

RESEARCH ARTICLE

Interdependencies between acoustic and high-speed videoendoscopy parameters

Patrick Schlegel^{1,2*}, Andreas M. Kist², Melda Kunduk³, Stephan Dürr², Michael Döllinger², Anne Schützenberger²

1 Department of Head & Neck Surgery, David Geffen School of Medicine, University of California Los Angeles (UCLA), Los Angeles, California, United States of America, **2** Dep. of Otorhinolaryngology, Div. of Phoniatics and Pediatric Audiology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany, **3** Dep. of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, Louisiana, United States of America

* patrickschlegel93@yahoo.de

OPEN ACCESS

Citation: Schlegel P, Kist AM, Kunduk M, Dürr S, Döllinger M, Schützenberger A (2021) Interdependencies between acoustic and high-speed videoendoscopy parameters. PLoS ONE 16(2): e0246136. <https://doi.org/10.1371/journal.pone.0246136>

Editor: Marie Jetté, University of Colorado School of Medicine, UNITED STATES

Received: August 8, 2020

Accepted: January 13, 2021

Published: February 2, 2021

Copyright: © 2021 Schlegel et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting information](#) files.

Funding: M.D. received funding from the Deutsche Forschungsgemeinschaft (DFG) under grants BO 4399/2-1 and DO 1247/8-1/2 and A.S. under grant SCHU 3441/3-2 (all combined under no. 323308998). The project description can be found online at <http://gepris.dfg.de/gepris/projekt/323308998>. A.M.K. was supported by a Joachim-Herz-Stiftung Add-on fellowship. The description of

Abstract

In voice research, uncovering relations between the oscillating vocal folds, being the sound source of phonation, and the resulting perceived acoustic signal are of great interest. This is especially the case in the context of voice disorders, such as functional dysphonia (FD). We investigated 250 high-speed videoendoscopy (HSV) recordings with simultaneously recorded acoustic signals (124 healthy females, 60 FD females, 44 healthy males, 22 FD males). 35 glottal area waveform (GAW) parameters and 14 acoustic parameters were calculated for each recording. Linear and non-linear relations between GAW and acoustic parameters were investigated using Pearson correlation coefficients (PCC) and distance correlation coefficients (DCC). Further, norm values for parameters obtained from 250 ms long sustained phonation data (vowel /i/) were provided. 26 PCCs in females (5.3%) and 8 in males (1.6%) were found to be statistically significant ($|\text{corr.}| \geq 0.3$). Only minor differences were found between PCCs and DCCs, indicating presence of weak non-linear dependencies between parameters. Fundamental frequency was involved in the majority of all relevant PCCs between GAW and acoustic parameters (19 in females and 7 in males). The most distinct difference between correlations in females and males was found for the parameter *Period Variability Index*. The study shows only weak relations between investigated acoustic and GAW-parameters. This indicates that the reduction of the complex 3D glottal dynamics to the 1D-GAW may erase laryngeal dynamic characteristics that are reflected within the acoustic signal. Hence, other GAW parameters, 2D-, 3D-laryngeal dynamics and vocal tract parameters should be further investigated towards potential correlations to the acoustic signal.

Introduction

Phonation begins with an airstream, rising from the lungs, setting the vocal folds located in the larynx in motion. The vocal folds subdivide this airstream in a series of flow pulses which are

the funding can be found online at: <http://portugueslab.com/andreas-kist-received-a-fellowship-from-the-joachim-herz-foundation.html>. The authors acknowledge support by Deutsche Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) within the funding programme Open Access Publishing. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

further modulated in the vocal tract until exiting through the mouth and being perceived as acoustic signal [1, 2]. It is logical to assume that relations between vocal fold oscillation characteristics and acoustic sound quality should exist. Uncovering such relations would highly improve treatment possibilities of voice disorders, since this knowledge will guide physicians in deciding what specific oscillation characteristic needs to be addressed in order to improve certain acoustic quality features.

Due to different underlying disorders the process of voice production can be impaired in a variety of ways. In this work, we divide voice disorders in two groups: organic dysphonias (OD) and functional dysphonias (FD) [3]. Whilst signs of ODs are always (visible) laryngeal anatomical changes, FD is a diagnosis of exclusion due to no underlying anatomical/tissue related (visible) changes are ascertainable [4]. A voice disorder classified as FD may also have purely psychological etiology [5]. It is important to note that some uncertainty surrounds the term FD. First, the exact boundary between ODs and FDs is not always absolute, since organic pathologies may eventually result in functional disorders [3], being named a secondary functional dysphonia. Second, subcategories of FD are not entirely standardized and often reflect clinician's supposition and bias in practice [6]. However, in this study the subjects with FD diagnosis had no organic pathologies at the time of recording, i.e. only the so called primary functional dysphonia was considered.

In patients with voice disorders, the acoustic signal is altered. In many cases, this is due to impairments in the vocal fold oscillations [7, 8]. It is assumed that there are three main vocal fold dynamical characteristics that foster healthy voice quality [9–11]: vocal fold oscillations are assumed to be (A) symmetric, (B) periodic and (C) exhibit a closed state during oscillations.

For instance, vocal fold asymmetry [12, 13] and aperiodicity [14] have been linked to perceived audible roughness; incomplete glottis closure is associated with vocal fatigue and a breathy voice [7, 8]. Better understanding the relations between features of vocal fold oscillations and their effects on the acoustic signal could be of great benefit in clinic settings: If auditory-perceptual symptoms can be traced back to specific vocal folds disorders or specific patterns of vocal fold oscillations, this may lead to improvement in patient's voice by directly treating underlying cause. Hence, finding relations between acoustic signal quality and vocal fold oscillation characteristics would provide further insight into fundamental connections in voice production and would eventually allow treatments tailored to the individual patient's needs.

One powerful tool for investigating vocal fold oscillations is high-speed videoendoscopy (HSV) [15–17]. As illustrated in Fig 1, during rigid-endoscope HSV data collection, as performed in this study, an endoscope is inserted in the mouth of the subject, to record the vocal fold oscillations. The oscillation frequency of the vocal folds lies between 80 and 400 Hz during normal phonation [3]. With HSV recording frame rates between 4,000 fps and 20,000 fps these oscillation frequencies are easily captured [7, 18], leading to a thorough recording of oscillation characteristics during each glottis cycle.

From the resulting HSV data, different types of signals can be extracted, such as vocal fold trajectories [19], Phonovibrograms [20] and the "Glottal Area Waveform" (GAW) [21]. The GAW describes the changing area between the vocal folds, i.e. the glottal area, over time. The GAW reaches maxima during maximum opening of the glottis and minima during the closed phase. Also, synchronous recording of the acoustic signal is possible and often put into practice [22–24] as it was done in this work.

Based on the extracted acoustic and GAW signals various parameters can be calculated, describing different features of the signals reflecting different features of the voice production process. A great number of parameters have been introduced [9, 25], but norm-values for

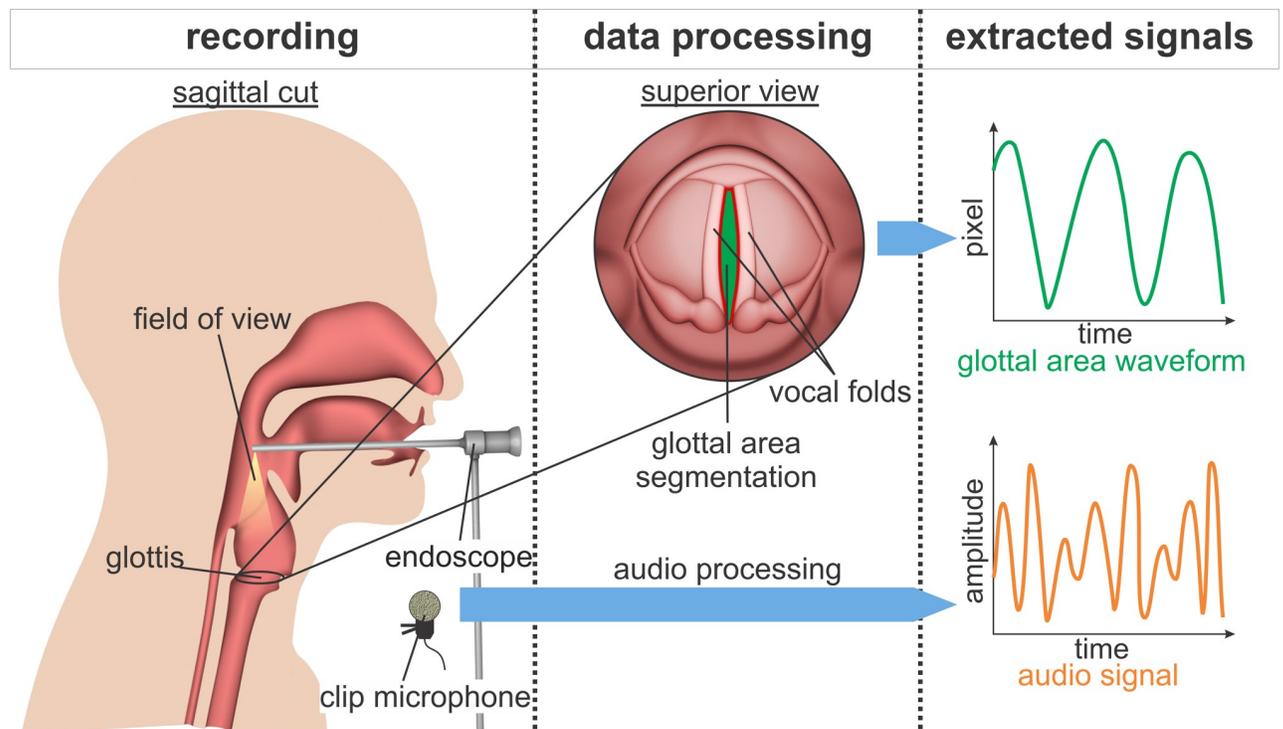


Fig 1. Parallel recording of acoustic and HSV data with subsequent extraction of signals.

<https://doi.org/10.1371/journal.pone.0246136.g001>

many parameters are still missing due to a variety of reasons [26–29]. Widely used parameters such as Jitter and Shimmer describe period irregularity in fundamental frequency and amplitude in the signal. Increased values of these parameters are e.g. associated with hoarseness if they were calculated on acoustic signals [30]. However, given norm values for Jitter (in this case Jitter Percent) differ, with one study stating "healthy" values of around 0.25% for females and males while producing the vowel /a/ [31] whereas another study considers values as high as 0.53% for younger and 0.84% for older males phonating the vowel /a/ as healthy [32]. Such differences may be related to inadequate subject recruitment in these studies or other variabilities in the data collection process. Also in studies employing HSV data different factors appeared to be influencing these parameters such as recording frame rate [27], camera resolution [28] or sequence length [29]. Hence, norm value tables for HSV parameters to aid in objective separation of healthy and disordered voices are needed.

To this date, various works have investigated relations between vocal fold movements and resulting acoustics. However, often only linear relations were explored [33–36] or data from a small number of subjects ($N \leq 20$) was used [33, 35–37]. Some relations between vocal fold oscillations and resulting acoustic signal are known with the most obvious one being the strong correlation between fundamental frequency of the vocal fold oscillations and the fundamental frequency of the resulting acoustic signal in sustained phonation. Other examples include connections between insufficient closure of the vocal folds during phonation and perceived hoarseness in the acoustic signal or the "force" with which the vocal folds collide and the acoustic amplitude [7, 8]. The fundamental frequency (F_0) at which the subject phonates is another factor that may influence acoustic and GAW parameters. For instance, period perturbation measurements in the GAW may be influenced by F_0 due to the lower sampling rate of

GAW signals and a changing F_0 may affect more complex parameters such as noise measurements [30].

This study investigated linear and non-linear relations between GAW and acoustic data for a large number of subjects and parameters. Female and male subjects with normal voices formed the healthy voice group, and subjects who had been diagnosed with FD formed the voice disordered group. The influence of F_0 on the other parameters considered in this work was of particular interest, since parameters that are strongly affected by F_0 may require a correction of this influence. Further, we used our collected data to provide preliminary norm values of parameters obtained from 250 ms long sustained phonation data (vowel /i/). The aims of this work are:

1. Create a set of norm-values for all investigated parameters that differentiate females and males with normal voices from subjects with diagnosis of FD for the given recording settings.
2. Find parameters that are influenced by F_0
3. Determine the linear and non-linear relations between GAW and acoustic parameters

Methods

HSV recordings (N: 351) with simultaneously recorded acoustic signal (time-synchronized) were used for data evaluation. This data (without the acoustic recordings) was already used in a previous study applying machine learning approaches for classification purposes [38]. All 351 acoustic recordings were unanimously rated by three experts on ordinal scales (0 to 2) for signal noise and background noise: 0 was chosen as the best rating (no signal noise / background noise) and 2 as the worst (strong signal noise / background noise). Only recordings that had signal noise and background noise rated at 1 or 0, were used in further analysis, leading to final set of 250 combined HSV-acoustic recordings from female and male subjects for further analyses.

The 250 combined HSV-acoustic recordings were divided into four groups depending on their gender and health status, Table 1. All recordings were taken under clinical conditions using a Photron Fastcam MC2 camera (frame rate: 4000 fps, resolution: 512×256 pixels, 70° rigid endoscope). The acoustic signal was simultaneously recorded using a clip microphone (pentax model #7175–6000, Lapel Microphone, Audio Technica ASP-0091, sampling rate: 40 kHz). All subjects phonated the vowel /i/ at their habitual pitch and loudness level (sustained phonation). From each combined HSV-acoustic recording, a section of 250 ms of sustained phonation was selected.

All disordered patients were diagnosed by our clinicians with FD and no concurrent OD during regular clinical routine (i.e. only primary functional dysphonia was considered). Healthy subjects were recruited separately but examined analogous to disordered subjects. Only healthy subjects were included that did not show signs of any voice disorder. This study was approved by the ethic committee of the Medical School at Friedrich-Alexander-University Erlangen-Nürnberg (no. 290_13B); written consent was obtained by all subjects.

Table 1. Number of combined HSV-acoustic datasets and subject range of age for each group (healthy females (N_F), females with FD (FD_F), healthy males (N_M) and males with FD (FD_M)).

	Healthy	Disordered
Females	124 (N_F) age range 18–64 (mean: 22)	60 (FD_F) age range 20–79 (mean: 49)
Males	44 (N_M) age range 21–30 (mean: 25)	22 (FD_M) age 26–83 range (mean: 53)

<https://doi.org/10.1371/journal.pone.0246136.t001>

Signal extraction and parameter calculation

High-speed video data were processed using a preliminary version of the in house developed software Glottis Analysis Tools (GAT-2020), being freely available upon request. It is the next version of GAT-2018, and includes several bug fixes and an improved cycle detection algorithm. The process of segmentation and parameter calculation is illustrated in Fig 2. For a detailed explanation of the segmentation process see [38]. GAWs describing the total glottal area (GAW_T) and the left and right half of this glottal area (GAW_L and GAW_R) were extracted from HSV videos. The acoustic signal was synchronously recorded using a clip microphone. Maximum based cycles (i.e. each cycle starts at a sufficiently distinct local maximum and ends before the next one) were detected in GAWs and acoustic signals. From all parameters featured in the GAT-software a set of relevant parameters, based on previous work [28, 29, 38, 39], was selected. Only parameters were included that were previously found to be resistant towards certain influencing factors (spatial resolution and sequence length) [28, 29], mathematically sound [39] and not strongly redundant [38]: 35 GAW- and 14 acoustic-based parameters were considered [40–54].

In Table 2 the parameters used in this study are summarized. "Signal" describes if the parameter was calculated exclusively for GAW or acoustic signal or for both signals. "Averaged" describes if only a single parameter value per signal was calculated or if multiple values were calculated (i.e. mean and standard deviation). Further, abbreviation, parameter unit and source are given. This means that a single row in this table can result in up to four parameters (e.g. for Fundamental Frequency acoustic and GAW based F_0 [Mean] and F_0 [Std] were

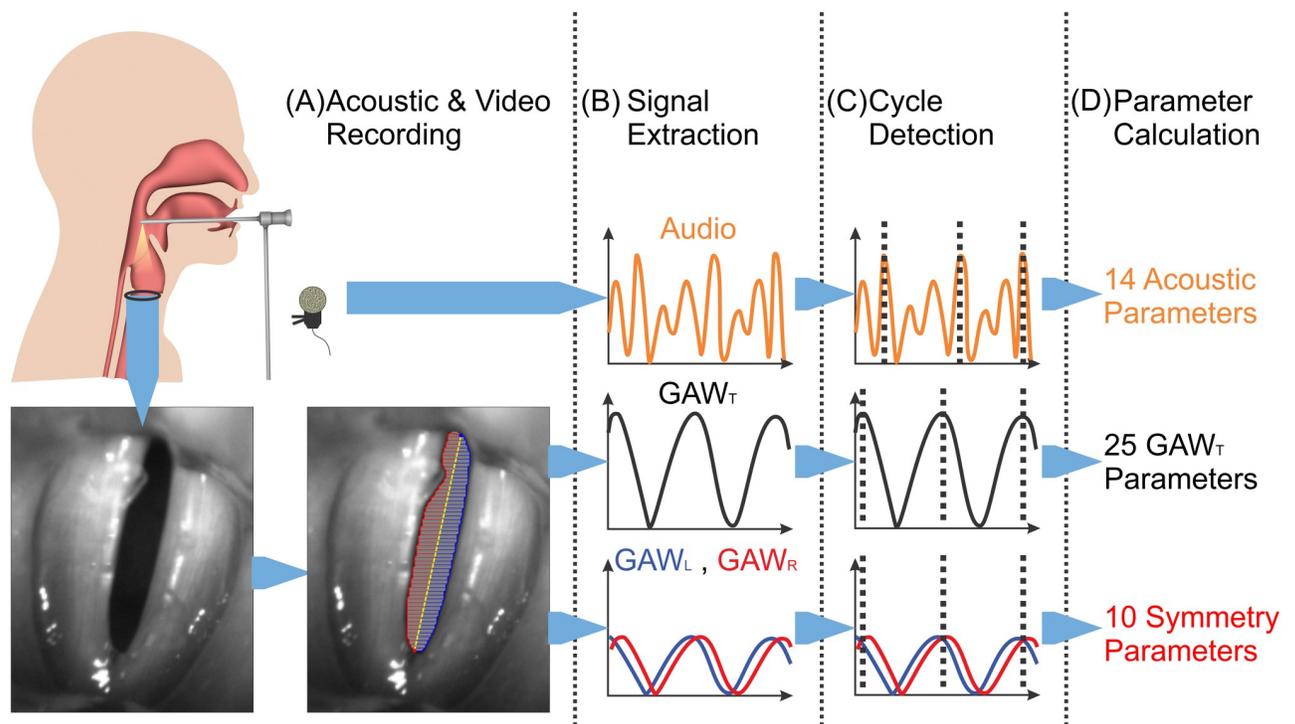


Fig 2. Parameter calculation in four steps. (A) Segmentation of the glottal area between the vocal folds and subdivision in left and right half. (B) Extraction of GAW_R (blue), GAW_L (red) and GAW_T (black) and synchronous audio recording. (C) Detection of maximum based cycles in all signals (all GAWs use cycles based on GAW_T). (D) Calculation of 14 acoustic parameters, 25 GAW parameters and 10 symmetry, i.e. GAW_L and GAW_R , based parameters.

<https://doi.org/10.1371/journal.pone.0246136.g002>

Table 2. Summarized parameter information.

Name	Signal	Averaged	Abbreviation	Unit	Source
Fundamental period measures					
Fundamental Frequency	both	yes	F ₀	Hz	-
Perturbation measures					
Mean Jitter	both	no	MJit	ms	[40]
Jitter (%)	both	no	Jit(%)	a.u.	[40]
Period Variability Index	both	no	PVI	a.u.	[41]
Mean Shimmer	both	no	MShim	dB	[40]
Amplitude Variability Index	acoustic	no	AVI	dB	[41]
Energy Perturbation Factor	GAW	no	EPF	a.u.	[42]
Symmetry measures					
Phase Asymmetry Index	GAW	yes	PhAI	a.u.	[43]
Phase Asymmetry	GAW	yes	PhA	a.u.	[43]
Spatial Symmetry Index	GAW	yes	SpSI	a.u.	[43]
Spatial Symmetry	GAW	yes	SpS	a.u.	[43]
Amplitude Symmetry Index	GAW	yes	AmSI	a.u.	[43]
Amplitude Symmetry	GAW	yes	AmS	a.u.	[43]
Waveform Symmetry Index	GAW	yes	WaSI	a.u.	[43]
Glottal dynamic characteristics					
Closing Quotient	GAW	yes	CQ	a.u.	[44]
Speed Quotient	GAW	yes	SQ	a.u.	[45]
Glottis Gap Index	GAW	yes	GGI	a.u.	[46]
Plateau Quotient	GAW	yes	PQ	a.u.	[47]
Glottal Area Index	GAW	yes	GAI	a.u.	[48]
Noise measures					
Cepstral Peak Prominence	both	no	CPP	dB	[49]
Harmonics-to-Noise Ratio	both	no	HNR	dB	[50]
Max. Waveform Matching Coeff.	both	no	WMC _{Max}	a.u.	[51]
Mean Waveform Matching Coeff.	both	no	WMC _{Mean}	a.u.	[51]
Normalized Noise Energy	both	no	NNE	dB	[52]
Signal-to-Noise Ratio (by Klingholz)	both	yes	SNR _K	dB	[53]
Signal-to-Noise Ratio (by Qi)	GAW	no	SNR _Q	dB	[54]

Signal: indicates if the parameter was calculated for the GAW, for the acoustic signal or for both. Averaged: if "no": there was only a single value for each signal. If "yes": multiple values per signal were calculated resulting in mean ([Mean]) and standard deviation ([Std]) of this parameter. Further, abbreviation, unit and parameter source are given.

<https://doi.org/10.1371/journal.pone.0246136.t002>

calculated). In [S1 Table](#) a more detailed version of [Table 2](#) is given containing names, abbreviations, sources and descriptions of all 49 parameters and, if feasible, formulas.

By definition, the GAW-parameters PhA [Mean], PhAI [Mean] and PhAI [Std] were calculated for minimum based cycles [43]. Custom scripts in Python 3.7 were used to analyze the data and to prepare the figures.

HSV-acoustic correlations

Linear and non-linear relations between HSV and acoustic parameters were considered separately for females and males. For each gender healthy and disordered groups were merged, since parameters are expected to scatter between healthy and disordered voice subjects; i.e. female group (N_F & $FD_F = 184$ subjects) and male group (N_M & $FD_M = 66$ subjects).

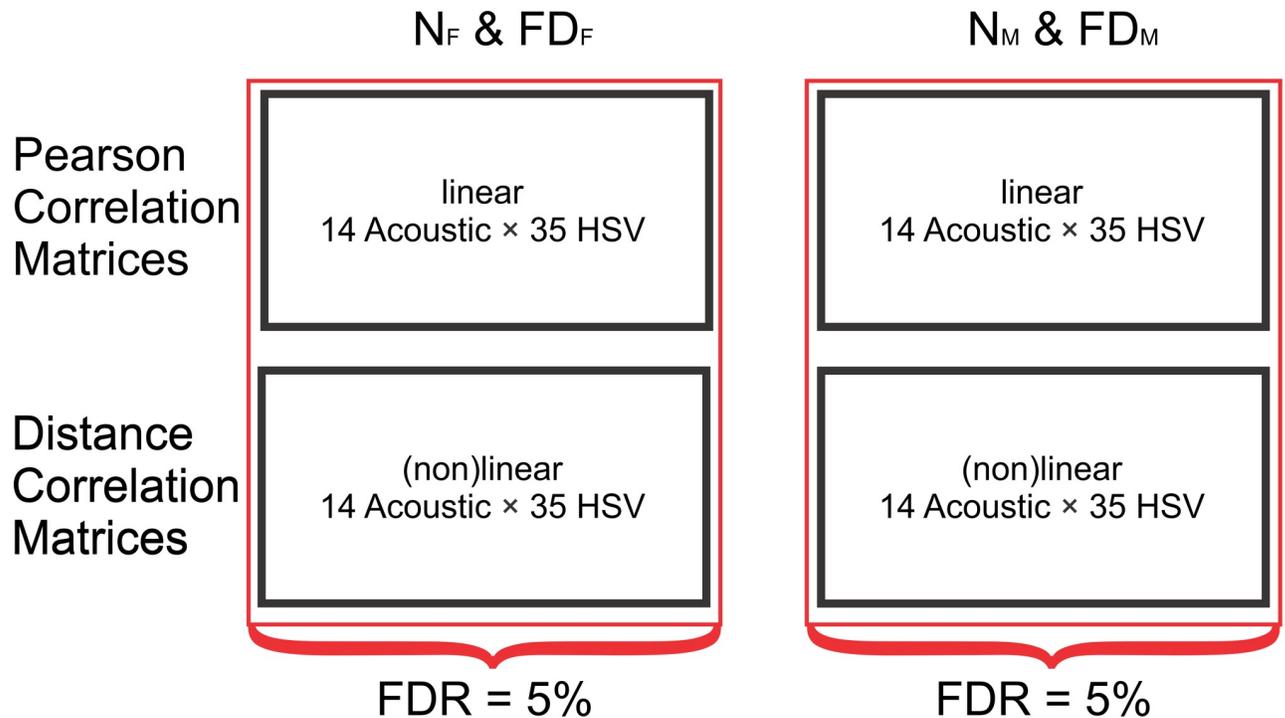


Fig 3. Calculation of Pearson Correlation Coefficients (PCC) and Distance Correlation Coefficients (DCC) between parameters forming Pearson / distance correlation matrices. False discovery rate (FDR) is set to 5% for males and females independently; i.e. FDR is set to 5% for combined subject groups $N_F \& FD_F$ and $N_M \& FD_M$.

<https://doi.org/10.1371/journal.pone.0246136.g003>

To investigate the linear relations, Pearson correlation coefficients (PCC) and p-values were calculated between all HSV and acoustic parameters. For investigation of general relations, distance correlation coefficients (DCC) and p-values were calculated. Distance correlation is a measure of dependence between random vectors that is only zero when the vectors are independent and 1 when the vectors are identical. Therefore DCC measures linear and non-linear associations between vectors and, contrary to PCC, cannot obtain negative values. For more information see the work by Székely, Rizzo and Bakirov [55]. The p-values calculated for the DCCs are, analogous to PCC p-values, the probability of a correlation being equal or greater than the observed DCC, if the null hypothesis (both parameters are uncorrelated) is true.

This approach yielded two sets of PCCs and two sets of DCCs with respective p-values. We controlled the false discovery rate (FDR), i.e. the expected percentage of false positive tests at 5% using the Benjamini-Yekutieli procedure, since there may be unknown interdependencies between the tests [56]. The p-values were adjusted accordingly. The entire process is illustrated in Fig 3.

Results

Three main topics were of interest in this work: (A) Determining the ranges of values for healthy subjects (i.e. females and males with normal voices) and subjects with diagnosis of FD for the investigated parameters, (B) investigating influence of F_0 on other parameters and (C) detecting relations between parameters not related to the fundamental vocal fold oscillation frequency F_0 .

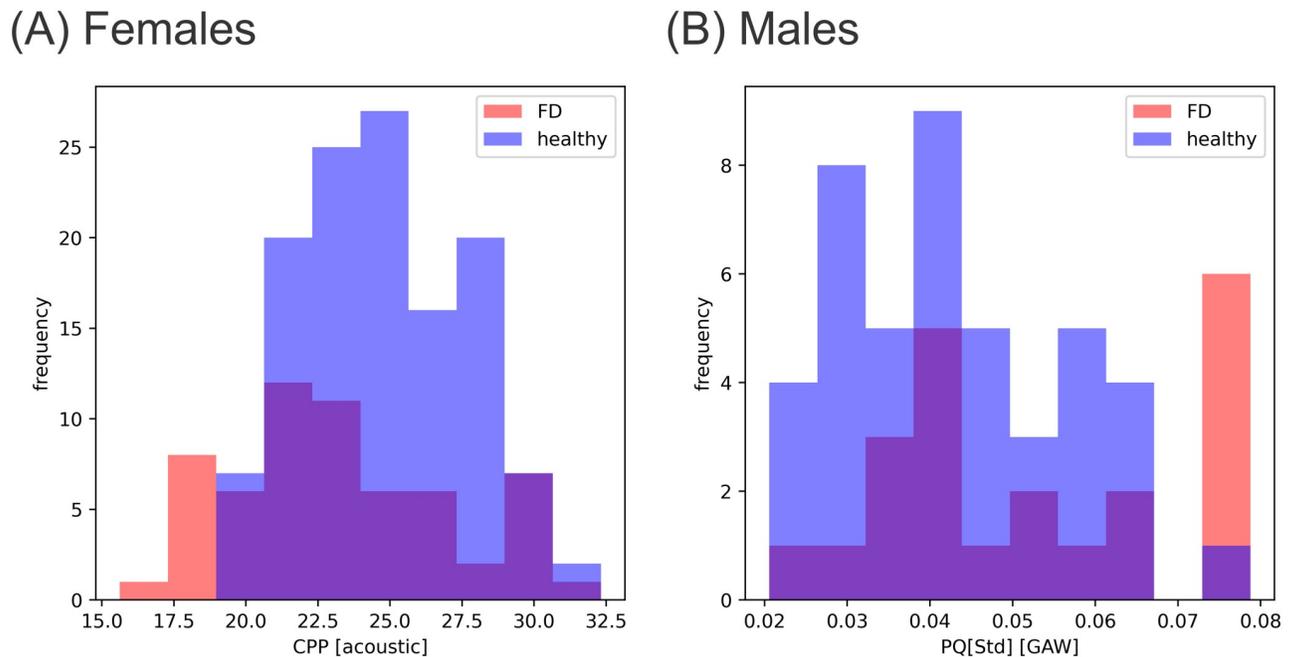


Fig 4. Distribution of parameter values for (A) acoustic based CPP in females and (B) GAW-based PQ[Std] in males. In the violet sections of the histograms blue and red histograms are overlapping.

<https://doi.org/10.1371/journal.pone.0246136.g004>

Ranges of values for healthy and FD subjects

Statistical values for all four groups (N_F , FD_F , N_M , FD_M) are collected in [S2 Table](#). This table contains Minimum, Maximum, mean and median-values for these groups as well as the standard deviations, skewness and kurtosis. Further, below this table, distributions of parameter values for all parameters investigated in this study are plotted (similar to [Fig 4](#)). Parameter values scattered severely and outliers were common. In [Fig 4](#), exemplary the distributions of two parameters, acoustic based CPP in females and GAW-based PQ [Std] in males, are depicted. Albeit some shift towards lower / higher values may be subjectively identifiable, no strong differences between healthy and FD groups are observable in low order statistical measures like means and medians. However, for some parameters, like GGI [Mean], high order statistical measures (skewness and kurtosis) deviate considerably (see [S2 Table](#)). Analogously differences were either similarly small or undetectable in all other GAW- and acoustic-based parameters for both females and males.

Parameters influenced by F_0

We used the rule-of-thumb limits proposed by Mukaka [\[57\]](#) to rate the size of the correlation coefficient (i.e. **absolute value** of PCC or DCC):

- $0.0 \leq x < 0.3$: negligible
- $0.3 \leq x < 0.5$: low
- $0.5 \leq x < 0.7$: moderate
- $0.7 \leq x < 0.9$: high
- $0.9 \leq x \leq 1.0$: very high

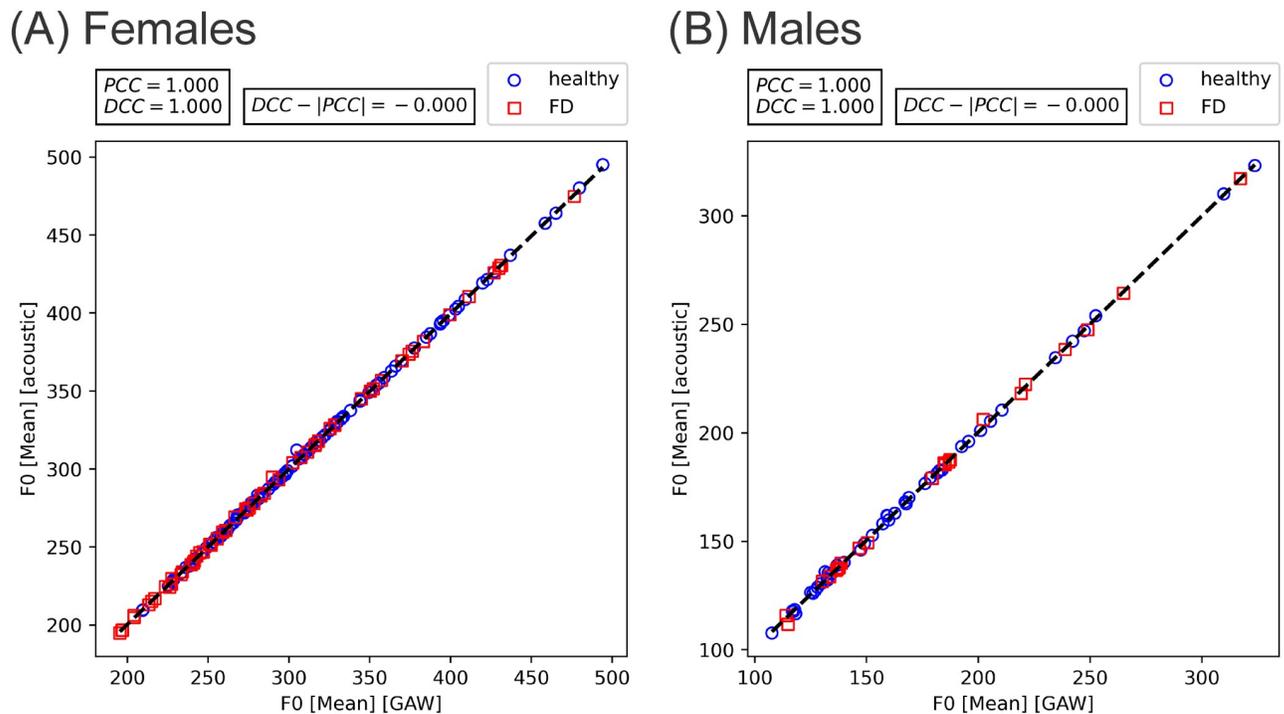


Fig 5. Correlation between GAW and acoustic F_0 [Mean] in (A) females and (B) males with fitted line (black).

<https://doi.org/10.1371/journal.pone.0246136.g005>

Mukaka only discussed linear relations; however, we also used this limit for distance correlation since it has (in absolutes) the same value range as Pearson correlation. This also leads to better comparability between PCCs and DCCs.

Further, we imposed two conditions that had to be fulfilled to determine a PCC or DCC between two parameters as relevant. (A) The PCC or DCC had to be statistically significant after FDR correction. (B) The PCC or DCC had to be above the rule-of-thumb limit of negligibility for correlation coefficients; i.e. an absolute value greater than or equal to 0.3.

The following relevant correlations were observed: The only parameters that correlated very high (≥ 0.9) were GAW- and acoustic-based F_0 [Mean], as depicted in Fig 5, for females and males. GAW-based but not acoustic based F_0 [Std] was highly associated with F_0 [Mean]. Two parameters were moderately associated with F_0 [Mean] (PCC or DCC between 0.5 and 0.7). Four parameters showed low and moderate correlations. 15 parameters showed only low correlations (between 0.3 and 0.5). A list of parameters that were associated with F_0 [Mean], as well as relevant PCCs and DCCs, is provided in Table 3.

Differences in PCCs (linear correlation) and DCCs (general correlation including linear and non-linear) for the same comparisons were small; i.e. linear correlations are dominant and non-linear relations seem to be small to negligible. For pairings with at least one, PCC or DCC, statistically significant and in absolute values ≥ 0.3 , the highest difference in females was 0.111 between GAW-based PhAI [Std] and acoustic-based F_0 [Mean]. In males the largest difference was 0.081 between GAW-based F_0 [Mean] and acoustic-based WMC_{mean} . In Fig 6, scatter plots for these parameter pairings are depicted, including a fitted regression line and second degree polynomial. Further, a regression line, applying the random sample consensus (RANSAC) algorithm [58], to exclude outlier data points is fitted. As shown in Fig 6, non-linear dependencies between parameters PhAI [Std] / F_0 [Mean] and WMC_{Mean} / F_0 [Mean] may exist, but are only weak with large scatter.

Table 3. Parameters correlated with GAW-based or acoustic based F_0 [Mean].

Name	source	PCC females N: 184	DCC females N: 184	PCC males N: 66	DCC males N: 66
F_0 [Mean]	GAW	1.000	1.000	1.000	1.000
CQ [Mean]	GAW	0.323	0.352	0.498	0.531
SQ [Mean]	GAW	-0.380	0.394		
PQ [Mean]	GAW	0.363	0.366		
HNR	GAW	-0.393	0.396		
SNR _Q	GAW	0.342	0.414		
WMC _{max}	GAW	-0.333	0.307	-0.458	
WMC _{mean}	GAW	-0.385	0.349		
Mean Shim	GAW	0.375	0.363		
EPF	GAW	0.315	0.303		
Jit(%)	GAW	0.400	0.381		
PVI	GAW	0.663	0.661		
F_0 [Std]	GAW	0.877	0.860	0.826	0.801
CQ [Std]	GAW	0.319	0.413		
PQ [Std]	GAW	0.412	0.435		
PhAI [Std]	GAW	0.524	0.635	0.598	0.579
HNR	acoustic		0.319		
NNE	acoustic	-0.420	0.461		
WMC _{max}	acoustic	0.403	0.502	0.571	0.614
WMC _{mean}	acoustic		0.383		0.518
Mean Shim	acoustic		0.344		
Mean Jitt	acoustic	-0.412	0.461	-0.474	0.552
PVI	acoustic		0.304		

All statistically significant [PCCs] or DCCs ≥ 0.3 for females and males are given. Values for these correlations can be found in [S3 Table](#). Empty fields are no relevant correlations (below 0.3 or not statistically significant).

<https://doi.org/10.1371/journal.pone.0246136.t003>

No notable differences in PCCs and DCCs between females and males were detected with the exception of GAW-based PVI, related on a moderate level (PCC = 0.663 and DCC = 0.661) with acoustic based F_0 in females but not in males (no statistically significant PCC or DCC). In general, if for a certain parameter relation a statistically significant PCC or DCC was found in males, the same parameter relation was also statistically significant in females, but not vice versa.

Correlations excluding mean F_0

Correlations between GAW- and acoustic-based parameters (excluding F_0 [Mean]) were in most cases negligible. As shown in [Table 4](#), only 17 low and one barely moderate PCCs or DCCs could be observed. The highest correlations were found between acoustic-based WMC_{Max} and GAW-based F_0 [Std], which were both also correlated to F_0 [Mean], [Table 3](#). Analogously to F_0 [Mean] associated correlations, no distinct differences between PCCs and DCCs in females and males were observable. In [S3 Table](#), all PCCs and DCCs and respective p-values (after FDR-correction) calculated in this study are given.

Discussion

For none of the investigated parameters healthy and disordered groups are clearly separable by parameter values, as shown in [Fig 4](#) for two example parameters. However, by inspecting high order statistical measures like skewness and kurtosis that describe the shape of the distribution

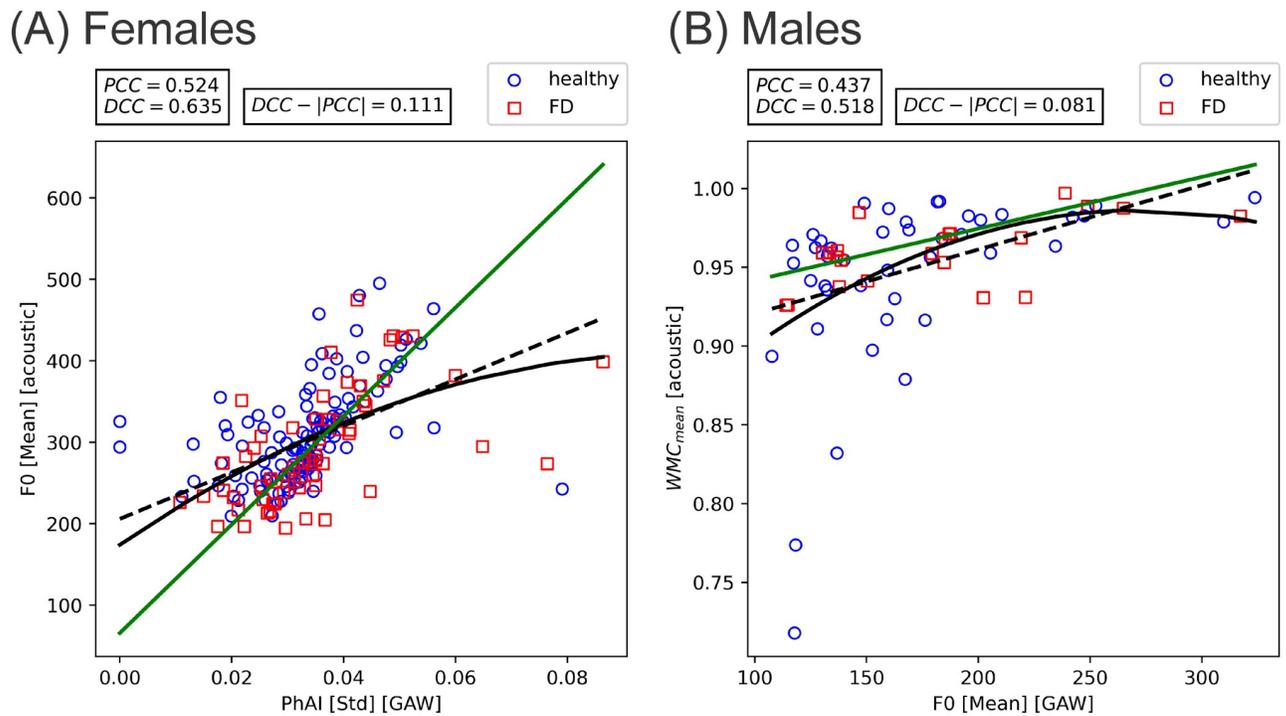


Fig 6. Relevant parameter relations with highest difference between PCC and DCC in (A) females (acoustic based F_0 [Mean] versus GAW-based PhAI [Std]) and (B) males (acoustic based WMC_{mean} versus GAW-based F_0 [Mean]). Fitted are the linear regression line (black, dashed), a second degree polynomial (continuous, black) and a RANSAC regression line (green).

<https://doi.org/10.1371/journal.pone.0246136.g006>

Table 4. GAW-based parameters correlated with the given acoustic based parameter (NNE, CPP, WMC_{max} , WMC_{mean} , AVI and MJit).

Name	source	PCC females N: 184	DCC females N: 184	PCC males N: 66	DCC males N: 66
NNE (acoustic based)					
F_0 [Std]	GAW		0.309		
CPP (acoustic based)					
GGI [Mean]	GAW	-0.322	0.319		
GAI [Mean]	GAW	0.342	0.336		
WMC_{max} (acoustic based)					
F_0 [Std]	GAW		0.373	0.489	0.511
WMC_{mean} (acoustic based)					
NNE	GAW	-0.345	0.339		
SNR_Q	GAW	0.332	0.311		
AVI (acoustic based)					
NNE	GAW	0.412	0.396		
CQ [Std]	GAW	0.320			
GGI [Std]			0.310		
MJit (acoustic based)					
PhAI [Mean]	GAW	0.309			
F_0 [Std]	GAW		0.335		

All statistically significant |PCCs| or DCCs ≥ 0.3 for females and males are given. Empty fields are not relevant correlations (below 0.3 or not statistically significant). Values for these correlations can be found in S3 Table.

<https://doi.org/10.1371/journal.pone.0246136.t004>

of parameter values for groups N_F , N_M , FD_F and FD_M , several differences between subject groups become apparent (see [S2 Table](#) for a comparison of statistical values of all parameters). This is not surprising, since FD is an umbrella-term for a variety of voice disorders [6]. Therefore parameters that describe a certain feature of the phonation process may be expressive for certain subcategories of FD, but may not for others. This and high individual physiological variability [7] may lead to the observed outliers and high variability of parameter values in the data. Since the female and male FD groups consist out of subjects with varying conditions, specific parameters may differ from normal values for only some of the FD subjects. This may then lead to changes in the shape of the parameter distribution in comparison to healthy subjects. In summary, single parameters are not suitable for differentiating healthy from FD subjects and multi-parametric approaches are needed as suggested before [38, 59, 60]. However, if not FD in general but subcategories of FD (e.g. psychogenic dysphonia, conversion dysphonia or tension-fatigue syndrome [6, 61]) are investigated, there could be single parameters or smaller sets of parameters that are able to differentiate these subcategories of FD from healthy voices. Therefore, the collected values for FD subjects, as provided in [S2 Table](#), should be considered preliminary (see shortcomings).

As expected [1, 2], GAW and acoustic F_0 [Mean] are highly correlated, additionally other parameters are also, to some degree, correlated to F_0 [Mean], see [Table 3](#). Albeit most of these correlations were only low (0.3 to 0.5), this still implies that these parameters change to a small degree with changing F_0 . Exceptions are GAW-based F_0 [Std] and PhAI [Std], showing no PCC or DCC below "high" (0.7 to 0.9) respectively "moderate" level (0.5 to 0.7) in females and males (see [Table 3](#)).

Only GAW-based F_0 [Std] but not acoustic based F_0 [Std] showed the aforementioned strong correlation with F_0 [Mean]. F_0 [Std] is calculated from the inverse cycle lengths ($\frac{1}{\text{cycle length}}$) which vary more for the acoustic signal than in the GAW due to noise and the more complex waveform shape of the acoustic signal which complicates the determination of the exact beginning and ending position of cycles. This effect may mask a potential existing correlation between acoustic based F_0 [Std] and F_0 [Mean].

PhAI describes the relative phase shift between GAW_L and GAW_R in one vocal fold oscillation cycle and PhAI [Std] respectively the standard deviation of this parameter, calculated for all oscillation cycles. Therefore the comparatively high positive correlation of this parameter with F_0 is expected, since with shorter cycles (higher F_0), the deviation of PhAI relative to cycle length increases. Regarding such effects, it may be needed to correct for the influence of F_0 during further use of the affected parameters.

The found, small differences between PCCs and DCCs indicate weak non-linear relations between the investigated GAW and acoustic features, since this implies that the "general association" between parameters that are measured by DCCs are almost completely explainable by "linear association" that are measured by PCCs. As shown in [Fig 6](#), in the parameter pairings with the highest difference between PCC and DCC, no obvious or only weak non-linear dependencies are observable.

Higher values of PCCs and DCCs and simultaneously a lower number of statistically significant PCCs and DCCs in males than in females may be attributable to the smaller number of available male subjects. PCC and DCC between GAW-based PVI and acoustic F_0 [Mean] differs the most between females and males. This can be attributed to males phonating at lower fundamental frequencies than females [30] and that the higher the F_0 [Mean], the stronger the association between GAW-based PVI and F_0 [Mean].

This relation may be to some degree an artefact attributable to the, in comparison to the speed of vocal fold oscillations, limited sampling rate of the GAW. Even though for 4000 fps

recording frame rate and vocal fold oscillation frequencies between 80 and 400 Hz [3], each cycle is represented by 27 to 10 data points; i.e. a single data point shift results in up to 10% change of the cycle length. In female GAWs, less data points are contained in each cycle and hence period perturbation measures such as Jit(%) and PVI are artificially increased. MJit is an exception, since it is not normalized and hence would be expected to be higher in males, however, this effect and the one mentioned before level each other out.

Only 11 pairings of parameters in females and 1 pairing of parameters in males that did not include F_0 [Mean] had statistically significant correlations and none of these correlations exceeded 0.5 (see Table 4). Therefore, the direct relation between investigated features of the GAW and the acoustic signal excluding F_0 is only low at best. However, there may be still some relations for subcategories of subjects that could not be detected. Further, the influences due to modulation of the airflow / acoustic signal in the vocal tract are not reflected by the GAW. Also, the actually 3-dimensional vocal fold oscillations are not entirely reflected by the one dimensional GAW. This means that 2D and 3D oscillatory characteristics of the vocal folds may be better suited to reflect changes in the acoustic signal than 1D-GAW features do [62, 63]. This also aligns with previous findings, that GAW-based parameters are less important for healthy / FD classification tasks than parameters based on a more complex signal describing the vocal fold oscillation pattern (i.e. Phonovibrogram-based parameters) [38].

To summarize, the main gains from this investigation are as follows:

1. Values of investigated parameters for healthy and FD subjects were not clearly separable. A table containing norm values (Minimum, maximum, Mean, median and standard deviation) for all parameters in all four investigated groups are provided (S2 Table). All parameters were obtained from 250 ms long sustained phonation data (vowel /i/).
2. In many cases parameters are correlated with F_0 , which may require a correction for the influence of F_0 on these parameters in future studies. We provide a comprehensive list of parameters statistically significantly associated with F_0 (Table 2).
3. Mostly, linear relations were found between GAW and acoustic parameters. Non-linear relations were only subjectively observable and weak. Further, no strong relations between GAW and acoustic signals, excluding F_0 , were found in females or males. This implies that no clear redundancy exists between both signals but also suggests that the GAW may be a too simplified one dimensional representation of the vocal fold oscillations.

Shortcomings

In this study more females than males have been investigated which influences the comparisons as explained in the discussion section. This imbalance was not avoidable without excluding many female subjects, since the vast majority of our clinical referrals are females, being similar to other clinics [64]. Also, voice pathologies are more common in females than males [65]. Further, subject age differed between healthy and disordered groups. Albeit we found no strong influence of subject age in a previous study [38], the influence of age on voice parameters is well documented in literature [66–68] and may have influenced the results.

FD is a diagnosis of exclusion and hence a broad term uniting a vast amount of different voice disorders that all have varying symptoms and causes [4, 6]. This means that a table of norm values for FD subjects is only of limited utility, since many parameter values describing only certain features of the voice may also be in the normal range for most of the subcategories of FD. Only for specific subcategories of FD, certain parameter values may deviate. In addition, the analyzed phonatory condition was limited to sustained phonation on vowel /i/. Other

paradigms as pitch raise or phonating other vowels will have to be investigated in order to analyze if they are more suitable to differentiate between healthy and FD subjects. However, since we only looked for more general relations between parameters and only limited data was available, the distinction of a large number of FD subcategories was not feasible.

The acoustic signal was recorded in a clinical setting using a clip-microphone and hence was often noisy. We addressed this problem by rating all acoustic signals in regard of signal and background noise and only used data with acceptable external noise levels.

The GAW is only a 1-dimensional representation of the vocal fold oscillation process and hence does not describe the whole information contained in the 2D-HSV recordings [20, 69] or the 3D vocal fold oscillations [62]. For further investigations in vocal fold—acoustic relations, Phonovibrogram-based parameters could be also considered, since the Phonovibrogram is a more complex, 2-dimensional representation of the vocal fold oscillations [63].

More signals, parameters and alternating definitions of parameters exist [25] that were not investigated in this study. Also, exact parameter definitions may differ between software tools [70].

Conclusion

In this study healthy and FD subjects were not separable by single parameter values. Still, we presented [S2 Table](#) containing values for male and female, healthy and FD subjects obtained from 250 ms long sustained phonation data (vowel /i/). This table does not rest upon a sufficiently large and diverse number of subjects to be used as a reference for clinical parameter value ranges. However, it can be expanded and supplemented in future studies to eventually lay the fundamentals for the development of software tools that may allow for objective clinical voice assessment and assisting clinicians.

About half of all 49 investigated parameters were found to be correlated statistically significantly with acoustic or GAW-based F_0 [Mean]. Albeit most correlations were low (between 0.3 and 0.5) this still implies a measurable influence of F_0 on the affected parameters. We suggest that, if the parameters affected by F_0 are used in the future, it may be required to correct for the influence of F_0 , at least for the stronger affected parameters PhAI [Std] and F_0 [Std].

Only low (and in one case barely moderate) correlations between not F_0 -related GAW- and acoustic-based parameters were found in females and males. Although no strong relations between features of the GAW and acoustic signal besides F_0 could be found in this work, these findings show the gain of synchronous HSV and acoustic recordings, since not much redundancy is present in both signals. Also, based on these only weak relations between acoustic and GAW-parameters, we conclude that other features besides the glottal area (i.e. specific vocal fold oscillation patterns or the vocal tract) may play a more prominent role in determining acoustic characteristics than the GAW.

Supporting information

S1 Table. Names, abbreviations, sources and descriptions of all 49 parameters.
(DOCX)

S2 Table. Minimum, maximum, mean, median, standard deviation, skewness and kurtosis values for all four groups (N_F , FD_F , N_M , FD_M) as well as distributions plots for all parameter values.
(XLSX)

S3 Table. PCCs and DCCs and respective p-values (after FDR-correction) for all paired parameter correlations calculated in this study.

(XLSX)

Author Contributions**Conceptualization:** Patrick Schlegel, Michael Döllinger.**Data curation:** Patrick Schlegel, Andreas M. Kist, Stephan Dürr, Anne Schützenberger.**Formal analysis:** Patrick Schlegel.**Funding acquisition:** Michael Döllinger, Anne Schützenberger.**Investigation:** Patrick Schlegel.**Methodology:** Patrick Schlegel.**Project administration:** Michael Döllinger, Anne Schützenberger.**Resources:** Stephan Dürr, Michael Döllinger.**Software:** Patrick Schlegel, Andreas M. Kist.**Supervision:** Michael Döllinger, Anne Schützenberger.**Validation:** Andreas M. Kist, Melda Kunduk, Anne Schützenberger.**Visualization:** Patrick Schlegel, Andreas M. Kist.**Writing – original draft:** Patrick Schlegel.**Writing – review & editing:** Patrick Schlegel, Andreas M. Kist, Melda Kunduk, Michael Döllinger, Anne Schützenberger.**References**

1. Titze IR. Principles of voice production. 2nd ed.: National Center for Voice and Speech, Iowa City, Iowa; 2000.
2. Colton RH, Woo P. Diagnosis and Treatment of Voice Disorders. In Rubin JS, Sataloff RT, Korovin GS, editors.: Plural Publishing Inc.; 2014. p. 253–287.
3. Wendler J, Seidner W, Eysholdt U. Lehrbuch der Phoniatrie und Pädaudiologie. In.: Thieme, Stuttgart, Germany; 2005. p. 139–189.
4. Wilson JA, Deary IJ, Scott S, MacKenzie K. Functional dysphonia. *BMJ*. 1995; 311: p. 1039. <https://doi.org/10.1136/bmj.311.7012.1039> PMID: 7580648
5. Aronson AE. Importance of the psychosocial interview in the diagnosis and treatment of “functional” voice disorders. *Journal of Voice*. 1990; 4(4): p. 287–289. [https://doi.org/10.1016/S0892-1997\(05\)80043-8](https://doi.org/10.1016/S0892-1997(05)80043-8)
6. Roy N. Functional dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery*. 2003; 11(3): p. 144–148. <https://doi.org/10.1097/00020840-200306000-00002> PMID: 12923352
7. Deliyiski D. Laryngeal Evaluation Kendall K, Leonard R, editors.: Georg Thieme, New York city, New York; 2010.
8. Morris R, Harmo AB. The Handbook of Language and Speech Disorders. In Damico JS, Müller N, Ball MJ, editors.: Wiley-Blackwell, Chichester, England; 2013. p. 455–473.
9. 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. *Journal of Voice*. 2011; 25(5): p. 576–590. <https://doi.org/10.1016/j.jvoice.2010.04.004> PMID: 20728308
10. Unger J, Schuster M, Hecker DJ, Schick B, Lohscheller J. A generalized procedure for analyzing sustained and dynamic vocal fold vibrations from laryngeal high-speed videos using phonovibrograms. *Artificial Intelligence in Medicine*. 2016; 66: p. 15–28. <https://doi.org/10.1016/j.artmed.2015.10.002> PMID: 26597002

11. Uloza V, Vegienė A, Pribušienė R, Šaferis V. Quantitative evaluation of video laryngostroboscopy: reliability of the basic parameters. *Journal of Voice*. 2013; 27(3): p. 361–368. <https://doi.org/10.1016/j.jvoice.2012.12.007> PMID: 23465526
12. Hess MM, Herzel H, Köster O, Scheurich F, Gross M. Endoskopische Darstellung von Stimmlippenschwingungen Digitale Hochgeschwindigkeitsaufnahmen mit verschiedenen Systemen. *HNO*. 1996; 44(12): p. 685–693. PMID: 9081953
13. Niimi S, Miyaji M. Vocal fold vibration and voice quality. *Folia Phoniatrica et Logopaedica*. 2000; 52(1–3): p. 32–38. <https://doi.org/10.1159/000021510> PMID: 10474002
14. Kreiman J, Gerratt BR. Perception of aperiodicity in pathological voice. *The Journal of the Acoustical Society of America*. 2005; 117: p. 2201. <https://doi.org/10.1121/1.1858351> PMID: 15898661
15. Zacharias SRC, Deliyiski DD, Gerlach TT. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *Journal of Voice*. 2018; 32(2): p. 216–220. <https://doi.org/10.1016/j.jvoice.2017.05.002> PMID: 28596101
16. Stellan H. What have we learned about laryngeal physiology from high-speed digital videoendoscopy? *Current Opinion in Otolaryngology & Head and Neck Surgery*. 2005; 13(3): p. 152–156. <https://doi.org/10.1097/01.moo.0000163451.98079.ba> PMID: 15908812
17. Döllinger M. The next step in voice assessment: High-speed digital endoscopy and objective evaluation. *Current Bioinformatics*. 2009; 4(2): p. 101–111. <https://doi.org/10.2174/157489309788184774>
18. Echternach M, Döllinger M, Sundberg J, Traser L, Richter B. Vocal fold vibrations at high soprano fundamental frequencies. *The Journal of the Acoustical Society of America*. 2013; 133(2): p. 82–87. <https://doi.org/10.1121/1.4773200> PMID: 23363198
19. Braunschweig T, Flaschka J, Schelhorn-Neise P., Döllinger M. High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias. *Medical Engineering & Physics*. 2008; 30(1): p. 59–66. <https://doi.org/10.1016/j.medengphy.2006.12.007> PMID: 17317268
20. Döllinger M, Dubrovskiy D, Patel R. Spatiotemporal analysis of vocal fold vibrations between children and adults. *The Laryngoscope*. 2012; 122(11): p. 2511–2518. <https://doi.org/10.1002/lary.23568> PMID: 22965771
21. Bohr C, Kraeck A, Eysholdt U, Ziethe A, Döllinger M. Quantitative analysis of organic vocal fold pathologies in females by high-speed endoscopy. *The Laryngoscope*. 2013; 123(7): p. 1686–1693. <https://doi.org/10.1002/lary.23783> PMID: 23649746
22. Patel RR, Unnikrishnan H, Donohue KD. Effects of vocal fold nodules on glottal cycle measurements derived from high-speed videoendoscopy in children. *Plos one*. 2016; 11(4): p. e0154586. <https://doi.org/10.1371/journal.pone.0154586> PMID: 27124157
23. Petermann S, Döllinger M, Kniesburges S, Ziethe A. Analysis method for the neurological and physiological processes underlying the pitch-shift reflex. *Acta Acustica united with Acustica*. 2016; 102(2): p. 284–297. <https://doi.org/10.3813/AAA.918944>
24. Döllinger M, Kunduk M, Kaltenbacher M, Vondenhoff S, Ziethe A, Eysholdt U, et al. Analysis of vocal fold function from acoustic data simultaneously recorded with high-speed endoscopy. *Journal of Voice*. 2012; 26(6): p. 726–733. <https://doi.org/10.1016/j.jvoice.2012.02.001> PMID: 22632795
25. Pedersen M, Jønsson A, Mahmood S, Agersted A. Which mathematical and physiological formulas are describing voice pathology: an overview. *Journal of General Practice*. 2016; 4(3): p. 253. <https://doi.org/10.4172/2329-9126.1000253>
26. Hohm J, Döllinger M, Bohr C, Kniesburges S, Ziethe A. Influence of F₀ and sequence length of audio and electroglottographic signals on perturbation measures for voice assessment. *Journal of Voice*. 2015; 29(4): p. 517.e11–517.e21. <https://doi.org/10.1016/j.jvoice.2014.10.001> PMID: 25944290
27. Schützenberger A, Döllinger MK, Alexiou C, Dubrovskiy D, Semmler M, Seger A, et al. Laryngeal high-speed videoendoscopy: sensitivity of objective parameters towards recording frame rate. *BioMed Research International*. 2016; 2016: p. Article ID 4575437, 19 pages. <https://doi.org/10.1155/2016/4575437> PMID: 27990428
28. Schlegel P, Kunduk M, Stingl M, Semmler M, Döllinger M, Bohr C, et al. Influence of spatial camera resolution in high-speed videoendoscopy on laryngeal parameters. *PLoS ONE*. 2019; 14(4): p. e0215168. <https://doi.org/10.1371/journal.pone.0215168> PMID: 31009488
29. Schlegel P, Semmler M, Kunduk M, Döllinger M, Bohr C, Schützenberger A. Influence of Analyzed Sequence Length on Parameters in Laryngeal High-Speed Videoendoscopy. *Applied Sciences-Basel*. 2018; 8(12): p. 2666. <https://doi.org/10.3390/app8122666>
30. Baken RJ, Orlikoff RF. *Clinical measurement of speech & voice (Speech Science)*. Second edition ed.: Cengage Learning, Clifton Park, New York; 1999.

31. Werth K, Voigt D, Döllinger M, Eysholdt U, Lohscheller J. Clinical value of acoustic voice measures: a retrospective study. *European Archives of Oto-Rhino-Laryngology*. 2010; 267(8): p. 1261–1271. <https://doi.org/10.1007/s00405-010-1214-2> PMID: 20567980
32. Wilcox KA, Horii Y. Age and changes in vocal jitter. *Journal of Gerontology*. 1980; 35(2): p. 194–198. <https://doi.org/10.1093/geronj/35.2.194> PMID: 7410776
33. Hirai R, Yoshihashi H, Sakuma N, Ikeda M. Relationship between HSV imaging and acoustic parameters. *Otolaryngology—Head and Neck Surgery*. 2017; 143(2_suppl): p. 219–220. <https://doi.org/10.1016/j.otohns.2010.06.435>
34. Uloza V, Vegienė A, Šaferis V. Correlation between the basic video laryngostroboscopic parameters and multidimensional voice measurements. *Journal of Voice*. 2013; 27(6): p. 744–752. <https://doi.org/10.1016/j.jvoice.2013.06.008> PMID: 24128894
35. Mehta DD, Deliyski DD, Zeitels SM, Quatieri TF, Hillman RE. Voice production mechanisms following phonosurgical treatment of early glottic cancer. *Annals of Otolaryngology, Rhinology & Laryngology*. 2010; 119(1): p. 1–9. <https://doi.org/10.1177/000348941011900101> PMID: 20128179
36. Chen G, Kreiman J, Shue YL, Alwan A. Acoustic correlates of glottal gaps. In 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011), Florence, Italy; 2011: ISCA Archive. p. 2673–2676. https://www.isca-speech.org/archive/interspeech_2011/i11_2673.html
37. Popolo PS, Johnson AM. Relating Cepstral Peak Prominence to cyclical parameters of vocal fold vibration from high-speed videoendoscopy using machine learning: a pilot study. *Journal of Voice*. Accepted 2020; In Press. <https://doi.org/10.1016/j.jvoice.2020.01.026> PMID: 32173147
38. Schlegel P, Kniesburges S, Dürr S, Schützenberger A, Döllinger M. Machine learning based identification of relevant parameters for functional voice disorders derived from endoscopic high-speed recordings. *scientific reports*. 2020; 10: p. 10517. <https://doi.org/10.1038/s41598-020-66405-y> PMID: 32601277
39. Schlegel P, Stingl M, Kunduk M, Kniesburges S, Bohr C, Döllinger M. Dependencies and ill-designed parameters within high-speed videoendoscopy and acoustic signal analysis. *Journal of Voice*. 2018; 33(5): p. 811.e1–811.e12. <https://doi.org/10.1016/j.jvoice.2018.04.011> PMID: 29861291
40. Horii Y. Vocal shimmer in sustained phonation. *Journal of Speech, Language, and Hearing Research*. 1980; 23(1): p. 202–209. <https://doi.org/10.1044/jshr.2301.202> PMID: 7442177
41. Deal RE, Emanuel FW. Some waveform and spectral features of vowel roughness. *Journal of Speech, Language, and Hearing Research*. 1978; 21(2): p. 250–264. <https://doi.org/10.1044/jshr.2102.250> PMID: 703275
42. 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. https://www.isca-speech.org/archive/eurospeech_1993/e93_1973.html
43. Jesus Goncalves MH. Methodenvergleich zur Bestimmung der glottalen Mittelachse bei endoskopischen Hochgeschwindigkeitsvideoaufnahmen von organisch basierten pathologischen Stimmgebungsprozessen. phdthesis., Friedrich-Alexander-University Erlangen-Nürnberg; 2015. <https://d-nb.info/1076911994/34>
44. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America*. 1988; 84(2): p. 511–529. <https://doi.org/10.1121/1.396829> PMID: 3170944
45. Timcke R, Leden H, Moore P. Laryngeal vibrations: measurements of the glottic wave. *AMA Arch Otolaryngol*. 1958; 68(1): p. 1–19. <https://doi.org/10.1001/archotol.1958.00730020005001> PMID: 13544677
46. Kunduk M, Döllinger M, McWhorter AJ, Lohscheller J. Assessment of the variability of vocal fold dynamics within and between recordings with high-speed imaging and by phonovibrogram. *The Laryngoscope*. 2010; 120(5): p. 981–987. <https://doi.org/10.1002/lary.20832> PMID: 20422695
47. Mehta DD, Zañartu M, Quatieri TF, Deliyski DD, Hillman RE. Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy. *The Journal of the Acoustical Society of America*. 2011; 130(6): p. 3999–4009. <https://doi.org/10.1121/1.3658441> PMID: 22225054
48. Chen G, Kreiman J, Gerratt BR, Neubauer J, Shue YL, Alwan A. Development of a glottal area index that integrates glottal gap size and open quotient. *The Journal of the Acoustical Society of America*. 2013; 133: p. 1656–1666. <https://doi.org/10.1121/1.4789931> PMID: 23464035
49. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*. 1994; 37(4): p. 769–778. <https://doi.org/10.1044/jshr.3704.769> PMID: 7967562
50. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*. 1982; 71(6): p. 1544–1550. <https://doi.org/10.1121/1.387808> PMID: 7108029

51. Lessing J. Entwicklung einer Klassifikationsmethode zur akustischen Analyse fortlaufender Sprache unterschiedlicher Stimmgüte mittels Neuronaler Netze und deren Anwendung. phdthesis., Georg-August-Universität Göttingen, Mathematisch-Naturwissenschaftlich Fakultät; 2007. <https://ediss.uni-goettingen.de/bitstream/handle/11858/00-1735-0000-0006-B45D-7/lessing.pdf?sequence=1>
52. Kasuya H, Ogawa S, Mashima K, Ebihara S. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *The Journal of the Acoustical Society of America*. 1986; 80(5): p. 1329–1334. <https://doi.org/10.1121/1.394384> PMID: 3782609
53. Klingholz F. Acoustic representation of speaking-voice quality. *Journal of Voice*. 1990; 4(3): p. 213–219. [https://doi.org/10.1016/S0892-1997\(05\)80016-5](https://doi.org/10.1016/S0892-1997(05)80016-5)
54. Qi Y, Hillman RE, Milstein C. The estimation of signal-to-noise ratio in continuous speech for disordered voices. *Journal of the Acoustical Society of America*. 1999; 105(4): p. 2532–2535. <https://doi.org/10.1121/1.426860> PMID: 10212434
55. Székely GJ, Rizzo ML, Bakirov NK. Measuring and testing dependence by correlation of distances. *Annals of Statistics*. 2007; 35(6): p. 2769–2794. <https://doi.org/10.1214/009053607000000505>
56. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*. 2001; 29(4): p. 1165–1189. http://www.math.tau.ac.il/~ybenja/MyPapers/benjamini_yekutieli_ANNSTAT2001.pdf
57. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*. 2012; 24(3): p. 69–71. PMID: 23638278
58. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*. 1981; 24(6): p. 381–395. <https://doi.org/10.1145/358669.358692>
59. Voigt D, Döllinger M, Braunschweig T, Yang A, Eysholdt U, Lohscheller J. Classification of functional voice disorders based on phonovibrograms. *Artificial Intelligence in Medicine*. 2010; 49(1): p. 51–59. <https://doi.org/10.1016/j.artmed.2010.01.001> PMID: 20138486
60. Schlegel P, Kist A, Semmler M, Döllinger M, Kunduk M, Dürr S, et al. Determination of clinical parameters sensitive to functional voice disorders applying boosted decision stumps. *IEEE Journal of Translational Engineering in Health and Medicine*. 2020; 8: p. 1–11. <https://doi.org/10.1109/JTEHM.2020.2985026> PMID: 32518739
61. Hsiao TY, Liu CM, Hsu CJ, Lee SY, Lin KN. Vocal fold abnormalities in laryngeal tension-fatigue syndrome. *Journal of the Formosan Medical Association*. 2001; 100(12): p. 837–840. PMID: 11802526
62. Semmler M, Kniesburges S, Birk V, Ziethe A, Patel R, Döllinger M. 3D reconstruction of human laryngeal dynamics based on endoscopic high-speed recordings. *IEEE Transactions on Medical Imaging*. 2016; 35(7): p. 1615–1624. <https://doi.org/10.1109/TMI.2016.2521419> PMID: 26829782
63. 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 Transactions on Medical Imaging*. 2008; 27(3): p. 300–309. <https://doi.org/10.1109/TMI.2007.903690> PMID: 18334426
64. Morton V., Watson DR. The teaching voice: problems and perceptions. *Logopedics, phoniatrics, vocology*. 1998; 23 p. 133–139.
65. Van Houtte E, Van Lierde K, D'Haeseleer E, Claeys S. The prevalence of laryngeal pathology in a treatment-seeking population with dysphonia. *The Laryngoscope*. 2010; 120(2): p. 306–312. <https://doi.org/10.1002/lary.20696> PMID: 19957345
66. Honjo I, Isshiki N. Laryngoscopic and voice characteristics of aged persons. *Arch Otolaryngol*. 1980; 106(3): p. 149–150. <https://doi.org/10.1001/archotol.1980.00790270013003> PMID: 7356434
67. Winkler R, Sendlmeier W. EGG open quotient in aging voices—changes with increasing chronological age and its perception. *Logopedics Phoniatrics Vocology*. 2006. 31(2): p. 51–56. <https://doi.org/10.1080/14015430500445534> PMID: 16754276
68. Xue SA, Deliyiski D. Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology*. 2001. 27(2): p. 159–168. <https://doi.org/10.1080/03601270151075561>
69. Wurzbacher T, Döllinger M, Schwarz R. Spatiotemporal classification of vocal fold dynamics by a multi-mass model comprising time-dependent parameters. *The Journal of the Acoustical Society of America*. 2008; 123: p. 2324. <https://doi.org/10.1121/1.2835435> PMID: 18397036
70. Bielamowicz S, Kreiman J, Gerratt B, Dauer M, Berke G. Comparison of voice analysis systems for perturbation measurement. *Journal of Speech and Hearing Research*. 1996 Feb; 39(1): p. 126–134. <https://doi.org/10.1044/jshr.3901.126> PMID: 8820704