JASA ARTICLE

Are source-filter interactions detectable in classical singing during vowel glides?

Matthias Echternach,[1,a)] Christian T. Herbst,[2] Marie Köberlein,[1,b)] Brad Story,[3] Michael Döllinger,[4] and Donata Gellrich[1]

[1]*Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Munich University Hospital (LMU), Marchioninistrasse 15, Munich, 81377, Germany*

[2]*Antonio Salieri Department of Vocal Studies and Vocal Research in Music Education, University of Music and Performing Arts Vienna, Vienna, Austria*

[3]*Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85718, USA*

[4]*Division of Phoniatrics and Pediatric Audiology at the Department of Otorhinolaryngology Head and Neck Surgery, University Hospital Erlangen, Medical School Waldstrasse 1, Erlangen, 91054, Germany*

ABSTRACT:

In recent studies, it has been assumed that vocal tract formants (F_n) and the voice source could interact. However, there are only few studies analyzing this assumption *in vivo*. Here, the vowel transition /i/-/a/-/u/-/i/ of 12 professional classical singers (6 females, 6 males) when phonating on the pitch D4 [fundamental frequency (f_o) ca. 294 Hz] were analyzed using transnasal high speed videoendoscopy (20.000 fps), electroglottography (EGG), and audio recordings. F_n data were calculated using a cepstral method. Source-filter interaction candidates (SFICs) were determined by (a) algorithmic detection of major intersections of F_n/nf_o and (b) perceptual assessment of the EGG signal. Although the open quotient showed some increase for the /i-a/ and /u-i/ transitions, there were no clear effects at the expected F_n/nf_o intersections. In contrast, f_o adjustments and changes in the phonovibrogram occurred at perceptually derived SFICs, suggesting level-two interactions. In some cases, these were constituted by intersections between higher nf_o and F_n. The presented data partially corroborates that vowel transitions may result in level-two interactions also in professional singers. However, the lack of systematically detectable effects suggests either the absence of a strong interaction or existence of confounding factors, which may potentially counterbalance the level-two-interactions. https://doi.org/10.1121/10.0005432

(Received 4 December 2020; revised 1 June 2021; accepted 3 June 2021; published online 28 June 2021)

[Editor: Zhaoyan Zhang] Pages: 4565–4578

I. INTRODUCTION

The human voice production apparatus is composed of three subsystems (Sundberg, 1987; Titze, 1994; Herbst, 2017). The first subsystem, i.e., the breathing apparatus, including the lungs, provides the subglottal pressure, which, on the one hand, abducts the membranous part of the closed vocal folds and, on the other hand, defines the transglottal pressure difference, which—in conjunction with the resistance caused by vocal fold adduction (Sundberg, 1987; Titze, 1994)—determines the transglottal airflow. This subsystem, thus, acts as a power source because (a) energy is transferred to the vocal fold oscillation and (b) the resulting sound pressure level (SPL) is greatly dependent on the subglottal pressure, initiating the transglottal pressure difference (Bouhuys *et al.*, 1968). The second subsystem includes the oscillating vocal folds, which interrupt the transglottal airflow, producing air pulses that generate harmonic sound waves (Sundberg, 1987; Titze, 1994). The acoustic product has been frequently denoted as the voice source (Sundberg, 1987). In the third subsystem, vocal tract resonances modify the sound wave structure of the voice source.

In a first approximation, the human resonance system, the vocal tract, acts as a linear filter system (Fant, 1960; Sundberg, 1987). If the frequency of an individual voice source harmonic lies near the center frequency of a vocal tract resonance, that particular harmonic's contribution to the radiated sound is relative strong, producing peaks in the acoustic spectrum that are frequently denoted as formants (Titze *et al.*, 2015). In contrast, if a voice source harmonic does not match any vocal tract resonance in the frequency domain, the radiated amplitude of this harmonic would be relatively weaker. This assumption has been described through the source-filter theory (Fant, 1960; Sundberg, 1987). In many studies, it has been shown that for many kinds of voice production, i.e., in speech, the estimations of the voice source by removing the vocal tract resonances, for example, by inverse filtering, are very precise given that the fundamental frequency (f_o) is low (La and Sundberg, 2015; Echternach *et al.*, 2016; Sundberg *et al.*, 2016).

Originally, the source-filter theory was established as a linear system in which the voice source operated independently from the vocal tract. Decades ago, that theory was,

[a)]Electronic mail: matthias.echternach@med.uni-muenchen.de, ORCID: 0000-0003-0095-5360.

[b)]Also at: Freiburg Institute of Musicians' Medicine, Freiburg University, Freiburg, Germany

however, extended by the possibility for interactions between the source and filter (Flanagan, 1968; Ishizaka and Flanagan, 1972), i.e., that properties of the vocal tract could have an influence of the behaviour of the laryngeal voice source (Rothenberg, 1981). Using computational modelling approaches, it was suggested that these interactions could affect the voice source on two levels, i.e., influencing the transglottal airflow (level-one interaction) and the quality of vocal fold oscillation (level-two interaction; Titze, 2008; Titze *et al.*, 2008; Maxfield *et al.*, 2017). It has been assumed that such an interaction is dependent on the produced vowel and, therefore, the vocal tract resonances in relation to the frequencies of the individual voice source harmonics (Sundberg *et al.*, 2016).

The pattern of resonances for a given vocal tract configuration can be characterized by the frequency-dependent impedance. At frequencies below the peak of a resonance, the reactive part of the impedance is inertive and provides conditions that are theorized to facilitate the vocal fold oscillation (Titze, 2008). At the frequency of resonance (f_{Rn}), the reactive part of the impedance undergoes an abrupt shift from inertance to compliance and, consequently, from conditions that support vocal fold oscillation to conditions that may potentially produce vibratory instabilities or even cessation of phonation (Titze, 2008). Thus, it is of interest in current voice science to understand the nature of these interactions as the f_{o} or harmonic components pass through a vocal tract resonance.

There are few current physiologic studies in which this central notion is tested *in vivo*. Some authors argue that the importance might be considered to be rather high (Titze *et al.*, 2008; Tokuda *et al.*, 2010; Maxfield *et al.*, 2017), whereas others argue that such predicted instabilities are frequently not observable (Sundberg *et al.*, 2013). Also, Echternach *et al.* (2016), who calculated vocal tract resonance frequencies from magnetic resonance imaging, observed no clear instabilities in cases in which the frequency of a harmonic was slightly higher than f_{Rn}. In another study by Sundberg *et al.* (2016), differences of the closed quotient were observed between different vowel conditions during phonation at a relatively low f_{o}, which were interpreted as signs of vocal tract/voice source interaction. However, the observed differences appeared rather independent from the difference of a voice source's partial frequency and the frequency of the first excited resonance, the first formant ($F1$). To test the theory of the corresponding paper (Titze, 2008), in 2008, Titze *et al.* (2008) analyzed pitch glides and found that 30% of the instabilities were present when the f_{o} was within the bandwidth of f_{R1}. Furthermore, Maxfield *et al.* (2017) analyzed pitch glides in eight participants using audio and electroglottographical (EGG) recordings with and without phonation in a tube with varying vocal tract resonances. They observed stronger interactions in the region of crossings of lower harmonics and lower vocal tract resonances. In 2017, Wade *et al.* (2017) analyzed $f_{\mathrm{o}}/f_{\mathrm{R1}}$ crossings using pitch glides in eight soprano participants. They observed that instabilities occurred more frequently at the borders of the vocal f_{o} range but not with regard to $f_{\mathrm{o}}/f_{\mathrm{R1}}$ crossings.

In 2011, Zanartu *et al.* (2011) documented the source-filter coupling with high speed videolaryngoscopy (HSV) in a single participant. To study source induced and acoustic induced irregularities, these investigators used two different vowel conditions, i.e., /i/ with a low f_{R1} for the acoustic induced vocal fold irregularities and /ae/ with a higher f_{R1} for the source induced vocal fold irregularities. During descending pitch glides for the /i/ vowel, transnasal HSV was performed. In contrast, rigid endoscopy was used for ascending glides on /ae/. The authors observed stronger irregularities for the /i/ vowel condition, indicating an interaction of vocal tract resonances with the voice source. Finally, vocal tract/voice source interactions could also be considered as a reason for irregularities observable during the vocal register transition in female voices around 700–800 Hz, i.e., the upper *passaggio* in which perceptual changes of the vocal quality frequently occur in untrained voices. These results are comparable to those of Svec *et al.* (2008) and Echternach *et al.* (2017a), who found changes of laryngeal oscillation patterns in this f_{o} range, i.e., where f_{R1} is very close to f_{o}. However, in these studies, it has not been clarified if the observed changes of the laryngeal oscillation patterns are a consequence of a primary change of the laryngeal mechanism or a consequence of vocal tract/voice source interaction.

Instead of varying f_{o}, $f_{\mathrm{Rn}}/nf_{\mathrm{o}}$ crossings could also be provoked by changing the resonatory properties, i.e., by performing vowel quality changes at a constant phonatory pitch (attempting to produce a stable f_{o}). This was attempted in the aforementioned study by Titze *et al.* (2008), where the participants performed gradual vowel transitions on the pitches C4 and C5 for male participants and C5 and C6 for female participants, respectively. For their male participants, these authors found instabilities with respect to a $f_{\mathrm{R1}}/nf_{\mathrm{o}}$ crossings in 54% of all cases and only 23% without crossings (Titze *et al.*, 2008). Notably, the occurrence of instabilities that were found for female voices was considerably lower with 34% for the $f_{\mathrm{R1}}/nf_{\mathrm{o}}$ crossing conditions. In their 2017 publication, Wade *et al.* (2017) repeated the experiments by Titze *et al.* (2008) for their eight soprano participants and found a comparable amount of instabilities (47%) but with stronger disagreements of expected and observed instabilities of the voice signals with regard to the $f_{\mathrm{o}}/f_{\mathrm{R1}}$ crossings. Both studies estimated f_{Rn} by separate recordings of the vocal fry, assuming that the resonatory properties would be transferable. However, as also stated by Wade *et al.* (2017), the vocal fry does not reliably indicate the f_{R1} for singing because, usually, singers avoid a crossing of $f_{\mathrm{o}}/f_{\mathrm{R1}}$. Instead, singers frequently tune $f_{\mathrm{o}}/f_{\mathrm{R1}}$ for higher pitches (Sundberg, 1975; Joliveau *et al.*, 2004). Therefore, for soprano voices, such an $f_{\mathrm{o}}/f_{\mathrm{R1}}$ crossover is only found at very high pitches around 1000 Hz or higher (Garnier *et al.*, 2010; Echternach *et al.*, 2015), preventing analysis of $f_{\mathrm{o}}/f_{\mathrm{R1}}$ crossovers in other realistic vocalization settings. Furthermore, both studies included audio and EGG signals

https://doi.org/10.1121/10.0005432

only, and were not able to assess laryngoscopical vocal fold oscillation patterns. Most studies analyze irregularities in the f_o range where nf_o matches a bandwidth of f_{Rn}. According to the reasoning provided by Titze (2008), during a crossover of nf_o/f_{Rn}, the biodynamics of voice production should suddenly change from stable (nf_o is slightly lower than f_{Rn}) via the crossing (nf_o matches f_{Rn}) to a potentially unstable part (nf_o is slightly higher than f_{Rn}.

This study aims to analyze such gradual vocal tract resonance changes in professional singers. Because professional Western classically trained singers are trained to avoid irregularities of voice production, the confirmation of voice source/vocal tact interactions in this group could augment the importance of such interactions in voice physiology also. The study focuses not on the f_o/f_{R1} crossover but on crossings between the lowest two vocal tract resonances and higher voice source harmonics (nf_o/f_{Rn}). As was pointed out before, a gradual vowel change experiment, including a f_o/f_{R1} crossover appears—at least for female voices—problematic because of the need for very high f_o's. Furthermore, such f_o/f_{R1} crossovers are rare in the singers' reality. Therefore, the presented study focuses on higher nf_o/f_{Rn}'s, which are especially present when singing diphthongs. In agreement with Titze (2008) and Titze *et al.* (2008), it is hypothesized that during vowel changes, instabilities or changes in vibratory characteristics would be observable in vocal fold oscillatory characteristics (level-two interactions), and such instabilities or vibration changes would be greater when a partial of the voice source would be equal to or slightly higher than f_{Rn}. Furthermore, because of higher acoustic energy, it is hypothesized that crossings of lower frequency harmonic components with lower f_{Rn} exhibit more prominent interactions and are, thus, more likely to produce irregularities as compared to crossings of higher partials and higher f_{Rn}'s. Finally, following the reasoning of Sundberg *et al.* (2016), it is hypothesized that phonation on different vowels would result in different open quotients due to the source-filter interactions.

II. MATERIAL AND METHODS

After approval from the local ethical committee (Medical Ethics Committee of the University of Munich, Nr.18/769), 12 professional singers were initially included. Table I shows their ranking in the taxonomy proposed by Bunch and Chapman (2000). None of the participants reported self-perceived voice problems. Vocal fold pathologies were excluded via phoniatric evaluation using videostroboscopy or high speed laryngoscopy and the Voice Handicap Index (VHI; Nawka *et al.*, 2003).

A. Task

Regardless of gender, all participants were asked to sing a sustained note on pitch D4 ($f_o \approx 294$ Hz) with gradual vowel transitions from /i/ via /a/ and /u/ to /i/; see Fig. 1. The pitch of D4 was provided by the examiner prior to each phonation. Every vowel should be sustained for approximately 1 s, and the change of vowel quality should be performed as smoothly as possible, avoiding major vibrato. One single participant (B2) was permanently more than 60 Hz below the required f_o and was, consequently, excluded from further analysis.

As per the design of this vocal task, three source-filter interaction candidates (SFICs) were expected, occurring at major intersections of harmonics (nf_o) and vocal tract resonances (f_{Rn}) as indicated by the green boxes shown in Fig. 1.

The first SFIC occurs at the vowel transition from /i/ to /a/ for $2f_o$ and f_{R1}; the second intersection is located at the vowel transition from /a/ to /u/ for $2f_o$ and f_{R1}; and the third SFIC occurs at the vowel transition from /u/ to /a/ for $3f_o$ and f_{R2} or/and $4f_o$ and f_{R2}. According to Titze (2008), it was expected that during these events, irregularities of the vocal fold oscillations (level-two interactions) could occur when nf_o would match f_{Rn} or be slightly higher than f_{Rn}, i.e., for the /i/ to /a/ transition before, for /a/ to /u/ after, and for /u/ to /i/ before the crossover.

B. Recordings

The participants' phonations were simultaneously documented with transnasal high speed videoendoscopy, electroglottography (EGG), and acoustic recording. In contrast to HSV recordings, which provide a two-dimensional laryngoscopic view to the vocal folds, EGG reflects the

TABLE I. Participants and their classification according to the taxonomy proposed by Bunch and Chapman (2000).

Subject	Classification		Taxonomy
S1	Soprano	7.1/4.5	Fulltime voice student university–postgraduate/regional touring: Concert, oratorio, recital
S2	Soprano	7.1/4.5	Fulltime voice student university–postgraduate/regional touring: Concert, oratorio, recital
S3	Soprano	2.15b1/4.5	International professional chorister/regional touring: Concert, oratorio, recital
M1	Mezzosoprano	2.15b1/2.4/4.1b	International professional chorister/international concert, oratorio, recital/regional opera minor principal
M2	Mezzosoprano	2.1	International opera principal
M3	Mezzosoprano	3.1c	National big city opera chorus
T1	Tenor	3.1b/3.4/3.15b1	National big city opera minor principal/concert/oratorio/recital
T2	Tenor	4.1b/4.5	Regional Touring Opera Minor Principal /Concert, Oratorio, Recital
T3	Tenor	2.1	International Opera Principal
B1	Baritone	4.5/6.1	Regional touring concert, oratorio, recital/singing teacher university and school
B2	Baritone	3.1b	National big city opera minor principal
B3	Baritone	7.1/4.5	Fulltime voice student university–postgraduate/regional touring: Concert, oratorio, recital

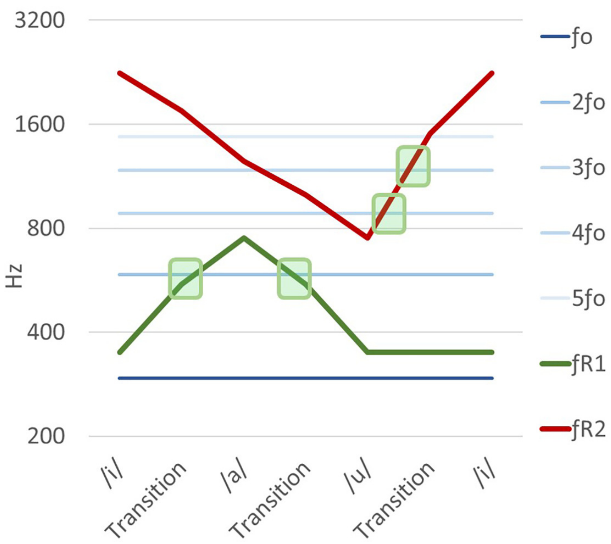


FIG. 1. (Color online) Schematic illustration of the experimental phonation task. The blue lines refer to the expected voice source partials and the red lines refer to the expected vocal tract resonance frequencies f_{R1} and f_{R2}. The green boxes show the expected nonlinear interaction candidates (SFICs), constituted by selected intersections of nf_o and f_{Rn}.

three-dimensional contact of the vocal folds by measuring the electric impedance changes. HSV recordings were performed using transnasal endoscopy with a Fastcam SA-X2 (Photron, Tokyo, Japan) and a flexible endoscope (ENF GP; Fa. Olympus, Hamburg, Germany) at a frame rate of 20 000 frames per second and a spatial resolution of 386 × 320 pixels, which was used in previous investigations (Echternach *et al.*, 2017a; Echternach *et al.*, 2017c). Audio signals were recorded with either a DPA IMK SC 4061 (DPA microphones, Alleroed, Denmark) or Sennheiser ME 62 microphone (Wedemark, Germany). EGG signals were captured with an EG2-PCX2 from Glottal Enterprises (Syracuse, NY). No anesthetic medication was applied for the transnasal endoscopic approach. The HSV videos were postprocessed by means of rotation, fast-Fourier-treatment, and cropping as described previously (Echternach *et al.*, 2017b). The calculation of the glottal area waveform (GAW) and phonovibrograms (PVGs) from the HSV images was performed based on Lohscheller and Eysholdt (2008) and Lohscheller *et al.* (2008). Furthermore, derivative electroglottogram (dEGG) wavegrams were constructed from the EGG signals as described by Herbst *et al.* (2010).

C. Construction of the SFIC crossing windows

To compare the oscillation characteristics of the sustained vowel conditions, a time window of 125 ms was constructed at the stable part of each sustained vowel, i.e., approximately at the temporal midpoint of each produced vowel. Concerning the SFIC, the vowel transition was analyzed with respect to a time window ("zero-window") of 25 ms in which the crossing of nf_o and f_{Rn} was expected. The crossing was constructed in the following way: First, an estimation of the first and second formant frequencies (F_1 and F_2)—as expression of the excited resonances f_{R1} and f_{R2}—was performed using a formant measurement technique based on cepstral analysis as described by Story and Bunton (2016). This method, designed specifically for analysis of formants when the f_o is high, such as in singing or children's speech, was shown to have errors of less than 10% when applied to synthetically generated speech signals with f_o ranging from 240 to 500 Hz (Story and Bunton, 2016). Then, the zero-window was constructed symmetrically around the point where the formant contour crossed the related partial (/i/ to /a/, $2f_o$ and F_1; /a/ to /u/, $2f_o$ and F_1; and /u/ to /a/, $3f_o$ and F_2 and $4f_o$ and F_2). The windows –2, –1, +1, and +2 refer to 25 ms windows before and after the zero-window, respectively.

Source-vocal tract interactions with the consequence of oscillatory irregularities are expected to be the greatest at the peak of an nf_o/f_{Rn} or F_n crossing (Titze, 2008; Titze *et al.*, 2008; Zanartu *et al.*, 2011). It has been shown before that irregularities of the EGG signals could be detectable using the EGG based sample entropy (Selamtzis and Ternstrom, 2014; Echternach *et al.*, 2017c; Wade *et al.*, 2017; Selamtzis *et al.*, 2018). Therefore, the sample entropy was measured using the algorithms introduced by Selamtzis and Ternstrom (2014) and Selamtzis *et al.* (2018) and was tested to determine if it could be used for the definition of the zero-window as performed in earlier studies (Echternach *et al.*, 2017b; Echternach *et al.*, 2017c; Echternach *et al.*, 2018; Echternach *et al.*, 2021). However, many transitions failed to show any rise of the EGG based sample entropy at all. Consequently, this measure was considered not meaningful for the definition of the zero-window.

D. Measures

The HSV video data were segmented with the Glottal Analysis Software (University Hospital at FAU Erlangen-Nürnberg, Germany), estimating the time-varying lateral deflection of the vocal fold edges along the anterior-posterior glottal dimension (Maryn *et al.*, 2020; Schlegel *et al.*, 2020). Based on the resulting GAW, the open quotient (OQ_{GAW}), closing quotient (ClQ; closing phase/period), and speed quotient were calculated for the mentioned windows using Glottal Analysis Tools (University Hospital at FAU Erlangen-Nürnberg, Germany) and Multi Signal Analyzer (University Hospital at FAU Erlangen-Nürnberg, Germany); see Table II. For the detection of OQ_{GAW}, a tolerance threshold of 5% was chosen. Consequently, all GAW values >5% from the baseline (i.e., pixel number of the fully open glottis) were considered to indicate an open glottis, whereas the condition GAW ≤5% was used as an indication for a closed glottis.

As noted previously, HSV segmentation data were used to create PVGs. Using custom scripts developed and written in the Python programming language by C.T.H., these PVG data were used to compute time-averaged glottovibrograms (GVGs; Karakozoglou *et al.*, 2012) before and after each SFIC to quantitatively compare the laryngeal vibration characteristics across each SFIC. This was achieved in the

TABLE II. Measures and origin.

HSV → Glottal area waveform	EGG	Acoustic signal
Closing quotient ($Q_{closing}$)		
Open quotient (OQ)	OQ	
Speed quotient (SQ)		
	Fundamental frequency (f_o)	Fundamental frequency (f_o) SPL
Relative average perturbation (RAP)		

following way: The glottal cycles of each PVG were segmented based on the corresponding GAW signals using the "To Point Process (periodic, cc)…" method of the Praat software (Boersma and Weenink, 2021). The glottal cycles at cycle indexes [−20..−10] and [10..20], respective to each SFIC, were further considered for the "pre" and "post" conditions, respectively. For both of these conditions, the PVG data of the relevant ten cycles were converted to GVGs through the simple addition of the left and right PVG parts and then normalized in time using two-dimensional cubic interpolation using the `imresize` function of Python's SciPy extension. The resulting time-normalized GVG data were then averaged for both the pre and post conditions, and the difference between the resulting averaged GVGs was computed—see the differential mean GVGs depicted in panels 3–5 in Fig. 2. From these GVGs, the differences concerning the opening and closing were calculated.

The EGG open quotient (OQ_{EGG} = contact $quotient_{EGG}$ −1) was calculated according to the method proposed by Howard (1995). The f_o was calculated from the EGG signal. From the audio signal, the SPL, indicated in dB(A), was estimated after a calibration with a sound level meter (Voltcraft, Conrad El., Hirschau, Germany) using the Sopran software (Svante Granqvist, Karolinska, Stockholm, Sweden). During the recordings, the background noise levels were measured at about 53 dB(A).

There was no time correction of the signals. The microphone signal has a natural delay with respect to the HSV and EEG signals. Given that the vocal tract would have a length of 17 cm added to a 3 cm microphone distance and the sound speed would equal 334 m/s, the signal delay would be 0.59 ms. Consequently, the error concerning the choice of time windows would equal 2.36% for the 25 ms time windows and 0.59% for the 100 ms windows. Because the zero-windows were constructed symmetrically around

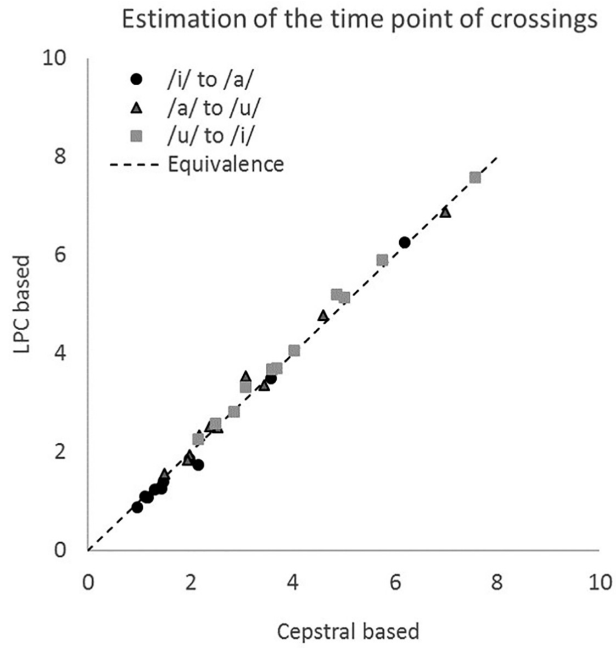


FIG. 3. The comparison of the nf_0 and F_n crossing points in the time domain by F_n calculation using the cepstral based method (Story and Bunton, 2016) and LPC.

the crossing point, the potential events should be part of the same window for all of the voice signals.

E. Perceptual analysis of the EGG signal

Besides this standardized analysis of the expected points of crossovers, the EGG signal was played through headphones as an audio signal and perceptually analyzed by three independent experts, concerning changes in the signal. The experts were able to watch the display of the signal during the evaluation. For indications of an occurrence by at least two of these experts, this point in the time domain was consecutively analyzed in more detail regarding audio and GAW signals as perceptive SFICs.

F. Statistical evaluation

Due to the small sample size, comparative statistics were considered problematic. However, statistical analysis was performed with SigmaStat (Jandel Corp., San Rafael, CA). For descriptive statistics, mean values with a standard deviation (SD) were used. As most data failed the normality testing, nonparametric tests have been used: The Friedman repeated measures analysis of variance on

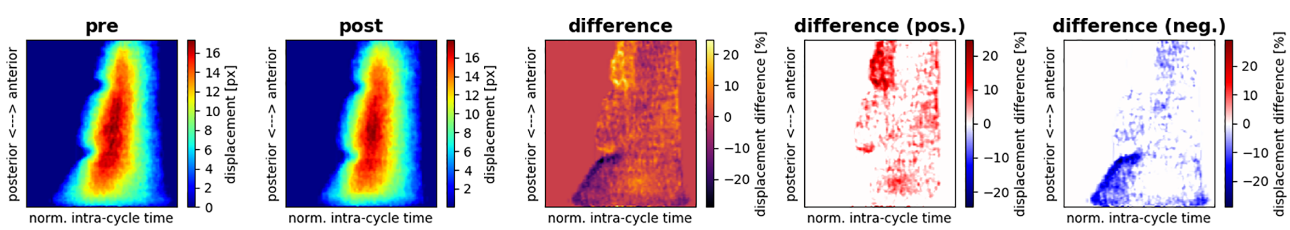


FIG. 2. (Color online) The exemplary illustration of the glottal cycle averaging based on the GVG data.

https://doi.org/10.1121/10.0005432

ranks was performed to compare the parameters at constant parts of different vowels (e.g., SPL at /i/ vs /a/ vs /u/ vs second /i/). The Wilcoxon-test was used to compare the parameters at the middle of the transition window and stable part of the preceding or subsequent vowel (e.g., SPL at /i/ vs $2f_o/F_1$).

G. Repeated measurement and validation

To check the accuracy of the formant estimations, the cepstrum based formant estimation was compared to a formant calculation by linear predicted coding (LPC; expectation of three formants below 3100 Hz) using the Praat software (University of Amsterdam, the Netherlands). The difference of the crossing points for all participants and transitions is shown in Fig. 3.

Furthermore, to validate the segmentation process, repeated measurements of OQ_{GAW} were performed by the same investigator for participant M1. The deviation for OQ_{GAW} for constant vowels was /i/, 0.31%; /a/, 1.73%; /u/, 1.56%; and for vowel transitions was /i/ to /a/, 1.73%; /a/ to /u/, 3.98%; and /u/ to /i/, 3.58.

III. RESULTS

A. Results concerning the expected crossing points

All except one participant (participant B2) were able to perform the experiment without any interruption and within

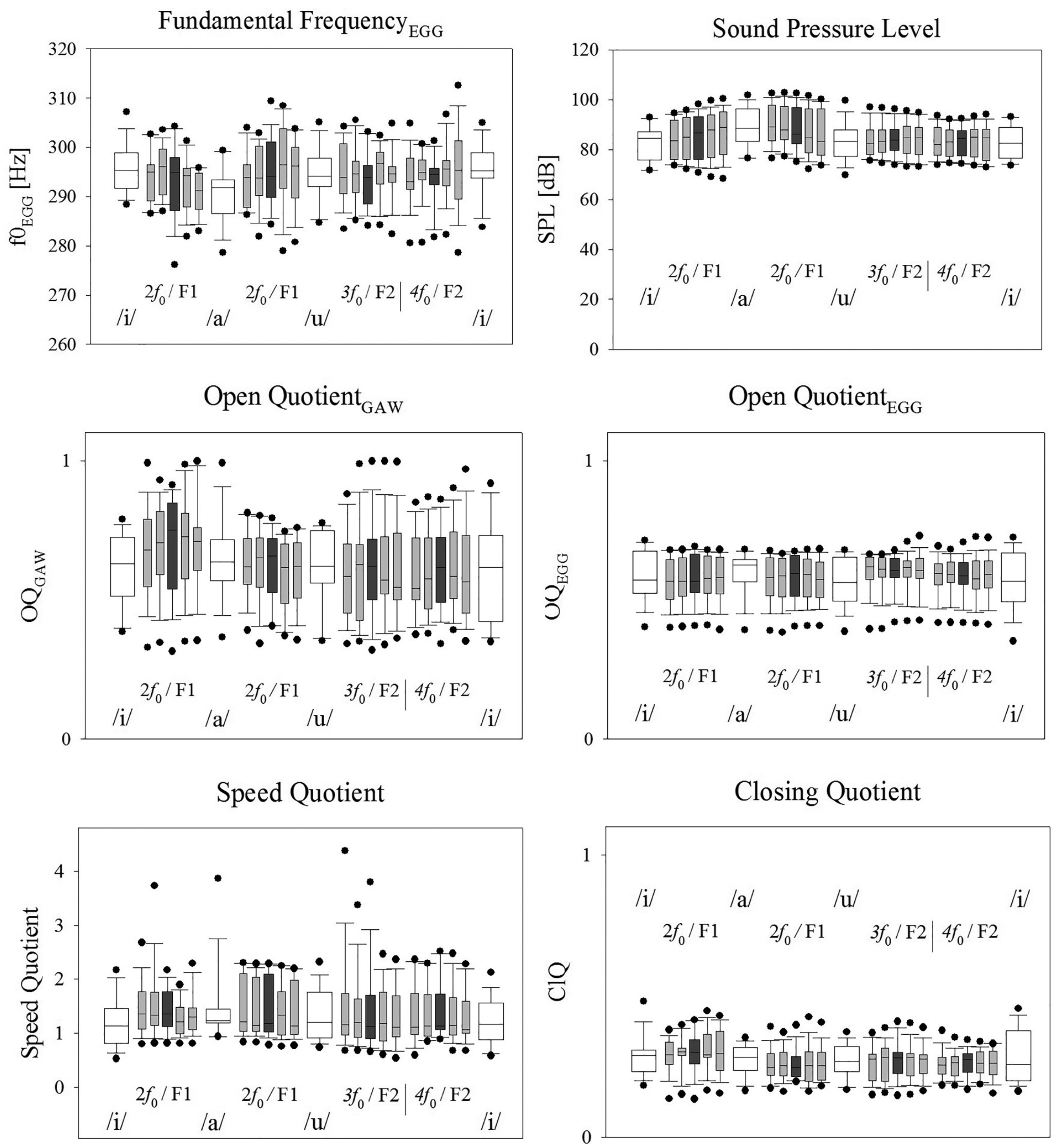


FIG. 4. Box-plots for the fundamental frequency, SPL, GAW derived open quotient (OQ_{GAW}), EGG derived open quotient (OQ_{EGG}), speed quotient, and closing quotient (ClQ). The white boxes refer to the stable vowel part, black boxes refer to the SFIC zero-window, and the gray boxes refer to the -2 to +2 25 ms transition windows relative to the SFIC zero-window.

https://doi.org/10.1121/10.0005432

TABLE III. Statistically significant comparisons between (a) the constant parts of vowels and expected neighboring SFICs, and (b) the different constant parts of vowels; see Fig. 4. All other comparisons failed to show any statistically significance.

Parameter	Time-points	p-value
(a)		
SPL	/i/ vs following SFIC	0.024
	/a/ vs preceding SFIC	0.007
	/u/ vs preceding SFIC	0.032
Fundamental frequency	/a/ vs following SFIC	0.014
Speed quotient	/i/ vs following SFIC	0.032
Open quotient	/i/ vs following SFIC	0.032
(b)		
SPL	/a/ vs all other vowels	<0.001
Fundamental frequency	/a/ vs all other vowels	<0.001
Speed quotient	First /i/ vs /a/	0.011
	First /i/ vs /u/	0.011

the required f_o range. As can be seen in Fig. 4, the constant parts of the vowel phonation showed median values of 296 Hz for the initial /i/, 294 Hz for /u/, and 295 Hz for the latest /i/. The /a/ vowel, however, was associated with an f_o drop of approximately 4–291 Hz and a greater SPL. During the transition windows, there was a rise in the SPL for the /i/ to /a/ transition and a lowering for the /a/ to /u/ transition. Table III shows the descriptive statistics for the statistically significant differences. All of the other comparisons failed to show statistical significance.

Regarding the constant vowel portions, the median of the vowel /a/ showed higher values for OQ_{EGG} but not OQ_{GAW} or ClQ_{GAW} as compared to the other vowel conditions. However, there were strong changes during the vowel transition /i/ to /a/ with a rise in both OQ_{GAW} and ClQ_{GAW}. Also, for the /u/ to /i/ transition ($4f_o$ and F_2 crossings), there was a tendency for a rise in OQ_{GAW} for the transition window. With respect to both the ClQ and the GVG differences (see Fig. 5), there was a slight decrease in the closing for the /a/ to /u/ transition. For the same transition, the opening was slightly increased; see Fig. 5.

As shown in the PVGs, there was no strong aperiodicity during these transition windows. Furthermore, in some participants, abnormalities, such as an anterior-posterior phase shift (participant S1), were present for all of the transition windows (see Fig. 6). There was a correlation of OQ_{EGG} and OQ_{GAW} (trendline, $y = 0.1753x + 0.4725$; $R^2 = 0.1188$); see Fig. 7. However, for OQ_{GAW} values lower than 0.7, the correlation was much stronger (trendline, $y = 0.613x + 0.2495$; $R^2 = 0.4728$).

As expected, the vowel transitions were associated with strong changes in the harmonics intensities. Figure 8 shows the spectrum, F_n, f_o, and the dEGG wavegram for participant T3 (graphs for all other participants are provided in the supplemental materials[1]). Some single participants exhibited adjustments with regard to the dEGG wavegram (see Fig. 8) and EGG derived sample entropy (see Figs. 8 and 9).

B. Perceptual evaluation of instabilities

For many vowel transitions, the perceptual analysis revealed instants at which the EGG signal exhibited distinct changes. It was found that for such instances, as estimated by at least two experts, nf_o/F_n intersections were detected. Figure 10 shows such crossing instances, which are in the neighborhood of the perceptual point (also recall Fig. 8).

For some of these instances shown in Fig. 10, f_o changes were observable. Interestingly, in participant S1,

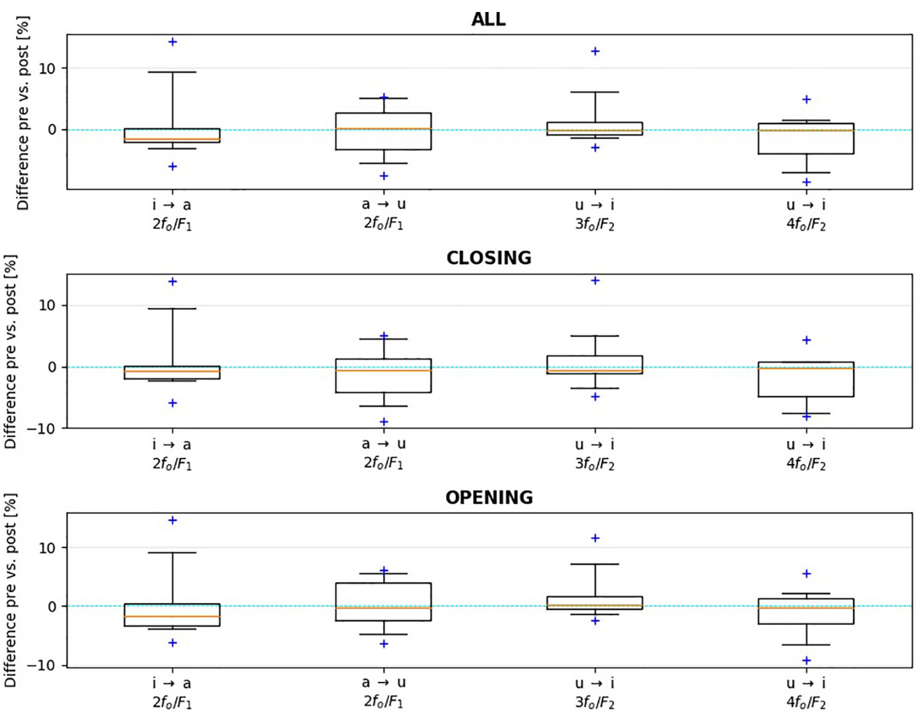


FIG. 5. (Color online) The differences between mean GVG data before (pre) and after (post) the SFICs averaged over all participants, grouped by vowel transition [/i/ to /a/, /a/ to /u/, and /u/ to /i/ (2×)]—recall Fig. 1. The middle and bottom panels describe only the changes in the closing and opening phases, respectively, as averaged over all analyzed glottal cycles.

during the crossing $4f_o/F_2$, there were not only f_o changes but a strong decrease in OQ_{EGG} at the instance at which the perceptual rating indicated a change. A corresponding *post hoc* perceptual rating of the audio signal, performed by the three experts, suggested that the observed sound changes were typical for a register shift (Echternach *et al.*, 2017c). Although all participants were asked to avoid vibrato, many participants performed the experiment with some extent of vibrato. Interestingly, for many participants, the vibrato stopped during the vowel transition as shown for participants M1, M2, T2, and B1.

IV. DISCUSSION

This investigation was focused on the analysis of gradual vowel quality changes on a given pitch in professional

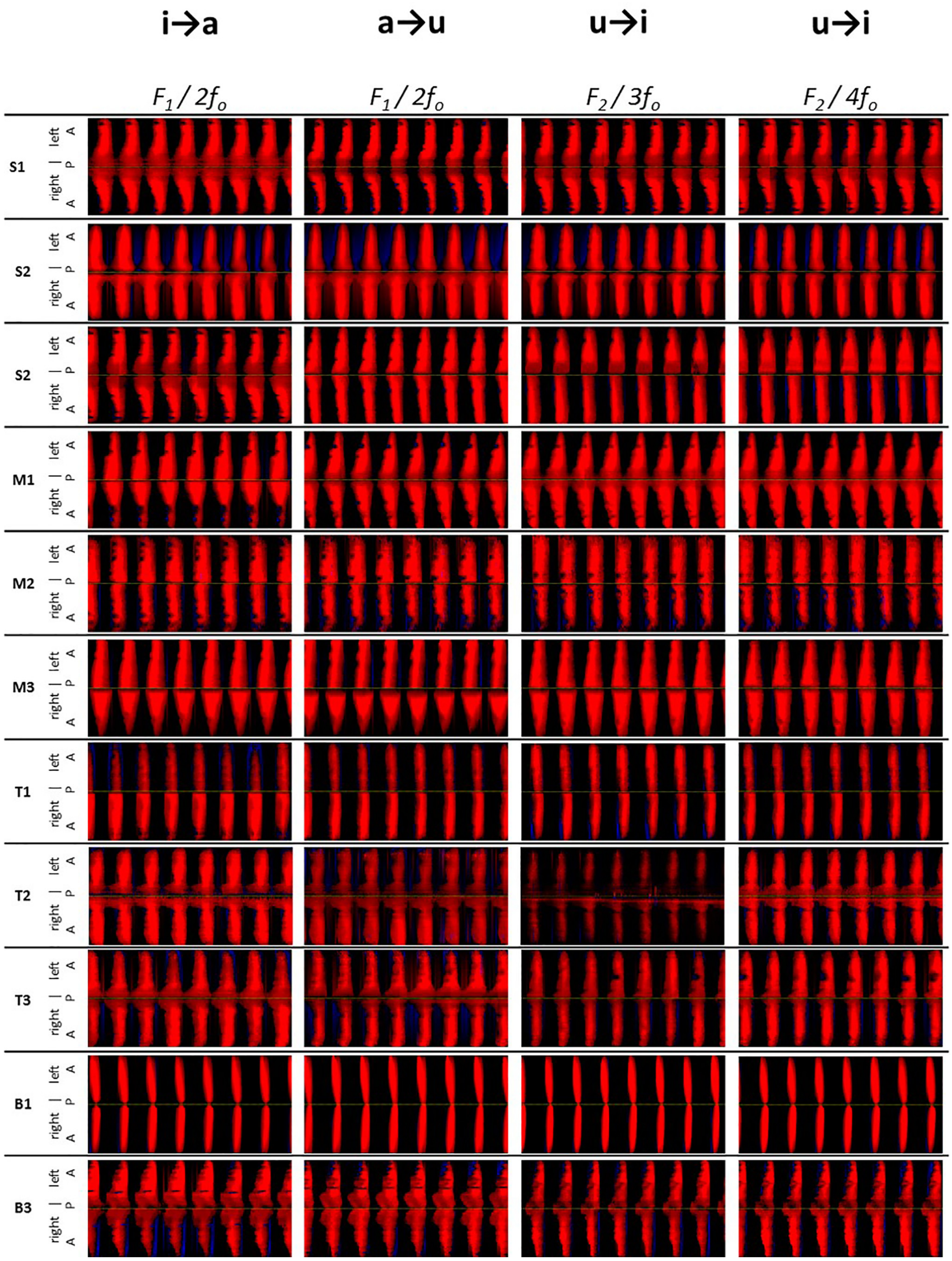


FIG. 6. (Color online) The PVGs for all participants and all SFIC zero-windows.

singers. It was found that in single cases, evidence of an interaction of the vocal tract and vocal fold oscillations (according to Titze, 2008, level-two) during a crossing of nf_o and F_n was detectable. For most of the vowel transitions, however, the acoustical (audio) and vocal fold oscillatory signals (EGG and GAW) changed rather smoothly, lacking clear evidence for a sudden vowel quality transition related interaction.

Vowel quality is an important aspect of voice production and is dependent on the vocal tract shape and its resonances (Fant, 1960; Sundberg, 1987; Titze, 1994; Stevens, 1998). Some previous investigations showed that f_{Rn} or F_n could interact with the voice source (Rothenberg, 1981; Titze, 2008; Titze *et al.*, 2008; Sundberg *et al.*, 2016). Especially during the last years, it has been stated by Titze and co-workers (Titze, 2008; Titze *et al.*, 2008; Titze and Worley, 2009) that such an interaction of vocal tract resonances influencing the air pulse (level-one) or vocal fold oscillations (level-two) could be nonlinear.

Such instabilities were not expected during the stable vowel phonation parts because many adjustments could be performed to avoid voice breaks. In the presented data, the stable vowel phonation exhibited no great differences concerning OQ_{GAW}, OQ_{EGG}, and ClQ_{GAW} among the vowel qualities.There was a significant increase in OQ_{GAW} only for the /a/ vowel. In contrast to the presented data, Sundberg *et al.* (2016) found that vowels differed with regard to the closed quotient and the maximum flow declination rate. Because these authors analyzed the inverse filtered flow pulse, this disagreement might, in part, be explained by a variation introduced through the different methods used. Furthermore, inverse filtering assumes a linear relation of the voice source and vocal tract resonances, which appears an assumption that might be detrimental to detecting nonlinear interactions.

The vowel transitions offered adjustments on the voice source level. In this respect, there was a higher f_o before the transition compared to the windows after the transition for /i/ to /a/, but the inverse relation occurred for the /a/ to /u/ transition. This is in agreement with Titze (2008), who predicted that for a harmonic component whose frequency is in the compliant portion of the impedance curve (where f_{Rn} is slightly lower than nf_o), the effective stiffness of the vocal folds may be increased, leading to a slight rise of $f_{o.}$ Furthermore, there was an increase in OQ_{GAW} for the SFIC zero-window during the /i/ to /a/ transition and a tendency for the /u/ to /i/ transition for the $4f_o$ and F_2 crossover. This could be in agreement with the hypothesis that irregularity could have caused the increased OQ_{GAW}, and stronger level-two interactions could be expected at the peak of the nf_o and F_n crossover. However, the OQ_{GAW} values for the -2 and $+2$ windows, as well as the -1 and $+1$ windows, appeared almost symmetrical. This observation is not in agreement with the notion of the inertive vs compliant reactance below/above any vocal tract resonance, which would rather postulate an asymmetric response.

With respect to the standardized time windows, the PVGs showed no clear instabilities, and there were only rare cases with perceptual changes of the EGG signal. It seems important to recapitulate that in the presented study, professional singers, who are trained to stabilize phonation for all sorts of challenging phonatory tasks, were examined. Therefore, it was, on the one hand, not surprising that the number of instabilities was low. On the other hand, it is noteworthy that even these trained singers showed such cases that could be interpreted as interactions. The absence of interactions in the majority of cases remains unclarified. In this respect, it could be possible that (1) there was no strong interaction or (2) an interaction was present, which was counteracted by stabilizing adjustments in the voice production system. Concerning the latter hypothesis, possible stabilizing factors have not yet been clarified in detail. However, for vocal fold oscillatory instabilities during the tenor's passaggio, it has been shown that an increase in nasality could contribute to stabilization (Echternach *et al.*, 2021). If the absence of interactions would be caused by stabilizing factors used by the trained singers, it may be speculated that the quantity of instabilities would be greater for untrained voices. In this context, it should be noted also that in the experimental data by Titze *et al.* (2008), not all of the predicted instabilities were measurable. For male voices, voice signals were found to be unstable in only 54% when a crossover of f_o and f_{Rn} was produced by a vowel transition. In addition, 23% of the instabilities occurred when no instabilities were predicted by means of a f_o/f_{R1} crossover. Female voices showed much lower values of instabilities, which were found in 34% of cases where instabilities were predicted by f_o/f_{R1} crossings. Furthermore, Wade *et al.* (2017) also found a greater disagreement of predicted and measured instabilities performing vowel glides on the pitches of C5 and C6. It should be mentioned here that in both studies, the prediction of a crossing of f_o and f_{Rn} was based on the assumption that the vocal tract with associated resonatory

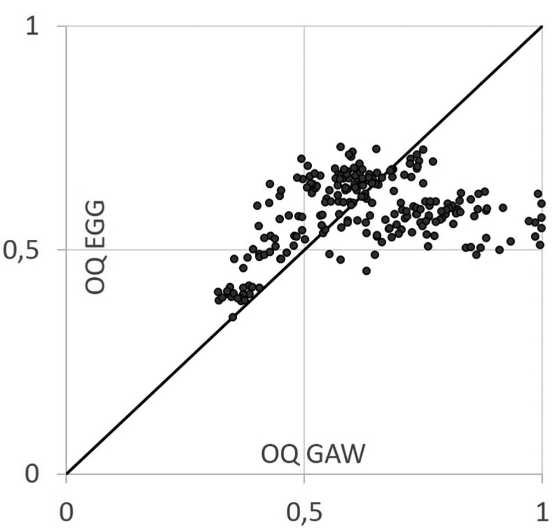


FIG. 7. The glottal area derived OQ_{GAW} vs OQ_{EGG}.

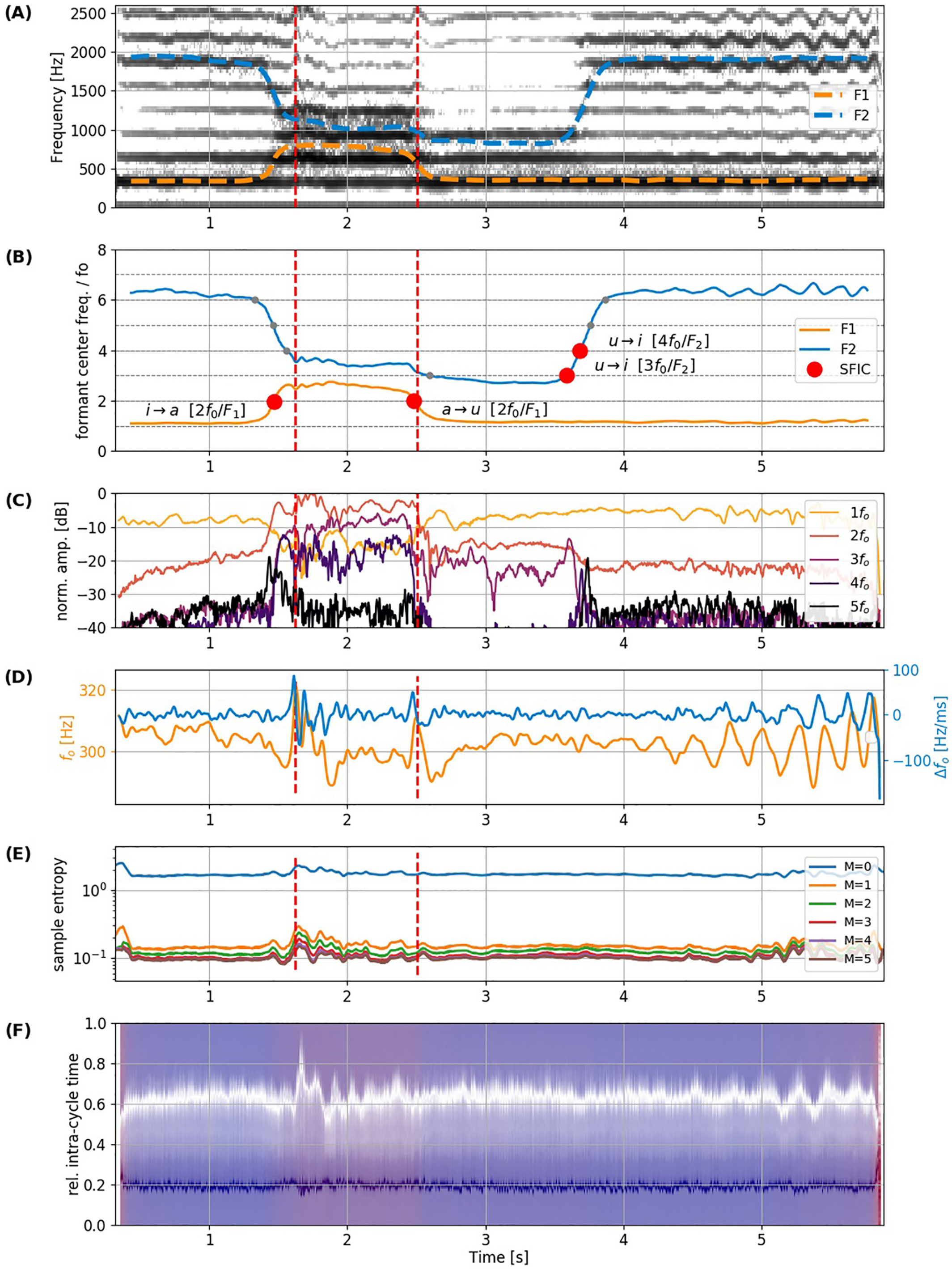


FIG. 8. (Color online) Documentation of the experimental phonation by participant T3. The vertical dashed lines in all of the panels indicate the temporal offsets of the perceptually determined SFICs by the percetual evaluation of the played electroglottographic signal as an audio signal. (A) A narrowband spectrogram of the acoustic data with the estimated formant frequencies (F_1 and F_2) superimposed. (B) The estimated formant frequencies relative to the time-varying f_o and their respective crossing of voice source harmonics. The four red dots indicate the four algorithmically determined SFICs [/i/ to /a/, /a/ to /u/, and /u/ to /i/ (2×)]—recall Fig. 1. (C) The relative amplitudes of the lowest five harmonics of the radiated acoustic signal. Note the abrupt amplitude changes of harmonics $2f_o$, $3f_o$, and $4f_o$ at the respective algorithmically determined SFIC. (D) f_o (orange) and its rate of change (blue) as a function of time. Note the abrupt f_o changes at the instance of the perceptually determined SFIC (vertical dashed lines). (E) The EGG derived sample entropy computed with different epoch lengths M (Lake *et al.*, 2002). (F) The dEGG wavegram corresponding to the phonation.

properties would remain stable between the vocal fry (where the formants as expression of vocal tract resonances were measured) and experiment (where the f_o measurement was performed). Estimations of vocal tract resonances examining the responses of the vocal tract to an external excitation by means of a vibrator and spectra (Sundberg, 1975), broadband acoustic excitations (Henrich *et al.*, 2011), or real-time magnetic imaging (Bresch and

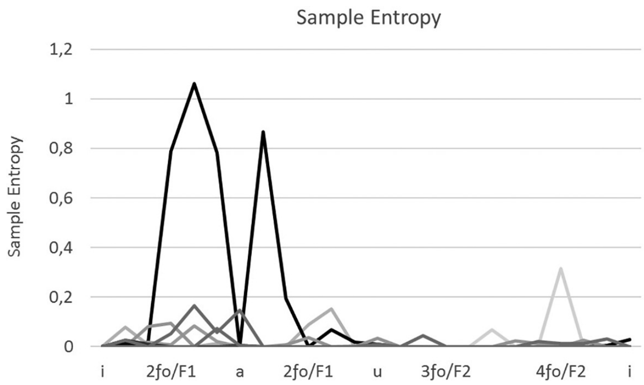


FIG. 9. The EGG derived sample entropy for all of the participants. The high peaks for transitions /i/ to /a/ and /a/ to /u/ refer to subject S1.

Narayanan 2010; Echternach *et al.*, 2016) have shown that the vocal tract with the associated resonances strongly changes with a rising f_o. Because female voices, especially, typically avoid a crossing of f_o and f_{R1} (Sundberg, 1987), the f_o for C5 (around 520 Hz) is already much greater than the f_{R1} for /u/ and /i/ (around 350 Hz; Sundberg, 1987), and it could be expected that f_{Rn} might be different from the estimations of f_{R1} from the vocal fry.

Although the medians showed no great general effects of interactions during the vowel transitions, there were some quite clear effects with regard to the perceptual rating of the EGG signal. In particular, a decrease in f_o was observed in cases in which F_n were rising for a crossover. Such effects were also predicted by Titze (2008). Furthermore, at these points in the time domain, there were abnormalities in some PVGs, i.e., left-right and/or anterior-posterior phase shifts (for example, participant B3). However, not all perceptual events in the EGG signal were accompanied by changes in the PVG. It should be noted here that EGG and GAW are not equal: As GAW is derived from a two-dimensional projection of the glottal opening, EGG represents the relative vocal fold contact. Therefore, part of the disagreement could be related to this methodical difference. Further, the identification of perceptive SFICs relied on the estimation of three experts, which is, on the one hand, subjective. On the other hand, however, many perceptive SFICs were accompanied by increases of the EGG derived sample entropy (recall subjects S1, M3, T2, T3, and T3 in Fig. 8 and supplementary Fig. 8[1]). For participant S1, a perceptual event was also detected, which was accompanied by such phase shifts in the PVG. Because of a perceptual rating of the audio signal, this event was concordant to a registration event. Such registration events between modal and middle registers frequently occur in this f_o region (Echternach *et al.*, 2017c). However, it is noticeable that this event occurred during a $4f_o/F_2$ intersection. Furthermore, it was a very sudden event, which makes it likely that a bifurcation (Svec *et al.*, 1999) might have caused this registration event.

Interestingly, some clear assignments of perceptual points to crossings were at higher partials than expected. Maxfield *et al.* (2017) observed stronger interactions in the region of crossings of lower harmonics and lower vocal tract resonances. However, this does not necessarily mean that no interaction could be present at higher nf_o/F_n crossings. Also, there were some perceptual events which could not be assigned to any nf_o/F_n crossover. Furthermore, the presented analysis is based on an estimation of the formant structure using a cepstral analysis approach (Story and Bunton, 2016). However, other approaches, such as inverse filtering (Sundberg *et al.*, 1993; Echternach *et al.*, 2011; Dong *et al.*, 2013; La and Sundberg 2015), comparison to the formant measurement of the vocal fry (Titze *et al.*, 2008; Wade *et al.*, 2017), broadband acoustic excitations (Joliveau *et al.*, 2004; Henrich *et al.*, 2008; Garnier *et al.*, 2010; Henrich Bernardoni *et al.*, 2014), or data derived from imaging, such as magnetic resonance imaging (Story *et al.*, 1998; Bresch and Narayanan 2010; Echternach *et al.*, 2011), also have been used for measurements of the vocal tract resonances and formants, which could lead to different results. In the presented study, the formant frequencies were also measured using the LPC technique provided in Praat. In some of the participants, there was a substantial difference concerning the time point at which the crossing could be calculated. Because the ground truth of the real f_{Rn} and, consequently, F_n is unknown, it cannot be excluded that the crossing of F_n and nf_o is at a slightly different point in the time domain.

Also, as specified earlier, only professional classical singers were included in the presented study as, on the one hand, they are likely to produce a rather stable phonation. On the other hand, however, the participants phonate frequently with vibrato. Although, the participants were asked to perform the task without any vibrato, not all participants were able to fulfill this requirement with the consequence of an instable f_o. Concerning the vibrato, there was also another noticeable effect: In many participants performing the task, the vibrato stopped during the vowel transition. The reason remains unclear, but it is speculated that maintaining vibrato while passing the f_o or nf_o through a vocal tract resonance may contribute to generating vocal instability. That is, in the vicinity of a resonance frequency, vibrato could rapidly move the vocal fold vibration back and forth from the inertive to compliant conditions. Ceasing the vibrato briefly to move through the resonance may be an acquired technique, which was learned by the singers to avoid instability of the vocal fold vibration. If this were the case, it would support the concept of source-tract interactions. Although noteworthy and interesting, systematically addressing these issues is beyond the scope of this study. However, these notions should be tested in a future investigation.

There are a number of further potential limitations associated with the presented study.

As noted previously, the first limitation is that this study only included professional singers without any control group. Beside other limitations associated with this circumstance, professional singers sometimes use special vowel conditions for their singing voice such as a covered voice

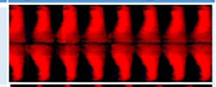
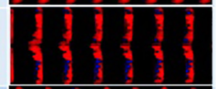
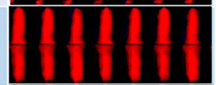
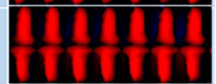
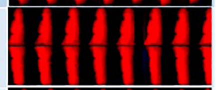
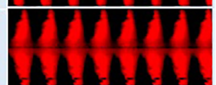
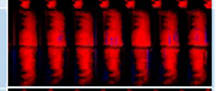
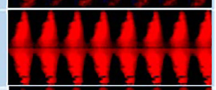
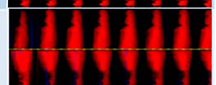
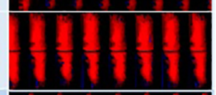
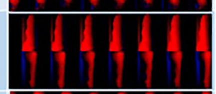
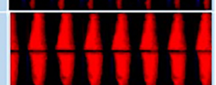
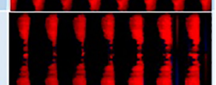
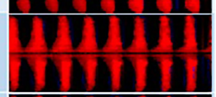
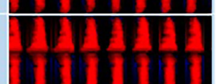
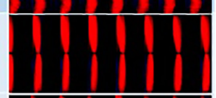
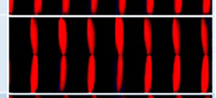
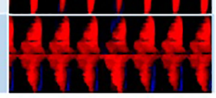
Subj.	sec.	Vowel	nf_o / F_n Crossing	Observation	phonovibrogram
S1	6.34	i→a	$5f_o$ / F_2	lowering of f_o	
	6.89	a→u	$4f_o$ / F_2	f_o strong instability: perceptual register shift	
S2	0.99	i→a	$2f_o$ / F_1 $5f_o$ and $6f_o$ / F_2	lowering of f_o	
	1.65	a→u	No clear crossing	lowering of f_o	
S3	2.1	a→u	$2f_o$ / F_1 $3f_o$ / F_2	rise of f_o	
	2.45	a	No clear crossing		
M1	1.1	i→a	$2f_o$ / F_1 $6f_o$ / F_2	lowering of f_o , Vibrato ↓	
	3.13	u→i	$7f_o$ / F_2	$2f_o$ intensity drop	
M2	3.8	i→a	$4f_o$ / F_2	Before: Non-vibrato	
	6.0	u→i	$5f_o$ and $4f_o$ / F_2	rise of f_o	
M3	1.4	a	No clear crossing	$4f_o$ intensity rise	
	2.6	u→i	$4f_o$ / F_2	lowering of f_o	
T2	2.22	i→a	$3f_o$ / F_2	For /a/ continuous $3f_o/F_2$ crossings, vibrato stops for vowel transition	
T3	1.6	a	No clear crossing	$4f_o$ intensity drop	
	2.5	a→u	$2f_o$ / F_1	general lowering of f_o but rise for the peak	
B1	1.29	i→a	$2f_o$ / F_1	lowering of f_o	
	2.65	a→u	$3f_o$ / F_2	rise of f_o	
B3	1.37	i→a	$2f_o$ / F_1 $4f_o$ / F_2	lowering of f_o , Vibrato stops	

FIG. 10. (Color online) Participants having the point in the time domain where at least two experts noticed a change of the EGG signal (perceptual SFIC), nf_o/F_n crossing in the neighborhood, the observations of the associated effects as also provided in Fig. 4, and the corresponding PVGs.

phenomenon (Pielke, 1912; Sonninen, 1962; Large, 1972). Therefore, it could be expected that the intended vowel conditions might differ from those of the untrained voices.

Although the stable vowel phonation was set as 1 s, the speed of the transition was not standardized. Maybe the speed of the transition could play a major role here. The participants were asked to perform the transition as smoothly as possible. As a consequence, the transition was rather long. Thus, it is possible that feedback corrections might have stabilized the vocal fold oscillations by a greater amount. It could be hypothesized that rather sudden vowel transitions with greater speed might show stronger irregularities.

Another limitation is that the number of participants was—due to the special collective of participants and extensive HSV data produced during the experiment—rather low, preventing a more detailed statistical analysis.

The participants were all asked to phonate at a given musical pitch of D4. This pitch is quite low for high soprano voices but rather high for basses. The pitch was chosen to avoid modulation of the vocal tract shape associated with an f_o/f_{R1} tuning. The rather high pitch for male voices and rather low pitch for female voices might also have influenced the vowel performance. However, this chosen approach facilitated the analysis of distinct and targeted crossings of voice source harmonics and vocal tract resonances. The question about the existence of any gender differences should be examined in future investigations analyzing the subjects in their relative pitch range. Further, the sustained pitch D4 is, for most of the singers, not precisely in the frequency range in which registration events typically occur, i.e., the passaggio. It has been shown before that bifurcations are more likely to occur in the passaggio regions (Svec *et al.*, 1999; Svec *et al.*, 2008; Echternach *et al.*, 2017a; Echternach *et al.*, 2017c), which could be potentially caused or modified by interactions (Titze, 2014).

The application transnasal laryngoscopy could have influenced the participants' phonation. In particular, the endoscope could have opened the velopharyngeal port, which could have an effect on the resonatory properties of the vocal tract. In this respect, it has been shown that a coupling of the nasal cavities could raise f_{R1} (Havel *et al.*, 2021). In such a case, it may be assumed that a crossover would occur at a slightly lower f_{R1} value. Further, nasalazition could contribute to stabilization of the vocal fold oscillation patterns as has been shown with respect to tenors' passaggio (Echternach *et al.*, 2021). Unfortunately, it is not possible to check if there was leakage because the endoscope is in the laryngeal view position. However, it could be expected that influences of a transoral approach would be greater than the chosen transnasal approach.

Finally, quantitative errors could have been introduced during the formant frequency computation. Given that the frequency spacing of harmonics linearly increases with a rising f_o, the vocal tract transfer function is "acoustically sampled" at fewer points, thus, increasing the possibility that a harmonic is interpreted as a formant at a higher f_o. Given that the actual f_o of the phonatory tasks was at about 300 Hz, a formant frequency estimation error of a certain magnitude has to be considered. In fact, the cepstral algorithm that was used was tested at comparable f_o (Story and Bunton, 2016), suggesting a maximum error of 10%. It is, however, conceivable that the actual error is lower given that a comparison with a state-of-the-art LPC approach suggested little discrepancy in most of the data (recall Fig. 3).

V. CONCLUSIONS

The data presented here partially corroborate the notion that interactions between the vocal tract and voice source may result in level-two interactions. However, the lack of systematic occurrence of these level-two interactions across all participants' phonations suggests that either the interactions are not strong in every case or the influence of confounding factors, such as anatomical differences or fine-adjustments within the voice production musculature, may play a crucial role for the emergence of interaction effects.

ACKNOWLEDGMENTS

M.E. (Grant No. Ec409/1–2) and M.D. (Grant No. DO 1247/8-1/2) are supported by the Deutsche Forschungsgemeinschaft (DFG). M.E. and C.T.H. contributed equally to this work.

[1]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0005432 for all other subject files of Fig. 8 (SuppPub1, subject S1; SuppPub2, subject S2; SuppPub3, subject S3; SuppPub4, subject M1; SuppPub5, subject M2; SuppPub6, subject M3; SuppPub7, subject T1; SuppPub8, subject T2; SuppPub9, subject B1; and SuppPub10, subject B3).

Boersma, P., and Weenink, D. (**2021**). "Praat: doing phonetics by computer (version 6.1.49) [Computer program]," http://www.praat.org (Last viewed June 12, 2021).

Bouhuys, A., Mead, J., Proctor, D., and Stevens, K. (**1968**). "Pressure flow events during singing," Ann. Am. Acad. Sci. **155**, 165–176.

Bresch, E., and Narayanan, S. (**2010**). "Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing," J. Acoust. Soc. Am. **128**(5), EL335–EL341.

Bunch, M., and Chapman, J. (**2000**). "Taxonomy of singers used as subjects in scientific research," J. Voice **14**(3), 363–369.

Dong, L., Sundberg, J., and Kong, J. (**2013**). "Formant and voice source properties in two male Kunqu Opera roles: A pilot study," Folia Phoniatr. Logop. **65**(6), 294–302.

Echternach, M., Birkholz, P., Traser, L., Flügge, T. V., Kamberger, R., Burk, F., Burdumy, M., and Richter, B. (**2015**). "Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies," J. Acoust. Soc. Am. **137**, 2586–2596.

Echternach, M., Birkholz, P., Sundberg, J., Traser, L., Kamberger, R., Korvink, J., and Richter, B. (**2016**). "Resonatory properties in professional tenors singing above the passaggio," Acta Acust. Acust. **137**, 198–206.

Echternach, M., Burk, F., Köberlein, M., Burdumy, M., Döllinger, M., and Richter, B. (**2017a**). "The influence of vowels on vocal fold dynamics in the tenor's passaggio," J. Voice **31**(4), 424–429.

Echternach, M., Burk, F., Köberlein, M., Herbst, C. T., Döllinger, M., Burdumy, M., and Richter, B. (**2017b**). "Oscillatory characteristics of the vocal folds across the tenor passaggio," J. Voice **31**, 381.e5–381.e14.

Echternach, M., Burk, F., Köberlein, M., Selamtzis, A., Döllinger, M., Burdumy, M., Richter, B., and Herbst, C. T. (**2017c**). "Laryngeal evidence for the first and second passaggio in professionally trained sopranos," PLoS One **12**(5), e0175865.

Echternach, M., Burk, F., Rose, F., Herbst, C. T., Burdumy, M., Döllinger, M., and Richter, B. (**2018**). "Impact of functional mass lesions in professional female singers: Biomechanics of vocal fold oscillation in the register transition regions," HNO **66**(4), 308–320.

Echternach, M., Högerle, C., Köberlein, M., Schlegel, P., Döllinger, M., Richter, B., and Kainz, M. A. (**2021**). "The effect of nasalance on vocal fold oscillation patterns during the male passaggio," J. Voice **35**, 500.e9–500.e16.

Echternach, M., Sundberg, J., Baumann, T., Markl, M., and Richter, B. (**2011**). "Vocal tract area functions and formant frequencies in opera tenors' modal and falsetto registers," J. Acoust. Soc. Am. **129**(6), 3955–3963.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Flanagan, J. (**1968**). "Source-system interaction in the vocal tract," Ann. N. Y. Acad. Sci. **155**(1), 9–17.
Garnier, M., Henrich, N., Smith, J., and Wolfe, J. (**2010**). "Vocal tract adjustments in the high soprano range," J. Acoust. Soc. Am. **127**(6), 3771–3780.
Havel, M., Sundberg, J., Traser, L., Burdumy, M., and Echternach, M. (**2021**). "Effects of nasalization on vocal tract response curve," J. Voice (published online).
Henrich, N., Smith, J., and Wolfe, J. (**2008**). "Resonance strategies and glottal behaviour in the two main laryngeal mechanisms for professional operatic singers," J. Acoust. Soc. Am. **123**(5), 3243.
Henrich, N., Smith, J., and Wolfe, J. (**2011**). "Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones," J. Acoust. Soc. Am. **129**(2), 1024–1035.
Henrich Bernardoni, N., Smith, J., and Wolfe, J. (**2014**). "Vocal tract resonances in singing: Variation with laryngeal mechanism for male operatic singers in chest and falsetto registers," J. Acoust. Soc. Am. **135**(1), 491–501.
Herbst, C. T. (**2017**). "A review of singing voice subsystem interactions—Toward an extended physiological model of 'support,'" J. Voice **31**(2), 249.E13–249.E19.
Herbst, C. T., Fitch, W. T., and Svec, J. G. (**2010**). "Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively," J. Acoust. Soc. Am. **128**(5), 3070–3078.
Howard, D. M. (**1995**). "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," J. Voice **9**(2), 163–172.
Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**(6), 1233–1268.
Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Acoustics: Tuning of vocal tract resonance by sopranos," Nature **427**(6970), 116.
Karakozoglou, S., Henrich, N., d'Alessandro, C., and Stylianou, Y. (**2012**). "Automatic glottal segmentation using local-based active contours and application to glottovibrography," Speech Commun. **54**, 641–654.
La, F. M., and Sundberg, J. (**2015**). "Contact quotient versus closed quotient: A comparative study on professional male singers," J. Voice **29**(2), 148–154.
Lake, D. E., Richman, J. S., Griffin, M. P., and Moorman, J. R. (**2002**). "Sample entropy analysis of neonatal heart rate variability," Am. J. Physiol. **283**(3), R789–R797.
Large, J. (**1972**). "Towards an integrated physiologic-acoustic theory of vocal registers," NATS Bull. **28**, 18–36.
Lohscheller, J., and Eysholdt, U. (**2008**). "Phonovibrogram visualization of entire vocal fold dynamics," Laryngoscope **118**(4), 753–758.
Lohscheller, J., Eysholdt, U., Toy, H., and Döllinger, M. (**2008**). "Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics," IEEE Trans. Med. Imag. **27**(3), 300–309.
Maryn, Y., Verguts, M., Demarsin, H., van Dinther, J., Gomez, P., Schlegel, P., and Döllinger, M. (**2020**). "Intersegmenter variability in high-speed laryngoscopy-based glottal area waveform measures," Laryngoscope **130**(1), E654–E661.
Maxfield, L., Palaparthi, A., and Titze, I. (**2017**). "New evidence that nonlinear source-filter coupling affects harmonic intensity and *fo* stability during instances of harmonics crossing formants," J. Voice **31**, 149–156.
Nawka, T., Wiesmann, U., and Gönnermann, U. (**2003**). "Validierung des Voice Handicap Index (VHI) in der deutschen Fassung" ("Validation of the German version of the Voice Handicap Index"), HNO **51**(11), 921–930.
Pielke, W. (**1912**). "Über 'offen' und 'gedeckt' gesungene Vokale" ("About 'open' and 'closed' sung vowels"), Passow-Schaefers Beitr. **5**, 215–231.
Rothenberg, M. (**1981**). "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo), pp. 305–328.
Schlegel, P., Kniesburges, S., Dürr, S., Schützenberger, A., and Döllinger, M. (**2020**). "Machine learning based identification of relevant parameters for functional voice disorders derived from endoscopic high-speed recordings," Sci. Rep. **10**(1), 10517.
Selamtzis, A., and Ternstrom, S. (**2014**). "Analysis of vibratory states in phonation using spectral features of the electroglottographic signal," J. Acoust. Soc. Am. **136**(5), 2773–2783.
Selamtzis, A., Ternstrom, S., Richter, B., Burk, F., Köberlein, M., and Echternach, M. (**2018**). "A comparison of electroglottographic and glottal area waveforms for phonation type differentiation in male professional singers," J. Acoust. Soc. Am. **144**(6), 3275–3288.
Sonninen, A. (**1962**). "Paratasis-gram of the vocal cords and the dimensions of the voice," in *Proceedings of the* Fourth International Congress *of* Phonetic Science (Mouton, The Hague).
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Story, B. H., and Bunton, K. (**2016**). "Formant measurement in children's speech based on spectral filtering," Speech Commun. **76**, 93–111.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). "Vocal tract area functions for an adult female speaker based on volumetric imaging," J. Acoust. Soc. Am. **104**(1), 471–487.
Sundberg, J. (**1975**). "Formant technique in a professional female singer," Acustica **32**, 89–96.
Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb, Illinois).
Sundberg, J., Gramming, P., and Lovetri, J. (**1993**). "Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing," J. Voice **7**(4), 301–310.
Sundberg, J., La, F. M., and Gill, B. P. (**2013**). "Formant tuning strategies in professional male opera singers," J. Voice **27**(3), 278–288.
Sundberg, J., La, F. M., and Gill, B. P. (**2016**). "Voice source variation between vowels in male opera singers," J. Voice **30**(5), 509–517.
Svec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," J. Acoust. Soc. Am. **106**(3 Pt 1), 1523–1531.
Svec, J. G., Sundberg, J., and Hertegard, S. (**2008**). "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," J. Acoust. Soc. Am. **123**(1), 347–353.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).
Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**(5), 2733–2749.
Titze, I. R. (**2014**). "Bi-stable vocal fold adduction: A mechanism of modal-falsetto register shifts and mixed registration," J. Acoust. Soc. Am. **135**(4), 2091–2101.
Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., Howard, D. M., Hunter, E. J., Kaelin, D., Kent, R. D., Kreiman, J., Kob, M., Löfqvist, A., McCoy, S., Miller, D. G., Noé, H., Scherer, R. C., Smith, J. R., Story, B. H., Švec, J. G., Ternström, S., and Wolfe, J. (**2015**). "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization," J. Acoust. Soc. Am. **137**(5), 3005–3007.
Titze, I. R., Riede, T., and Popolo, P. (**2008**). "Nonlinear source-filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**(4), 1902–1915.
Titze, I. R., and Worley, A. S. (**2009**). "Modeling source-filter interaction in belting and high-pitched operatic male singing," J. Acoust. Soc. Am. **126**(3), 1530–1540.
Tokuda, I. T., Zemke, M., Kob, M., and Herzel, H. (**2010**). "Biomechanical modeling of register transitions and the role of vocal tract resonators," J. Acoust. Soc. Am. **127**(3), 1528–1536.
Wade, L., Hanna, N., Smith, J., and Wolfe, J. (**2017**). "The role of vocal tract and subglottal resonances in producing vocal instabilities," J. Acoust. Soc. Am. **141**(3), 1546–1559.
Zanartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R., and Hillman, R. E. (**2011**). "Observation and analysis of *in vivo* vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study," J. Acoust. Soc. Am. **129**(1), 326–339.