (2015).
32. Umapathy, S., Rachel, S. & Thulasi, R. Automated speech signal analysis based on feature extraction and classification of spasmodic dysphonia: a performance comparison of different classifiers. *Int. J. Speech Technol.* **21**, 9–18 (2018).
33. Sama, A., Carding, P. N., Price, S., Kelly, P. & Wilson, J. A. The clinical features of functional dysphonia. *Laryngoscope.* **111**, 458–463 (2009).
34. Schlegel, P. *et al.* Dependencies and ill-designed parameters within high-speed videoendoscopy and acoustic signal analysis. *J. Voice.* **33**, 811.e1–811.e12, https://doi.org/10.1016/j.jvoice.2018.04.011 (2018).
35. Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U. & Döllinger, M. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Med. Image Anal.* **11**, 400–413 (2007).

36. Mukaka, M. M. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **24**, 69–71 (2012).
37. Bohr, C. *et al*. Spatiotemporal analysis of high-speed videolaryngoscopic imaging of organic pathologies in males. *J. Speech Lang. Hear. R.* **57**, 1148–1161 (2014).
38. Darwiche, G., Almér, L.-O., Björgell, O., Cederholm, C. & Nilsson, P. Measurement of gastric emptying by standardized real-time ultrasonography in healthy subjects and diabetic patients. *J. Ultrasound Med.* **18**, 673–682 (1999).
39. MathWorks, corrcoef, https://de.mathworks.com/help/matlab/ref/corrcoef.html#f80-999628-RL (2019).
40. Iba W. & Langley, P. Induction of one-level decision trees. In *Machine Learning Proceedings 1992, Aberdeen, Scotland* (1992).
41. Géron, A. Hands-on machine learning with Scikit-Learn & TensorFlow (vol. 1) 183-205 (Media, Inc., 2017).
42. Caruana R. & Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. In *ICML '06 Proceedings of the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania, USA*, 2006.
43. MathWorks, Ensemble Algorithms, https://de.mathworks.com/help/stats/ensemble-algorithms.html#btfwpd3 (2019).
44. Géron, A. Hands-on machine learning with Scikit-Learn & TensorFlow (vol. 1) 81-106 (Media, Inc., 2017).
45. MathWorks, predictorImportance, https://de.mathworks.com/help/stats/compactclassificationensemble.predictorimportance.html (2019).
46. Semmler, M. *et al*. 3D reconstruction of human laryngeal dynamics based on endoscopic high-speed recordings. *IEEE Trans. Med. Imaging* **35**, 1615–1624 (2016).
47. Luegmair, G. *et al*. Optical reconstruction of high-speed surface dynamics in an uncontrollable environment. *IEEE Trans. Med. Imaging* **29**, 1979–1991 (2010).
48. Coughlan, C. A. *et al*. *In vivo* cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography. *Sci. Rep.* **6**, 22792, https://doi.org/10.1038/srep22792 (2016).
49. Klingholz, F. Acoustic representation of speaking-voice quality. *J. Voice.* **4**, 213–219 (1990).
50. Timcke, R., Leden, H. & Moore, P. Laryngeal vibrations: measurements of the glottic wave. *Arch. Otolaryngol.* **68**, 1–19 (1958).
51. Mehta, D. D., Zañartu, M., Quatieri, T. F., Deliyski, D. D. & Hillman, R. E. Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy. *J. Acoust. Soc. Am.* **130**, 3999–4009 (2011).
52. Honjo, I. & Isshiki, N. Laryngoscopic and voice characteristics of aged persons. *Arch Otolaryngol.* **106**, 149–150 (1980).
53. Winkler, R. & Sendlmeier, W. EGG open quotient in aging voices—changes with increasing chronological age and its perception. *Logoped. Phoniatr. Vocol.* **31**, 51–56 (2006).
54. Xue, S. A. & Deliyski, D. Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educ. Gerontol.* **27**, 159–168 (2001).
55. Qiu, Q., Schutte, H. K., Gu, L. & Yu, Q. An automatic method to quantify the vibration properties of human vocal folds via videokymography. *Folia Phoniatr. Logop.* **55**, 128–136 (2003).
56. Horii, Y. Vocal shimmer in sustained phonation. *J. Speech Lang. Hear. R.* **23**, 202–209 (1980).
57. Kasuya, H., Endo, Y. & Saliu, S. Novel acoustic measurements of jitter and shimmer characteristics from pathological voice. In *3rd European Conference on Speech Communication and Technology, EUROSPEECH'93, Berlin, Germany* (1993).
58. Koike, Y. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Stud. Phon.* **7**, 17–23 (1973).
59. Deal, R. E. & Emanuel, F. W. Some waveform and spectral features of vowel roughness. *J. Speech Lang. Hear. R.* **21**, 250–264 (1978).
60. de Jesus Goncalves, M. H. Methodenvergleich zur Bestimmung der glottalen Mittelachse bei endoskopischen Hochgeschwindigkeitsvideoaufnahmen von organisch basierten pathologischen Stimmgebungsprozessen, https://d-nb.info/1076911994/34 (2015).
61. Holmberg, E. B., Hillman, R. E. & Perkell, J. S. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J. Acoust. Soc. Am.* **84**, 511–529 (1988).
62. Kunduk, M., Döllinger, M., McWhorter, A. J. & Lohscheller, J. Assessment of the variability of vocal fold dynamics within and between recordings with high-speed imaging and by phonovibrogram. *Laryngoscope.* **120**, 981–987 (2010).
63. Chen, G. *et al*. Development of a glottal area index that integrates glottal gap size and open quotient. *J. Acoust. Soc. Am.* **133**, 1656–1666 (2013).
64. Hillenbrand, J., Cleveland, R. A. & Erickson, R. L. Acoustic correlates of breathy vocal quality. *J. Speech Lang. Hear. R.* **37**, 769–778 (1994).
65. Yumoto, E., Gould, W. J. & Baer, T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *J. Acoust. Soc. Am.* **71**, 1544–1550 (1982).
66. Lessing, J. Entwicklung einer Klassifikationsmethode zur akustischen Analyse fortlaufender Sprache unterschiedlicher Stimmgüte mittels Neuronaler Netze und deren Anwendung, https://ediss.uni-goettingen.de/bitstream/handle/11858/00-1735-0000-0006-B45D-7/lessing.pdf?sequence=1 (2007).
67. Kasuya, H., Ogawa, S., Mashima, K. & Ebihara, S. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J. Acoust. Soc. Am.* **80**, 1329–1334 (1986).
68. Qi, Y., Hillman, R. E. & Milstein, C. The estimation of signal-to-noise ratio in continuous speech for disordered voices. *J. Acoust. Soc. Am.* **105**, 2532–2535 (1999).
69. Döllinger, M., Lohscheller, J., McWhorter, A. & Kunduk, M. Variability of normal vocal fold dynamics for different vocal loading in one healthy subject investigated by phonovibrograms. *J. Voice.* **23**, 175–181 (2009).
70. Döllinger, M., Dubrovskiy, D. & Patel, R. Spatiotemporal analysis of vocal fold vibrations between children and adults. *Laryngoscope.* **122**, 2511–2518 (2012).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grants BO4399/2-1 and DO1247/8-1 (no. 323308998) and Friedrich-Alexander-University Erlangen-Nürnberg (FAU) within the funding program Open Access Publishing.

Author contributions

Conceptualization, M.D. and P.S.; Data Curation, P.S., S.D. and A.S.; Formal Analysis, P.S.; Funding acquisition, M.D.; Investigation, P.S.; Project Administration, M.D. and A.S.; Resources, M.D. and A.S.; Software, P.S.; Supervision, M.D., S.K. and A.S.; Validation, M.D. and S.K.; Writing original draft, P.S; Writing - review & editing, S.K., M.D., P.S, S.D., A.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-66405-y.

Correspondence and requests for materials should be addressed to P.S.