Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies

Matthias Echternach[a)]
Institute of Musicians' Medicine, Freiburg University Medical Center, Breisacher Str. 60, 79106 Freiburg, Germany

Peter Birkholz
Institute of Acoustics and Speech Communication, Technische Universität Dresden, Dresden, 01062 Dresden, Germany

Louisa Traser
Institute of Musicians' Medicine, Freiburg University Medical Center, Breisacher Str. 60, 79106 Freiburg, Germany

Tabea V. Flügge
Department of Craniomaxillofacial Surgery, Freiburg University Medical Center, Hugstetterstr. 55, 79106 Freiburg, Germany

Robert Kamberger
Laboratory of Simulation, Department of Microsystems Engineering—IMTEK, University of Freiburg, Georges-Köhler-Allee 102, 79110 Freiburg, Germany

Fabian Burk
Institute of Musicians' Medicine, Freiburg University Medical Center, Breisacher Str. 60, 79106 Freiburg, Germany

Michael Burdumy
Department of Radiology, Medical Physics, Freiburg University Medical Center, Breisacher Str. 60a, 79106 Freiburg, Germany

Bernhard Richter
Institute of Musicians' Medicine, Freiburg University Medical Center, Breisacher Str. 60, 79106 Freiburg, Germany

(Received 19 October 2013; revised 11 July 2014; accepted 25 March 2015)

The role of the vocal tract for phonation at very high soprano fundamental frequencies (F0s) is not yet understood in detail. In this investigation, two experiments were carried out with a single professional high soprano subject. First, using two dimensional (2D) dynamic real-time magnetic resonance imaging (MRI) (24 fps) midsagittal and coronal vocal tract shapes were analyzed while the subject sang a scale from Bb5 (932 Hz) to G6 (1568 Hz). In a second experiment, volumetric vocal tract MRI data were recorded from sustained phonations (13 s) for the pitches C6 (1047 Hz) and G6 (1568 Hz). Formant frequencies were measured in physical models created by 3D printing, and calculated from area functions obtained from the 3D vocal tract shapes. The data showed that there were only minor modifications of the vocal tract shape. These changes involved a decrease of the piriform sinus as well as small changes of tongue position. Formant frequencies did not exhibit major differences between C6 and G6 for F1 and F3, respectively. Only F2 was slightly raised for G6. For G6, however, F2 is not excited by any voice source partial. Therefore, this investigation was not able to confirm that the analyzed professional soprano subject adjusted formants to voice source partials for the analyzed F0s.
[http://dx.doi.org/10.1121/1.4919356]

[JFL] Pages: 2586–2595

I. INTRODUCTION

Voice production at very high soprano fundamental frequencies above 1000 Hz has been a subject of scientific discussion for many decades; in particular, the voice source mechanism at these fundamental frequencies is not yet fully understood. In this respect, competing hypotheses have been postulated, including a sound production mechanism analogous to that found in whistles (Schultz, 1902; Van den Berg, 1963), turbulence produced by vocal tract/voice source interactions (Herzel and Reuter, 1997), a flageolet-like mechanism (Martienssen-Lohmann, 1963) and the modification of the airflow by oscillating vocal folds without complete closure (Keilmann and Michek, 1993; Svec *et al.*, 2008). In a recent study using trans-nasal high speed endoscopy with a

a)Electronic mail: matthias.echternach@uniklinik-freiburg.de

 0001-4966/2015/137(5)/2586/10/$30.00

frame rate of 20 000 fps, it was shown for a single professional singer that, up to fundamental frequencies of 1568 Hz, the vocal folds oscillated and closed completely during the glottal cycle (Echternach *et al.*, 2013). There was no major change in laryngeal oscillation patterns in the very high frequency range from C6 (1047 Hz) to G6 (1568 Hz). Using electroglottography (EGG), Garnier *et al.* (2012) observed that in the wide range of D#5 to D6 there was only a single change in the EGG patterns with regard to amplitude and open quotient, which corresponded to the transition from head to whistle register. However, as has been shown in a previous experiment (Echternach *et al.*, 2013), both the singer and the examiner perceived a register transition slightly higher in pitch between D6 (1175 Hz) and E6 (1319 Hz). For the pitches E6 (1319 Hz) and F6 (1397 Hz) there was a change of the acoustic spectrum with an increase in subharmonics and noise. In the endoscopy data this register shift was associated with a collapse of the piriform sinuses. Since the general oscillatory vocal fold pattern did not change greatly, the question arises whether the perceptual and spectral differences associated with the register shift were caused by a modification of vocal tract shape and associated formants, rather than by changes in the voice source.

Walker (1988) analyzed spectral differences between B5 (988 Hz) and G6 (1568 Hz) and found that the intensity of overtones was higher for B5 compared to G6. However, intensities of spectrum partials are very dependent on the formant structures of the vocal tract and therefore formant structures and adjustments of formants to voice source partials have been the subject of a number of experiments. Using broadband acoustic excitation Garnier *et al.* (2010) observed that the tuning strategy might change for very high F0s. In agreement with earlier studies by Sundberg (1975) and Joliveau *et al.* (2004), Garnier *et al.* (2010) showed that, for the upper register, the first formant (F1) was tuned to the first partial of the acoustic spectrum (H1) up to 1200 Hz, and in single cases even up to 1400 Hz. For greater H1 they found that the resonance strategy might be altered.

Miller and Schutte (1993) also found evidence for an F1 to H1 tuning strategy for soprano voices. However, as they pointed out, at some point in the fundamental frequency range F1 can no longer be raised to track H1. At this point the register is changed, which the authors denote as a change into flageolet register.

If formants are to be adjusted, the vocal tract has to be modified. The use of dynamic real-time magnetic resonance imaging (MRI) has recently attracted growing interest (Echternach *et al.*, 2012) in the study of modifications of the vocal tract. Using this technique, it has been shown in soprano voices that vocal tract modifications can be observed at F0s greater than 750 Hz (Sundberg, 2009; Echternach *et al.*, 2010a; Bresch and Narayanan, 2010). Until now, there has been no morphometric data for sopranos producing very high F0s, and estimates of formant frequencies from 3D MRI volumetric data have not yet been carried out in order to augment the findings of Garnier *et al.* (2010). The current study therefore aims to analyze vocal tract changes for professional singing at very high F0s of a single professional soprano subject. Since for the same subject there was no change in vocal fold oscillatory patterns but a register shift between D6 (1175 Hz) and E6 (1319 Hz) in a previous investigation (Echternach *et al.*, 2013), it was expected that vocal tract modifications occur in the frequency region of the register change.

II. MATERIAL AND METHODS

After approval from the local ethics committee we analyzed the same subject as described in a previous study (Echternach *et al.*, 2013). All magnetic resonance images were acquired on a 3 T MAGNETOM Trio scanner (Siemens HealthCare AG, Erlangen, Germany).

A. Dynamic real-time MRI experiment

In the first experiment a dynamic real-time sequence of the midsagittal plane of the vocal tract was measured in combination with an iterative reconstruction method that yielded a frame rate of 24 frames per second. Two dimensional midsagittal slices were acquired with a radio frequency (RF)-spoiled radial gradient echo sequence [repition time (TR) = 2.34 ms, echo time (TE) = 1.5 ms, flip angle (FA) = 6°, pixel size $1.64 \times 1.64\,mm^2$, slice thickness = 10 mm, matrix = 128×128, BW 1500 Hz/pixel]. Angular steps of 111° (golden angle) were chosen. This angle scheme provides a nearly optimal coverage of k-space for any given number of spokes (Winkelmann *et al.*, 2007). At this high bandwidth, radial sampling suffers from gradient delays, which were subsequently corrected with precalculated delays from phantom measurements. An iterative CG-SENSE reconstruction was performed on the phase corrected raw data. For one image, 50 succeeding echos were used to calculate low-resolution sensitivity maps. These provided the SENSE weights for a regularized conjugated gradient reconstruction. A moving window over 34 consecutive spokes was applied to achieve the temporal resolution of approximately 24 frames per second.

The subject was asked to sing a scale from Bflat5 (932 Hz) to G6 (1568 Hz) in supine position on the vowel /a/ at a comfortable loudness level, as she would perform on stage. She was asked to sustain each pitch for approximately 1 s. Since it seems possible that there could be different ways of tuning strategies for the same F0, the subject was asked before the experiment whether she was able to use different strategies. This was denied by the subject.

As in previous studies (Echternach *et al.*, 2008; Echternach *et al.*, 2010a; Echternach *et al.*, 2010b; Echternach *et al.*, 2011b) the midsagittal plane was chosen for measurements. However, since narrowing of the pharynx in the lateral/medial dimension could not be represented using this plane, a coronal plane was also used in a separate recording which showed the larynx. As in previous studies mentioned above, the subject could hear her own audio signal since it was recorded by an optical microphone system (CONFON HP-SI 01, sampling rate 8 kHz, MR confon GmbH, Magdeburg, Germany) and transmitted to headphones to provide acoustic feedback. The audio signal was recorded and F0 extracted using PRAAT software (University of Amsterdam, Amsterdam, The Netherlands). In the midsagittal MRI images different vocal tract distances

were measured, as described previously (Echternach *et al.*, 2010b; Echternach *et al.*, 2011b): (a) the lip opening defined by the lowest distance between the lips, (b) jaw opening defined as the distance between the spina of the upper jaw and the lower front edge of the mandible, and (c) the jaw protrusion defined as the distance between the lower front edge of the mandible and the mucosal cover of the spine at a 90-deg angle. In order to measure the cranio-caudal larynx position, an auxiliary line A was constructed which connects the cranial-most part of the dens axis (anatomical structure of the second cervical spine bone) and the caudo-anterior edge of the sixth vertebra. The larynx position was measured as the distance from the cranial-most part of the dens axis to the point where auxiliary line A crosses a line (auxiliary line B) from the anterior commissure rectangular to the helpline A. Concerning the accuracy of the measurement it should be noted that these measures are limited from a technical perspective by the pixel size of $1.64 \times 1.64\,mm^2$. In the worst case, the measured distances can have a deviation of as much as one full pixel width.

In order to check the accuracy of the measurement all distances in every MRI picture were measured by the same experimenter twice. The maximum difference concerning the 201 pictures was for the lip opening 1 mm, for the jaw opening 2 mm, for the jaw protrusion 1 mm, and for the larynx position 3 mm. Furthermore, a second experimenter performed the same measurements. The maximum inter-rater difference was for the lip opening 2 mm, for the jaw opening 3 mm, and for the jaw protrusion 2 mm. Only the larynx position showed a greater inter-rater difference with a maximum difference of 7 mm and a mean difference of 4 mm. The difference concerning the larynx position between both experimenters was, however, systematic (i.e., the same changes over time were observed for both measurements) and the correlation for the measures of larynx position between both experimenters was very high ($r = 0.95$).

Additionally, the tongue contour was segmented for all images over time by a vocal tract anatomy expert using ITK-SNAP (2.4.0, open source application, User-Guided Level Set Segmentation of Anatomical Structures with ITK-SNAP Paul Yushkevich). In order to measure the segmentation variability of the contours 10 random slices were segmented twice and the Dice-coefficient was calculated for the segmented anatomies of these slices (Dice, 1945). The mean of the Dice-coefficient was 0.967 with a standard deviation of 0.0061 for the 10 compared slices. This indicates a high similarity of the segmentations and underlines their validity.

Finally, the *cross-distance function* of the vocal tract was calculated at the temporal midpoint of each note. The cross-distance function specifies the cross-distance of the vocal tract in the midsagittal plane perpendicular to the centerline between the anterior-inferior and posterior-superior outline of the vocal tract. Therefore, it is similar to the area function in depicting the tube shape of the vocal tract. The calculation of the centerline based on the contour tracings of the corresponding real-time MRI frames followed the procedure detailed further below for the extraction of the area function from the 3D-MRI vocal tract models.

B. Three dimensional MRI experiment and formant calculations

In a second experiment, the 3D vocal tract was recorded while the subject sustained a note for 13 s, first for C6 (1047 Hz) and then for G6 (1568 Hz). MRI volumetry was performed with the following parameters, as previously described (Echternach *et al.*, 2011a): 3D gradient recalled echo (GRE) imaging, spatial resolution $= 1.0 \times 1.6 \times 1.3\,mm^3$, TE = 1.67 ms, TR = 4.85 ms, flip angle = 12°, band width = 300 Hz/Px, matrix $= 250 \times 142$, GRAPPA = 2. Images were acquired in a 3D slab. The vocal tract structure was segmented from the MRI data using ITK-SNAP (see above) using a region growing method (see Fig. 1). In order to include the teeth into the 3D dataset of the vocal tracts a laser scan of the subjects' teeth was performed in the dental department. Intraoral scans were acquired with the scanning device iTero (Align Technology, San Jose, CA). The technology uses a confocal laser light beam and an instant digital image processor to record every tooth from different angles. The fusion of single images results in a complete virtual model of the upper and lower jaw (Birnbaum and Aaronson, 2008). The relation of the jaws and their accurate dimensions are represented in the virtual models (Flugge *et al.*, 2013). The 3D models of the vocal tract and the teeth scan were superimposed using the original MRI data in the Program VoXim (Fa. IVS Solutions AG, Chemnitz, Germany). A wall thickness of 2 mm was assigned to the reconstructed 3D model using Rapidform XOR (Inus Technologies, Seoul, Korea). Later, a 3D printout was performed using the powder-based printer ZPrint 450 (Fa. 3D Systems, South Carolina, USA). After printing, the models were injected with the super-glue-like material Z-Bond 90 (Fa. 3D Systems, South Carolina, USA), which improves the mechanical strength and decreases the gas permeability of the model.

In order to estimate the formant frequencies of the printed 3D models, they were placed in a sound-treated room and excited with sine sweeps (20–20 000 Hz) and white noise introduced at the glottal end using an ear loudspeaker (Energy Urban 300 Black Metal, Energysistem, Alicante, Spain). The radiated sound was recorded 1 cm in front of the mouth of the model with an omnidirectional microphone (ME 62, Fa. Sennheiser, Wedemark, Germany). Formant frequencies were estimated as peaks in the magnitude spectrum calculated from the recorded audio signals. In order to check the validity of data the complete process of segmentation, printout, and formant estimation for the pitch G6 was performed twice. The formant frequencies differed for F1 by 2.6% (1218 vs 1250 Hz), F2 4.1% (2375 vs 2281 Hz), and F3 1.9% (3312 vs 3375 Hz). A perceptual check was made by using the measured formant frequencies in the audio synthesizer "madde" (Svante Granqvist, KTH Stockholm) synthesizing a signal 10 s long and comparing it to the original audio file. The perceptual comparison was performed by the authors M.E. and L.T. For the perceptual check, fundamental frequencies of 1047 and 1568 Hz (referring to pitch C6 and G6, respectively) were set for the voice source and the formants were set using the frequencies calculated for the printed models (for the C6 and both G6 models). The

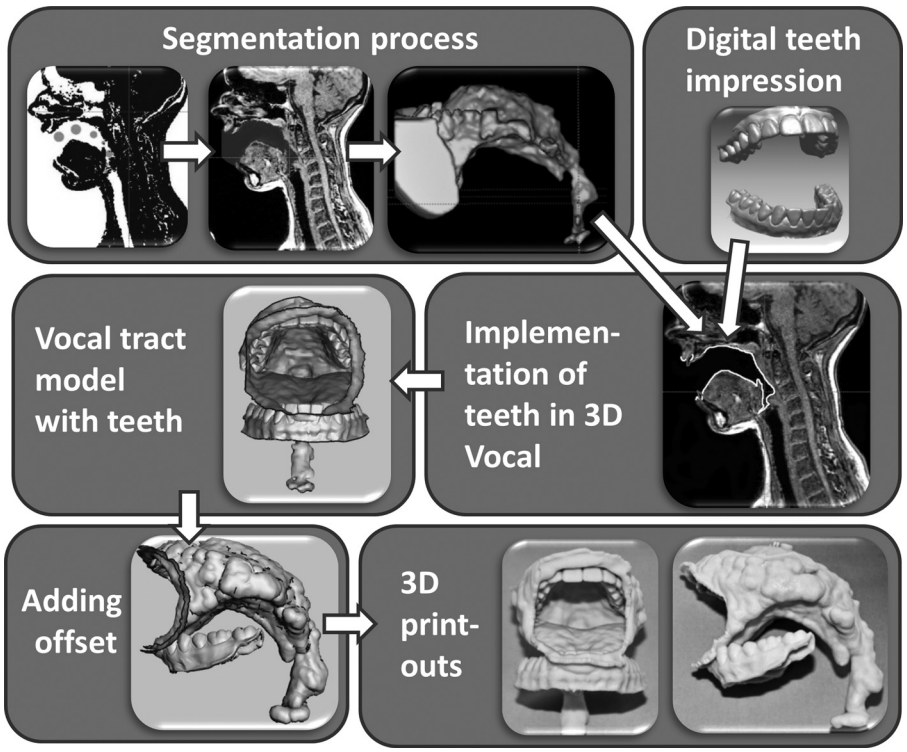

FIG. 1. The process of reconstruction of the physical vocal tract: The preprocessed MRI image with active contour evolution driven by intensity regions leads to a segmented vocal tract in two dimensions. All 2D vocal tract shapes are fused to a 3D vocal tract. In this 3D vocal tract teeth are implemented from teeth laser scan material using anatomical landmarks from the original MRI material.

investigators did not note a perceptual difference between both printed G6 models (the original and the second model for error estimation, respectively).

Finally, the relationship between the different vocal tract shapes (for the notes C6 and G6) and the resulting formant locations was tested by assuming 1D plane-wave propagation in the vocal tract. This assumption is commonly made when relating vocal tract shapes to acoustics, especially for the analysis of speech production. First, we extracted the area functions from both MRI scans (see Fig. 2 for the example note C6) and then used these to calculate the vocal tract transfer functions based on a one-dimensional transmission-line model of the vocal tract. Finally, the formant locations were identified from the transfer functions.

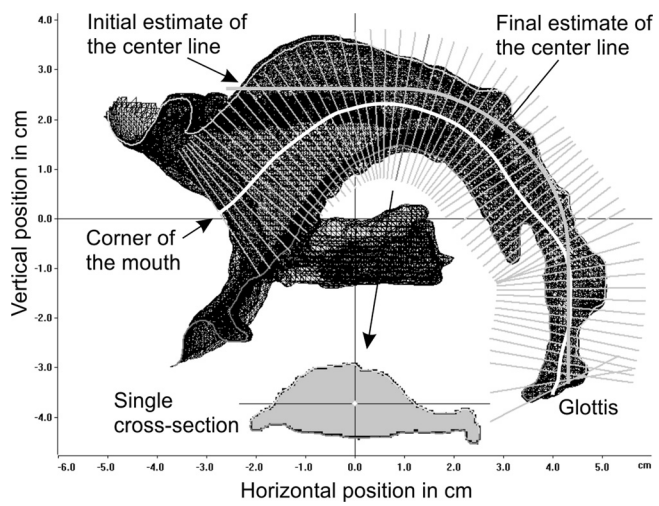

FIG. 2. Vocal tract centerline and cutting planes for the estimation of the area function for the wireframe mesh of model C6 (1047 Hz). The thick gray line is the initial estimate of the center line, and the thick white line the final estimate. The acoustic termination of the vocal tract was set to the corner of the mouth. The thin gray line segments are oriented normal to the final centerline and sagittal projections of the planes that cut the vocal tract to obtain its cross-sections.

Area functions were obtained from the same wire-frame meshes of the segmented vocal tract that were used for printing the physical models in 3D. The precondition for the calculation of the area function is the identification of the vocal tract centerline, along which the acoustic waves are assumed to propagate. Following Birkholz (2005), we applied the following procedure to obtain the centerline. First we extracted the superior-posterior and the inferior-anterior contours of the vocal tract mesh in the midsagittal plane (Fig. 2), then we constructed a rough initial estimate of the centerline composed of a horizontal line segment through the mouth cavity, a vertical line segment through the lower pharyngeal cavity, and a quarter of a circle connecting these. The center of the circle was set to the geometric center of the inferior-anterior outline, and the radius was set to 3.5 cm. The inferior end of this centerline was set to the position of the glottis, and the anterior end was set to the horizontal position of the corners of the mouth. The position of the mouth corners was therefore assumed to be the acoustic termination of the vocal tract in accordance with Lindblom and Sundberg (1971) and Mermelstein (1973). At each of 63 equally spaced points along this initial centerline, the two vocal tract outlines were intersected with a grid line normal to the center line at the respective point. Let the points of intersection with the two outlines for a gridline j be A_j and B_j. An improved estimate of the centerline was obtained as the sequence of straight-line segments joining the midpoints $(A_j+B_j)/2$ for all j. This line was subsequently smoothed with a 2 cm long moving average filter to obtain the final center line. At each of 65 equally spaced points along the final center line, the 3D wire-frame mesh was intersected

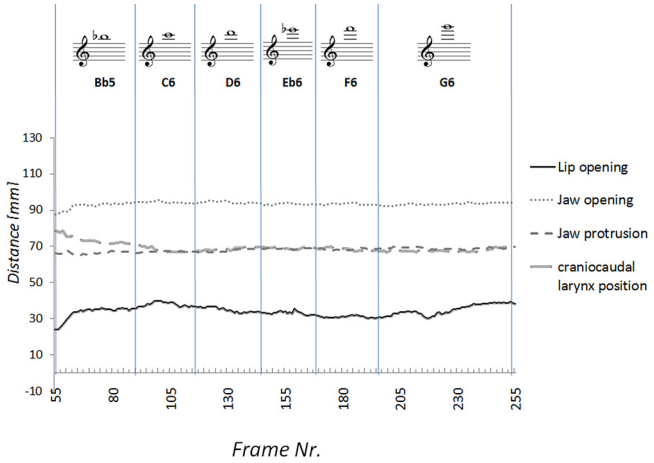

FIG. 3. (Color online) Different articulatory changes shown in the real-time MRI data for the different pitches Bflat5 (932 Hz), C6 (1047 Hz), D6 (1175), Eflat6 (1245 Hz), F6 (1397 Hz), and G6 (1568 Hz). Each frame is related to 1/24 s. The lines refer to the lip protrusion, jaw opening, jaw protrusion and larynx position for the indicated pitches.

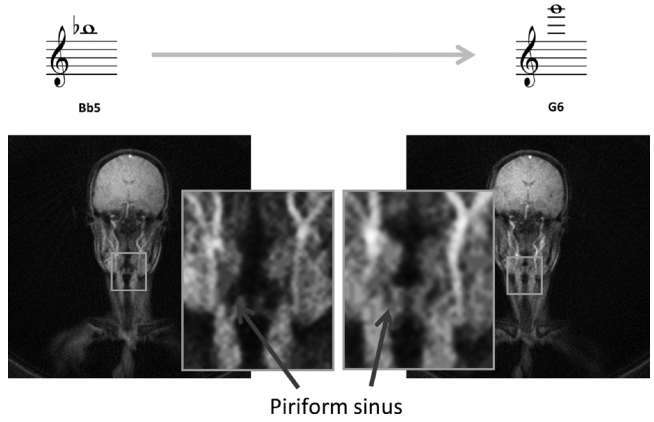

FIG. 5. Representative coronal MRI pictures from the dynamic real-time MRI experiment demonstrating the narrowing of the piriform sinus between Bflat5 (932 Hz) and G6 (1568 Hz).

with a plane perpendicular to the center line in the respective point. Sagittal projections of these planes are shown as gray lines in Fig. 2, and an example cross section is given by the gray-filled shape below the vocal tract. The cross-sectional area was calculated for each intersection.

For the acoustic simulation each area function was approximated as a sequence of 32 abutting cylindrical tube sections. An acoustic transmission-line model with lumped elements was simulated in the frequency-domain to obtain the volume-velocity transfer functions corresponding to the tube models according to Birkholz and Jackèl (2004). These simulations included realistic radiation impedance, and the effects of viscous friction, heat conduction, and compliant vocal tract walls. The radiation impedance was an approximation to that of a piston in an infinite baffle (Birkholz, 2005), where the area of the piston was set to the last cross-sectional area of the vocal tract (at the mouth corners). The formant frequencies were extracted from the transfer functions by peak-picking.

III. RESULTS

In the experiment with the dynamic MRI data at a midsagittal slice, a register transition was perceived between Eb6 (1245 Hz) and F6 (1397 Hz) by three experts when the subject performed the scale from Bb5 (932 Hz) to G6 (1568 Hz). However, there were only minimal changes of the vocal tract shape, as shown in Fig. 3. Here, in the midsagittal plane there were only small tongue movements including a small lowering of the tongue dorsum with a nearly stable larynx position. The changes of tongue shape are shown in detail in Fig. 4 representing the 2D tongue shape vs time. Similar to previous observations during laryngoscopy (Echternach *et al.*, 2013). Figure 5 shows that the piriform sinuses collapse at the highest pitches but this was a continuous rather than sudden process.

Figure 6 shows the cross-distances of the vocal tract perpendicular to the centerline from the glottis (at 0 cm) to the mouth corner for each of the six pitches measured with real-time MRI. For reference, the *average* cross-distance function for all six pitches is drawn as a dashed line in each subimage. This indicates small, quasi-monotonic changes in the vocal

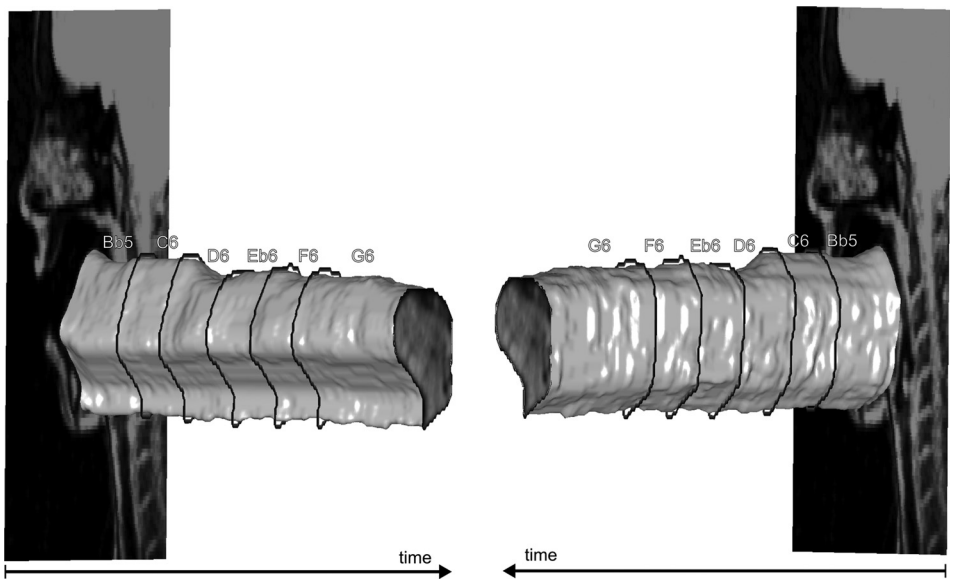

FIG. 4. Tongue contours of the semi-automatic tongue segmentation over time. The left picture is related to a frontal and the right picture to a back view. Additionally, an example MRI slice image is depicted to clarify the orientation. The labeled black lines indicate the times at which a pitch change occurred. Also, the pitches are indicated between these black lines.

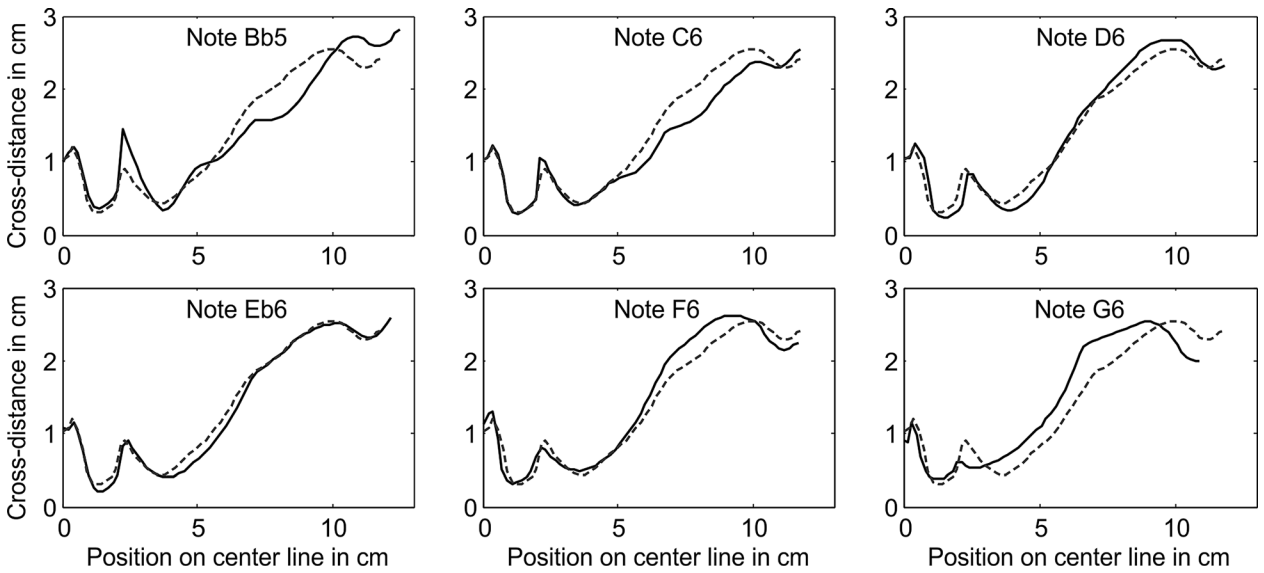

FIG. 6. Cross-distance functions of the vocal tract calculated from the temporal center frames of the real-time MRI measurements of the six notes (solid lines). The center line position of 0 cm corresponds to the position of the glottis. The dashed lines are the same in all six subimages and represent the average cross-distance function of all six notes.

tract shape from low to high pitches in all parts of the vocal tract. It is also evident that the vocal tract gets slightly shorter toward higher pitches.

The collapse of piriform sinuses was also visible for the 3D segmented models shown in Figs. 7 and 8. Nevertheless even at the pitch of C6 the piriform sinuses can be already considered very small. In contrast to the 2D-dynamic data reflecting the midsagittal plane, there was a very slight increase of the lip opening by 1.5 mm, laterally, showing more of the upper teeth (Figs. 7 and 8). The acoustic experiment with physical models demonstrates that the formant frequencies for F1 and F3 did not differ for the pitches C6 and G6 by a great amount, as shown in Fig. 9, only F2 was slightly higher for G6 (2.61%). The area functions calculated from the 3D data are shown in Fig. 10(a), and the transfer functions computed from the area functions are in Fig. 10(b). As for the transfer functions measured from the physical models shown in Fig. 9, F2 was 70 Hz higher for G6 also in the simulation.

In the synthesis experiment the signals for C6 (with related formant frequencies and F0 of 1047 Hz) and G6 (with related formant frequencies and F0 of 1568 Hz) were rated by two investigators as being in very good agreement with the real sound when a voice source spectrum with a spectral tilt of −12 dB/octave was used to excite the vocal tract. Simulated spectra of the radiated sound of the printout models and the computed models are shown in Fig. 11.

F1 and F3 showed good agreement between the computed model and the physical model data (for F1: C6 1264 Hz vs 1250 Hz and G6 1225 Hz vs 1218 Hz; for F3: C6 3330 Hz vs 3312 Hz and G6 3477 Hz vs 3312 Hz, Table I). However, the computed F2 was noticeably lower than that measured in the physical models for both C6 and G6.

IV. DISCUSSION

This study analyzed vocal tract articulatory data and formant frequencies for a single professional soprano voice singing at very high fundamental frequencies. In general, it was found that the vocal tract does not change by any great amount even at these high F0s and that formant frequencies therefore remain almost stable regardless of the sung pitch.

Resonance strategies play an important role in professional soprano singing. In previous investigations it has been shown that F1 is adjusted to H1 for the upper register below 1000 to 1200 Hz (Sundberg, 1975; Joliveau *et al.*, 2004; Garnier *et al.*, 2010). In this F0 region vocal tract changes have been observed in dynamic MRI studies (Sundberg,

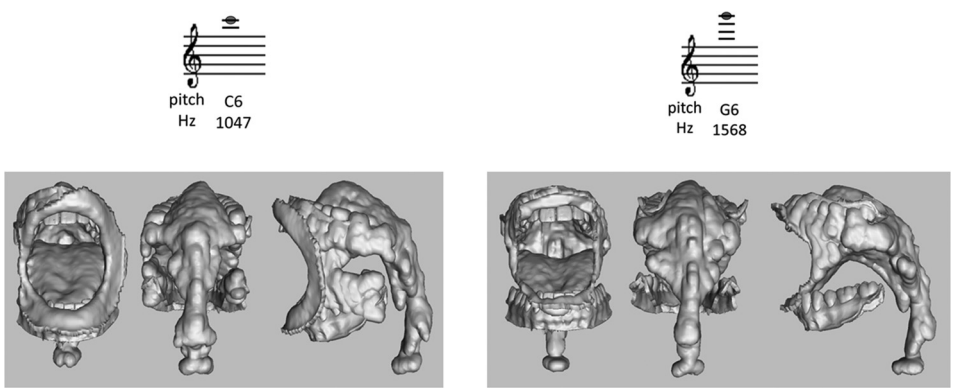

FIG. 7. Segmented models for C6 (1047 Hz, left) and G6 (1568 Hz, right).

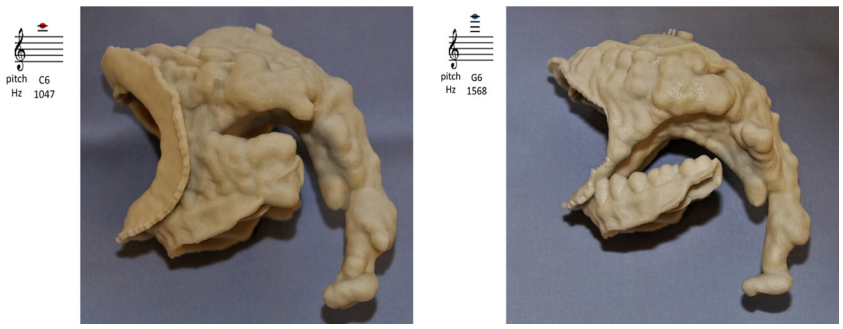

FIG. 8. (Color online) 3D printouts for C6 (1047 Hz) and G6 (1568 Hz).

2009; Echternach *et al.*, 2010a; Bresch and Narayanan, 2010). However, using broadband acoustic excitations, Garnier *et al.* (2010) observed for some subjects a change of tuning strategy where F2 is adjusted to H1 at fundamental frequencies above 1200 Hz. In the data of Garnier *et al.* (2010), however, for a single "non-expert" subject (NE1) the formants stayed almost stable for F0s from B5 (988 Hz) to F6 (1397 Hz). Although in the presented study a professional singer subject was analyzed, our data are in agreement with the latter observation and we found that the formant frequencies for C6 and G6 stay relatively stable. Only the frequency for F2 was slightly changed. Therefore, the agreement of the presented data of a professional opera singer subject with the findings of Garnier *et al.* concerning the non-expert subject NE1 suggest that this strategy of almost no change of formant frequencies is not related to special singing training.

Consistent with the stability of the formant frequencies, no large vocal tract adjustments were found in our experiment. In this respect there was no increase of lip opening in the midsagittal plane, but a very small increase of the lip opening at the lateral part of the mouth for the 3D data. Again, this observation is in good agreement with the findings of Garnier *et al.* (2010) concerning detection of lip area for the mentioned subject NE1. The very small increased lip opening might raise F1 (Ungeheuer, 1962). However, since the jaw opening and teeth position remains stable this effect would be considered minor and F1 was found almost stable. Furthermore, the tongue position was only slightly changed. Since tongue position might be very relevant for F2 (Ungeheuer, 1962) the change between both pitches (C6 and G6, respectively) might be in part be explained by this phenomenon. Last, a small collapse of the piriform sinuses could be observed (small since the piriform sinuses were already quite narrow for C6). Dang and Honda (1997) have shown that the piriform sinuses might influence anti-resonances associated in the region of 4–5 kHz. As a consequence, small changes in the volume of the piriform sinuses might contribute to small formant structure changes for F4 and F5 in the present study.

But what could be the explanation of the register shift between D6 and E6 observed in this and the previous study? The data presented show that for C6 (H1 ≈ 1046 Hz and H2 ≈ 2092 Hz) H1 is boosted by F1 and H2 by F2. The formant peak of F1 was estimated to be around 1250 Hz by both the physical model and the computer model. Therefore, if the formants remain stable, D6 would also be on the ascending

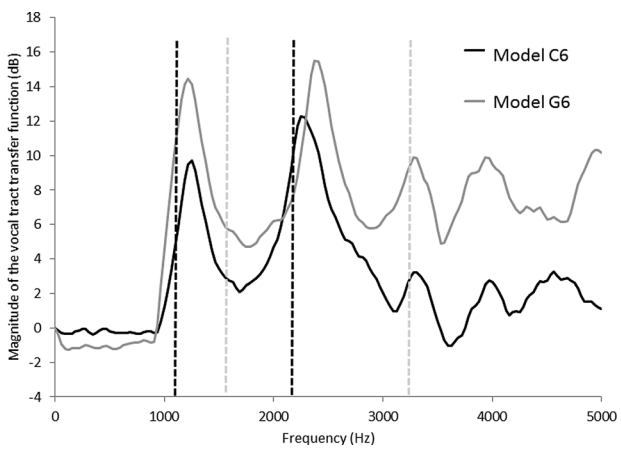

FIG. 9. Transfer functions for the physical models calculated for models C6 (1047 Hz) and G6 (1568 Hz). The dashed lines refer to H1 and H2 for pitches of C6 and G6, respectively.

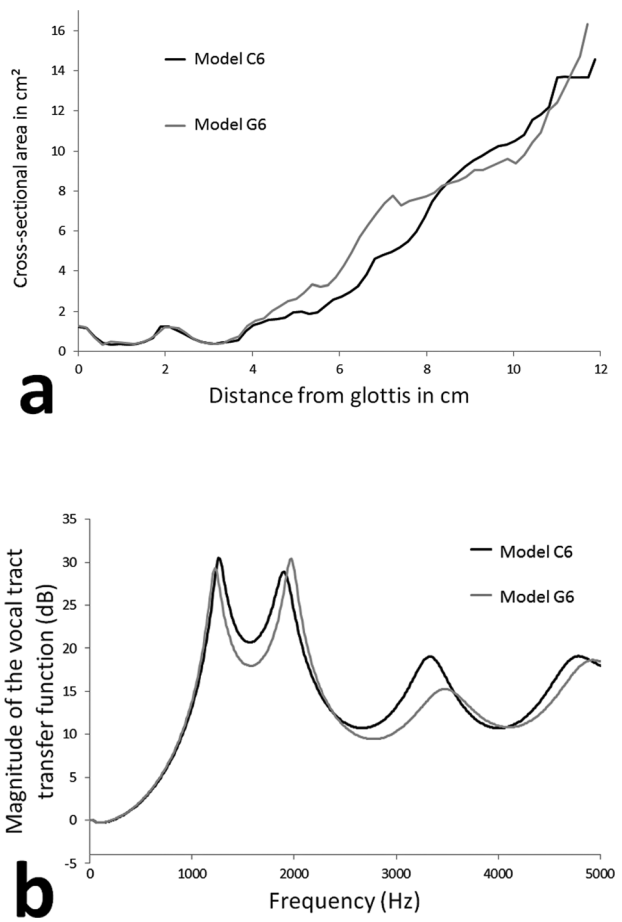

FIG. 10. (a) Area functions calculated for models C6 (1047 Hz) and G6 (1568 Hz) and (b) volume velocity transfer functions calculated for models C6 and G6.

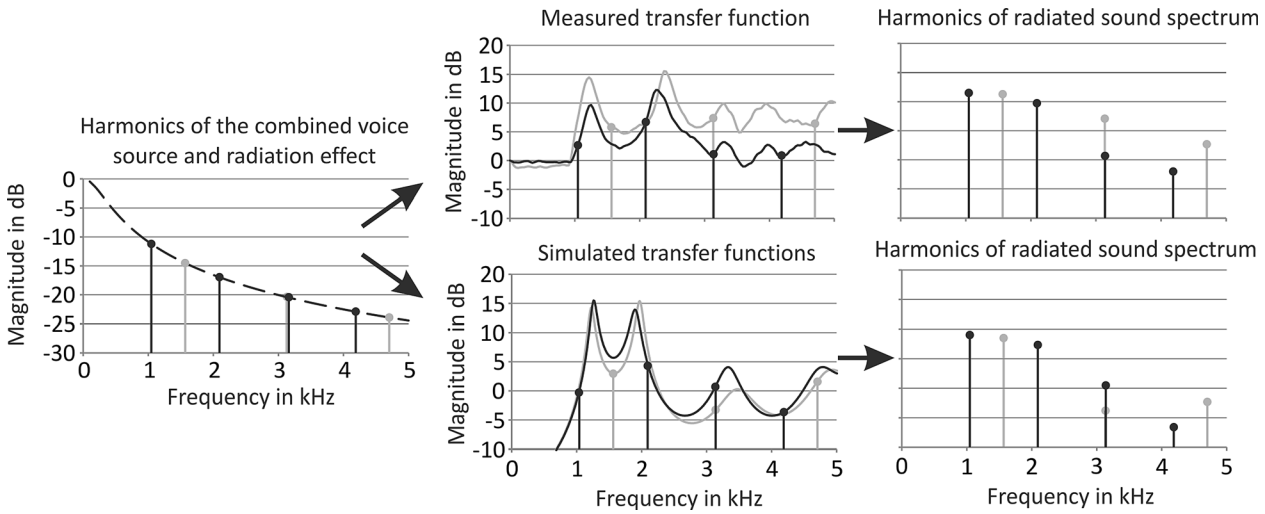

FIG. 11. Illustration of the combined spectrum of the voice source and the radiation effect (spectral slope of approximately -6 dB/oct) at the left, the measured and simulated transfer functions of the notes C6 (1047 Hz, black) and G6 (1568 Hz, gray) in the middle, and the corresponding radiated sound pressure harmonics at the right. The partials are indicated by the vertical black and gray lines with the round tops.

part of the formant peak. However, for E6, H1 would cross the formant peak. Interestingly, in the dynamic 2D-MRI experiments a register shift was also perceived by three experts. Since a Bflat major scale was sung there was no E6 but rather Eflat6 (≈1245 Hz) and the register transition was between Eb6 and F6. If the first formant is assumed at 1250 Hz, the pitch of Eb6 would be still on the ascending part and F6 would be the first pitch on the descending part of the formant. In the synthesis using the "madde" software, perceptual differences were also observed when F0 was raised above F1. Such an explanation of register difference concurs with a previous hypothesis by Miller and Schutte (1993). However, it should be mentioned that Miller and Schutte (1993) expected this phenomenon at much lower F0s than in the present study.

The fact that H1 crosses F1 might also provide an explanation of the occurrence of subharmonics in the previous study: Expecting a non-linear interaction between vocal tract and voice source, Titze (2008; Titze *et al.*, 2008) predicts that frequencies at the peak of the formant or on the descending part of the peaks might lead to an increase of instabilities in vocal fold oscillations or modulation of the trans-glottal airflow. The pitches of E6 and F6 would be on the descending part of the first formant and so the occurrence of subharmonics in the acoustic spectrum for the pitches E6 and F6 in the previous experiment (Echternach *et al.*, 2013) could be related to this phenomenon. For G6, the situation is different.

TABLE I. Formant frequencies for the formants 1–3 for the physical and the computed model.

Pitch	Formant	Physical model (Hz)	Computed model (Hz)	Deviation (%)
C6	F1	1250	1264	1,12
	F2	2312	1898	−17,91
	F3	3312	3330	0,54
G6	F1	1218	1225	0,57
	F2	2375	1968	−17,14
	F3	3312	3477	4,98

Since H1 is expected around 1568 Hz and H2 around 3136 Hz, H1 would be in the "formant valley" between F1 and F2 for all formant measurements. On the other hand, F2 is an "empty formant" which, according to our estimations with the physical model and the computer model, is not excited by any partial. Given a F0 of 1568 Hz, H1 would be 807 Hz and H2 761 Hz away from the formant peak, according to the physical model formant measurement. It should be therefore stated once again that spectrum peaks are not equal to vocal tract resonances. The authors would thus recommend using the term "formant" only for the latter. From a theoretical perspective it could be possible that our single subject would use another tuning strategy (i.e., F2 is adjusted to H1) for even higher F0s. Given that the formants would remain almost stable, it can be assumed that H1 would be boosted by F2 above the pitch of C7 (2093 Hz). However, the pitch of G6 was at the end of the subject's voice range profile [without the experimental setup ending at A6 (1760 Hz)], thus preventing further possibilities of analysis.

The formant calculations for F1 and F3 based on the numerical simulations agreed with the physical printout model measurements to a very high degree. However, the amplitudes differed and F2 was estimated at a much lower frequency by the computer model. As a consequence, the simulated radiated sound spectrum (Fig. 11) differed between the numerical model and the print out model for G6. Comparing the audio sound spectrum for G6 of the same subject during a high speed imaging experiment (Echternach *et al.*, 2013) the harmonic structure (i.e., that H2 showed a greater intensity than H3) was comparable to the simulated harmonic structure for the printed model of the recent experiment but not for the numerical model. There are a number of conceivable causes for the deviation of the formant calculations of the numerical model to the physical model. One reason might be that the vocal tract centerline deviates from the "true" path of acoustic propagation. Many other methods to compute the centerline were proposed (e.g., Goldstein, 1980; Beautemps *et al.*, 2001), but to our knowledge there is no systematic comparison of these

approaches with respect to their realism. Another point of debate is where to put the vocal tract termination, because the spread lips form a tube with a notch at the mouth end instead of a tube with a straight-cut end, as in the tube model. Despite some initial efforts to resolve this issue (Lindblom *et al.*, 2007; Motoki *et al.*, 1994), more research is needed, especially for such wide-open mouth cavities as in our data. Furthermore, we did not produce a model of the acoustic effect of the airspace between the epiglottis and the tongue root, which is a side branch to the main vocal tract tube and might affect formant locations. Finally, the 3D vocal tract shapes that were analyzed in this study are special in terms of their extreme articulations. Usually, cross-sectional areas as large as 15 cm^2 do not occur during normal speech production, for which the transmission-line model was developed. For such extreme articulations, standing acoustic waves in the vocal tract occur not only in the longitudinal dimension, but also in the cross-dimension. They cannot be accounted for by acoustic simulations based on the plane wave assumption. Instead, full 3D acoustic simulations would be needed here. However, existing 3D acoustic simulations of the vocal tract have other limitations and have not yet reached the quality of formant measurements with physical models. For example, in a study by Takemoto *et al.* (2010), formant estimations based on 3D sound field simulations deviated up to 10% from the corresponding formants measured with equivalent 3D-printed plastic models.

Despite these limitations, it should be pointed out that for both kinds of formant calculation the formants stayed nearly stable for the analyzed pitches C6 and G6. Also, for both calculations the situation that F2 is not excited by any spectrum partial holds true for the pitch of G6.

Our formant estimations do not include boundary conditions such as voice source production by vocal fold oscillations. It has been shown before that the first and second formant frequencies might increase with increasing glottal width (Barney *et al.*, 2007). Also, whispering was found to increase formant frequencies (Swerdlin *et al.*, 2010). Therefore, it could be expected that acoustic evaluations including both voice source and vocal tract might offer more valid data compared to our isolated vocal tract formant analysis. In our previous study (Echternach *et al.*, 2013) we analyzed the same professional soprano subject using high speed digital imaging. Simultaneously, we recorded audio signals and constructed the audio spectrum. During phonation on E6 there was increased noise and appearance of subharmonics. In order to check our present formant results the audio spectrum for E6 was re-analyzed. It was predicted that if noise is increased and formants are not excited by any partial, such formant peaks would also be visible in the audio spectrum. For the pitch of E6 and the presented formant calculations it was assumed that this prediction would be true for the formants 2 and 3. As shown in Fig. 12, the formants 2 and 3 could be easily identified in the audio spectrum. Furthermore, the formant peaks with F2 of 2015 Hz and F3 3375 Hz agreed with our calculations from the MRI data. Therefore, the authors believe that the presented formant calculations also offered valid data although no boundary conditions were included.

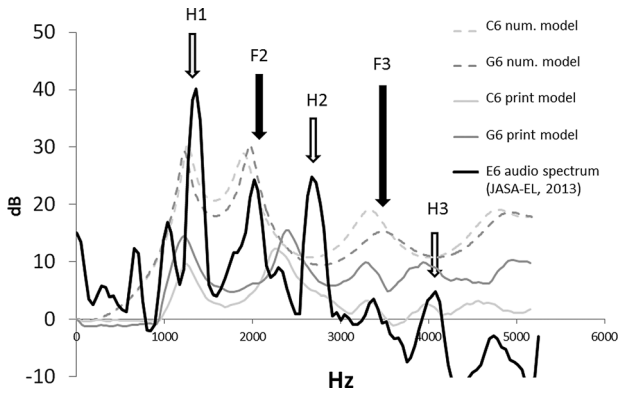

FIG. 12. Audio spectrum for the pitch E6 (1319 Hz) from the high speed digital imaging experiment (Echternach *et al.*, 2013) of the same subject in relation to the transfer functions of the recent study for both pitches C6 and G6, respectively. The arrows indicate the formants 2 and 3 (F2 and F3) and the harmonics 1−3 (H1−3), respectively.

There are certain limitations of the presented study. Only a single trained professional soprano subject was included and so it is possible that different sopranos use different strategies in order to reach high pitches. Also, the highest pitch examined was G6 which is at the upper limit of the subject's voice range profile. Since some human beings can reach much higher pitches it might be possible that resonance strategies change for even higher pitches, as described by Garnier *et al.* (2010). The MRI recordings were performed in supine position, but in a recent study analyzing professional tenors, Traser *et al.* (2013) observed in professional tenors that supine vs upright positions do not greatly affect the measurements. However, the larynx position for the tenors in the study of Traser *et al.* was higher and the jaw protrusion greater for the supine position. Thus, it cannot be excluded that the larynx would be somehow lower and the jaw more retracted in a more realistic upright singing position for our soprano subject. Last, our printout models had stiff walls. It was shown before that yielding walls raise F1 (Fant, 1972). Thus, it cannot be excluded that our F1 measurements for the printout models show slightly lower values compared to real human tissue.

V. CONCLUSIONS

This study analyzed vocal tract articulation and vocal tract acoustics for very high F0s in a professional soprano subject by means of MRI technology. It was found that the vocal tract shows only very minor modifications while singing a scale from the pitch of Bflat5 up to G6. Here, only small modifications of tongue position and the piriform sinuses were observed. Three dimensional models revealed a small increase of the lip opening at the lateral part of the mouth. Since there was no great change in articulation, the formant frequencies also did not exhibit major differences between C6 and G6 for F1 and F3, respectively. Only F2 was found greater for G6, which might be explained by the small change of tongue position. For the pitch of G6, the formant F2, however, would not be excited by any voice

source partial. As a consequence, we were not able to verify adjustments of formants to voice source partials for the analyzed F0s in this particular professional soprano subject.

Since there were neither great vocal tract adjustments in the recent study nor changes of vocal fold oscillatory patterns in the previous study, the previously perceived register difference between D6 (1175 Hz) and E6 (1319 Hz) (Echternach *et al*., 2013) might be caused by F0 being below F1 for D6 and above it for E6.

ACKNOWLEDGMENTS

The authors thank Horst Urbach, M.D., and Hansjörg Mast, Department of Neuroradiology, Freiburg University Medical Center, for their help to perform the MRI scans and Jan G. Korvink, Ph.D., Laboratory of Simulation, Department of Microsystems Engineering–IMTEK, University of Freiburg, for help in realizing the 3D prints. Furthermore, the authors thank Johan Sundberg, Ph.D., KTH Stockholm, for editorial help and for helpful discussions, Nathalie Henrich, Ph.D., GIPSA Lab, Grenoble for the idea of the *post hoc* analysis of the audio spectrum, and Jude Brereton, Ph.D., for native corrections. Also, the authors thank the subject for her willingness to take part in this study. M.E.'s and B.R.'s work is supported by the Deutsche Forschungsgemeinschaft (DFG), Grant No. EC 409/1-1 and RI 1050/4-1.

Barney, A., De Stefano, A., and Henrich, N. (**2007**). "The effect of glottal opening on the acoustic response of the vocal tract," Acta. Acust. Acust. **93**, 1046–1056.

Beautemps, D., Badin, P., and Bailly, G. (**2001**). "Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling," J. Acoust. Soc. Am. **109**, 2165–2180.

Birkholz, P. (**2005**). *3D-Artikulatorische Sprachsynthese* (*3D Articulatory Speech Synthesis*) (Logos Verlag, Berlin), pp. 1–171.

Birkholz, P., and Jackèl, D. (**2004**). "Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system," in *Proceedings of the Interspeech 2004-ICSLP*, 2004, Jeju, Korea, pp. 1125–1128.

Birnbaum, N. S., and Aaronson, H. B. (**2008**). "Dental impressions using 3D digital scanners: Virtual becomes reality," Compend. Contin. Educ. Dent. **29**, 494–498, 505.

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.

Dang, J., and Honda, K. (**1997**). "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. **101**, 456–465.

Dice, L. R. (**1945**). "Measures of the amount of ecologic association between species," Ecology **26**, 297–302.

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., Markl, M., and Richter, B. (**2012**). "Dynamic real-time magnetic resonance imaging for the analysis of voice physiology," Curr. Opin. Otolaryngol. Head Neck Surg. **20**, 450–457.

Echternach, M., Sundberg, J., Arndt, S., Breyer, T., Markl, M., Schumacher, M., and Richter, B. (**2008**). "Vocal tract and register changes analysed by real time MRI in male professional singers—A pilot study," Logoped. Phoniatr. Vocol. **33**, 67–73.

Echternach, M., Sundberg, J., Arndt, S., Markl, M., Schumacher, M., and Richter, B. (**2010a**). "Vocal tract in female registers—a dynamic real-time MRI study," J. Voice **24**, 133–139.

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

Echternach, M., Sundberg, J., Markl, M., and Richter, B. (**2010b**). "Professional opera tenor's vocal tract configurations in registers," Folia Phoniatr. Logop. **62**, 278–287.

Echternach, M., Traser, L., Markl, M., and Richter, B. (**2011b**). "Vocal tract configurations in male alto register functions," J. Voice **25**, 670–677.

Fant, G. (**1972**). "Vocal tract wall effects, losses, and resonance bandwidths," *STL-QPSR, Department of Speech Communication and Music Acoustics, Royal Institute of Technology Stockholm*, Vol. 2-3/1972, pp. 28–52.

Flugge, T. V., Schlager, S., Nelson, K., Nahles, S., and Metzger, M. C. (**2013**). "Precision of intraoral digital dental impressions with iTero and extraoral digitization with the iTero and a model scanner," Am. J. Orthod. Dentofacial Orthop. **144**, 471–478.

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.

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tracts of growing children," dissertation, Massachusetts Institute of Technology, Cambridge, MA, pp. 1–542.

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

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

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

Lindblom, B., Branderud, P., Sundberg, J., and Djamshidpey, H. (**2007**). "On the acoustics of spread lips," in *Proceedings of Fonetik*, TMH-QPSR (2007), pp. 13–16.

Lindblom, B. E., and Sundberg, J. E. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179.

Martienssen-Lohmann, F. (**1963**). *Der Wissende Sänger* (*The Knowing Singer*) (Atlantisverlag, Zürich), pp. 1–456.

Mermelstein, P. (**1973**). "Articulatory model for the study of speech production," J. Acoust. Soc. Am. **53**, 1070–1082.

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

Motoki, K., Badin, P., and Miki, N. (**1994**). "Measurement of acoustic impedance density distribution in the near field of the labial horn," in *Proceedings of the ICSLP1994*, Yokohama, Japan.

Schultz, P. (**1902**). "Über einen Fall von willkürlichem laryngealen Pfeifen beim Menschen" ("About a case of voluntary laryngeal whistle in a human"), Arch. F. Physiol. Suppl. 523.

Sundberg, J. (**1975**). "Formant technique in a professional female singer," Acustica **32**, 89–96.

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

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

Swerdlin, Y., Smith, J., and Wolfe, J. (**2010**). "The effect of whisper and creak vocal mechanisms on vocal tract resonances," J. Acoust. Soc. Am. **127**, 2590–2598.

Takemoto, H., Mokhtari, P., and Kitamura, T. (**2010**). "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," J. Acoust. Soc. Am. **128**, 3724–3738.

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

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

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.

Ungeheuer, G. (**1962**). *Elemente einer Akustischen Theorie der Vokalartikulation* (*Elements of an Acoustic Theory of Vowel Articulation*) (Springer-Verlag, Berlin, Germany).

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

Walker, J. S. (**1988**). "An investigation of the whistle register in the female voice," J. Voice **2**, 140–150.

Winkelmann, S., Schaeffter, T., Koehler, T., Eggers, H., and Doessel, O. (**2007**). "An optimal radial profile order based on the Golden Ratio for time-resolved MRI," IEEE Trans. Med. Imaging **26**, 68–76.