DECEMBER 19 2025

Investigation of singing with and without vibrato, using high-speed videolaryngoscopy and electroglottography

Isabella Gantner; Sophia Gantner; Marie Köberlein; Jonas Kirsch; Christian Kleiner; Peter Birkholz; Michael Döllinger; Matthias Echternach

J. Acoust. Soc. Am. 158, 4936–4946 (2025)
https://doi.org/10.1121/10.0041893

JASA ARTICLE

Investigation of singing with and without vibrato, using high-speed videolaryngoscopy and electroglottography

Isabella Gantner,[1] Sophia Gantner,[1] Marie Köberlein,[1,2] Jonas Kirsch,[1] Christian Kleiner,[3] Peter Birkholz,[3] Michael Döllinger,[4] and Matthias Echternach[1,a)]

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

[2]*Antonio Salieri Department of Vocal Studies and Vocal Research in Music Education, mdw University of Music and Performing Arts Vienna, Anton-von-Webern-Platz, 1030 Vienna, Austria*

[3]*Institute of Acoustics and Speech Communication, Technische Universität Dresden, Helmholtzstraße 18, 01069 Dresden, Germany*

[4]*Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head & Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Waldstraße 1, 91054 Erlangen, Germany*

ABSTRACT:
This study aimed to characterize the biomechanics of vibrato and non-vibrato phonation in professional singers, focusing on vocal fold oscillation patterns and cuneiform cartilage movements. Ten vocally healthy, professionally-trained, western classical singers were instructed to sustain the vowel [i:] without vibrato, with vibrato, and again without vibrato, for approximately two seconds for each quality. Data were collected via acoustic recordings, electroglottography (EGG), and transnasal high-speed videoendoscopy. After segmentation, glottal area waveforms (GAW) were analyzed, and temporal trajectories of fundamental frequency (f_o) and glottal parameters were compared. Cuneiform cartilage movement was tracked in three spatial dimensions using the videoendoscopic image sequences. The participants maintained a relatively stable mean f_o across conditions, with no significant differences in open quotient (OQ_{EGG} and OQ_{GAW}) and closing quotient (ClQ) regarding vibrato and non-vibrato. During vibrato, most participants showed a higher OQ in vibrato f_o minima than in maxima. Vibrato was associated with lower cepstral peak prominence and higher sound pressure level (SPL), and SPL courses inverse to f_o modulation. The cuneiform tubercle tracking correlated to f_o showed vertical and horizontal movements which followed, in one subject, an epiglottis forward positioning, presumably due to a tilting by cricothyroid muscle activity linked to the vibrato phenomenon.
 https://doi.org/10.1121/10.0041893

(Received 18 July 2025; revised 6 November 2025; accepted 3 December 2025; published online 19 December 2025)

[Editor: James F. Lynch] Pages: 4936–4946

I. INTRODUCTION

Human voice production is not only characterized by its fundamental frequency (f_o), sound pressure level (SPL), and spectral structure but also by low frequency-modulations of f_o, known as vibrato. Vocal vibrato is frequently regarded as an important stylistic element in many singing styles. Furthermore, the sound quality of an cultivated vibrato has been associated with characteristics of a healthy, flexible voice and is described in terms of richness and vibrancy.[1,2]

In general, vibrato is a physiological, aerodynamic, and acoustic phenomenon.[3] The term vibrato appears in several music traditions, styles, and cultures but differs in acoustical and physiological characteristics. In western classical singing, vibrato is defined by an undulation of f_o.[4] The vibrato corresponds to a periodic, approximate sinusoidal modulation of f_o and is described with parameters such as rate (undulations per second), amplitude (f_o extrema distance in reference to the mean or median f_o), regularity (similarity of frequency excursion to another), waveform (similarity to a sine wave), and onset (delay from onset of phonation until the first vibrato cycle).[1,4] A vibrato amplitude of approximately one semitone with an f_o increase in 6% is considered common for singers.[5] In general, professional singers aspire to having a regular and consistent vibrato with a rate of around 4–8 Hz.[6,7] Vibrato amplitude can increase with raising SPL.[4,8]

Achieving a balanced and free vibrato is an element of vocal training.[9] However, there is a high variability for vibrato parameters depending on sex, age, emotion, singing style, or repertoire.[4,9] Furthermore, over the years of commercial voice recordings, current singing styles and technique have influenced vibrato parameters.[10] Today, with informed performance practice corresponding to different repertoire (e.g., early music, concert or opera, solo or ensemble), adaptability of vocal vibrato depending on the required style is essential for a singer.[1,11] These features require both high controllability and flexibility of the vocal apparatus. It is assumed that vibrato tones need less glottal adduction than non-vibrato phonation and consume less air flow.[4]

A lot of analyses of vibrato parameters with different settings have been performed, predominantly using audio recordings.[10] From an acoustic point of view, vibrato can be

[a)]Email: matthias.echternach@med.uni-muenchen.de

https://doi.org/10.1121/10.0041893

described as a set of harmonics whose frequency and amplitude varies periodically, resulting in the perceived timbre, pitch, and loudness variation.[4,12] Synchronous SPL variations are often appearing with the fluctuation of f_o because of the altering relation between the voice-source partials and the vocal tract resonances, and synchronous alterations of the subglottic pressure (p_{sub}).[3,13]

It has been postulated that vibrato can be classified into two principal mechanisms of production. The first could rely upon rhythmic pulsations in p_{sub} produced by abdominal muscle contraction, which also influences f_o. The second could rely on laryngeal muscle activities.[14] In general, glottal adjustments are carried out by a set of intrinsic and extrinsic muscles, such as the cricothyroid (CT) muscle stretching the vocal folds by a cricothyroid tilting, thus raising f_o. Further, the adductor muscles may play an important role.[4] It has been discussed, if nonlinear source-filter interactions could be the source of vibrato. However, currently it seems like interactions between vocal tract and voice source can modulate vibrato but not be its primary source.[15] Whenever vocal folds are oscillating, the glottis alternates between an open and closed phase,[16] while their amplitude is affected by the adduction level. In general, with decreasing adduction in the sense of flow phonation, the vocal fold oscillatory amplitude increases, while the closed phase of the glottal cycle is decreasing, even though a certain minimum adduction is needed for the self-sustained oscillation.[5,17]

The generation of vibrato and the biomechanic and dynamic processes are not yet entirely understood. In particular, the changes of vocal fold oscillation patterns due to vibrato singing are not clarified in detail. To get a detailed view into vibrato associated modifications of the vocal apparatus, transnasal high speed videolaryngoscopy was used in this study in addition to audio and electroglottographic signal recordings. It was hypothesized that oscillatory characteristics would differ within a vibrato cycle and that laryngeal movements would show features of CT muscle activity for vibrato generation. The present study focuses on acoustic and physiological differences of non-vibrato and vibrato phonation in male and female professional singers of western classical education. Furthermore, this study investigates modulations of cuneiform cartilage movements connected to vibrato. These cuneiform cartilages are small, elongated pieces of cartilage located within the aryepiglottic folds of the larynx und associated with great visibility during laryngoscopy, thus being more suitable for tracking than the arytenoid cartilages. Due to the close positional relation, it was expected that position changes of the cuneiform cartilages would represent arytenoid cartilage movement.

II. MATERIAL AND METHODS

After approval from the local ethics committee (Medical Ethics Committee of the University of Munich, No. 18/769), ten professional singers (five female, five male; aged 20–30 years) were included in the study. The taxonomy of all subjects is shown in Table I. All participants ($n = 10$) were at least full time post-graduate students of Western classical singing. In addition to the negation of any self-perceived voice problems by the subjects, vocal fold pathologies were excluded by clinical evaluation performed by an experienced phoniatrician and laryngologist as well as by consideration of the dysphonia severity index.[18]

A. Task

All subjects were asked to sing a sustained tone on vowel [i:] and pitch B2 ($f_o \approx 124$ Hz) for male and B3 ($f_o \approx 247$ Hz) for female subjects, respectively, for around 6 s in total and at a comfortable loudness. Within this total period, participants were asked to begin the phonation with around 2 s of singing the tone straight/non-vibrato, followed by the singers' individual vibrato for around 2 s and a concluding phase of 2 s of non-vibrato. The intended pitch was provided by the examiner prior to each phonation. In order not to influence the participants and to potentially identify fundamental similarities, the only instruction given was to perform with or without vibrato.

Only subjects with quasi regular periodic vibrato as visually controlled by the audio and electroglottographic signals by the examiner and a minimal amplitude of ± 0.25 semitone were included for further analysis. Half a semitone corresponds to a 3% frequency change, setting the minimum vibrato extent (f_o min-to-max) threshold at 3.6 Hz for males and 7.2 Hz for females.

TABLE I. Classification and taxonomy of all subjects according to Bunch and Chapman (Ref. 19).

Subject	Classification	Taxonomy	
S1	Soprano	7.1	Fulltime voice student university–postgraduate
		4.5	Regional touring
S2	Soprano	7.1	Fulltime voice student university–postgraduate
		4.5	Regional touring
S3	Soprano	4.5	Regional touring
M1	Mezzosoprano	7.1	Fulltime voice student university–postgraduate
		4.5	Regional touring
M2	Mezzosoprano	7.1	Fulltime voice student university–postgraduate
		4.5	Regional touring
T1	Tenor	7.1	Fulltime voice student university–postgraduate
		4.5	Regional touring
B1	Baritone	7.1	Fulltime voice student university–postgraduate
		4.5	Regional touring
B2	Baritone	7.1	Fulltime voice student university–postgraduate
		4.5	Regional touring
B3	Baritone	4.5	Regional touring
B4	Baritone	4.5	Regional touring

B. Recordings

During the phonation, the participants' performances were recorded by a microphone (5 cm distance to mouth's edge), electroglottography (EGG), and transnasal high-speed videolaryngoscopy (HSV), as performed in previous investigations.[20–22] The audio signals [microphone model 4061 (DPA Microphones, Kokkedal, Denmark) and microphone model ME 62 (Sennheiser, Wedemark, Germany)] were calibrated using a sound level meter (DSL 331, Tecpel, New Taipei City, Taiwan) and the Sopran software (Svante Granqvist, Karolinska, Stockholm, Sweden). EGG (EG2-PCX2, Glottal Enterprises, Syracuse, NY) was used to evaluate impedance changes during vocal fold oscillations. HSV was performed with a flexible endoscope [HSV, Fastcam SA-X2 (Photron, Tokyo, Japan) and ENF GP (Olympus, Hamburg, Germany)] with a framerate of 20.000 frames per second and a spatial resolution of 386×320 pixels, generating a two-dimensional grayscale view of the glottal and supraglottal area.

C. Data processing

HSV images were processed by a 3×3 binomial filter (an approximated Gauß filter to reduce random noise) and Fourier filter to minimize the noise due to the honeycomb structure produced by endoscopic fibers.[23] The spectral peaks corresponding to the comb were picked manually. Rotation to the glottal midline and cropping was performed as described previously.[20] Segmentation with the Glottal Analysis Tools Software (University Hospital at FAU Erlangen-Nürnberg, Germany)[24] was followed by glottal area waveform (GAW) analysis. Based on the GAW, parameters such as the open quotient (OQ_{GAW}) and closing quotient (ClQ_{GAW}) were calculated for 500 ms of the stable part of every phase of the task (non-vibrato, vibrato, and again non-vibrato) using Glottis Analysis Tools[24] and Multi Signal Analyser (University Hospital at FAU Erlangen-Nürnberg, Germany). An increase in OQ would be expected for raising f_o,[25] and ClQ would be expected to negatively correlate with SPL.[26] From the electroglottogram, the vocal folds' open quotient (OQ_{EGG}) was computed with the Howard method.[27]

D. Vibrato extrema analysis

For further processing, f_o was extracted from GAW data using the Praat software (University of Amsterdam, Netherlands). The f_o extraction was a two-pass operation with an algorithm manipulating floor and ceiling f_o values. After f_o object extraction in the first pass (50–700 Hz) and calculation of f_o, the values of the first and third quartiles of the f_o distribution (q1, q3) were taken, and new optimal values for the f_o floor and f_o ceiling were calculated using the following formulas: $floor = 0.7 \cdot q1$ and $ceiling = 1.5 \cdot q3$(Ref. 28) in the second pass, creating another f_o object extraction. The modulation of f_o over time was further processed using *islocalmin* and *islocalmax* (MATLAB) extrema recognition. To find regions with vibrato in the f_o *course*, minima and maxima were recognized by a minimal distance of 1/10 s, expecting a vibrato rate below 10 Hz, i.e., 4–8 Hz.[6,7] Resulting regions were then further selected by a minimal prominence of 0.5 semitones between minimum and maximum f_o during vibrato, i.e., 0.25 semitones above and below center f_o. The prominence of a local extremum as a peak indicates how the peak stands out with respect to its height and location relative to other peaks and is measured by the vertical distance from a neighboring valley to the peak. In the next step, mean values of minima-maxima pairs in the f_o line were calculated and used as a center point for the identification of the upper and lower 50% of a half-wave representing the region of interest in a valley or peak (see Fig. 1). These f_o regions were then used as a reference for GAW cycle identification and selection. Similar as for the f_o, *islocalmin* and *islocalmax* were used for extrema recognition in the selected parts of the GAW. Between minima ($Vibf_{omin}$) as well as maxima ($Vibf_{omax}$), a minimal time gap of $(1/max(f_o))/2.5$ was assumed, and for maxima a prominence of $0.5 \cdot max(GAW)$ was implemented, excluding smaller artifacts present in the signal. Finally, in a time frame of 500 ms within the vibrato phase of the task, the OQ_{GAW} and ClQ_{GAW} of present minima and maxima were calculated.

E. Comparison to synthetic signals

Since vibrato adds constant modulations of f_o and amplitude to a signal, it was suspected that it might affect voice stability measures. In order to check the influences of the vibrato on such outcome parameters and to create a basic expectation, synthetic vibrato signals were generated according to the measured vibrato parameters (frequency, rate and f_o) of each singer. To generate the synthetic vibrato signal, frequency modulation was used. Frequency modulation is a technique that expresses the amplitude of a slow varying signal by the change of frequency of a high-frequency signal, the so-called carrier frequency. A sine wave signal with the observed vibrato frequency was modulated onto the carrier signal corresponding to requested f_o for each singer using MATLAB's *fmmod* (frequency modulation) function and its frequency deviation parameter set to the vibrato extent.

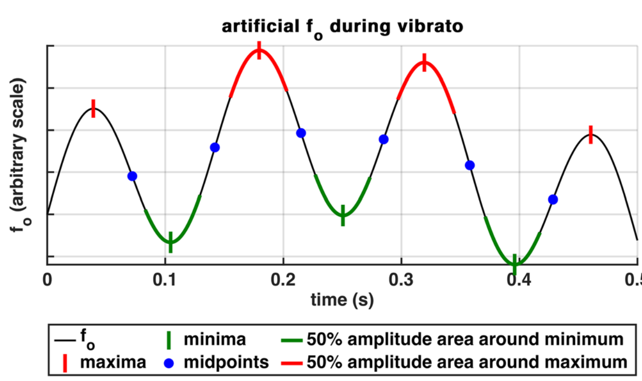

FIG. 1. Exemplary extraction of areas around f_o maxima and minima using a 50% halfwave amplitude threshold referred to the extrema midpoints. Since the f_o course was extracted from the GAW data, this allows to compare ClQ_{GAW} and OQ_{GAW} values between positive and negative halfwaves.

https://doi.org/10.1121/10.0041893

Synthetic sines of 0.5 s length with up to 30 harmonics were generated, including a spectral roll-off of 12 dB/octave to simulate a complex voice spectrum. Those signals were generated as "absolutely regular" versions without vibrato and with vibrato (with parameters corresponding to the subjects' vibrato) and a sample rate of 20 kHz (analog to the recorded signals). Then, the cepstral peak prominence (CPP)[29,30] was calculated and compared to the CPP from the human recordings. CPP is a measure of stability based on the cepstrum and captures periodicity perturbations in amplitude, frequency, or noise.

F. Cartilage tracking

In order to analyze adduction and abduction of the arytenoid cartilages indirectly, the laryngoscopic image series were further processed by tracking the cuneiform tubercles. As these are part of the aryepiglottic folds and close to the arytenoid cartilages, it was expected that motion of the cuneiform cartilages would indirectly reflect the vocal fold adduction and abduction, as performed in studies on articulation.[31] However, due to the fact that the cuneiform tubercles are not fixed in their position, supraglottic contractions could also have an effect on the position, i.e., by stretching of the aryepiglottic fold by a forward tilt of the thyroid cartilage through CT activity. If necessary, brightness and contrast of the image material were increased. For the tracking, the same 500 ms windows as for the other calculations were used.

One of the reflection points visible on the right lateral cuneiform tubercle was manually marked as a region of interest for tracking in one of the frames (see Fig. 2). The selected region was then tracked in the following images within the 500 ms frame in order to detect movements of the cartilage in x and y. Since laser triangulation[32–34] is neither available for flexible endoscopy nor yet approved for clinical use, the movement in the z direction was estimated by intensity measurements of the reflection point, which allows only qualitative approximations.

The tracking of the reflection in the x, y, z direction was then performed automatically based on Kleiner *et al.*[31] with a frame step size of 100 frames corresponding to 5 ms and a pixel search range of 2. To correct the angle of the cuneiform tubercles and glottis to the vertical axis, in the case of image series with rotated visualization of these structures, the results of the x and y directional movement of the cuneiform tubercles were rotated around the z-axis.[23,35]

Results of the x, y, z directional movement of the cuneiform tubercles were compared to the corresponding f_o frequency and SPL variation. The resulting graphs were smoothed using a median filter of length 5 followed by exponential smoothing, and a scaling factor was added to enable visualization of small movements as well.

In the results, positive x-value-directions would mean increasing adduction, positive y-directions would indicate forward movement, and positive z-direction would mean an upward movement.

G. Statistics

Statistical analyses were done with Excel 365 (Microsoft, Redmond, Washington). Furthermore, Wilcoxon's signed-rank test for paired samples was used for the direct comparisons of non-vibrato and vibrato results (using RealStatistics Toolbox[36] with continuity and ties correction and exact test statistics p-values). Comparative analysis (Pearson correlation) of the collected parameters of vocal non-vibrato and vibrato tones was performed. A significance level of $\alpha = 0.05$ was set, and Bonferroni correction was applied to compensate for multiple comparisons. Reported p-values are already adjusted accordingly.

III. RESULTS

A. Analysis of vibrato extrema

Three of the subjects exhibited either an irregular vibrato (B4) or an amplitude smaller ± 0.25 semitones (M2, T1) and were therefore excluded from further analysis. The remaining subjects exhibited vibrato frequencies of

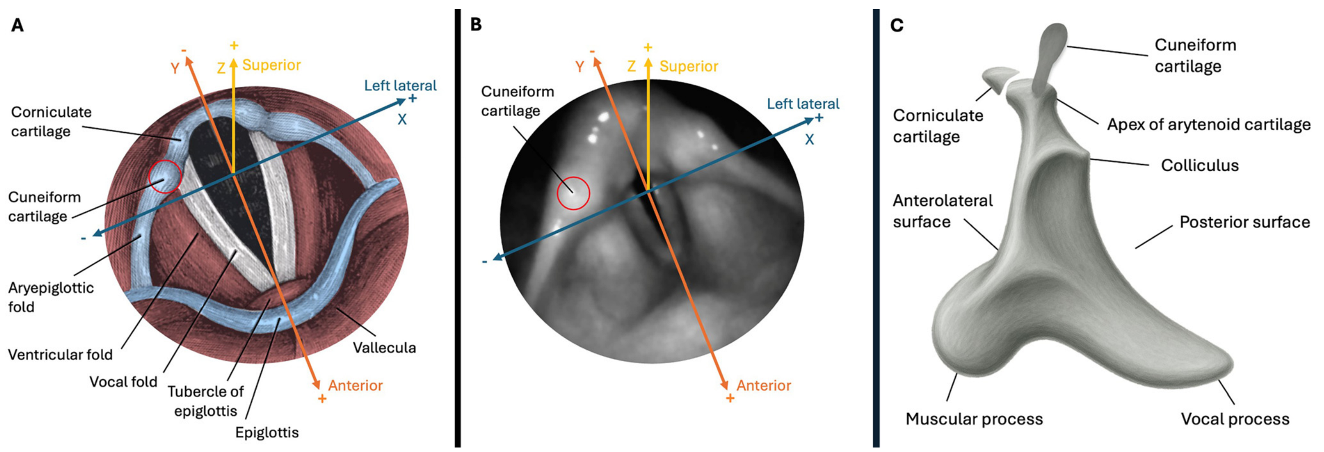

FIG. 2. Schematic illustration of the cartilage movement along x, y, and z-axes for an anatomical image including anatomical landmarks (A) and the laryngoscopic image (B). The red circle marks the tracked light reflex. (C) shows the arytenoid and the cuneiform cartilage and their relation.

https://doi.org/10.1121/10.0041893

TABLE II. Vibrato frequency (the rate of f_o change) and amplitude (deviation from mid frequency) of the analyzed subjects, together with the f_o to which the values refer.

Subject	Vibrato Frequency [Hz]	Vibrato Amplitude [Hz]	Vibrato Amplitude [semitones]	f_o [Hz]
S1	4.9	11.0	0.73	254.95
S2	6.6	5.4	0.35	249.58
S3	5.8	5.9	0.38	250.79
M1	5.3	6.5	0.43	250.13
B1	6.1	5.7	0.75	127.19
B2	6.6	4.8	0.65	125.48
B3	4.1	4.1	0.55	126.3

4.1–6.6 Hz and amplitudes of 0.35–0.75 semitones (reported in Table II). Figure 3 shows the mean difference of Vibf_{omin} from Vibf_{omax} for OQ_{GAW} and ClQ_{GAW}. The values of both quotients are widely spread, while more participants showed a higher OQ_{GAW} for Vibf_{omin} than Vibf_{omax}.

B. Comparative analysis of vocal non-vibrato and vibrato tones

Analysis of $f_{o,GAW}$ showed that singers performed close to the target pitches and associated f_o (see SuppPub1 in the supplementary material). On average, female singers sang at 245 Hz and male singers at 122 Hz, maintaining an almost constant f_o level for non-vibrato and vibrato tones.

SPL was statistically significantly higher during vibrato (median 81.2 dB) compared to non-vibrato (78.6 dB, $p = 0.031$ and 76.8 dB, $p = 0.031$) (see Fig. 4). Figure 5 shows that there are no homogenous and statistically significant differences in ClQ_{GAW} and OQs regarding vibrato and non-vibrato. However, for single subjects (S2, B2) ClQ_{GAW} and OQ_{GAW} decreased for vibrato tones.

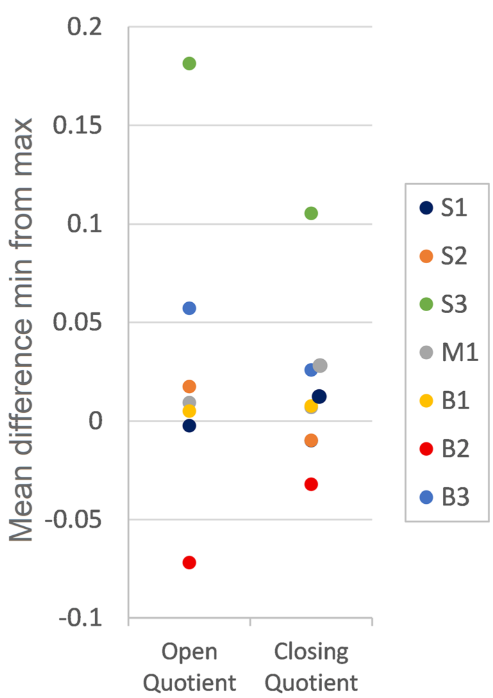

FIG. 3. Mean difference of OQ_{GAW} and ClQ_{GAW} compared between f_o minimum (Vibf_{omin}) and maximum (Vibf_{omax}). Values > 0 indicate a higher OQ or ClQ during Vibf_{omin} compared to Vibf_{omax}. (Formula: $-(\text{Quotient}_{\text{Vib}f_{omax}} - \text{Quotient}_{\text{Vib}f_{omin}})$.

CPP_{EGG} was lower for vibrato tones compared to non-vibrato tones ($p = 0.031$ and 0.031), with median values of 27.7 dB for vibrato and 37.5 dB and 34 dB for the non-vibrato phases. The CPP_{Audio} showed no systematic difference when comparing non-vibrato to vibrato ($p = 0.063$ and 0.16). The artificial vibrato signals' CPP values were higher compared to the human vibrato ($p = 0.031$), but the artificial signals showed comparable values ($p = 0.44$) regardless of the existence of vibrato (Figs. 6 and 7).

C. Cuneiform tubercles' movements during non-vibrato and vibrato

Subjects B1, B2, and B3 had to be excluded from analysis of the cuneiform tubercles' movements due to artefacts from light reflections. Thus, only subjects S1, S2, S3, and M1 were selected for automatic cartilage tracking. All non-vibrato tones showed only small, steady and non-periodic x, y, and z movements of the cuneiform cartilages (see Fig. 8 and SuppPub2 in the supplementary material). In contrast, during vibrato, systematic movements of the cuneiform cartilages with respect to f_o were found (summarized in Table III). At the same time, there were no great changes of arytenoid cartilages or vocal fold position by visual observation of the videos performed by an experienced laryngologist. In relation to f_o, there were inverse movements in the x direction, suggesting cuneiform cartilage adduction during Vibf_{omin} and abduction during Vibf_{omax} (recall Fig. 2: negative x direction means lateralization). All subjects exhibited a forward movement for Vibf_{omax}. After reinspection, the detected positional change in forward direction for M1 appeared to be an artefact produced by laryngeal elevation.

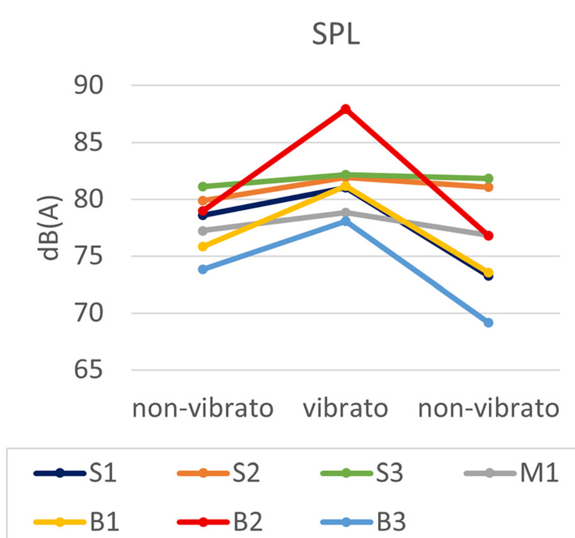

FIG. 4. SPL comparison of non-vibrato and vibrato phonation. SPL was statistically significantly higher during vibrato ($p = 0.031$).

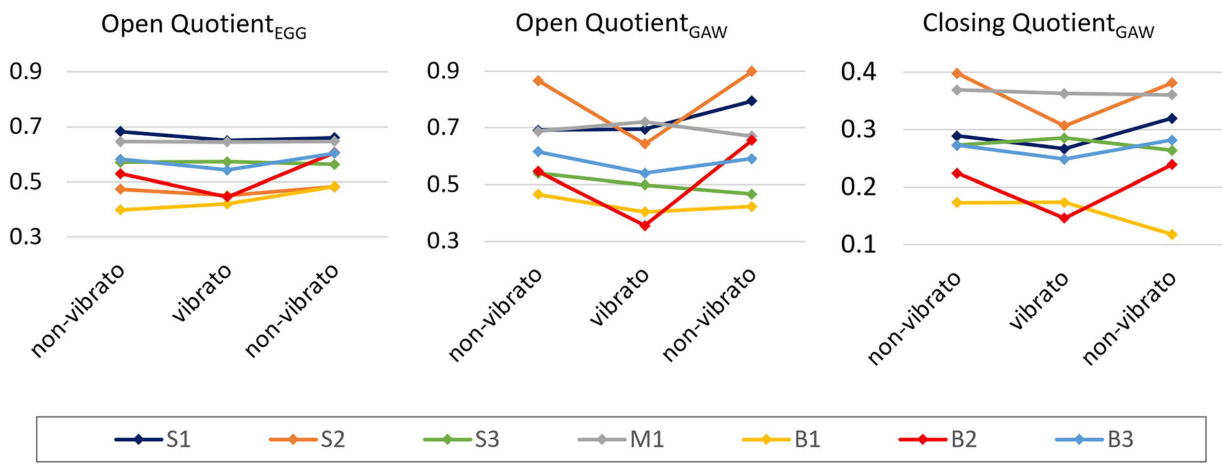

FIG. 5. All subjects' OQ and ClQ values of vibrato and non-vibrato phonation based on GAW and EGG data. Note that GAW and EGG represent different aspects of glottal opening, which can lead to discrepancies in OQ (Ref. 37).

For subject S1, however, the positional change was different: Due to the rather stable arytenoid position and motion of the cuneiform cartilages, the relation of the supraglottis was analyzed using a kymography through the cuneiform tuberculum right to the epiglottis. As can be seen in Fig. 9 for subject S1, there was a strong anterior movement of the epiglottis during Vibf_{omax}, which was synchronized with the small forward displacement of the cuneiform tubercle.

All subjects showed parallel z courses with respect to f_o. SPL was unsystematically connected to f_o being inversely related in subjects S2 and M1 but parallel in S1.

IV. DISCUSSION

The presented study examined acoustic and physiological parameters during vibrato and non-vibrato phonation in male and female professional singers. In general, it has been found that during the vibrato, Vibf_{omin} was frequently associated with a higher OQ compared to Vibf_{omax}. Although SPL was found to be significantly higher during vibrato compared to non-vibrato, there were no statistically significant differences in overall OQ and ClQ, indicating that vibrato does not affect these glottal parameters on average. Furthermore, visible cuneiform cartilage movement to the front during vibrato was not necessarily associated with arytenoid movement but a consequence of forward movement of the epiglottis, which could be explained by increased CT muscle activity or laryngeal elevation for Vibf_{omax}.

A. SPL or glottal parameters related to vibrato

In previous studies, an increase in vibrato amplitude has been associated with an increase in SPL or loudness.[38] Furthermore, as vibrato is probably linked to variations in muscle activity and, therefore, ad/abduction forces, it can be assumed that there are also variations in subglottal pressure and SPL.[5] All participants in the presented study showed a significantly higher and fluctuating SPL during the vibrato phase in contrast to the non-vibrato phases. Sundberg *et al.*[5] suggested that non-vibrato tones are produced with higher vocal fold adduction force and, therefore, higher glottal resistance than vibrato tones. Since vocal fold ad- and abduction usually contributes to the maximum flow declination rate and can consecutively help increase loudness and/or SPL,[5] the presented SPL data could represent a rather hyperadducted and pressed phonation with lower SPL for the non-vibrato. Apart from the fact that hyperadduction is rather unlikely in trained professional singers, it is

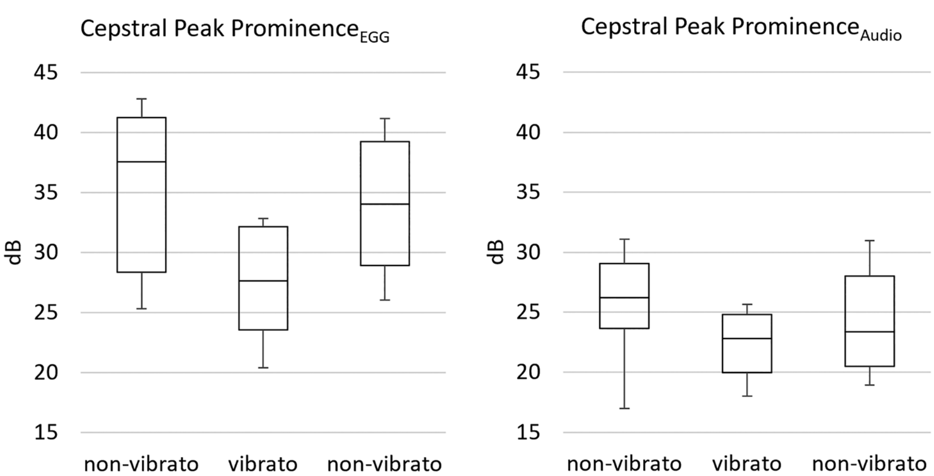

FIG. 6. Comparative analysis of CPP results of the three different phases of the task, based on EGG and audio data.

JASA

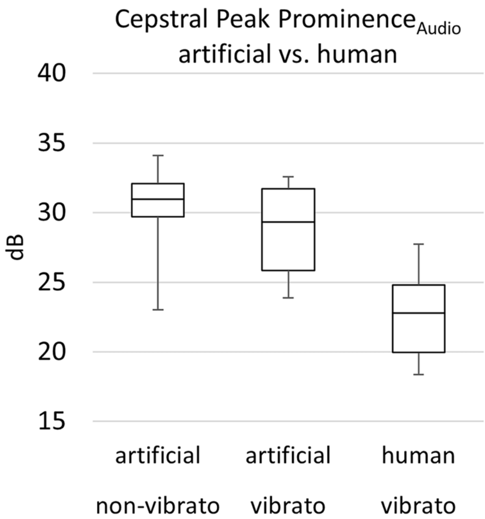

FIG. 7. Comparative analysis of CPP results of the artificial generated signal without vibrato, with vibrato, and the human vibrato. CPP values of artificial signals were higher compared to the human vibrato.

TABLE III. Subjects' courses of *x*, *y*, *z* movement and SPL in relation to f_o and respective correlation (*r*) and significance (*p*). If two parameters change into the same direction at the same time, the courses are tagged as parallel. Cells with $p < 0.05$ are marked gray.

	S1	S2	S3	M1
x (lateral-medial)	Inverse $r = -0.74$ **p < 0.001**	Inverse $r = -0.85$ **p < 0.001**	— $r = +0.12$ p = 0.899	Inverse $r = -0.40$ **p < 0.001**
y (posterior-anterior)	Parallel $r = +0.42$ **p < 0.001**	Parallel $r = +0.48$ **p < 0.001**	Parallel $r = +0.41$ **p < 0.001**	Parallel $r = +0.33$ **p < 0.003**
z (cranial-caudal)	Parallel $r = +0.69$ **p < 0.001**	Parallel $r = +0.67$ **p < 0.001**	Parallel $r = +0.34$ **p < 0.001**	Parallel $r = +0.51$ **p < 0.001**
SPL	Parallel $r = +0.58$ **p < 0.001**	Inverse $r = -0.49$ **p < 0.001**	— $r = +0.17$ p = 0.376	Inverse $r = -0.65$ **p < 0.001**

contradicted by visual analysis and no reduced OQ_{GAW} and OQ_{EGG} during the non-vibrato compared to vibrato in the presented data set. High SPL requires a higher air compression, frequently associated with a small OQ during phonation.[25] Beside the glottal resistance, SPL is also greatly depending on the transglottic pressure difference mainly controlled by the subglottal pressure.[39] In order to prevent a collapse of vocal fold oscillation for higher subglottic pressure, vocal fold tension frequently has to be increased, i.e., by the vocalis (VOC) muscle activity, which could, in turn, reduce OQ for the higher SPL in vibrato phonation, as for example in subject B2.

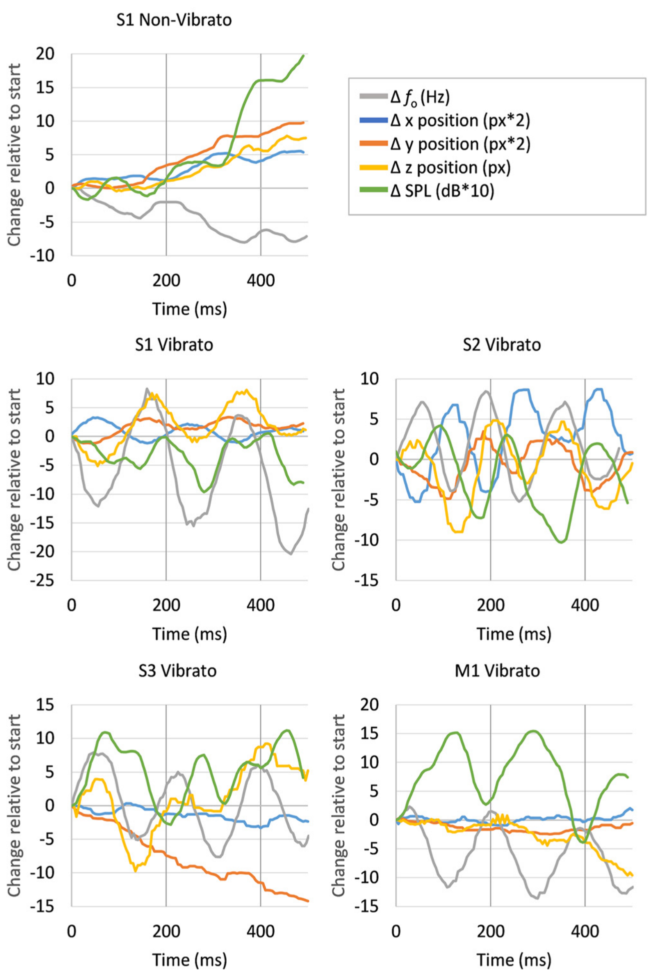

FIG. 8. Qualitative cuneiform cartilage tracking in *x*, *y*, *z*-direction within 500 ms of non-vibrato and vibrato in comparison to f_o and SPL. Values are relative to the start. Scaling factors are indicated in the legend.

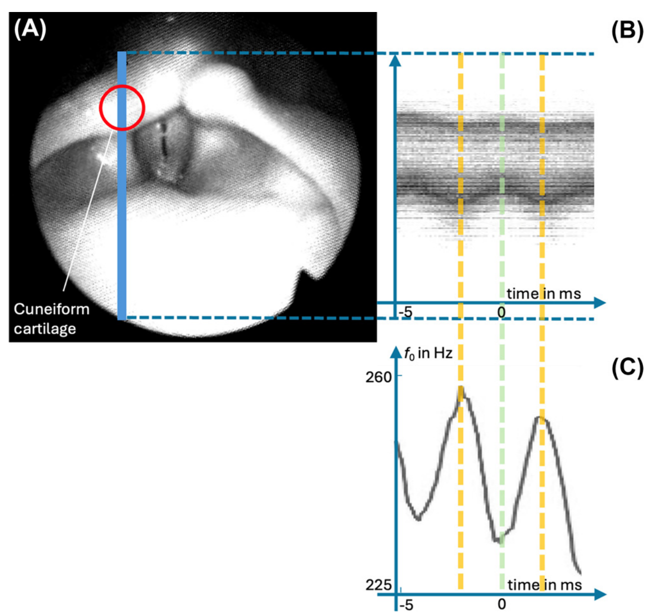

FIG. 9. Kymogram of cuneiform tubercle and epiglottic cartilage (B) in relation to vibrato's f_o modulation (C). (A) The kymography line; the red circle marks the cuneiform tubercle.

In a majority of participants, the OQ_{GAW} was higher in Vibf_{omin} compared to Vibf_{omax} within the vibrato cycle. With respect to SPL changes produced by subglottic pressure, it could be expected that a greater SPL would be present for a lower OQ, thus for Vibf_{omax}.[25] However, the relation of f_o and SPL was rather unsystematic. It should be mentioned that other factors could also explain such missing relations. In addition to control by subglottic pressure, singers frequently use a flow phonation to increase SPL, which is associated with reduced glottal resistance by abduction, and therefore increased OQ.[17,40] Further, Henrich *et al.*[25] and Yokonishi *et al.*[41] reported no clear relation between OQ and f_o for modal phonation but a relation of lower f_o and higher OQ for falsetto. Even though the participants in the present study phonated rather in modal voice quality, i.e., within speaking voice range, classically trained singers might use some kind of mixed voice[42] showing the previously mentioned feature of falsetto quality. Last, from an acoustic point of view, vibrato can be described as a set of harmonics or partials whose frequency and amplitude varies regularly along time, resulting in the perceived timbre, frequency, and SPL variations.[12,38] However, due to the varying distance of voice source partials to a resonance, both Vibf_{omax} and Vibf_{omin} could exhibit a greater SPL due to individual vocal tract shapes and resonances. Synchronous SPL or loudness variations often appear with the fluctuation of f_o because of the altering relation or intersection between the source harmonics and the vocal tract resonances and synchronous oscillations in the SPL.[3,13]

B. Vibrato and CPP

In the presented study, subjects produced lower CPP for vibrato tones than for non-vibrato tones in both audio and EGG signals. Different CPP values reflect how pure and dominant the periodic structure is in the respective signal. A comparison of these values provides insights into how sensitive and robust CPP is as a measure of periodicity in different measurement contexts, and how strongly external factors (modulation, harmonics, physiological fluctuations) influence the evaluation. For vibrato, it could be suggested that the constant alteration of the proximity between voice source partials and vocal tract resonances resulted in a small decrease in periodicity, the latter being represented by lower CPP. Still, all data for the vibrato and non-vibrato exhibited no values considered as pathological.[43] Interestingly, the CPP_{EGG} values, which reflect only vocal fold oscillations, showed much higher values—representing higher periodicity—than the CPP_{audio}, which is possibly affected more strongly by vocal tract resonances. In order to check the influence of other contributing factors on the CPP values, a synthetic signal was constructed based on the individual f_o, vibrato frequency, and amplitude. The synthetic signals exhibited a higher periodicity than the human signal. In addition to the fact that synthetic signals are, in principal, more periodic than natural utterances, other factors such as interactions, entrainment, etc. might influence CPP in human voice production, reducing vocal periodicity as expressed by the CPP in the presented study.

C. Cuneiform cartilage movement during non-vibrato and vibrato

In general, voice production relies on factors like subglottal pressure, vocal tract shaping, and the adjustment of vocal fold position and tension by muscles such as the interarytenoid (IA), thyroarytenoid (TA), lateral cricoarytenoid, VOC, and CT muscles, each influencing cartilage movements.[44] It has been suggested before that vibrato is associated with CT activation.[14] According to Shipp *et al.*,[14] CT activity precedes the f_o modulation in vibrato, which leads to a high activity when f_o is in the lower phase and less activity when f_o is in a higher phase. However, these EMG results may include a time shift between the measured activation impulse of the muscle and the actual muscle tensioning, which is expected to occur simultaneously with f_o modification. Consecutively, the presented study aimed to find a visual measure avoiding a time delay.

It was hypothesized that tracking of the cuneiform cartilage would reflect the movement of the arytenoid cartilage and, consequently, the vocal process. It has been shown before that an arytenoid adduction increases the medial surface[45] and, thus, raises f_o.[46] An increase on medial surface is associated with a reduction of the phonation threshold pressure, an effect on the subglottal pressure and a better energy transfer.[45–47] Further, a forward positioning of the arytenoids shortens and thickens the vocal folds, thus increasing the mass in action and, therefore, primarily decreasing f_o. However, the changes of f_o are greatly dependent on TA and VOC muscle activity.[48]

In general, it could be expected that during CT activation the arytenoids would not move, or would move very slightly posteriorly and medially.[49] Indeed, by means of observation, there was no great movement of the arytenoids or vocal folds. For Vibf_{omax}, however, there was an anterior movement of the cuneiform cartilages in the presented study for all subjects. Especially, for subject M1 the very small forward motion should be regarded as artefact due to a laryngeal elevation. Further, SuppPub 3 in the supplementary material and Fig. 9 show the movement of both, the epiglottic cartilage and the (tracked) cuneiform cartilage for subject S1. Here, it appears obvious that there is a very strong anterior displacement of the epiglottic cartilage, which should be produced by the tilting effect of the CT. As shown in the kymogram, the position of the cuneiform cartilage follows the epiglottic cartilage movement, albeit by a much smaller extent.

There are several consequences of this observation. The first is that the cuneiform cartilage movement does not necessarily reflect the arytenoid movement on the vocal fold level. It should be mentioned that, in contrast to investigations on articulation,[31] the motions were rather small. Also, during studies on articulation concerning speaking f_o, there is no great tilting due to a high amount of CT activation expected.[48] As a consequence, the motions of the cuneiform cartilage observed in studies on articulatory behaviors could very well reflect arytenoid motions.

With respect to CT activation, it appears plausible for the presented data, however, that such tracking could be used as an indirect measure for CT tilting but does not necessarily reflect arytenoid movement. Second, this finding would suggest that either the cuneiform cartilages, which are not fixed with the arytenoids, are movable by the general tilting force or that there were slight problems in adjusting the endoscope. Because the images themselves appear rather stable and the changes are so fast, the second possibility appears unlikely. Last, due to the variations in the supraglottic area, there could be an effect on the resonatory properties with an influence on SPL and nonlinear interactions, demonstrating the complexity of the system during vibrato production.

All analyzed participants performed an upwards lift of the cuneiform cartilage during f_o maxima and vice versa, visible to the examiners in the video material and also statistically significant, which should be an expression of a larynx elevation. Further, the measured small anterior position in subject M1 was also associated with such an elevation without a clear laryngeal tilt, as shown for subject S1. It has been found in previous studies that a laryngeal elevation is consistently associated with an increase in f_o, compared to neutral or lowered positions.[50,51] However, larynx elevation does not reliably alter SPL.[49] Honda *et al.*[52] found that when decreasing f_o, the hyoid bone tended to move back- and downward, accompanied by a cricoid cartilage rotation along the cervical lordosis and a downwards movement of the laryngeal cartilages mainly due to contraction of the infrahyoid extrinsic laryngeal muscles. As a result, larynx lowering contributed to vocal fold shortening and relaxation.[52] For high f_o, a small upward and horizontal forward movement was documented in the same study. These findings underline that elevation of f_o is not necessarily only caused by CT activation but that other factors like activation of strap muscles, hyoid bone position, or tongue position might contribute. Furthermore, it could be discussed if the effects are in line with the tongue-pull theory for high vowels.[53,54] In general, strong tongue and tongue root movements are frequently visible during endoscopy or influence the endoscope position by a great amount. Such influences were not detectable in the presented study. However, it should be noted that there are other forms of vibrato, especially for Musical Theater singers, which include hyoid bone position and tongue position changes and which are different from classically western trained singers.[55] Furthermore, adjustments of the vibrato concerning amplitude and frequency, which sometimes involved these structures, have also been observed for classical dramatic repertoire.[56] However, the vowel quality was constant during the presented experiment, thus limiting the possible explanation by the tongue-pull theory for the observed modifications.

D. General aspects

The results of the present study on f_o vibrato are in accordance with other studies' findings on airflow vibrato.[3] As Scherer and colleagues state, the causative factors are likely quite different between airflow vibrato and f_o vibrato. f_o vibrato depends on changes in vocal fold effective stiffness, length, and mass throughout the vibrato cycle, whereas airflow vibrato is most sensitive to changes in laryngeal airflow resistance (more airflow with less resistance due to greater glottal area) and subglottal pressure (more airflow with greater subglottal pressure). Both f_o vibrato and airflow vibrato may depend on vocal fold length changes, suggesting that a subtle elongation could raise f_o and also increase the glottal area and, consequently, the airflow value. Thus, f_o vibrato and airflow vibrato may be in phase with each other.[3] Because airflow vibrato and f_o vibrato are often synchronized, it is further hypothesized that as the cricothyroid (CT) and vocalis (VOC) muscle contraction alter to govern vocal fold tissue tension to change the f_o, the IA muscles are also changing in a synchronized manner with the CT and VOC for the subjects of this study. Possible phase delays between airflow vibrato and f_o vibrato constitute another area of needed explanation but may relate to the phasing between the vocal fold lengthening mechanism (CT and VOC) and the posterior adductory system (the IA muscles).[3]

E. Limitations

Several limitations are associated with this study. First, the small number of participants constrained detailed statistical analysis due to the demanding requirements and extensive amount of HSV data produced.

Second, SPL might be influenced by psychological factors of the participants, as vibrato often accompanies dramatic singing styles,[56] while non-vibrato is associated with a more solemn and clear sound. Moreover, vibrato is often produced automatically and unconsciously in professional Western trained classical singers. The transformation of a non-vibrato tone to a vibrato and reverse in a conscious way requires a high controllability of the personal voice and might have been unintuitive for the participants, since the task had not been practiced before the measurements.

Third, variations in vibrato parameters related to sex, age, emotion, and predominant singing style introduce inherent variability among subjects, which was not evaluated in this study. The intra-subject variability observed in the differences between the initial and final non-vibrato segments could not be accounted for sufficiently in the present study design. It raises the question of the reproducibility of the individual vibrato sections. Future studies should include several repetitions of the same task. Also, in this context, the participants were not asked about their individual technique. Therefore, no statements can be made about the intended quality or the maneuvers performed.

Additionally, discomfort due to transnasal laryngoscopy might have affected phonation. This could be a co-factor preventing a vibrato for subjects M2, T1, and B4.

Moreover, slight body movements might have caused variability in light reflections on cuneiform tubercles, influencing cartilage tracking. As pointed out before, the detected z-directional cartilage movements may also be influenced by endoscope movements, for example, due to soft palate activity rather than direct cartilage motion. Vertical movements, i.e., changes in distance to the camera, might have influenced the values of the *x*/*y* directions. The three-dimensioal tracking of laryngeal components *in vivo* is a challenging task for which no robust method has been established so far. As mentioned before, the applied cuneiform cartilage tracking cannot show absolute values, and the results must be regarded with caution. Furthermore, as pointed out before, the cuneiform tracking does not necessarily reflect arytenoid movement, at least if there is a stronger supraglottic movement, i.e., during CT activation. Thus, it appears desirable to find better indicators for positional changes reflecting framework changes for the voice source production.

V. CONCLUSION

This study examined acoustic and physiological parameters during non-vibrato versus vibrato phonation of male and female professional singers. During vibrato, most participants showed a higher OQ in Vibf_{omin} than in Vibf_{omax}. Vibrato was associated with lower CPP and higher SPL, and SPL courses inverse to f_o modulation. The cuneiform tubercle tracking showed an abduction and a laryngeal elevation for a rising f_o during the vibrato. A laryngeal tilting for rising f_o could indicate CT muscle activation. The relative forward motion for rising f_o was, at least in part, artificially influenced by the elevation. The cuneiform cartilage tracking does not necessarily reflect arytenoid and vocal fold movements.

SUPPLEMENTARY MATERIAL

See the supplementary material for fundamental frequencies of female and male subjects during the experimental phases of the task (SuppPub1), an illustration of the principle of reflection position tracking (SuppPub2), and a video of the parallel movement of epiglottic cartilage and cuneiform tubercle (SuppPub3).

ACKNOWLEDGMENTS

Part of the material is part of a bachelor thesis by I.G. The contribution of M.D. was supported by Deutsche Forschungsgemeinschaft under Grant No. DO 1247/25-1 and M.E. under Grant Nos. Ec 409/1-4 and Ec 409/5-1. The authors thank Rebekka Hoppermann for help concerning Figs. 2 and 9. Author contributions are as follows: I.G., performing experiments, analysis, manuscript draft; S.G., performing experiments, analysis, manuscript draft; P.B., manuscript draft; M.K., analysis, manuscript draft; J.K., analysis, manuscript draft; C.K., analysis, manuscript draft; M.D., analysis, manuscript draft; M.E., study design, performing experiments, analysis, manuscript draft, supervision.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

Ethical approval for this study was obtained from the Medical Ethics Committee of the University of Munich (No. 18/769). Written informed consent was obtained from all participants.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1]H. F. Mitchell and D. T. Kenny, "Change in vibrato rate and extent during tertiary training in classical singing students," J. Voice **24**(4), 427–434 (2010).
[2]C. Dromey, N. Carter, and A. Hopkin, "Vibrato rate adjustment," J. Voice **17**(2), 168–178 (2003).
[3]S. Nandamudi and R. C. Scherer, "Airflow vibrato: Dependence on pitch and loudness," J. Voice **33**(6), 815–830 (2019).
[4]J. Sundberg, "Quarterly progress and status report: Acoustic and psychoacoustic aspects of vocal vibrato," StL-QPSR **35**, 45–68 (1994).
[5]J. Sundberg, *Die Wissenschaft Von Der Singstimme* (*The Science of the Singing Voice*) (Wißner-Verlag, Augsburg, Germany, 2014), p. 85.
[6]C. Manfredi, D. Barbagallo, G. Baracca, S. Orlandi, A. Bandini, and P. H. Dejonckere, "Automatic assessment of acoustic parameters of the singing voice: Application to professional western operatic and jazz singers," J. Voice **29**(4), 517.e1–517.e9 (2015).
[7]M. Müller, "New aspects in objective voice analysis for generating knowledge relevant to vocal pedagogy," Dissertation, Charité – Universitätsmedizin Berlin, Berlin, Germany, 2022.
[8]H. J. Schultz-Coulon and R. D. Battmer, "Die quantitative Bewertung des Sängervibratos" ("Quantitative evaluation of the vibrato of singers"), Folia Phoniatr. Logop. **33**(1), 1–14 (1981).
[9]M. Muller, T. Schulz, T. Ermakova, and P. P. Caffier, "Lyric or dramatic - Vibrato analysis for voice type classification in professional opera singers," IEEE/ACM Trans. Audio. Speech. Lang. Process. **29**, 943–955 (2021).
[10]I. Ferrante, "Vibrato rate and extent in soprano voice: A survey on one century of singing," J. Acoust. Soc. Am. **130**(3), 1683–1688 (2011).
[11]H. Daffern, "Blend in singing ensemble performance: Vibrato production in a vocal quartet," J. Voice **31**(3), 385.e23–385.e29 (2017).
[12]I. Arroabarren and A. Carlosena, "Voice production mechanisms of vocal vibrato in male singers," IEEE Trans. Audio. Speech Lang. Process. **15**(1), 320–332 (2007).
[13]M. Rothenberg and D. Miller, "Aerodynamic investigation of sources of vibrato," Folia Phoniatr. Logop. **40**, 244–260 (1988).
[14]T. Shipp, E. T. Doherty, and S. Haglund, "Physiologic factors in vocal vibrato production," J. Voice **4**(4), 300–304 (1990).
[15]Z. Zhang, "The influence of source-filter interaction on the voice source in a three-dimensional computational model of voice production," J. Acoust. Soc. Am. **154**(4), 2462–2475 (2023).
[16]M. Echternach, M. Döllinger, J. Sundberg, L. Traser, and B. Richter, "Vocal fold vibrations at high soprano fundamental frequencies," J. Acoust. Soc. Am. **133**(2), EL82–EL87 (2013).
[17]C. T. Herbst, M. Hess, F. Müller, J. G. Švec, and J. Sundberg, "Glottal adduction and subglottal pressure in singing," J. Voice **29**(4), 391–402 (2015).
[18]F. L. Wuyts, M. S. De Bodt, G. Molenberghs, M. Remacle, L. Heylen, B. Millet, K. Van Lierde, J. Raes, and P. H. Van de Heyning, "The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach," J. Speech Lang. Hear. Res. **43**(3), 796–809 (2000).
[19]M. Bunch and J. Chapman, "Taxonomy of singers used as subjects in scientific research," J. Voice **14**(3), 363–369 (2000).
[20]M. Echternach, F. Burk, M. Köberlein, C. T. Herbst, M. Döllinger, M. Burdumy, and B. Richter, "Oscillatory characteristics of the vocal folds across the tenor passaggio," J. Voice **31**(3), 381.e5–381.e14 (2017).
[21]M. Echternach, F. Burk, M. Köberlein, M. Döllinger, M. Burdumy, B. Richter, I. R. Titze, C. P. H. Elemans, and C. T. Herbst, "Biomechanics of

sound production in high-pitched classical singing," Sci. Rep. **14**(1), 13132 (2024).

[22]M. Echternach, C. T. Herbst, M. Köberlein, B. Story, M. Döllinger, and D. Gellrich, "Are source-filter interactions detectable in classical singing during vowel glides?," J. Acoust. Soc. Am. **149**(6), 4565–4578 (2021).

[23]R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. (Pearson Education Limited, London, UK, 2018).

[24]A. M. Kist, P. Gómez, D. Dubrovskiy, P. Schlegel, M. Kunduk, M. Echternach, R. Patel, M. Semmler, C. Bohr, S. Dürr, A. Schützenberger, and M. Döllinger, "A deep learning enhanced novel software tool for laryngeal dynamics analysis," J. Speech. Lang. Hear. Res. **64**(6), 1889–1903 (2021).

[25]N. Henrich, C. D'Alessandro, B. Doval, and M. Castellengo, "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency," J. Acoust. Soc. Am. **117**(3), 1417–1430 (2005).

[26]J. Kirsch, M. Köberlein, M. Döllinger, and M. Echternach, "Influence of Messa di Voce speed on vocal stability of untrained, healthy subjects," PLoS One **20**(1), e0314457 (2025).

[27]D. M. Howard, "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," J. Voice **9**(2), 163–172 (1995).

[28]D. Hirst, "The analysis by synthesis of speech melody," J. Speech Sci. **1**(1), 55–83 (2011).

[29]Glottis Analysis Tools 2020: User guide, https://pdfs.semanticscholar.org/1a06/525cadb2993727b4ef360f59001ca43ff045.pdf (Last viewed December 12, 2025).

[30]R. Fraile and J. I. Godino-Llorente, "Cepstral peak prominence: A comprehensive analysis," Biomed. Signal Process. Control **14**, 42–54 (2014).

[31]C. Kleiner, M.-A. Kainz, M. Echternach, and P. Birkholz, "Velocity differences in laryngeal adduction and abduction gestures," J. Acoust. Soc. Am. **151**(1), 45–55 (2022).

[32]R. R. Patel, M. Döllinger, B. Jakubaß, H. Pinhack, U. Katz, and M. Semmler, "Analyzing vocal fold frequency dynamics using high-speed 3D laser video endoscopy," Laryngoscope **134**(7), 3267–3276 (2024).

[33]R. R. Patel, M. Döllinger, and M. Semmler, "3D reconstruction of vocal fold dynamics with laser high-speed videoendoscopy in children," Laryngoscope Investig. Oto. **9**(5), e70024 (2024).

[34]M. Semmler, M. Döllinger, R. R. Patel, A. Ziethe, and A. Schützenberger, "Clinical relevance of endoscopic three-dimensional imaging for quantitative assessment of phonation," Laryngoscope **128**(10), 2367–2374 (2018).

[35]G. Strang, *Linear Algebra and Its Applications*, 2nd ed. (Elsevier Science & Technology, Amsterdam, the Netherlands, 1980).

[36]C. Zaiontz, "Real Statistics Excel add-in, version 9.5.5," www.real-statistics.com (Last viewed December 12, 2025).

[37]M. Echternach, M. Döllinger, M. Köberlein, L. Kuranova, D. Gellrich, and M. A. Kainz, "Vocal fold oscillation pattern changes related to loudness in patients with vocal fold mass lesions," J. Otolaryngol. Head Neck Surg. **49**(1), 80 (2020).

[38]J. Sundberg, J. Iwarsson, and H. Hagegård, "Quarterly progress and status report: A singer's expression of emotions in sung performance," StL-QPSR **35**, 81–92 (1994).

[39]A. Bouhuys, J. Mead, D. F. Proctor, and K. N. Stevens, "Pressure-flow events during singing," Ann. N.Y. Acad. Sci. **155**(1), 165–176 (1968).

[40]Z. Zhang, "Laryngeal strategies to minimize vocal fold contact pressure and their effect on voice production," J. Acoust. Soc. Am. **148**(2), 1039–1050 (2020).

[41]H. Yokonishi, H. Imagawa, K.-I. Sakakibara, A. Yamauchi, T. Nito, T. Yamasoba, and N. Tayama, "Relationship of various open quotients with acoustic property, phonation types, fundamental frequency, and intensity," J. Voice **30**(2), 145–157 (2016).

[42]S. Lamesch, R. Expert, M. Castellengo, N. Henrich Bernardoni, and B. Chuberre, "Investigating voix mixte: A scientific challenge towards a renewed vocal pedagogy," in *Proceedings of the ICA 2007 - 19th International Congress on Acoustics*, Madrid, Spain (2007).

[43]O. Murton, R. Hillman, and D. Mehta, "Cepstral peak prominence values for clinical voice evaluation," Am. J. Speech. Lang. Pathol. **29**(3), 1596–1607 (2020).

[44]T. Dimon and G. D. Brown, *Anatomy of the Voice: An Illustrated Guide For Singers, Vocal Coaches, and Speech Therapists* (North Atlantic Books, Berkeley, CA, 2018).

[45]M. Döllinger, D. A. Berry, and S. Kniesburges, "Dynamic vocal fold parameters with changing adduction in *ex-vivo* hemilarynx experiments," J. Acoust. Soc. Am. **139**(5), 2372–2385 (2016).

[46]Z. Zhang, "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model," J. Acoust. Soc. Am. **139**(4), 1493–1507 (2016).

[47]Z. Zhang, "Interaction effects in laryngeal and respiratory control of the voice source and vocal fold contact pressure," J. Acoust. Soc. Am. **156**(6), 4326–4335 (2024).

[48]I. R. Titze, *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ, 1994).

[49]F. Unteregger, F. Honegger, S. Potthast, S. Zwicky, J. Schiwowa, and C. Storck, "3D analysis of the movements of the laryngeal cartilages during singing," Laryngoscope **127**(7), 1639–1643 (2017).

[50]Y. T. Hong, K. H. Hong, J.-P. J. Jun, and P. H. Hwang, "The effects of dynamic laryngeal movements on pitch control," Am. J. Otolaryngol. **36**(5), 660–665 (2015).

[51]C. Madill and D. D. Nguyen, "Impact of instructed laryngeal manipulation on acoustic measures of voice — Preliminary results," J. Voice **37**(5), 723.e1–723.e7 (2023).

[52]K. Honda, H. Hirai, S. Masaki, and Y. Shimada, "Role of vertical larynx movement and cervical lordosis in F0 control," Lang. Speech **42**(4), 401–411 (1999).

[53]W. R. Chen, D. H. Whalen, and M. K. Tiede, "A dual mechanism for intrinsic f0," J. Phon. **87**, 101063 (2021).

[54]D. H. Whalen and A. G. Levitt, "The universality of intrinsic F0 of vowels," J. Phon. **23**(3), 349–366 (1995).

[55]M. Echternach, L. Popeil, L. Traser, S. Wienhausen, and B. Richter, "Vocal tract shapes in different singing functions used in Musical Theater singing – A pilot study," J. Voice **28**, 653.e1–653.e7 (2014).

[56]M. Echternach, F. Burk, J. Kirsch, L. Traser, P. Birkholz, M. Burdumy, and B. Richter, "Articulatory and acoustic differences between lyric and dramatic singing in western classical music," J. Acoust. Soc. Am. **155**, 2659–2669 (2024).