# Investigation of resonance strategies of high pitch singing sopranos using dynamic three-dimensional magnetic resonance imaging

Marie Köberlein, Peter Birkholz, Michael Burdumy, et al.

Citation: The Journal of the Acoustical Society of America **150**, 4191 (2021); doi: 10.1121/10.0008903
View online: https://doi.org/10.1121/10.0008903
View Table of Contents: https://asa.scitation.org/toc/jas/150/6
Published by the Acoustical Society of America

JASA ARTICLE

# Investigation of resonance strategies of high pitch singing sopranos using dynamic three-dimensional magnetic resonance imaging

Marie Köberlein,[1,a)] Peter Birkholz,[2] Michael Burdumy,[3] Bernhard Richter,[1] Fabian Burk,[4] Louisa Traser,[1] and Matthias Echternach[5,b)]

[1]*Medical Faculty of the Albert-Ludwigs-University Freiburg, Freiburg Institute for Musicians' Medicine, University Medical Center Freiburg, University of Music Freiburg, Elsässer Straße 2m, 79110, Freiburg, Germany*

[2]*Institute of Acoustics and Speech Communication, Technische Universität Dresden, Germany*

[3]*Department of Medical Physics, Radiology, Freiburg University Medical Center, Germany*

[4]*Department of Otorhinolaryngology, Head and Neck Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[5]*Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, University Hospital, LMU Munich, Germany*

**ABSTRACT:**

Resonance-strategies with respect to vocal registers, i.e., frequency-ranges of uniform, demarcated voice quality, for the highest part of the female voice are still not completely understood. The first and second vocal tract resonances usually determine vowels. If the fundamental frequency exceeds the vowel-shaping resonance frequencies of speech, vocal tract resonances are tuned to voice source partials. It has not yet been clarified if such tuning is applicable for the entire voice-range, particularly for the top pitches. We investigated professional sopranos who regularly sing pitches above C6 (1047 Hz). Dynamic three-dimensional (3D) magnetic resonance imaging was used to calculate resonances for pitches from C5 (523 Hz) to C7 (2093 Hz) with different vowel configurations ([a:], [i:], [u:]), and different contexts (scales or octave jumps). A spectral analysis and an acoustic analysis of 3D-printed vocal tract models were conducted. The results suggest that there is no exclusive register-defining resonance-strategy. The intersection of fundamental frequency and first vocal tract resonance was not found to necessarily indicate a register shift. The articulators and the vocal tract resonances were either kept without significant adjustments, or the $f_{R1}$:$f_o$-tuning, wherein the first vocal tract resonance enhances the fundamental frequency, was applied until F6 (1396 Hz). An $f_{R2}$:$f_o$-tuning was not observed.

 https://doi.org/10.1121/10.0008903

(Received 29 July 2021; revised 12 October 2021; accepted 10 November 2021; published online 8 December 2021)

[Editor: Zhaoyan Zhang] Pages: 4191–4202

## I. INTRODUCTION

The underlying physiological principles of the production of the highest part of the female voice range are still not completely understood. For professional singing, several hypotheses previously tried to explain the existence of a separate register, i.e., a pitch region of homogenous sound quality, for fundamental frequencies ($f_o$) above 1000 Hz; this register is inconsistently called whistle register, flageolet, flute, Pfeifregister, M3, etc. (Herbst, 2021, 2020; Vormann-Sauer, 2017) Some of the hypotheses concern voice source modifications by laryngeal mechanisms, such as a "whistle mechanism" (van den Berg, 1963), vocal tract/voice source interactions (Herzel and Reuter, 1997), flageolet-like mechanisms where the oscillating parts of the vocal folds are shortened (Martienssen-Lohmann, 1993) and the modification of the airflow by oscillating vocal folds without complete closure using video stroboscopy or video kymography (Keilmann and Michek, 1993; Švec *et al.*, 2008). However, in a more recent study by Echternach *et al.* (2013) using an endoscopic highspeed camera with a rate of 20 000 frames per second, it was observed that the vocal folds continued oscillating up to the fundamental frequency of 1568 Hz (pitch G6) with complete closure of the glottis. Concerning the oscillation patterns of the vocal folds, Garnier *et al.* (2012) observed a decrease in the electroglottographical (EGG) amplitude during a possible transition to the highest register, and a rising larynx, as well as a stretched glottis during a glissando for the highest pitch range of one subject, using an endoscopic high-speed camera at a rate of 2000 fps. Additionally, the study observed changes in the open quotient (OQ) of the glottis around the possible register transition points, which could be either decreased and then increased around the transition, or the other way round. Pitch instability was mostly detected in the pitch range D5–F#5. The values of the sound pressure level decreased with increasing pitch after having reached a maximum in the range of A5-C6.

Other studies observed changes in the acoustic spectrum, such as overtones of less intensity for high pitches, resulting in a fluty sound (Garnier *et al.*, 2012, 2010; Story, 2015) or changes in the formant tuning strategies (Garnier *et al.*, 2010), such as a shift from $f_{R1}$: $f_o$ to $f_{R2}$: $f_o$

a)Electronic mail: marie.christine.koeberlein@uniklinik-freiburg.de, ORCID: 0000-0001-5827-0781.
b)ORCID: 0000-0003-0095-5360.

tuning, wherein a vocal tract resonance $f_{Rn}$ enhances a partial $nf_o$. Also, the rising of $f_o$ above $f_{R1}$ was observed and pointed out as a possible reason for a perceptual register change (Echternach *et al.*, 2015; Miller and Schutte, 1993).

Such acoustic phenomena are caused by the vocal tract shape and the resulting resonances and formants (e.g., Bresch and Narayanan, 2010; Joliveau *et al.*, 2004a; Sundberg, 2009). Garnier *et al.* (2010) figured out different tuning strategies for singing in the highest register and located a passaggio, i.e., the transition between two vocal registers, in the octave range between D#5 and D6.

The studies of Garnier *et al.* (2012) and Garnier *et al.* (2010) investigated subjects of all levels of voice education, including non-professional singers, *via* broadband excitation of the vocal tract and resulted in very inconsistent observations. They stand in contrast to the single-subject studies by Echternach *et al.* (2015) and Echternach *et al.* (2013), who located the acoustically perceived passaggio between D6 and E6, in accordance with the professional subject's own statement. The study by Echternach *et al.* (2015) combined dynamic two-dimensional (2D) magnetic resonance imaging (MRI) (24 fps) and static three-dimensional (3D) MRI with a recording time of 13 s per pitch. The 3D-models of the vocal tract were printed and their resonances were obtained *via* broadband excitation. The results did not show major differences or adjustments between the pitches of C6 and G6, which led to the hypotheses that no specific tuning strategy is applied for the highest register and that the rising of $f_o$ above $f_{R1}$ around E6 might result in a change of sound quality.

A study by Vos *et al.* (2018) tried to characterize the voice technique considering the vowels [a:], [i:], and [u:] in the high pitch range of a single subject, who was a mezzo-soprano with a top pitch of C6. The 2D- and 3D-MRI and acoustic excitation of the vocal tract were combined. The study showed that the vocal tract shape was similar for the vowels in the high range and that a $f_{R1:}\ f_o$ was applied up until C6. The study was not able to figure out linear relationships between the MRI data and the resonance tuning obtained from vocal tract excitation. In the mentioned study, the MRI measurements and broadband excitation could not be performed at the same time and were combined later. As Bresch and Narayanan (2010) suggest, sopranos might not use a generalizable technique; furthermore, it might be the case that one subject applies different techniques depending on the situation. Hence, it can be doubted that the measurements in the study of Vos *et al.* (2018) fit together reliably. Although the subject was a professional singer, she was a mezzo-soprano with an upper limit of C6 who might not be able to deliver data comparable to coloratura sopranos who often sing above C6, where the highest register is assumed to be located. The 3D-MRI measurements in the study were static and the subject was asked to sustain phonation for 16 s, which is unusually long.

Since it is quite hard even to characterize the sound of a whistle register uniformly, because untrained singers typically struggle in this seldom-used area, the present study concentrates on established, professional sopranos who have mastered the upper end of their voice range and regularly perform these pitches on stage. It can be assumed that these subjects have reliable voice control as well as conscious and effective strategies to use this very high pitch area, as they are used to sing on opera stages in many different kinds of positions and movements (see also Table I in Sec. II A).

Due to technical limitations, there has not yet been an establishment of a gold standard for the detection of vocal tract resonances: Purely acoustic analyses are not sufficient for showing vocal tract adjustments which have only barely detectable impact on the spectrum, in case no partial is near to stimulate a vocal tract resonance. In case of *in vivo* excitations with sine sweeps, where subjects must keep their vocal tract setting constant without singing, while sine waves are passing through the vocal tract, the immobility of unconscious and reflex-controlled parts of the throat cannot be monitored. In some broadband-excitation procedures where the subjects are singing at the same time no data about the vocal tract shape are gathered. In other broadband excitation procedures, the acoustic signal is brought into the vocal tract during vocal stops. For both cases, however, it cannot be ensured that the vocal folds are totally closed. In cases of a glottal opening, even the opening during the glottal cycle, an opening to subglottal spaces could possibly influence the resonance properties.

Two-dimensional MRI can only show a single sagittal slice of a volume, and, consequently, neglects the lateral expansion of the vocal tract. Static 3D-MRI is frequently used to depict or print vocal tract models for acoustic excitation, but requires sustained sounds and vocal tract configurations of more than 10 s (Birkholz *et al.*, 2020), which is doubly challenging as it is difficult for the highest pitch area to be sung in supine position. Burdumy (2016) and Burdumy *et al.* (2017) developed a method that allows for 3D-MRI volume images of the vocal tract at higher temporal resolutions. The presented sequence continuously samples k-space using a "stack-of-stars" trajectory with golden angle rotation in the "stars" plains and continuous incrementation in "stack" direction. Thereby, the k-space data windows used for image reconstruction can be arbitrarily chosen. Even though highly under-sampled, image quality is retained by using an iterative offline reconstruction method that employs compressed sensing theory. With a software tool programmed by Peter Birkholz (Echternach *et al.*, 2015), the resonances of the virtual vocal tract shapes obtained from the 3D-MRI-data can be calculated. This innovative method permits the observation of vocal tract adjustments and their acoustic impact in a realistic musical context, and not only for isolated pitches.

With regard to the resonance strategies applied in the whistle register, two hypotheses shall be investigated:

(1) The specific formant tuning ends with the crossing of $f_{R1}$ and $f_0$ (abbreviated notation: $f_{R1}/f_o$) and a constant vocal tract shape is maintained for higher pitches (Echternach *et al.*, 2015).
(2) The tuning strategy is intentionally changed, e.g., from $f_{R1:}\ f_o$ to $f_{R2:}\ f_o$ tuning through targeted changes in the articulators (Garnier *et al.*, 2010).

https://doi.org/10.1121/10.0008903

TABLE I. All subjects' classification according to the taxonomy of Bunch and Chapman (2000) and the performed tasks.

| Subject | Taxonomy according to Bunch-Chapman | | Performed scales | Performed octave jumps |
|---|---|---|---|---|
| 1 | 3.1a | Big City Opera Major Principal | C5–C6 | G5–G6: [a], [i] |
| | | | F5–F6 | |
| 2 | 4.1a / 4.5 | Touring Opera Major Principal / Concert, Oratorio, Recital | C5–C6 | |
| 3 | 3.1a | Big City Opera Major Principal | C5–C6 | |
| | | | D#5–D#6 | |
| 4 | 2.4 | International Concert, Oratorio, Recital, Chorus | C5–C6 | |
| | | | D5–D6 | |
| 5 | 3.1a | Big City Opera Major Principal | C5–C6 | |
| | | | D5–D6 | |
| 6 | 3.1a / 3.15ab* | Big City Opera Major Principal / *International acknowledged Radio Choir, Soloist and Chorister | C5–C6 | E5–E6: [a], [i], [u] |
| | | | D#5–D#6 | F5–F6: [a] |
| | | | | F#5–F#6: [a] |
| 7 | 4.5 / 6.2 | Touring Concert, Oratorio, Recital / Singing Teacher Freelancer | C5–C6 | |
| 8 | 7.2 / 4.1a | Full-time Voice Student University / Touring Opera Major Principal | C5–C6 | G5–G6: [a], [i] |
| | | | | A5–A6: [a] |
| | | | F5–F6 | H5–H6: [a] |
| | | | | C6–C7: [a] |
| 9 | 3.1a | Big City Opera Major Principal | C5–C6 | |
| | | | D#5–D#6 | |

Regarding the current level of knowledge, there is no uniform picture and there are also discrepancies between the results for the same subject for different types of measurements and a combination of methods does not necessarily deliver combinable data (Echternach *et al.*, 2018; Vos *et al.*, 2018). In this sense, the present study aims to add further elements to the overall picture.

## II. MATERIAL AND METHODS

This study was approved by the local ethical committee (Medical Ethics Committee of the University of Freiburg, 206/09). The nomenclature and acronyms for voice acoustics in this paper follow the general consensus summarized by Titze (2016).

### A. Subjects and tasks

Nine professional soprano subjects were asked to sing a scale from C5 (ca. 528 Hz) to C6 (ca. 1047 Hz), which was subsequently transposed, so that the highest pitch of the scale would meet the highest pitch of their individual range in the form of the day (Fig. 1). Each pitch had to last two seconds, guided by an optical metronome, and each scale was repeated with three different vowel configurations: [a:], [i:], and [u:]. The sopranos were asked to sing at a comfortable level of loudness. Three of the subjects were additionally asked to do several octave jumps to their highest possible pitch, partly on different vowels. Table I gives an overview of all subjects' taxonomy according to Bunch and Chapman (2000) and the performed tasks.

### B. Dynamic 3D-MRI data acquisition

The magnetic resonance imaging data were collected with a SIEMENS 3 T MAGNETOM Trio scanner (Siemens Healthcare GmbH, Erlangen, Germany). The audio signal was recorded simultaneously with the audio-feedback system (MR confon GmbH, Magdeburg, Germany).

The MRI sequence parameters are: radio frequency-spoiled gradient echo sequence Stack-of-Stars: repetition time = 2.9 ms, echo time = 1.4 ms, flip angle = 6°, Field of view $200 \times 200 \times 62\,\text{mm}^3$, voxel size $1.6 \times 1.6 \times 1.3\,\text{mm}^3$, bandwidth 1500 Hz/pixel. The accelerated 3D recording technique used with a selective "stack-of-stars" trajectory requires a stable vocal tract configuration in time windows of at least 1.8 s. In our case, the time windows were manually set to start after pitch changes. The suggested 1.8 ms

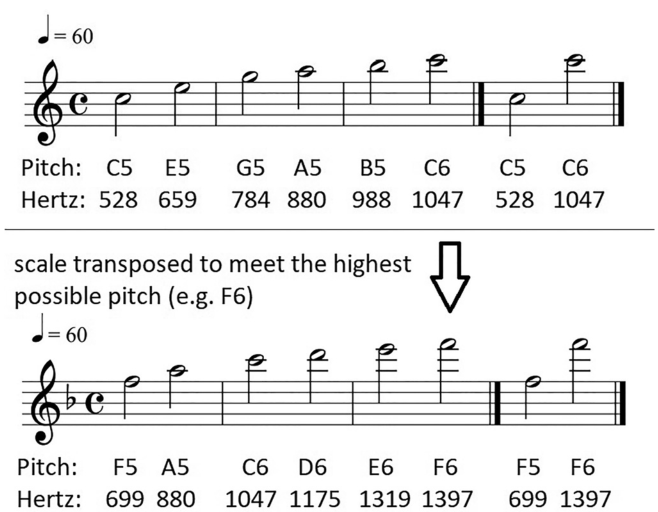


FIG. 1. The scale C5–C6 (top left) was sung by all subjects on the vowels [a:], [i:], [u:] and repeated transposed (bottom left), so that the highest pitch of the scale would coincide with the highest pitch of their individual voice range. The octave jump on the right side was performed by three subjects from different starting pitches in the 5th octave and on the vowels [a:] and [i:]. Each tone lasted two seconds.

coincide with the time the constant pitches were held in our study (2 s). More details on the MRI sequence and reconstruction can be found in Burdumy *et al.* (2017). From the reconstructed data the vocal tracts for the pitches C5, A5, C6, and every pitch above C6 were segmented with a semi-automatic toolchain programmed by the authors in MeVisLab (Version 2.8.2, 26-10-2016). At first, a bounding box around the vocal tract cavities is manually defined. Next, gray values were defined, so that the program could distinguish between tissue and air. Then, these air-tissue boundaries were found by an automatic constrained region-growing. Mistakes occurring for reasons of artefacts were corrected manually. Questionable points were discussed in the group. Screenshots of the segmentation process are shown in Fig. 2. The segmented vocal tracts were exported as 3D mesh tubes and opened at both ends using Meshmixer (Autodesk Inc., Version 3.5.474).

### C. Derivation of vocal tract transfer function and resonances

The area functions and transfer functions of the mesh tubes were calculated in the application vocal tract transfer function (VTTF) by Peter Birkholz (described in Echternach *et al.*, 2015). A vocal tract centreline was automatically calculated from the glottis to the corners of the mouth level. Then the tube was sliced into 64 segments, whose cross-sectional areas were measured to obtain the vocal tract area function. The volume velocity transfer function between the glottis and the lips was calculated based on this area function using the one-dimensional (1D) transmission-line model described in Birkholz and Jackel (2004). The calculation took into account yielding walls (with the parameters for the relaxed cheek provided by Ishizaka *et al.* (1975), as well as energy losses due to sound radiation from the mouth opening, viscous friction, and heat conduction. The radiation loss was an approximation of the case of a vibrating piston in an infinite wall, and the glottis was assumed to be closed (no glottal losses). The peaks of the magnitude of the transfer function mark the vocal tract resonances $f_{Rn}$ (Fig. 3). To check the accuracy of the measured values, all instances of C5 and random instances of higher pitches were reproduced in the software Madde Synthesizer (Svante Granqvist, Karolinska, Stockholm, Sweden) and perceived as representative when compared with the original audio recording by two experts, i.e., (1) university professor in voice research and singing, experience >45 years and (2) voice researcher and professional singer, experience >10 years.

The software developers meanwhile recommended lengthening the vocal tract tube by half the depth of the corner of the mouth notch with the circumference of the corner of the mouth plane (Birkholz and Venus, 2017). The attachment of the tube was performed on all vocal tract models. There were no significant deviations in the comparison of the first two vocal tract resonances; the Pearson correlation coefficient was $r = 0.98$ for $f_{R1,2}$, $r = 0.91$ for $f_{R3}$ and $r = 0.76$ for $f_{R4}$, where $r = 1$ means perfect agreement.

The first four peaks of the transfer functions, which mark the vocal tract resonances ($f_{R1\text{–}4}$), were extracted and compared separately for each vowel, as well as their medians. Since the Hertz-value of a pitch can vary between different subjects, the pitches were normalized and set in ratio to the spectrum partials using the formula $\boldsymbol{TQ} = \boldsymbol{f_{Rn}}/\boldsymbol{f_o}$, where TQ stands for Tuning Quotient. The resulting number in the place in front of the decimal point is the partial and the decimal places the normalized distance to the partial. It follows $\boldsymbol{fo} = 1$.

### D. Spectral analysis

In order to check the validity, a sample spectral analysis was carried out for four pitches sung by subject 1, i.e., C5, A5, C6, and F6 on vowel [a:]. Herein, the noise-reduced audio file from the original MR measurement, as well as the calculated transfer function and its acoustic reproduction, were plotted against each other (for illustration, see Fig. 13 in the result Sec. III C). Noise Reduction was performed in Audacity, version 2.3.0 (Noise reduction standard settings: 30 dB, sensitivity: 12, frequency smoothing: 2). In this step, the pure noise is captured in zones without voice and then subtracted from the whole audio track with the expectation of not influencing the recorded voice signal. Acoustic reproduction was executed using the peak-values of the transfer function as formant settings in Madde Synthesizer (tilt: –12 dB), recorded with the microphone Realtek High Definition SST combined with Audacity in a silent room.

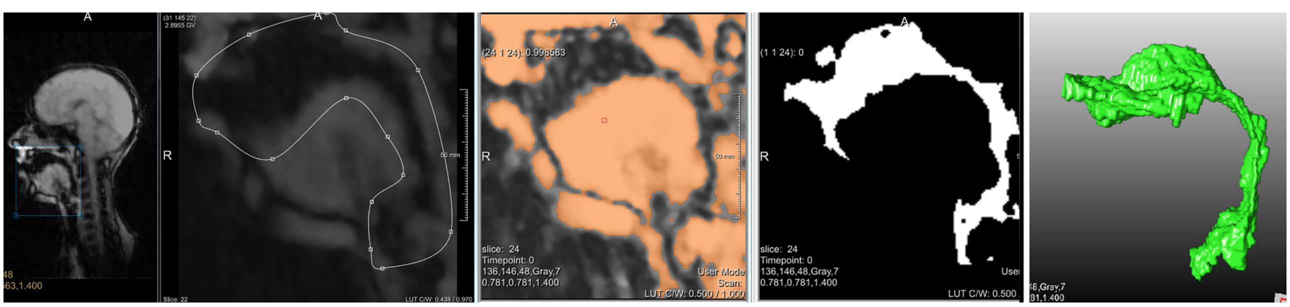

FIG. 2. (Color online) Screenshots of the semi-automatic segmentation process of a 3D vocal tract tube, demonstrating the sagittal plane, from the left to the right: The original MRI picture of the skull; followed by the circumscribing of the area where the vocal tract should be detected within the white line; the third panel shows the definition of gray values for the tissues and the fourth panel shows the derived negative image of the vocal tract cavities; finally, a 3D model of a segmented vocal tract tube after manual correction.

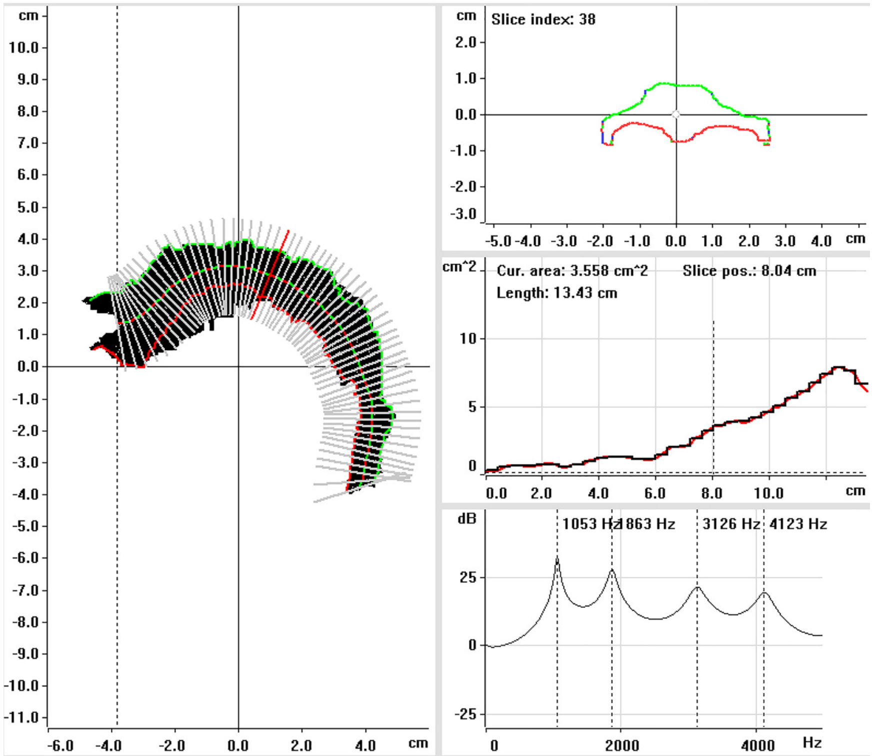


FIG. 3. (Color online) Screenshot of the application VTTF (Echternach *et al.*, 2015), which was used to measure the cross-sectional areas of the 64 slices (upper right side) of the vocal tract tube (left side) in order to gain the area function (middle right side) from which the transfer function (lower right side) is derived.

The spectra were obtained in Audacity (version 2.3.0, window type: Hanning, window size: 1024).

### E. Acoustic analysis of two 3D-printed models

To further validate the results, two vocal tract models, i.e., C5 and F6 on vowel [a:] of subject 1, were printed out of polylactic acid (wall thickness 4 mm) and excited in a sound-proofed chamber with logarithmic sine sweeps (100–10 000 Hz) as described by Birkholz *et al.* (2020) and Fleischer *et al.* (2018) to obtain their volume velocity transfer functions (VVTF). The measurement procedure is illustrated in the video Birkholz *et al.* (2021). Teeth were not added, since the virtual vocal tracts that were used for resonance calculation in the application VTTF did not have teeth either. The omission of teeth is due to a study by Traser *et al.* (2017), which found the acoustic relevance of teeth in vocal tract models to be insignificant for higher pitches.

## III. RESULTS

### A. Scales

#### *1. Area functions*

The area functions show that the vocal tracts were generally shortened successively with increasing $f_o$, even though the highest pitch does not necessarily have the shortest length, see for example S8 [a:] and [i:] in Fig. 4.

Further, the cross-sectional areas of the oral cavity increased with increasing pitches. While there were vowel-related differences for the lower pitches C5 and A5, the vocal tract shapes became similar for the higher pitches of all vowels, even though for S1 [i:] the mouth shape was kept in order to form the vowel.

#### *2. First vocal tract resonance $f_{R1}$*

The medians of the $f_{R1}$ of the different vowels, depicted in Fig. 5, converged between A5 and C6 and met the fundamental frequency $f_o$ at F6 (1397 Hz), due to the lower slope value compared to $f_o$. For the resonance–partial relations shown in Fig. 6, a similar picture emerged, which shows large differences for C5 and an overall approximation towards $f_o$ for higher frequencies.

#### *3. Second vocal tract resonance $f_{R2}$*

The medians of the $f_{R2}$ of the different vowels, depicted in Fig. 7, approached $2f_o$ between A5 and C6. The development was slightly different for each vowel: For [a:], the median fell and stayed under $2f_o$ from between A5 and C6; for [i:], $f_{R2}$ approached $2f_o$ but stayed parallel above until it fell below after D6; for [u:] $f_{R2}$ was below $2f_o$ from A5 on. A break of the $f_{R2}$ frequencies was observed for the $f_{R2}$ of [a:] and [i:] on pitch D#6. Also, a large individual spread of the resonance frequencies can be seen in Fig. 7. The $f_{R2}$ frequencies approached each other at 1397 Hz. In the normalized relations of $f_{R2}$ to the partials in Fig. 8, a clear vowel difference can be seen for C5 and a fall below the second partial from D6-D#6 for [a:], D#6 for [i:], and D6 for [u:].

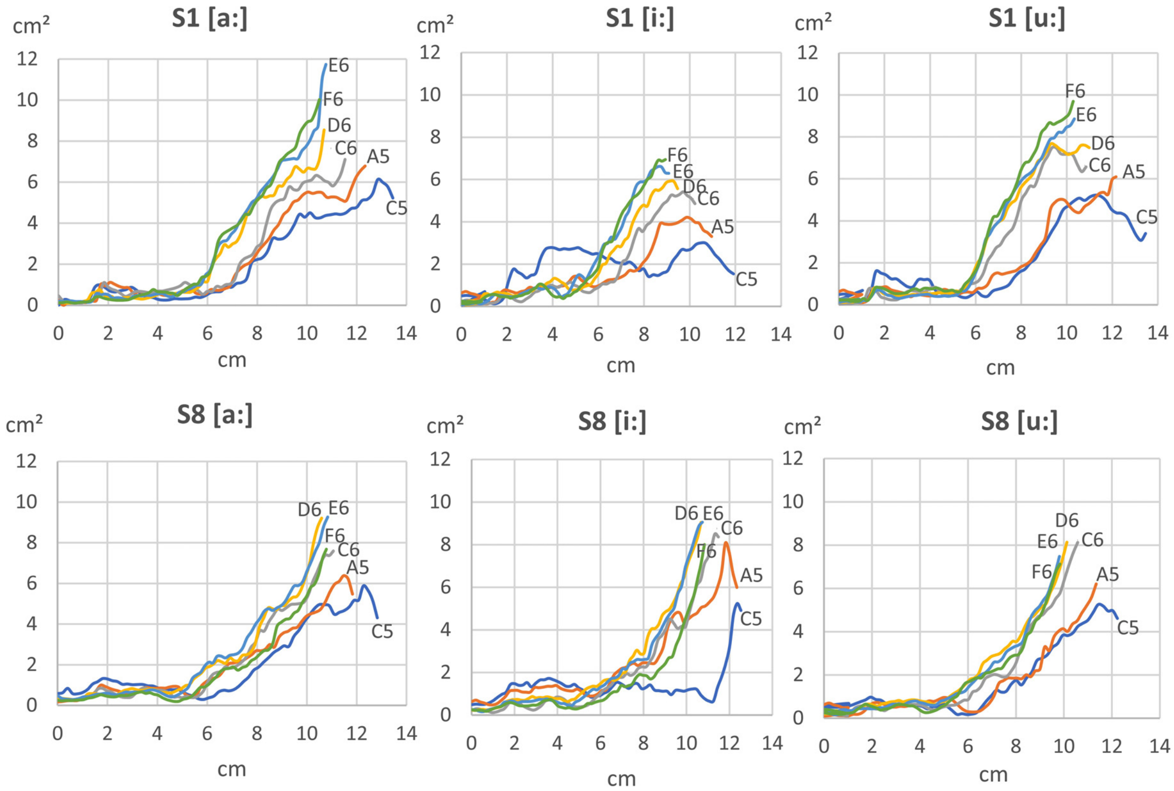


FIG. 4. (Color online) Shorter vocal tracts, larger mouth opening, and vowel assimilation for increasing pitches shown by the area functions of all pitch-related vocal tract configurations from the scales of subjects 1 and 8 for the vowels [a:], [i:], and [u:].

#### 4. Third and fourth vocal tract resonances $f_{R3}$ and $f_{R4}$

The medians of the $f_{R3}$ of all measured vowels converged on pitch D6, approaching $3f_o$ from above. On D#6 $f_{R3}$ of [i:] and [u:] were slightly below $3f_o$, climbing above the partial again for E6.The medians of the $f_{R4}$ of all measured vowels converged on the pitch A5, where they were above $5f_o$. On C6 the medians of $f_{R4}$ were right in between $4f_o$ and $5f_o$, but below $4f_o$ on D#6, rising again closer to $4f_o$ or even exceeding this partial for vowel [i:] on E6 (Fig. 9).

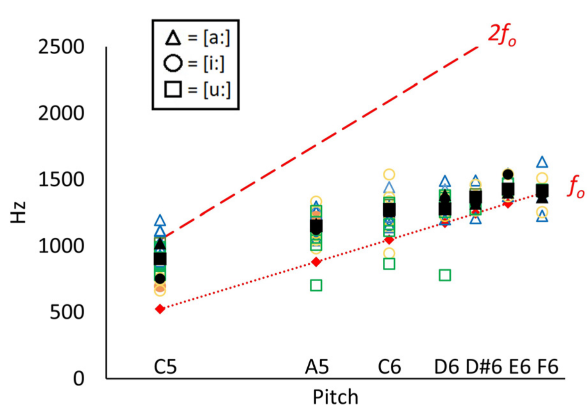


FIG. 5. (Color online) Medians of first vocal tract resonances are converging with increasing pitch: The $f_{R1}$ of all subjects (empty symbols) and vowels and their medians (filled symbols) in comparison to $f_o$ (dotted line) and $2f_o$ (dashed line). Vowel symbols: Triangle, [a:]; circle, [i:]; square, [u:]. Connecting lines between the pitches mark the vowels' medians.

### B. Octave jumps

For the octave jumps of S8 on the vowel [a:] (Fig. 10) there was no distinct adjustment up until C7, the resonance frequencies stayed within a range of 379 Hz ($f_{R1}$: 1157–1536 Hz) and 391 Hz ($f_{R2}$: 2145–2536 Hz), respectively. The measured resonances from the same subject's scales were also located in this range.

As shown in Fig. 11, the octave jumps of subject 8 on G5-G6 and A5-A6 on the vowels [a:] and [i:] also stayed in the above-mentioned range and showed no distinct adjustment, neither vowel-specific, nor octave-specific. The absence of adjustment was also observed for subject 1's octave jumps on G5–G6 on the vowels [a:] and [i:], though the resonances were placed differently ($f_{R1}$: 982–1359 Hz and $f_{R2}$: 1833–2082 Hz). The octave jumps of subject 6 started on E5, where there was a clear vowel distinction between [a:], [i:], and [u:], and each setting was kept for the higher octave. For a row of jumps on [a:] (E5–E6, F5–F6, F#5–F#6), the resonances stayed in the same range ($f_{R1}$: 1143–1298 Hz and $f_{R2}$: 1718–1960 Hz, see Fig. 13).

### C. Spectral analysis

As shown in Fig. 12, the spectral peaks of the original audio signal and the reproduction using the transfer function showed good agreement, positioned at the partials. The transfer function itself matched the spectra or the harmonics only partially. The spectrum of the original F6 showed peaks in the middle between the partials, which

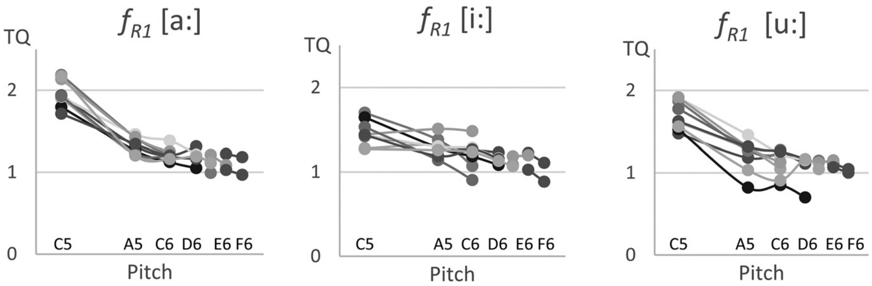


FIG. 6. The Tuning Quotients (for the definition of TQ see Sec. II C) of the first vocal tract resonances of the vowels [a:], [i:], and [u:] for all subjects (different shades of gray) tend towards the first partial with increasing pitch.

were found neither in the transfer function nor in its reproduction.

### D. Acoustic analysis of two 3D-printed models

The vocal tract resonances determined by the broadband excitation using sine-sweeps showed a deviation of 12.9% (concerning the position of the peaks in Hz) on average compared to the calculated transfer function (Fig. 13). The curves above C5 showed a similar structure, where $f_{R1}$ differed by 133.14 Hz and $f_{R2}$ by 189.77 Hz. The curves for F6 were only similar in the fact that, above the respective $f_{R1}$ (difference of 290.1 Hz), all frequencies were reinforced, while, otherwise, peaks and valleys even tended to be opposite. The calculated transfer function met the second and third partial of C5 ($f_0 \approx 523$ Hz), whereas the excited VVTF met the first partial of F6.

## IV. DISCUSSION

### A. Scales

#### 1. Vowel shaping

The evaluation of the pitch C5 (523 Hz) was chosen because many vowel differences can still be shaped at this pitch, therefore it could represent a comparative value for higher pitches. Clear vowel differences on C5 were confirmed in the area functions and, consequently, the vocal tract resonances $f_{R1\text{–}2}$. However, the results do not necessarily correspond to the vowel formant expectations from speech: While the first vowel formant for [a:] ≈ 750 Hz could still be set at a fundamental frequency of C5 (523 Hz), albeit without an obvious partial, the first resonance in the present results is already at a median of 1021 Hz, which suggests a $f_{R1}$: $2f_o$ tuning. The vowel formants in speech of [i:] ≈ 330 Hz and [u:] ≈ 350 Hz (Pätzold and Simpson, 1997) are already exceeded at C5 and are set significantly higher by the subjects: the medians are [i:]: 752 Hz and [u:]: 903 Hz. Since these resonance frequencies are distant from the partials, it can be assumed that the first resonance neither tunes a partial nor shapes a distinct vowel to a great amount, except in the sense of an approximation towards [a:].

For the second resonances of C5, the placements of [a:] and [i:] correspond more to the expectations from speech, i.e., [a:] ≈ 1460 Hz and [i:] ≈ 2300 Hz. The medians in the present study were [a:]: 1640 Hz and [i:]: 2239 Hz. For the vowel [u:] in speech, the expectation would be 1050 Hz, but the median of the sung vowels was at 1591 Hz. Accordingly, an acoustic approximation of the vowels [a:] and [u:] towards [ɔ], in the sense of vowel equalization (Vos *et al.*, 2018), was confirmed. The phenomenon that all vowels become more and more similar for increasing pitches was confirmed by the area functions and consequently, the calculated vocal tract resonances for all measured vowels, as well as the application of a $f_{R1}$: $f_o$ tuning (Joliveau *et al.*, 2004b; Sundberg, 1977; Vos *et al.*, 2018), reached through a shortening of the vocal tract and a wider mouth cavity.

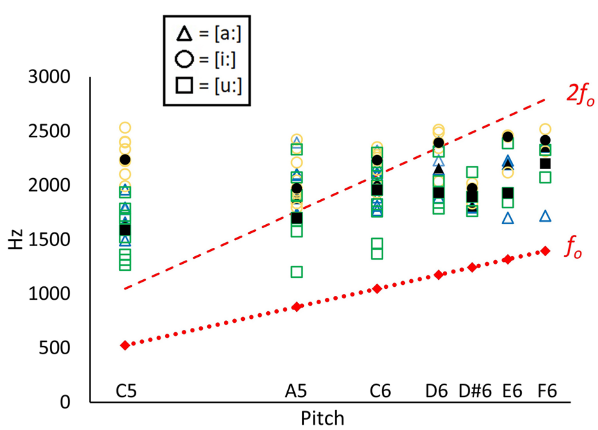


FIG. 7. (Color online) Medians of the second vocal tract resonances crossing the second partial: The $f_{R2}$ of all subjects (empty symbols) and vowels and their medians (filled symbols) in comparison to $f_o$ (dotted line) and $2f_o$ (dashed line). Triangle, [a:]; circle, [i:]; square, [u:].

#### 2. First and second vocal tract resonances

In the scales, the same pattern of approximation of $f_{R1}$ towards the rising $f_o$ occurred for all measured vowels, which can be interpreted as $f_{R1}$: $f_o$ tuning and which was partly maintained until F6. This corresponds to one of the observations made by Garnier *et al.* (2010). It does, however, not correspond to the hypothesis that the crossing of these two frequencies or the end of formant tuning marks a register transition into a whistle register (Echternach *et al.*, 2015). An assessment of whether an $f_{R1}$: $f_o$ tuning actually takes place is made more difficult because, based on the results of the additional tests and the existing limitations of the method, it is to be expected that the position of the calculated resonance frequencies might be imprecise to a small degree (see Sec. IV E). According to Hanna *et al.* (2016), formant bandwidths in women's speech have a range of 70–90 Hz for resonances up to 4 kHz, derived from vowel

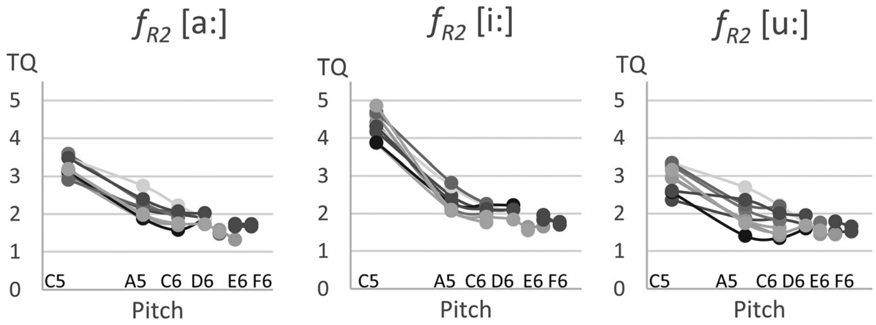


FIG. 8. Tuning quotients (for the definition of TQ see Sec. II C) of the second vocal tract resonances of the vowels [a:], [i:], and [u:] for all subjects (different shades of gray) show a crossing of the second partial around C6–D6.

[ɜ:] and recommended especially for closed vowels as [i:] and [u:]. In a study by Hawks and Miller (1995) investigating formants in speech for different vowels, the first three formants' bandwidths range from approximately 40 to 130 Hz for the vowels [a:], [i:], and [u:]. Titze (2004) described an interaction between resonances and partials which were 100 Hz apart in a study with a vocal tract model of male dimensions. However, these studies investigated only very low fundamental frequencies and speech circumstances. To our knowledge, values for formant bandwidths for the high fundamental frequency range in female singing voices have not yet been investigated, and might differ due to different pressure ratios and vowel unspecificity.

The calculated resonances of $f_{R2}$ show, especially for the setting of the vowel [i:], indications of an $f_{R2}$: $2f_o$ tuning up to D6, as also described by Garnier (2010). In contrast, for [a:] and [u:] $f_{R2}$ falls below $2f_o$. However, the large spread of individual data does not allow a clear localization. It can be said that from D#6 on all $f_{R2}$ are below $2f_o$. Echternach *et al.* (2015) observed that no partial was enhanced by $f_{R2}$ anymore between C6 and G6. The perceptual and proprioceptive register transition in the same study was located between D6 and E6, which is why a new assumption arises from the data of the present work that the perceived passaggio to the whistle register could be related to an $f_{R2}/2f_o$ crossing. It should be noted, however, that D#6 was the top pitch for some subjects and that at this level in an upward sung line, fatigue could already have an impact. A transition of the formant tuning strategy to an $f_{R2}$: $f_o$ tuning, as described by Garnier *et al.* (2010), could not be observed in the presented ascending scales.

### 3. Third and fourth vocal tract resonances

No $f_o$ related changes of $f_{R3}$ and $f_{R4}$ were observed. Thus, with rising $f_o$ and its harmonics, collocations that can be interpreted as $f_{R3}$: $4f_o$ and $f_{R4}$: $5f_o$ tunings automatically became $f_{R3}$: $3f_o$ and $f_{R4}$: $4f_o$ tunings, respectively, from above C6. Further, no vowel-specific differences could be observed.

## B. Octave jumps

In the octave jumps the case of $f_{R1}$ being exceeded by $f_o$ occurred regularly, as well as $f_{R2}$ being exceeded by $2f_o$. The octave jumps suggest that there is no specific adjustment of the vocal tract shape for the highest part of the range, and that the setting is taken over from what had been in place before. This is in agreement with Echternach *et al.* (2015), and further, it matches the comments of all the subjects concerning their adjustments for singing above C6, stating that there was no specific change in their vocal tracts.

Story (2015) explains the flute-like sound of falsetto with the fact that the acoustic energy is concentrated on the first few partials. Considering the great distance of the partials from C6 upwards and the particularly sensitive acoustic perception of extremely high frequencies from 1000 Hz

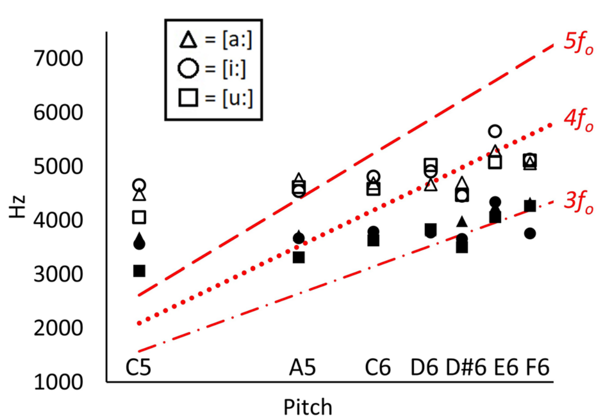


FIG. 9. (Color online) Third and fourth vocal tract resonances crossing several partials: The medians of $f_{R3}$ (filled symbols) and $f_{R4}$ (empty symbols) of all vowels in comparison to $3f_o$ (dashed-dotted line), $4f_o$ (dotted), and $5f_o$ (dashed). Vowel symbols: Triangle, [a:]; circle, [i:]; square, [u:].

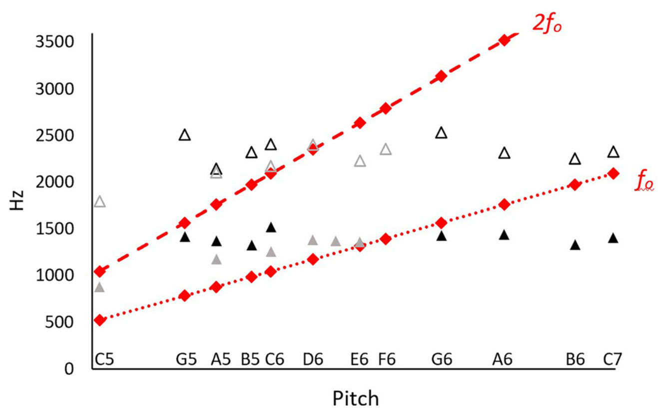


FIG. 10. (Color online) Rather constant configuration despite rising pitches: Vocal tract resonances $f_{R1}$ (black filled triangles) and $f_{R2}$ (black empty triangles) for the octave jumps of subject 8 on the vowel [a:] in comparison to $f_o$ (dotted line) and $2f_o$ (dashed line). The gray triangles are vocal tract resonances from the same subject's scales.

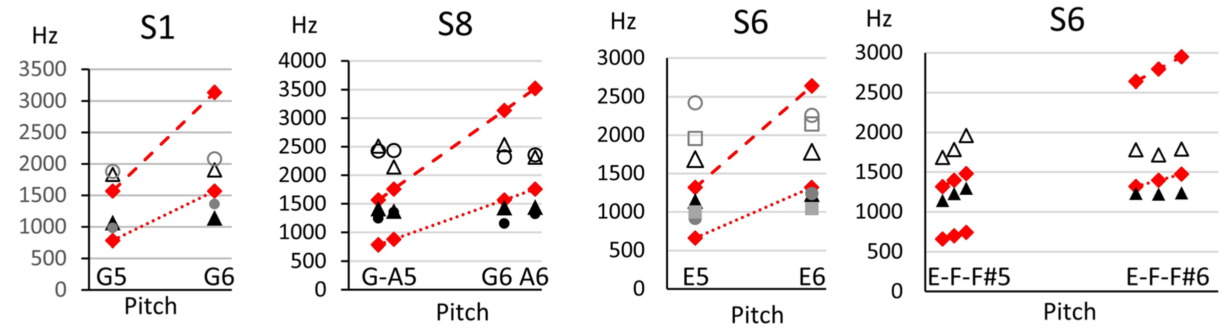


FIG. 11. (Color online) Rather constant configuration despite rising pitches: First and second vocal tract resonances ($f_{R1}$, filled symbols; $f_{R2}$, empty symbols) of the octave jumps of subjects 1, 6, and 8 on the vowels [a:] (triangles), [i:] (circles), and [u:] (squares). Rhombi mark $f_o$ (dotted line) and $2f_o$ (dashed line).

upwards, as well as the observation that the vocal tract is not changed intentionally, it can be imagined that there is no targeted tuning in the whistle register and also the whistling sound in female voices is the result of the pure partial tones. This is in contrast to the statement by Garnier *et al.* (2010) that changes in the settings of the articulators would have to take place, which would lead to adjustments in the tuning and, thus, to a flute-like sound.

### C. Spectral analysis

The diagrams in Fig. 12 show that the spectral peaks of the original audio signal and the reproduction are matching well in the sense that spectral peaks occur at the partial tones. The cause for the spectral peaks roughly in the middle between the partial tones of F6 in the original signal remains unexplained. Echternach *et al.* (2015) and Titze (1994) describe a noise between the partial tones which was so strong that formants could be recognized in it. The peak just above 2000 Hz could therefore be explained by the fact that the setting of $f_{R2}$ was maintained from C6, but there is no explanation for the peak at approximately 3500 Hz. It is conceivable that there is another acoustic phenomenon, which has not yet been adequately addressed. Noise could be explained by the register theories of air turbulence and incomplete vocal closure, which is, however, contradicted by the finding of complete vocal fold closure up until G6 by Echternach *et al.* (2013). In addition, there was no obvious broad band noise in the spectrum. It could also be speculated that the phenomenon is due to period doubling at the source.

Another aspect worth considering for the deviations between the spectra and the transfer function could be nonlinear interactions, which are not included in the transfer function and the reproduction in the Madde software, but which could influence the sound source. In the present study, acoustic resonances of the subglottic system could not be taken into account. According to Titze (1994), however, there is a hypothesis with regard to the falsetto of male voices that these subglottic resonances are involved in the occurrence of register breaks. In a study on various

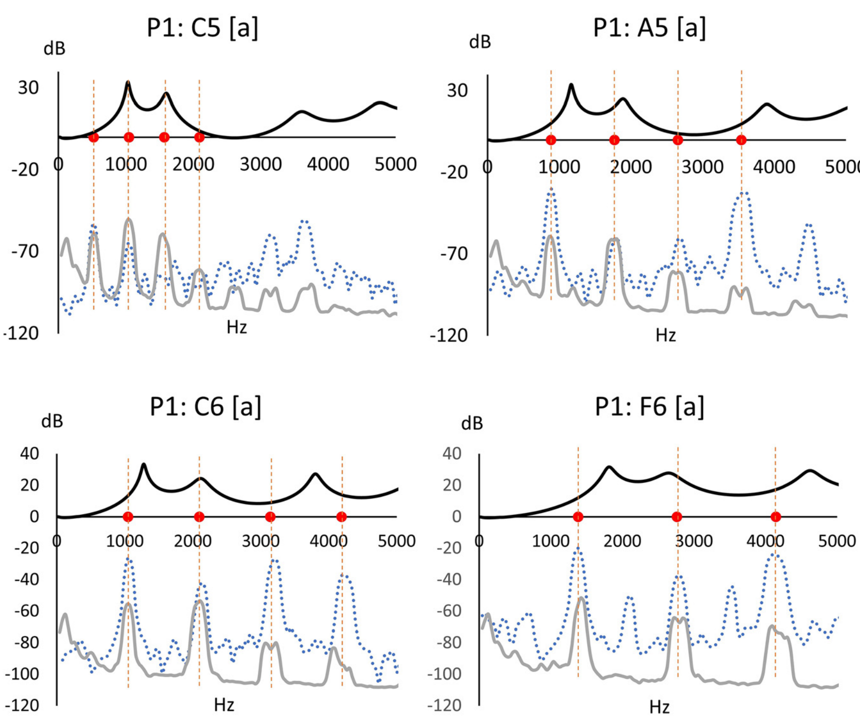


FIG. 12. (Color online) Diagrams of the spectral analyses of the pitches C5, A5, C6, and F6 on [a:] of subject 1. The dots on the $X$ axis and the vertical dashed lines mark the positions of the partials, the solid curve above the $X$ axis shows the transfer function calculated in VTTF. The dotted line represents the spectrum of the original recording during magnetic resonance measurement, and the gray solid line below shows the acoustic reproduction of the transfer function values in Madde Synthesizer, recorded with the computer microphone (Realtek High Definition Smart Sound Technology (SST)) and Audacity in a silent room.

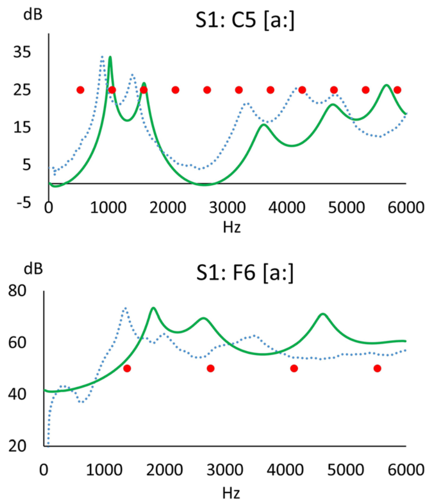


FIG. 13. (Color online) Comparison of the transfer functions of the pitches C5 (above) and F6 (below) obtained by excitation by sine-sweeps (dotted line) and calculation (solid line). Dots mark the position of the partials.

analytical approaches to determine resonance, it was observed that the coupling of a trachea and lung model had effects on the transfer function (Echternach *et al.*, 2018).

### D. Volume velocity transfer function of the printed models

The deviation of the peaks in the transfer functions by means of excited vocal tract models or numerical calculation of an average of 12.91% is considerable, but does not allow a determination of a "more correct" method, since there is no gold standard. Although there has been a comparatively extensive test of vocal tract model excitation, the question arises whether the hard polylactic acid (PLA) material of the 3D-printed models represents the impedance properties of the real vocal tract walls precisely enough. A recent study by Birkholz *et al.* (2022) indicates that softer walls shift $f_{R1}$ to slightly higher values, compared to harder materials, which could partly explain the different transfer functions. Additionally, the calculated frequencies might appear slightly higher because the transfer functions calculated with the software VTTF were based on area functions that assume the end of the vocal tract at the corners of the lips. This somewhat underestimates the length of the vocal tract. According to Birkholz and Venus (2017), a better approximation of the position of the "termination plane" of the vocal tract is at a position halfway between the corners of the lips and the most anterior points of the lips, i.e., in the middle of the notch formed by the lips. A longer tube lowers all resonance frequencies, and an enlarged radiation surface in the area of the lip notch increases losses that lead to increased formant bandwidths. Moreover, the transfer functions of the 3D-printed models were measured without baffling (to consider the acoustic effect of the face), while the simulated transfer functions assumed a radiation impedance of a piston in an infinite baffle. For the big mouth openings of the considered vocal tract shapes, this factor may also contribute to the observed differences (Dalmont *et al.*, 2001).

Further tests with and without teeth, with different wall materials, and with a baffling of the 3D-printed models would be required in order to determine what causes the deviations and how they are to be classified.

Regarding a specific partial tone enhancement by vocal tract resonances, the calculated transfer function seems to fit the lower sample tone C5, while the excited transfer function fits better for the extremely high tone F6.

Nevertheless, the data of the print model also indicate that, at least in this subject, $f_{R1}$ is increased together with $f_o$ and no crossing has taken place until F6. Furthermore, it can be interpreted that in the extremely high ambit, no specific spectrum partial is enhanced apart from the fundamental frequency, or that at the same time all overtones are enhanced and no frequency range is attenuated anymore.

### E. Limitations

The data set contains only a limited number of subjects, so no statistical analysis was meaningful. The spread of the resonances was relatively large, e.g., above the tone C6 on the vowel [a:] the particular resonances of $f_{R2}$ extend over a range of 692 Hz. Previous studies already mentioned the possibility that sopranos did not use a generalizable tuning strategy for singing extremely high (e.g., Bresch and Narayanan, 2010; Garnier *et al.*, 2010). A larger data set is needed to identify patterns.

It must be remembered that singing extremely high, ascending tones at the slow pace of two seconds per pitch while motionless and supine is an enormous challenge even for trained singers. It cannot be ruled out that this might be a reason why some subjects have not been able to achieve the top notes that they could sing under normal circumstances. Furthermore, it cannot be ruled out that the loud noise of the MRI evokes a Lombard effect, namely, that the phonation technique changes if one can hear oneself less and receives little acoustic feedback (Lombard, 1911). However, the use of the headset attenuates the MRI noise and it can be assumed that the professional activity of the singers guarantees a stable singing technique regardless of the background noise, and that there is even a certain habit of standing next to an orchestra or in a choir, resulting in only reduced self-perception, mainly weighted to the inner sound conduction.

The effect of the supine position on phonation has already been the subject of various MRI studies. Perry *et al.* (2014) describe an insignificant gravity effect on the velum, especially for [i]. The slightest modification of the vocal tract can, however, have a major effect on resonances, and the velum, in particular, has a decisive influence on the inertia of the entire vocal tract in its sluice function between the throat and nasal spaces (Sundberg *et al.*, 2015). In vocal education and practice, the velum is particularly important for high notes, as it determines the nasalance components and helps stabilize phonation in frequency regions in which instabilities frequently occur (Echternach *et al.*, 2021b; Havel *et al.*, 2021; Havel *et al.*, 2016; Havel *et al.*, 2014). The supine position may also have affected the height of the larynx and the protrusion of the jaw. However, the larynx rises for singing in the highest registers and in a comparative study with professional tenors regarding supine or upright position, the supine position was without any significant relation to register or pitch (Traser *et al.*, 2013). Another consideration on this topic is that the entire vocal apparatus is under greater tension for higher fundamental frequencies and therefore the gravitational effects of the lying position have less of an effect than would be the case with lower fundamental frequencies. Even with consideration of these problems, we cannot ultimately determine the reasons why some subjects were not able to reach their usual top pitches, much less as they might differ from person to person. However, in order to ensure the meaningfulness of the measurements, we only took into account the pitches that were sung in a reliable quality depending on the form of the day of each participant.

The low time resolution of 1.8 s for a 3D vocal tract resulted in image noise, which is why in some of the cases the boundaries between tissue and cavity were depicted imprecisely but still detectably during the manual segmentation process. The method is a trade-off, especially with regard to the imaging of the glottic surface, as it opens and closes approximately 2100 times in two seconds at pitch C6. For this reason, no statement can be made concerning the configuration of the glottis, which was, however, also not in the focus of this study.

There are several possibilities for imprecisions in the calculated resonances and their bandwidths with the software VTTF:

No teeth were implemented in the models. However, Traser *et al.* (2017) found only a slight influence of the teeth on the sound properties for higher pitches and wide-open mouth, as well as the consequent absence of lateral cavities.

The VTTF program does not consider the influence of subglottic structures and possible side branches such as the piriform sinus, nor the nasal coupling nor the specific 3D radiating surface at the mouth opening. The latter has already been addressed in Sec. IV D, where the preliminary experiment with extended mouth tubes showed no significant differences.

Even though Honda *et al.* (2010) showed effects of the piriform sinus and the vallecula on acoustic spectra above 2 kHz, these side cavities collapsed in the upper pitch area of all subjects of this study, and thus it could be assumed that they can be neglected in the resonance calculations.

As mentioned earlier, nasalance has an influence on the inertia of the vocal tract and is used specifically in register transitions (Echternach *et al.*, 2021b; Havel *et al.*, 2021; Sundberg *et al.*, 2007). It is a deficiency that in this calculation model there is no possibility of coupling/uncoupling the nasal cavities or calculating the opening width to the nasopharynx.

The 3D magnetic resonance imaging has the great advantage that vocal tract configurations can be represented in their vocal context, which contributes to a differentiated assessment of the results. This becomes clear in the differences between the results of the scales and the octave leaps. However, rapid adjustment processes between the static tones are not visible. Especially with regard to the topic of non-linear source-filter interactions to stabilize the voice production in register transitions (Echternach *et al.*, 2017; Echternach *et al.*, 2021a; Titze *et al.*, 2008; Titze, 2008), the knowledge of what happens between two pitches, where intersections of partials and vocal tract resonances might occur or a transition might be audible, would be another important component.

Despite some attempts to validate the method, there are still discrepancies between the numerically determined resonance frequencies and the spectral analysis methods or acoustic measurement methods. The relationships and reasons for the deviations should also be further researched in terms of the critical review of the procedures that have already been established.

## V. CONCLUSION

The present study applied the method of 3D-MRI to expand the knowledge regarding resonance strategies in the area of the so-called whistle register. Although the results were found to be consistent within the method, there are still discrepancies compared to other methods, which must be explored further. With regard to the resonance strategies, the subjects did not behave uniformly. Despite the limitations, it can be stated that, contrary to the hypothesis of Garnier *et al.* (2010), in the area of the whistle register, there was no clear register-related vocal tract adjustment and correlated change in the tuning strategy. Constant settings without adjustments could lead to intersections of resonances and partials and thus to an end of any tuning, which is in accordance with the hypothesis of Echternach *et al.* (2015). The $f_{R1}$: $f_o$ tuning turned out to be, in agreement with Garnier *et al.* (2010) applicable up to F6 (1397 Hz).

## ACKNOWLEDGMENTS

This study was supported by the German Research Foundation (DFG RI 1050/4-3, ZA 422/3-3).

Birkholz, P., and Jackel, D. (**2004**). "Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system," in *8th International Conference on Spoken Language Processing*, October 4–8, Jeju, Korea, pp. 1125–1128.

Birkholz, P., Kürbis, S., and Häsner, P. (**2021**). "Physical tube models of the vocal tract with soft versus hard walls," youtube video available at https://youtu.be/9AoRS9X2BNY.

Birkholz, P., Kürbis, S., and Häsner, P. (**2022**). "Physical tube models of the vocal tract with soft versus hard walls," in *Proceedings of the* ICASSP 2022, May 22–27, Singapore.

Birkholz, P., Kürbis, S., Stone, S., Häsner, P., Blandin, R., and Fleischer, M. (**2020**). "Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties," Sci. Data **7**, 255, pp. 1–16.

Birkholz, P., and Venus, E. (**2017**). "Considering lip geometry in one-dimensional tube models of the vocal tract," in *Proceedings of ISSP*, October 16–19, Tianjin, China, pp. 78–86.

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

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

Burdumy, M. (**2016**). *Dynamic Imaging of Singers Using Magnetic Resonance Tomography* (Verlag Dr. Hut, München, Germany).

Burdumy, M., Traser, L., Burk, F., Richter, B., Echternach, M., Korvink, J. G., Hennig, J., and Zaitsev, M. (**2017**). "One-second MRI of a three-dimensional vocal tract to measure dynamic articulator modifications," J. Magn. Reson. Imaging **46**, 94–101.

Dalmont, J. P., Nederveen, C. J., and Joly, N. (**2001**). "Radiation impedance of tubes with different flanges: Numerical and experimental investigations," J. Sound Vib. **244**, 505–534.

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**, 2586–2595.

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

Echternach, M., Döllinger, M., Sundberg, J., Traser, L., and Richter, B. (**2013**). "Vocal fold vibrations at high soprano fundamental frequencies," J. Acoust. Soc. Am. **133**, EL82–EL87.

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

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

Echternach, M., Kob, M., Sundberg, J., Traser, L., Birkholz, P., Köberlein, M., and Richter, B. (**2018**). "Bestimmung von Vokaltraktresonanzen und Formanten mittels verschiedener Analyseansätze" "Determination of vocal tract resonances and ßormants using various analytical approaches", in *Wissenschaftliche Jahrestagung Der Deutschen Gesellschaft Für Phoniatrie Und Pädaudiologie (DGPP)*.

Fleischer, M., Mainka, A., Kürbis, S., and Birkholz, P. (**2018**). "How to precisely measure the volume velocity transfer function of physical vocal tract models by external excitation," PLoS One **13**, e0193708.

Garnier, M., Henrich, N., Crevier-Buchman, L., Vincent, C., Smith, J., and Wolfe, J. (**2012**). "Glottal behavior in the high soprano range and the transition to the whistle register," J. Acoust. Soc. Am. **131**, 951–962.

Garnier, M., Henrich, N., Smith, J., and Wolfe, J. (**2010**). "Vocal tract adjustments in the high soprano range," J. Acoust. Soc. Am. **127**, 3771–3780.

Hanna, N., Smith, J., and Wolfe, J. (**2016**). "Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation," J. Acoust. Soc. Am. **139**, 2924–2936.

Havel, M., Hofmann, G., Mürbe, D., and Sundberg, J. (**2014**). "Contribution of paranasal sinuses to the acoustic properties of the nasal tract," Folia Phoniatr. Logop. **66**, 109–114.

Havel, M., Kornes, T., Weitzberg, E., Lundberg, J. O., and Sundberg, J. (**2016**). "Eliminating paranasal sinus resonance and its effects on acoustic properties of the nasal tract," Logop. Phoniatr. Vocology **41**, 33–40.

Havel, M., Sundberg, J., Traser, L., Burdumy, M., and Echternach, M. (**2021**). "Effects of nasalization on vocal tract response curve," J. Voice.

Hawks, J. W., and Miller, J. D. (**1995**). "A formant bandwidth estimation procedure for vowel synthesis [43.72.Ja]," J. Acoust. Soc. Am. **97**, 1343–1344.

Herbst, C. T. (**2020**). "Registers—The snake pit of voice pedagogy. Part 1: Proprioception, perception, and laryngeal mechanisms," J. Sing. **77**, 175–190.

Herbst, C. T. (**2021**). "Registers—The snake pit of voice pedagogy. Part 2: Mixed voice, vocal tract influences, individual teaching systems," J. Sing. **77**, 345–358.

Herzel, H., and Reuter, R. (**1997**). "Whistle register and biphonation in a child's voice," Folia Phoniatr. Logop. **49**, 216–224.

Honda, K., Kitamura, T., Takemoto, H., Adachi, S., Mokhtari, P., Takano, S., Nota, Y., Hirata, H., Fujimoto, I., Shimada, Y., Masaki, S., Fujita, S., and Dang, J. (**2010**). "Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling," Comput. Methods Biomech. Biomed. Eng. **13**, 443–453.

Ishizaka, K., French, J. C., and Flanagan, J. L. (**1975**). "Direct Determination of Vocal Tract Wall Impedance," IEEE Trans. Acoust. **23**, 370–373.

Joliveau, E., Smith, J., and Wolfe, J. (**2004a**). "Tuning of vocal tract resonance by sopranos," Nature **427**, 116.

Joliveau, E., Smith, J., and Wolfe, J. (**2004b**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2434–2439.

Keilmann, A., and Michek, F. (**1993**). "Physiologie und akustische Analysen der Pfeifstimme der Frau" "Physiology and acoustic analyzes of woman's whistling voice," Folia Phoniatr. (Basel) **45**, 247–255.

Lombard, É. (**1911**). "Le signe de l'élévation de la voix," "The indication of the raising of the voice", Ann. des Mal. l'oreille, du larynx, du nez du pharynx **37**, 101–119.

Martienssen-Lohmann, F. (**1993**). *Der Wissende Sänger: Gesangslexikon in Skizzen* "The cognisant singer: Singing lexicon in sketches" (Atlantis Musikbuch-Verlag, Zürich, Switzerland).

Miller, D. G., and Schutte, H. K. (**1993**). "Physical definition of the 'flageolet register,'" J. Voice **7**, 206–212.

Pätzold, M., and Simpson, A. (**1997**). "Acoustic analysis of German vowels in the Kiel Corpus of read speech," **32**, 215–247; available at https://www.isfas.uni-kiel.de/de/linguistik/forschung/arbeitsberichte/aipuk/aipuk32.

Perry, J. L., Sutton, B. P., Kuehn, D. P., and Gamage, J. K. (**2014**). "Using MRI for assessing velopharyngeal structures and function," Cleft Palate-Craniofacial J. **51**, 476–485.

Story, B. H. (**2015**). "Mechanisms of voice production," in *The Handbook of Speech Production*, edited by M. A. Redford (John Wiley & Sons, Inc., Hoboken, NJ), pp. 34–58.

Sundberg, J. (**1977**). "The acoustics of the singing voice," Sci. Am. **236**, 82–84.

Sundberg, J. (**2009**). "Articulatory configuration and pitch in a classically trained soprano singer," J. Voice **23**, 546–551.

Sundberg, J., Birch, P., Gümoes, B., Stavad, H., Prytz, S., and Karle, A. (**2007**). "Experimental findings on the nasal tract resonator in singing," J. Voice **21**, 127–137.

Sundberg, J., Havel, M., Gill, B. P., Lã, F. M. B., and Lee, J. (**2015**). "Round table on nasal resonance in singing," in *Pan European Voice Conference (PEVOC): Firenze, Italy, August 31 – September 2, 2015*, p. 154; available https://fupress.com/archivio/pdf/2999_8023.pdf.

Švec, J. G., Sundberg, J., and Hertegård, S. (**2008**). "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," J. Acoust. Soc. Am. **123**, 347–353.

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

Titze, I. R. (**2004**). "A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice," J. Voice **18**, 292–298.

Titze, I. R. (**2008**). "Nonlinear source–filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.

Titze, I. R. (**2016**). "Some consensus has been reached on the labeling of harmonics, formants, and resonances," J. Voice **30**, P129.

Titze, I., Riede, T., and Popolo, P. (**2008**). "Nonlinear source–filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**, 1902–1915.

Traser, L., Birkholz, P., Flügge, T. V., Kamberger, R., Burdumy, M., Richter, B., Korvink, J. G., and Echternach, M. (**2017**). "Relevance of the implementation of teeth in three-dimensional vocal tract models," J. Speech, Lang. Hear. Res. **60**, 2379–2393.

Traser, L., Burdumy, M., Richter, B., Vicari, M., and Echternach, M. (**2013**). "The effect of supine and upright position on vocal tract configurations during singing—A comparative study in professional tenors," J. Voice **27**, 141–148.

Van den Berg, J. W. (**1963**). "Vocal ligaments versus registers," NATS Bull. **20**, 16–21

Vormann-Sauer, M. (**2017**). *Die Singstimme Der Frau: Anatomie Und Physiologie – Technik Und Strategien* "Woman's singing voice: Anatomy and physiology – technique and strategies" (Wißner, Augsburg, Germany).

Vos, R. R., Murphy, D. T., Howard, D. M., and Daffern, H. (**2018**). "Determining the relevant criteria for three-dimensional vocal tract characterization," J. Voice **32**, 130–142.