

Article

# Acoustic and Aerodynamic Coupling during Phonation in MRI-Based Vocal Tract Replicas

Judith Probst <sup>1,\*</sup>, Alexander Lodermeier <sup>1</sup>, Sahar Fattoum <sup>2</sup>, Stefan Becker <sup>1</sup>,  
Matthias Echternach <sup>3</sup>, Bernhard Richter <sup>4</sup>, Michael Döllinger <sup>2</sup> and Stefan Kniesburges <sup>2</sup>

<sup>1</sup> Department of Process Machinery and Systems Engineering, Friedrich-Alexander-University Erlangen-Nürnberg, Cauerstrasse 4, 91058 Erlangen, Germany

<sup>2</sup> Division of Phoniatics and Pediatric Audiology at the Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Medical School at Friedrich-Alexander-University Erlangen-Nürnberg, Waldstraße 1, 91054 Erlangen, Germany

<sup>3</sup> Division of Phoniatics and Pediatric Audiology at the Department of Otorhinolaryngology, Head and Neck Surgery, University of Munich, Campus Grosshadern, Marchioninistrasse 15, 81377 München, Germany

<sup>4</sup> Institute of Musicians' Medicine, Freiburg University, Medical Faculty, Elsässer Strasse 2 m, 79110 Freiburg, Germany

\* Correspondence: pro@ipat.uni-erlangen.de

Received: 9 March 2019; Accepted: 23 August 2019; Published: 30 August 2019



**Abstract:** Voiced speech is the result of a fluid-structure-acoustic interaction in larynx and vocal tract (VT). Previous studies show a strong influence of the VT on this interaction process, but are limited to individually obtained VT geometries. In order to overcome this restriction and to provide a more general VT replica, we computed a simplified, averaged VT geometry for the vowel /a/. The basis for that were MRI-derived cross-sections along the straightened VT centerline of six professional tenors. The resulting mean VT replica, as well as realistic and simplified VT replicas of each tenor were 3D-printed for experiments with silicone vocal folds that show flow-induced oscillations. Our results reveal that all replicas, including the mean VT, reproduce the characteristic formants with mean deviations of 12% when compared with the subjects' audio recordings. The overall formant structure neither is impaired by the averaging process, nor by the simplified geometry. Nonetheless, alterations in the broadband, non-harmonic portions of the sound spectrum indicate changed aerodynamic characteristics within the simplified VT. In conclusion, our mean VT replica shows similar formant properties as found in vivo. This indicates that the mean VT geometry is suitable for further investigations of the fluid-structure-acoustic interaction during phonation.

**Keywords:** voice production modeling; vocal tract acoustics; artificial vocal fold models

## 1. Introduction

The human voice results from the flow-induced oscillations of the vocal folds, as described by Titze [1]. This so-called phonatory process is the result of a fluid-structure-acoustic interaction between the laryngeal airflow and the vocal fold tissue. By this oscillation, the basic sound of the human voice is generated which is further transformed in the vocal tract producing the voice signal that radiates from the mouth. Hereby, the vocal tract does not only serve as a downstream resonator, but influences the pressure distribution in the glottal duct or even enhances the vocal fold oscillations acoustically during singing [2].

As the flow field inside the larynx cannot be observed in vivo, experimental and computational models have been developed to investigate the entire phonatory process. Overviews of the existing experimental and computational larynx models are given elsewhere [3–7].

Beside the aerodynamic and acoustic coupling effects, the main function of the vocal tract (VT) is to serve as an acoustic filter or resonator, respectively. Acoustic resonances are excited in the VT and amplify the sound pressure of the basic sound signal at specific frequency bands, which are called formants. The formant frequencies vary depending on the geometry of the VT, which conditions the acoustical properties of the voice signal for articulated voiced speech [8]. Although up to five formants have been detected for vowels [9], the first two formants F1 and F2 are sufficient to differentiate vowels [10].

In order to investigate the acoustic resonance properties of the VT, it is necessary to extract the geometric shape of the VT. This can be done using medical imaging techniques, which allow a detailed examination of the VT shape during phonation. Fant [11] and Mermelstein [12] determined the geometry of the VT for different vowels by means of X-ray. In recent studies, Magnetic Resonance Imaging (MRI) was used, which provides much higher contrast and resolution. Based on MRI, Story et al. [13], Kitamura et al. [14] and Aalto et al. [15] presented VT shapes corresponding to different vowels obtained from single subjects. Echternach et al. analyzed the VT shape of four professional sopranos [16] and ten professional tenors [17] who sang the vowel /a/ during register transitions.

To generate simplified VT models with similar acoustic properties, Story et al. [13] determined the cross-sectional area as a function of the position along the centerline of the VT. This area function was subsequently used to design straight VT models with circular cross section. Evaluating these simplified models, numerical simulations of the acoustic behavior showed good agreement with small sound samples recorded from the subject. In a recent study, Arnela et al. [18] compared VT models of the vowels /a/, /i/, and /u/ based on the data of Aalto et al. [15] for different levels of simplification in an acoustic simulation to determine the influence of geometry simplifications on the sound radiation. The results showed only small deviations of the formants in the transfer function being below 5% in the frequency range up to 5 kHz. For higher frequencies, the simplification produces more relevant deviations due to higher order modes [19].

After the development and computational evaluation of the VT shapes, 3D-printed VT replicas were manufactured to examine the acoustic properties of the VT in experiments. Kitamura et al. [14] used rapid prototyping to build replicas of the VT for five Japanese vowels based on MRI-data. They determined the transfer function of the VT replicas by exciting them with a time-stretched acoustic pulse played by a horn driver in order to provide a database for the testing of numerical analysis methods. The results showed a good agreement with recorded speech, although the formants were shifted to lower frequencies due to the solid walls of the VT replicas [14]. In a similar setup, Takemoto et al. [20] compared the transfer functions of replicas with computational MRI-based VT models and found good agreement, as well.

In recent years, the influence of anatomical details like the lips or the sinus piriformi have been analyzed with regard to their acoustical impact. Whereas the lips [21,22], the piriform fossae and the vallecule [20] only showed negligible or little impact on the transfer function with deviations less than 5% below 4 kHz, the acoustic influence of the teeth depends on the vowel that is phonated and the mouth opening. In the study of Traser et al., the effect of teeth on the VT transfer function was investigated. Although they found a significant effect of the teeth on the resulting resonance frequencies in several individual VTs, these frequencies changed by less than 150 cents (i.e., 1.5 semi-tones) for VT shapes that exhibit no side branches or side cavities. This is the case for high-pitched singing or vowels with a wide mouth opening like the vowel /a/. Another focus was set on the effect of the body position during phonation as MRI-imaging is commonly performed in supine position [23,24]. The reports showed that the effect of the supine position is statistically insignificant for different registers of professional tenors, whereas 5 out of 10 untrained singers phonated statistically significantly different being unable to completely compensate the different direction of the gravitation force in the supine position. Hence, Echternach et al. [25] recommended to include only professional singers in research studies concerning VT acoustics.

In this context, Echternach et al. [17] analyzed the VT shape during the register transition between modal and falsetto register sung by ten professional tenors, who performed the scale from C4 to A4 in the modal and the falsetto register. An overview of the different singing registers is given by Sundberg [26]. The results show that the shape of the VT varies negligibly when the subject changes from the modal register to the falsetto register. However, if the subject remains in the modal register during the scale C4-A4, the shape changes significantly. This shows that for higher pitches the natural form of the VT is represented by the falsetto register.

Lucero et al. [27] observed sudden jumps of the vocal fold oscillation frequency originating from an acoustic coupling between VT and vocal folds at oscillation frequencies near resonances of the VT. Similar results were reported by Titze et al. [28] in a study with 18 subjects. They also found an acoustic coupling between the VT and the vocal folds, producing frequency jumps, subharmonic tones and chaotic vibration patterns of the vocal folds.

Research of the VT impact on the glottal aerodynamics and structural dynamics is mainly focused on the immediate supraglottal region downstream of the vocal folds, especially on the ventricular folds impact [29–32]. In this context, effects such as the reduction of the subglottal phonation threshold pressure [29], the decrease of laryngeal [30] and increase of glottal flow resistance [31,32] and the stabilization of the acoustic output signal [29] were reported. A comprehensive study was presented by Horáček et al. [33], who performed qualitative flow visualization of the airflow in simplified, straightened VT models of the vowels /a/, /u/, and /i/ within a physical larynx model with self-oscillating vocal folds. They showed that for all observed vowels, large vortices appear in the supraglottal region, which disperse in the narrow pharyngeal part. The formant frequencies deviated from physiological values for the vowels /a/ and /i/, but showed good agreement for the vowel /u/.

The results of previous studies indicate that experimental models of the VT allow a reproduction of the acoustic properties of the human VT. Furthermore, they reveal that a simplification of the VT as proposed by Story et al. [13] (straightened VT with circular cross section) has no significant effect on the formant frequencies below 5 kHz. However, this conclusion was drawn only on the basis of experiments with an acoustic excitation by a loudspeaker at the glottis. Although Horáček et al. [33] investigated the flow in a simplified VT replica excited with vocal fold replicas that show flow-induced oscillations, they did not perform a comparison with realistic VT shapes. A second restriction is that previous studies were performed with VT models that exclusively rely on the geometry of single subjects. The influence of individual characteristics were not considered or evaluated, respectively.

Thus, our hypotheses are that (1) the resulting sound signal includes additional sound components generated by the unsteady flow field inside the VT when the VT is excited by a pulsatile jet flow and (2) a generalized geometry of the VT shows the relevant formants with reduced individual sound characteristics of the unique subjects' VT geometries.

Hence, the first aim of our study is to investigate the influence of the VT simplification on the radiated sound in a setup with coupled acoustics and aerodynamics of the VT replicas and self-oscillating vocal folds. Thus, we manufactured VT replicas with realistic and simplified shape, according to Story et al. [13], based on the VT geometries from the MRI data obtained from six professional tenors who phonated the vowel /a/ at same pitch and register [34]. These VT replicas were included in our experimental larynx replica with self-oscillating vocal folds [35]. To decrease the influence of individual geometric features, the second aim of this study is to generate and evaluate a simplified, mean VT replica that is based on the simplified VT geometries of the six tenor subjects. To evaluate the quality, we evaluated the first four formant frequencies of the different VT replicas.

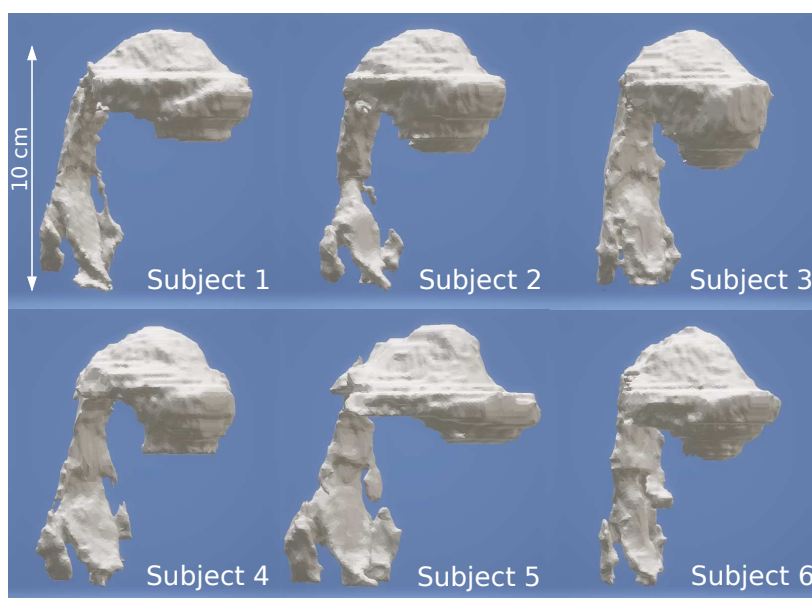
## 2. Materials and Methods

### 2.1. Vocal Tract Replicas

In this study, the VT MRI data of six professional tenors as published in a previous study by Echternach et al. [34] were used. The acquisition of the MRI data and the analysis of the acoustic

properties were performed at the Institute Musicians' Medicine of the University Hospital Freiburg by Echternach et al. [17,34]. The images were acquired with a 3.0-tesla TIM TRIO (Siemens, Munich, Germany) MRI device in sagittal planes in the center of the head. Further information regarding the MRI acquisition can be found in [34]. The professional tenors sustained the vowel /a/ on a pitch of F4 (349 Hz) in falsetto register for 20 s in a stable manner. In two subjects, the velopharyngeal area was not entirely closed; hence, the aerated nasopharyngeal cavity was removed at the upper uvula. Additionally to the MRI imaging, the sound was recorded in a separate session. The formants for each subject were determined based on inverse filtering [34].

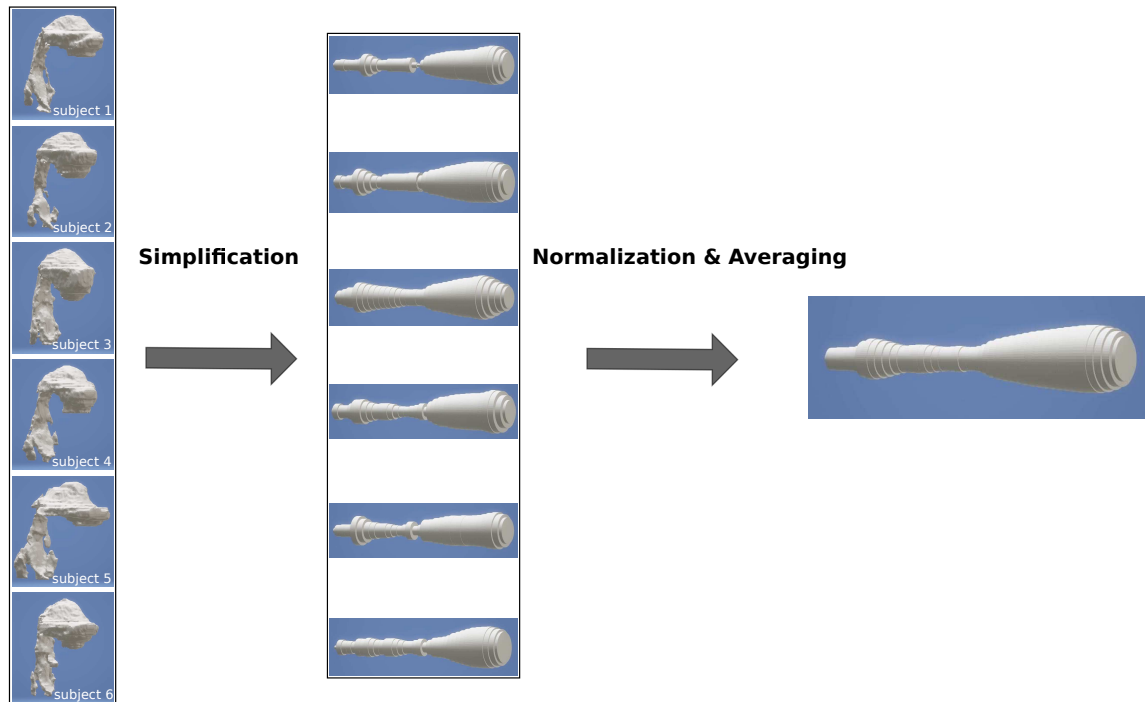
The VTs were segmented manually using the open-source software *3D-Slicer* [36,37]. Since the teeth are not resolved in standard MRI procedures, they were neglected in this study. This is a limitation of the presented model; however, according to Traser [38], the change of the cross-sectional area due to the teeth is not sufficient to yield a statistically significant change of the formant frequency for the vowel /a/. The lips were also neglected in the model, as their influence on the first resonance frequencies is not perceptually relevant according to Arnela et al. [21]. The VT replicas that directly rely on the segmentation of the MRI data represent the most realistic VT shape used in this study. They are depicted in Figure 1.



**Figure 1.** Segmented VTs of the six subjects.

The influence of the VT is often investigated in numerical and analytical models using simplified VT geometries. Thus, the segmented VT shapes have been further processed to analyze the effects of this simplification in an experimental setup with self-oscillating vocal folds. The resulting simplified VTs represent straight replicas with circular cross section that varies along the centerline, similar to the work by Story et al. [13]. They are based on the area functions that were determined by Echternach et al. [34].

A simplified, averaged model of the VT was developed for the application with larynx replicas to understand the basic mechanisms of the VT filtering and the contribution of the aerodynamics. We determined a mean VT geometry based on the area functions of the simplified models. As Figure 2 shows, we averaged the six area functions of the tenors' VTs to achieve a mean VT. Therefore, the length of each simplified VT was normalized to their mean length and the cross section area was averaged at each axial position to yield the mean VT replica, as depicted in Figure 2.

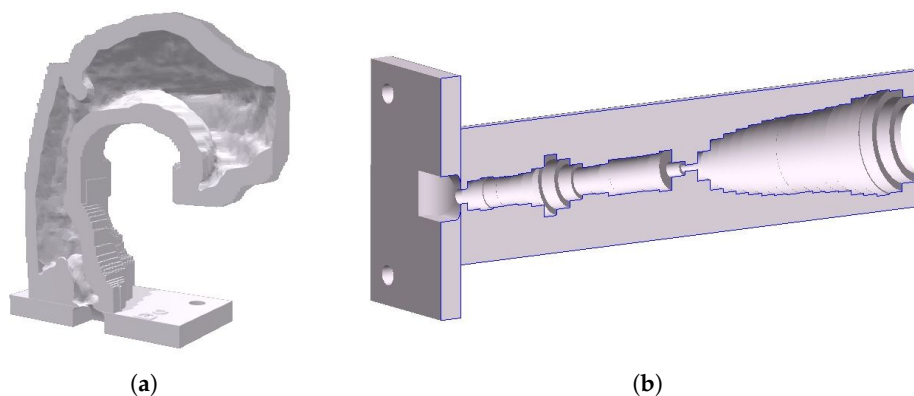


**Figure 2.** Simplification according to the method proposed by Story et al. [13] and averaging procedure of the VT geometries obtained from six tenors based on the MRI data by Echernach et al. [34].

The generation of standing waves in a tube like the VT depends on the gradient of the reflection coefficient. Hence, the reflection coefficients along the VT have been determined to evaluate the resonance properties of the VT replicas. Based on the corresponding area function  $A_i$ , the reflection coefficient  $r = (A_i - A_{i+1}) / (A_i + A_{i+1})$  was calculated at each cross sectional jump along the centerline.

## 2.2. Experimental Setup

To analyze the resonance properties, physical VT replicas were 3D printed using selective laser melting for the realistic VTs and fused deposition modeling for the simplified replicas. The connection of the VT replicas to the excitation source was realized with a mounting adapter, as shown in Figure 3.



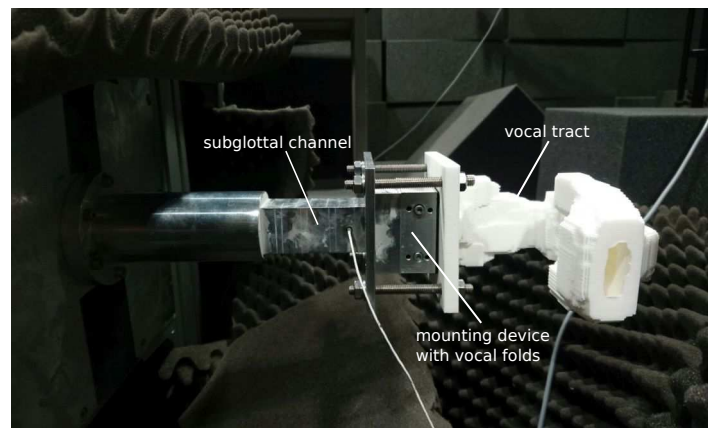
**Figure 3.** Cut through the 3D replica of (a) a realistic VT and (b) a simplified VT.

As an excitation source, a synthetic larynx model was used that includes two synthetic vocal folds made of silicone rubber as used in other studies [32,39–41]. Their geometry was derived from the M5 model as proposed by Scherer et al. [42] and Thomson et al. [43]. Each vocal fold was casted from

the silicone rubber compound *Ecoflex 30* (Smooth-On Inc., Macungie, PA, USA) and had a Young's modulus of  $E = 4.4 \text{ kPa}$  [44].

The experimental setup is based on the setup described by Kniesburges et al. [35] and Lodermeier et al. [39]. The flow is produced by a mass flow generator that applies a hypercritical valve [45]. This provides an adjustable, constant mass flow between 0 and 180 L/min, which is used as the constant driving parameter as recommended by Howe and McGowan [46]. Between the mass flow generator and the synthetic larynx model, a silencer is integrated in order to damp acoustic fluctuations in the inflow. The mounting device for the silicone vocal folds is assembled to the subglottal channel that has a rectangular cross section of 15 mm  $\times$  18 mm. All measurements were performed with the same pair of vocal folds. Upstream of the synthetic larynx, the VT replica is mounted. Figure 4 shows the setup with the subglottal channel, the synthetic vocal folds and a realistic VT replica.

Corresponding to Lodermeier et al. [39], the vocal folds oscillated without glottal closure just after the oscillation onset and turned into a mode with glottis closure after further increase of the subglottal pressure. As the periodic glottis closure is an important characteristic for physiological phonation, the measurements were carried out at the lowest subglottal pressure that provided stable oscillations with periodic closure. Therefore, the volume flow was decreased after oscillations when closure was established.



**Figure 4.** Experimental setup with subglottal channel, synthetic vocal folds and realistic VT replica.

### 2.3. Measuring Setup and Evaluation Methods

To investigate the formants of the VTs, the sound pressure was measured in an anechoic chamber by a 1/2" free-field microphone of type 4189 (Brüel & Kjaer, Nærum, Denmark). The sound signals were amplified by a *Nexus conditioning amplifier* (Brüel & Kjaer, Denmark) and sampled by the multifunctional module *NI PXIe-6356* (National Instruments, Austin, TX, USA) with a sample rate of 44.1 kHz. The microphone was located at a distance of 90 cm from the mouth exit of the VT in the sagittal plane with an inclination angle of 45°.

The sound pressure level (SPL) was calculated for each VT replica using a *Matlab* routine (Mathworks, USA) that is based on the Matlab function *pwelch*. The resulting power spectral density of the sound pressure was further converted into SPL. Thereby, a window length of 1 s was applied and the different windows were averaged in a subsequent step. The formant frequencies were detected using the *Aalto Aparat* Software (Aalto University, FI-00076 Aalto, Finland) [47,48]. This software tool has been applied in voice research before [49–51]. It is based on an automatic inverse filtering method to obtain the formant frequencies, which was also applied by Echternach et al. [34]. For formant detection, a partition of 1 s in the middle of the recorded sound signal was used.

### 3. Results and Discussion

#### 3.1. Oscillation Frequency and Mean Subglottal Pressure

The mean subglottal pressure  $p_{sub}$ , the volume flow rate  $\dot{V}$ , and the fundamental oscillation frequency  $f_0$  for the different VT replicas are listed in Table 1. The values were acquired for stable vocal fold oscillation at a subglottal pressure slightly above the physiological oscillation threshold. According to Table 1, the three parameters vary for the different VT replicas. The fundamental frequency is increased for the configurations with VT replicas in comparison to the configuration without VT. The shift of the fundamental frequency is less than 10% for all VT replicas, with the exception of the simplified replicas of subject 1 and subject 2. The simplified replica of subject 1 exhibits a decreased fundamental frequency by 7% and the fundamental frequency of the simplified subject 2 replica is increased by 51%. Between the realistic and simplified VT replicas no systematic variation of the fundamental frequency could be identified. Lucero et al. [27] reported that acoustic coupling between the vocal folds and the VT leads to frequency jumps when the oscillation frequency crosses a resonance frequency of the VT. However, they induced this effect with a variation of the VT length between 1.6 and 245.6 cm. As the length of the VT replicas used in this study differs by less than 5%, we assume that the variation of the fundamental frequency cannot be explained by this parameter. The deviation of the fundamental frequency for the simplified replica of subject 2 may be due to a change of the oscillation mode of the vocal folds. The back-coupling due to the specific VT geometry seems to be beneficial for that different modal behavior. However, the present measurement data is insufficient to prove that assumption.

**Table 1.** Physical parameters of the mean subglottal pressure  $p_{sub}$ , the volume flow rate  $\dot{V}$ , and the fundamental oscillation frequency  $f_0$  as measured in our experimental setup.

Configurations	$p_{sub}/\text{Pa}$	$\dot{V}/\text{L}/\text{min}$	$f_0/\text{Hz}$
without VT	4691	147	149
subject 1—realistic	4179	56	149
subject 2—realistic	4781	68	155
subject 3—realistic	3960	115	156
subject 4—realistic	4047	65	159
subject 5—realistic	4481	68	156
subject 6—realistic	3923	60	157
subject 1—simplified	2769	15	139
subject 2—simplified	5638	130	225
subject 3—simplified	4043	105	160
subject 4—simplified	3313	86	162
subject 5—simplified	2823	41	154
subject 6—simplified	2493	25	151
mean VT	3881	78	164

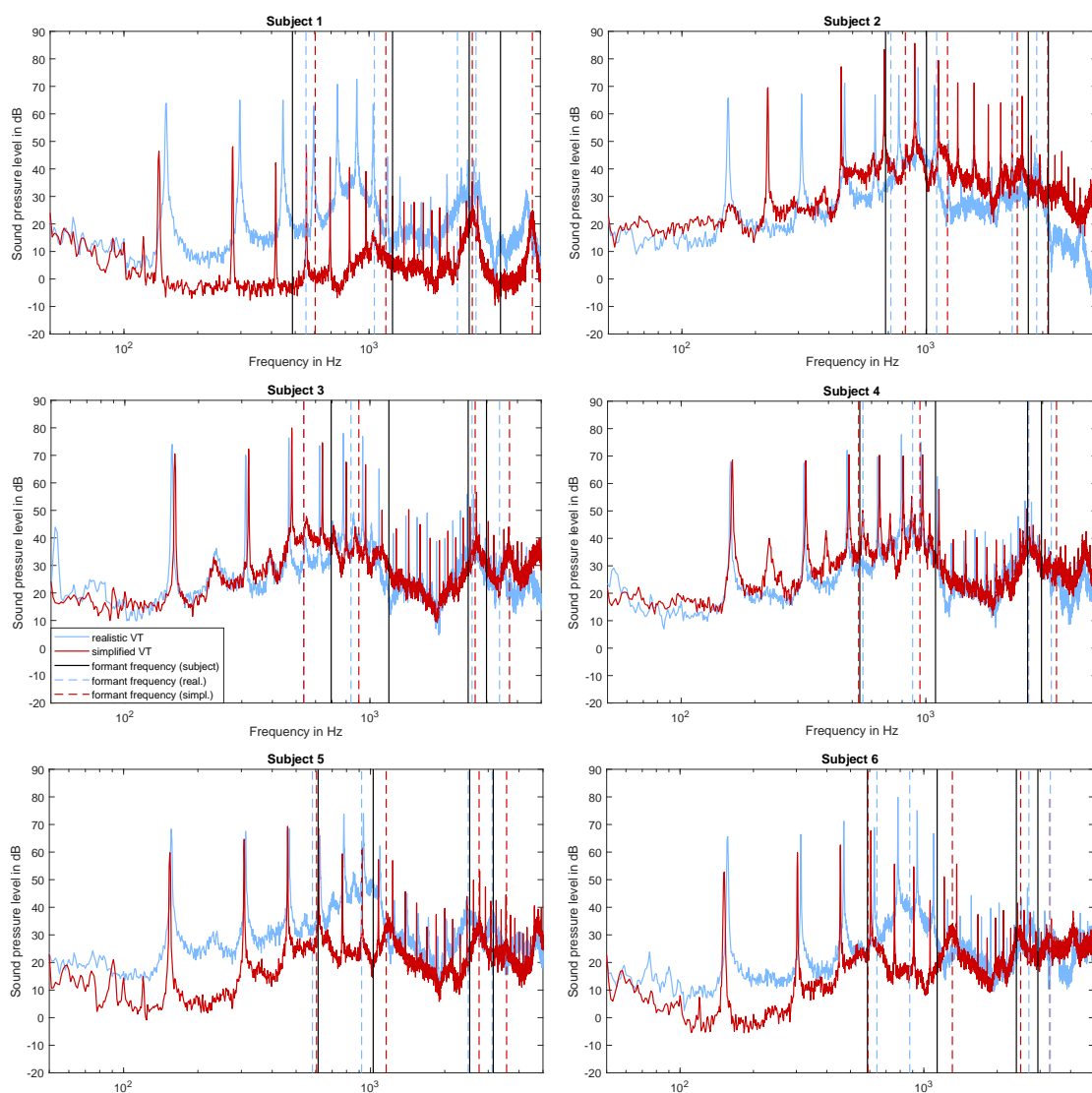
The volume flow rate and the subglottal pressure are reduced by the VT in comparison to the configuration without VT for all replicas with the exception of the realistic and simplified replica of subject 2, which exhibit an increase of 2% and 20%, respectively, for the subglottal pressure  $p_{sub}$  when related to the configuration without VT. The comparison of the subglottal pressure and the volume flow rate between the realistic and simplified replica shows no consistent behavior. For subject 1, subject 5 and subject 6, the volume flow rate and the subglottal pressure are reduced up to 73% and up to 37%, respectively, for the simplified replica. For the simplified replica of subject 2, both volume flow and subglottal pressure are increased by 18% and 91%, when respectively, compared to the realistic replica. For subject 3 and subject 4, the influence of the simplification on subglottal pressure and volume flow rate is different. Whereas for subject 3 the volume flow rate is decreased by 9% and the

subglottal pressure is slightly increased by 2%, for subject 4 the subglottal pressure is decreased by 18% and the volume flow rate is increased by 32%.

The detected reduction of the subglottal pressure and the volume flow rate for the configurations with VT replica in comparison to the configuration without VT shows that the phonation is facilitated by the VT. The reason for the reduced oscillation thresholds between configurations with and without VT is the decrease of pressure immediately downstream of vocal folds due to the channel effect as already reported in Kniesburges et al. [32,40].

### 3.2. Spectral Analysis of the Radiated Sound Pressure

Figure 5 depicts the SPL for the realistic and simplified VT replicas that were measured with the experimental setup for each subject. Each of the spectra shows peaks corresponding to the fundamental frequency, the corresponding harmonics and the broadband sound, which is the part of the spectrum without the energy of the harmonics.



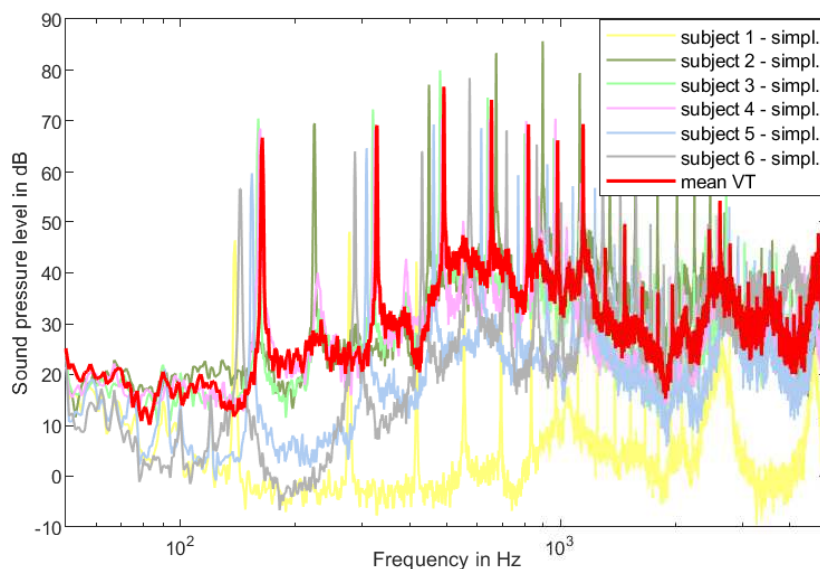
**Figure 5.** SPL of the sound emitted by vocal fold replicas coupled with MRI-based VT replicas from six subjects. Realistic geometries (blue curve) and simplified geometries with a concentric straight tube approximation (red curve) are shown for each individual. The formant frequencies are marked with vertical lines. The applied detection method for all formants is based on inverse filtering, as described in Section 2.3.



The comparison of the spectra of the realistic and the simplified replicas reveals that the SPL is changed by the simplification for all subjects, but to a different extent. Whereas the broadband sound level of the simplified replicas remains on a comparable level with their realistic counterpart for subject 2, subject 3, and subject 4, the broadband sound level of the simplified subject 1, subject 5, and subject 6 replicas is reduced. Hence, on average the simplified replicas exhibit a smaller broadband sound level than the realistic ones. This reduced broadband level correlates with a smaller volume flow rate or subglottal pressure in the experiments for these simplified VTs. As a result thereof, the global turbulence intensity may be expected smaller, thus creating less turbulence-induced broadband sound.

Investigations on the influence of a geometry simplification on the radiation using models with pure acoustic excitation have so far led to the result that the influence up to a frequency of 5 kHz is perceptually not relevant [18,52]. However, since we found that the broadband sound was changed for the simplified replica for three out of six subjects, it can be assumed that this presumption is only conditionally valid, although it has to be mentioned that the studies [18,52] only analyzed the transfer function of the VT. While there is proof of negligible influence of the simplification on the acoustic resonance behavior [18,52], there may be a significant change of the flow field that creates spectral differences.

The spectral response of the mean VT is displayed in Figure 6 in addition to the spectra of the simplified VT replicas. The spectra show different levels of broadband sound, which correlate with differences of 115 L/min in the required flow rates for the different simplified VT replicas. The spectrum of the mean VT replica is in the range of the replicas showing a higher broadband sound. As the required volume flow rate of the mean VT replica is also in the range of these replicas, the shift of the broadband sound of the mean VT to a higher level in comparison to the average of the broadband sound levels can be attributed to the higher volume flow rate.



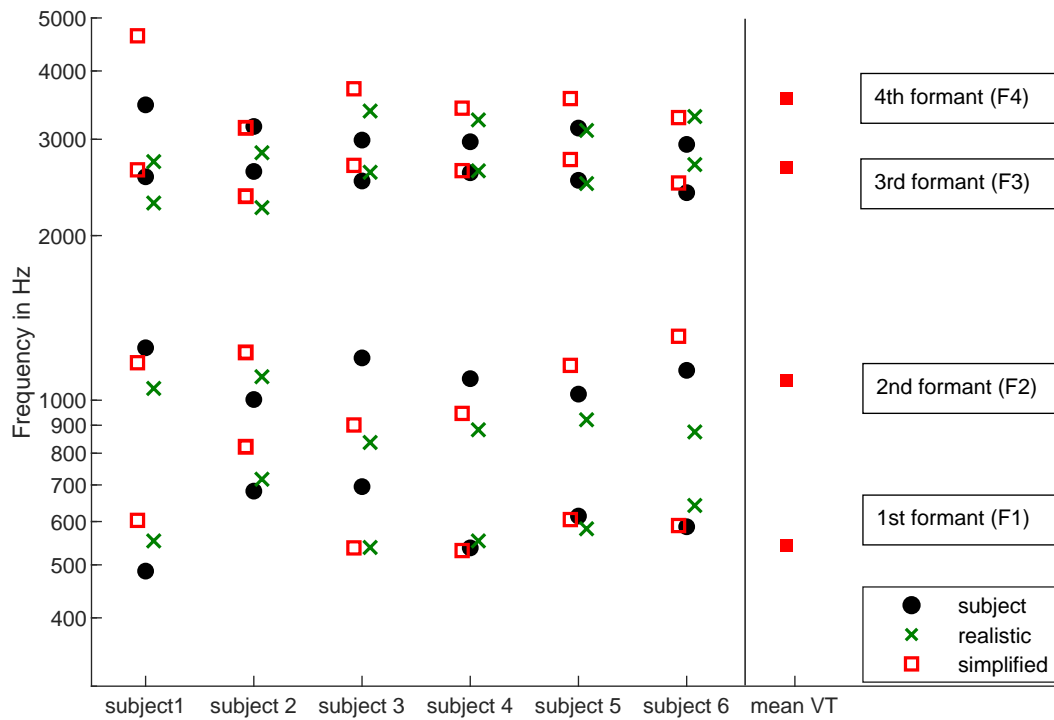
**Figure 6.** SPL of the sound emitted by vocal fold replicas coupled with simplified VT replicas from six subjects, based on the MRI-data by Echternach et al. [34] in comparison with the mean VT replica. The geometries of the simplified VTs are based on the cross-sectional area function of the realistic MRI-derived geometries. The simplified VT was straightened and the cross-section was designed with a circular form.

### 3.3. Analysis of the Formant Frequencies

The formants as detected with *Aalto Aparat* are marked in Figure 5 for both replica types with dashed lines. For comparison, the subjects' formant frequencies, as detected from the tenors' audio signals with inverse filtering by Echternach et al. [34], are added with black lines. It can be seen that for each VT replica four formant frequencies were detected in the examined frequency range from 50 Hz to

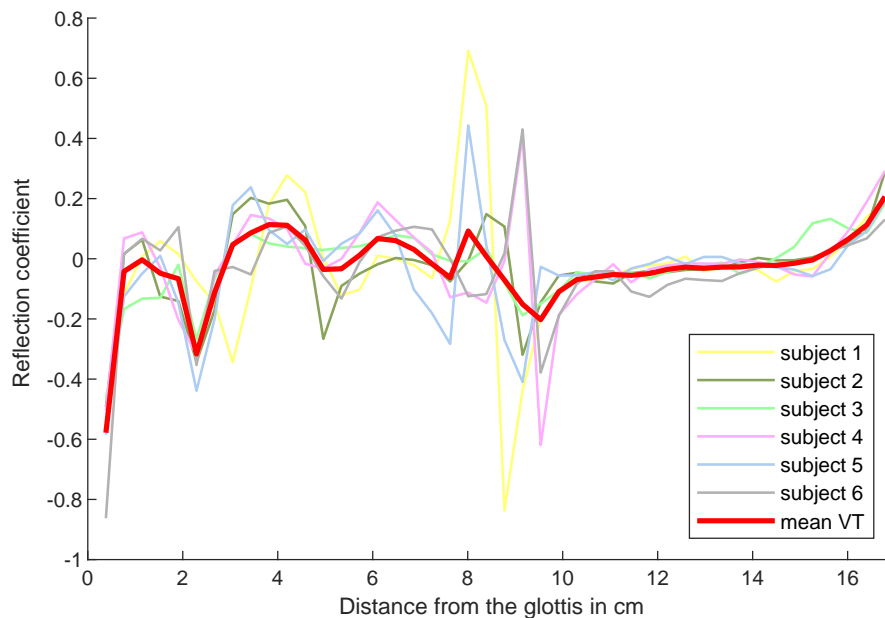
5 kHz that can be assigned to the subjects' formant frequencies. An overview of the detected formant frequencies in comparison to the subjects' formant frequencies is given in Figure 7. It shows that, in general, the replicas of all six tenors exhibit similar trends. The comparison between the realistic and the simplified replicas reveals that for five out of six subjects, the formant frequencies of the simplified VT replica are shifted to higher frequencies. In the previous study of Echternach et al. [34], which yielded the geometries of the VTs applied in our study, the subjects' formants frequencies were detected by inverse filtering of audio recordings. In addition, they determined the area function of the VTs and, again, computed the formant frequencies with the custom made software *FORMFLEK* [53,54], which is based on a mathematical model calculating the transfer function. They found that these formant frequencies were shifted to lower values in comparison to the subjects' formant frequencies detected with inverse filtering. Similar observations were made in a computational model with acoustical excitation by Arnela et al. [18] who found that the straightening of the VT led to a formant shift to lower frequencies of less than 5% in the frequency range below 4 kHz. These previous studies with acoustic excitation only indicate that the bend of the VT is acoustically not relevant for frequencies below 4 kHz. This is contrary to the results obtained in this study, which exhibit a mean deviation of 13% between the formant frequencies of the simplified and realistic VT replicas. By including the internal flow, we assume that a change of the flow field in the VT (realistic vs. simplified) leads to a variation of the flow acoustics, which also influences the formant frequencies. Additionally, the opening of the vocal folds, which is induced by the flow, may decrease the reflection coefficient at the glottis and periodically change the acoustic properties of the VT. This effect has so far not been taken into account in acoustic simulations. Another possible explanation for the difference between our results and the results of Arnela et al. [18] are differences in the simplification process. Whereas the simplified VT replicas in this study contain between 34 and 37 cross sections, the number of cross sections used by Arnela et al. [18] is 80. Furthermore, the differences between the formant frequencies of the realistic and simplified VT replicas of the subjects 3 and 4 are in the same order of magnitude as those found by Arnela et al. [18], which are based on a single subject. The difference would arguably be greater if the VT of an additional subject would have been included in that study.

In order to elaborate the difference between the resonance properties of the VT replicas and the tenor's VTs, the relative deviations of the four replica formants from the subjects' formant frequencies are depicted in Figure 9. Therein, the mean deviation of the formant frequencies for both, realistic and simplified VT replicas, is 12%, with a range of 1% to 30% for the realistic VTs and a range of 1% to 35% for the simplified VT replicas. Thereby, the deviations of F1 are less than 5% for both replica versions of subject 4, subject 5, and subject 6 whereas subject 1, subject 2, and subject 3 show larger deviations. On average, the formant frequencies of the realistic VT replicas are shifted to lower frequencies in comparison to the subjects' formant frequencies, as depicted in Figure 7. We suppose that the deviations of the formant frequencies of the realistic replicas in comparison to the subjects' formant frequencies are due to the rigid walls of the VT replicas. This effect of a frequency shift to lower frequencies at rigid walls compared with reflection at tissue was also reported by Fleischer et al. [55] and Kitamura et al. [14]. Another possible explanation for the appearing deviations are potential inaccuracies of the VT air volume due to the segmentation process. Nevertheless, the filter properties of both VT types are reproduced in an acceptable range, since a vowel formant is not characterized by a discrete frequency, but by a frequency band [56]. Hence, the simplification of the VT in our replica setup shows a good validity regarding this basic formant characteristics. This finding is similar to other studies [13,18].



**Figure 7.** Formant frequencies of the realistic and the simplified VT replicas based on the MRI-data by Echtertnach et al. [34] originating from six tenors singing the vowel /a/. Additionally, the subjects' formant frequencies are plotted as detected by Echtertnach et al. [34] from audio recordings. The applied detection method for all formants is based on inverse filtering, as described in Section 2.3.

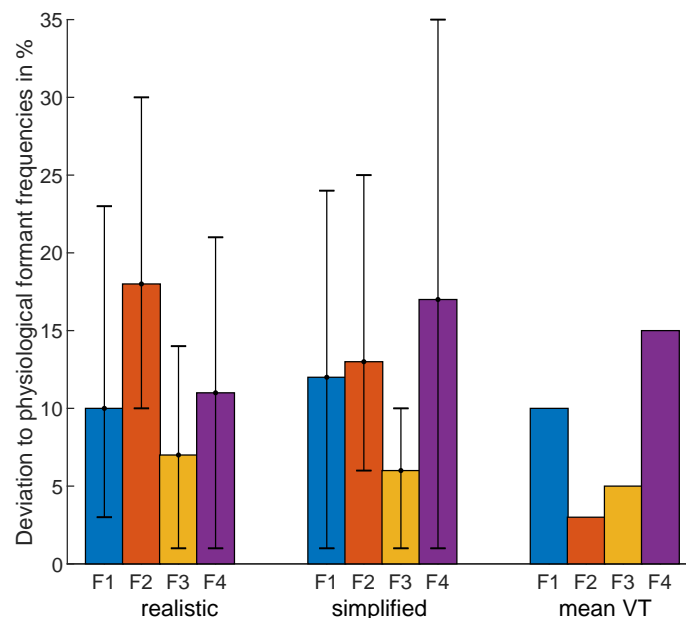
As the generation of the formant frequencies depends on the gradient of the reflection coefficient, we analyzed the reflection coefficients of the simplified and the mean VT replica, see Figure 8.



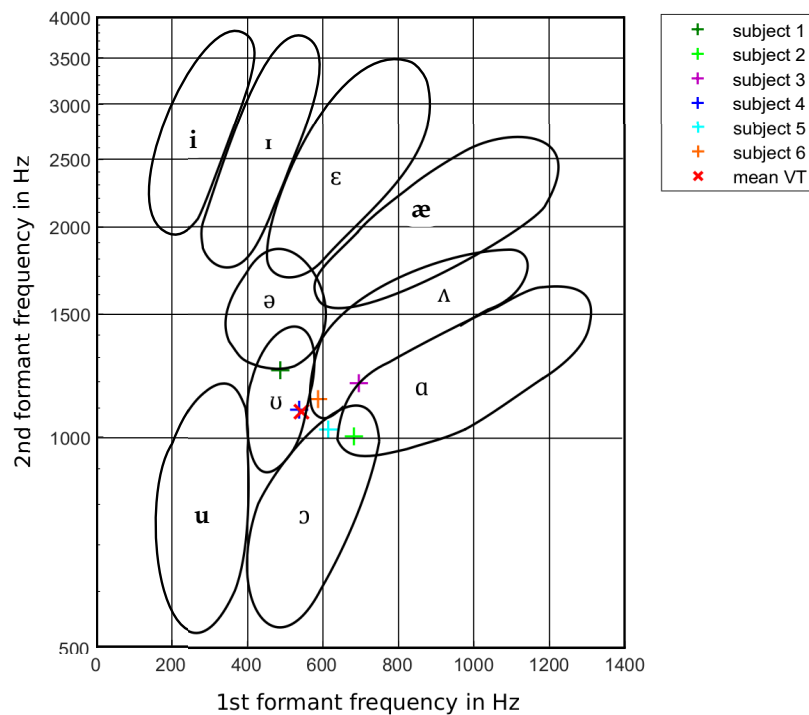
**Figure 8.** Reflection coefficients of the mean and individual simplified VTs as a function of distance to glottis. The geometries for the VT replicas were obtained according to the method described in Figure 2 and Section 2.1.

It shows that the slopes of the reflection coefficient are similar for the simplified replicas of all subjects. Furthermore, the reflection coefficients of the individual VT replicas are well reproduced by the reflection coefficient of the mean VT replica with increasing distance to the VT inlet. Only in the region between 8 and 10 cm, deviations up to 100% occur. These deviations are produced by the uvula, which is subject to large geometrical inter-individual differences. It is visible that for subject 1, subject 4, subject 5, and subject 6, the reflection coefficient in this area is larger than for subject 2, subject 3, and the mean VT replica due to a more pronounced constriction of the VT. However, no correlation between the degree of the VT narrowing and the formant frequencies of the simplified VT replicas can be observed.

The formants of the mean VT replica exhibit a mean deviation of 8% from the average of the subjects' formant frequencies, with a deviation of 10% for F1 and 3% for F2, as depicted in Figure 9. The detected formant frequencies of the mean VT are plotted in the formant chart in Figure 10, as proposed by Peterson and Barney [56], in addition to the subjects' formant frequencies. The chart shows F1 and F2 for different vowels, since the speech identification of a certain vowel depends mainly on the first two formant frequencies [10]. The subjects' formant frequencies are all located within the same frequency range, although F1 is shifted to lower frequencies towards the vowel /u/. This shift originates from the fact that professional singers tune their vowels to a darker voice quality, as described by Sundberg [57]. A classification of F3 and F4 is done based on literature values for men by Flanagan [58], Story et al. [13] and Sundberg [9]. The values of F3 are in the range of 2440 Hz for a spoken /a/ [58] and in the range of 2700 Hz for a sung /a/ [9]. According to Sundberg [9], F4 is in the order of 2750 Hz for a dark voice quality and about 3150 Hz for a light voice quality. F1 and F2 of the mean VT replica are located in the same region as the subjects' formant frequencies, especially considering the fact that a formant corresponds to a frequency band. F3 and F4 of the mean VT replica were found to be 2660 Hz and 3558 Hz, respectively, which match the reference values for a sung /a/ with light voice quality as typical for tenor singers.



**Figure 9.** Absolute mean relative deviations of the formant frequencies (F1, F2, F3, F4) of the realistic and simplified VT replicas to the subjects' formant frequencies [34] based on the MRI-data and audio recordings by Echternach et al. [34] of six tenors singing the vowel /a/. The errorbars mark the minimum and maximum relative deviation appearing among the subjects for the particular formant. The applied detection method for all formants is based on inverse filtering, as described in Section 2.3.



**Figure 10.** The formant chart as proposed by Peterson and Barney [56], shows the formant frequencies for the mean VT, which is based on the MRI-data of six professional tenors. In addition, formant frequencies of the individual tenors are shown, as detected by Echternach et al. [34] in the audio recordings.

#### 4. Conclusions

The aim of this study was to examine the influence of the VT on the phonatory process and to analyze if the influence of a geometry simplification can also be neglected considering the flow field in the VT. Based on the MRI-data of six professional tenors, 3D-printed replicas of the VT with realistic and simplified geometry were generated for each subject. The latter are simplified in terms of a straightened centerline and a conversion into a circular cross section based on the subjects' area functions. Additionally, we computed an averaged VT replica based on the simplified geometries to reduce effects due to individual characteristics and analyzed all VT replicas in an experimental setup including auto-oscillating vocal folds.

The results of the aerodynamical investigations show that the inclusion of the VT leads to a reduced phonation threshold pressure in combination with a decreased flow rate. This confirmed the already reported facilitation of the vocal fold oscillation by the VT [32,40].

The results of the acoustical investigation show that both the realistic and the simplified VT replicas reproduce the typical formant structure of the human VT. However, the broadband sound level is changed for the simplified VT. We assume that this is due to a change of the flow field in the VT and the resulting broadband sound due to the simplification of the geometry. This is contrary to the results of previous studies with purely acoustically excited VT models that did not include oscillating vocal folds and reported a negligible effect due to the simplification [18,52].

The SPL spectrum of the mean VT reproduces the basic characteristics of the individual simplified VT replicas. The deviation of the formant frequencies of the mean VT replica from the average subjects' formant frequencies exhibits a mean deviation of 8%. Appearing deviations from the individuals spectra in terms of the broadband sound level are attributed to the differences between the subjects in the VT constriction in the region of the uvula. Hence, averaging individual tenor VT geometries preserves the basic formant distribution and reduces individual acoustic characteristics. However,

according to previous literature, this only holds for VT shapes singing in the falsetto register, as used here [17].

The comparison of the occurring formant frequencies with the subjects' formant frequencies shows that the agreement is better for the realistic VT replicas than for the simplified VT replicas. Occurring deviations of the formant frequencies, which exhibit an average of 12%, are attributed to the rigid walls of the replica. However, as a vowel is not characterized by discrete frequencies, but by frequency bands, the spectral properties are preserved for both the realistic and the simplified VT replicas.

The comparison of the formant frequencies of the mean VT with subjects' formant frequencies shows that the formant characteristics were maintained despite the averaging. This confirms that by averaging the simplified individual VT geometries, a VT model is created that preserves the essential vowel characteristics of the VT without exhibiting strong individual characteristics.

The influence of the VT on the oscillation of the vocal folds in terms of volume flow rate and subglottal pressure reduction shows that the VT not only filters the basic sound generated by the vocal fold oscillation, but also influences the generation of the basic sound itself. This shows that the already reported coupling [2,28] between the VT and the vocal folds can also be reproduced in an experimental setup. Hence, the aim for future work is to investigate potential nonlinear coupling effects between VT and vocal folds.

However, our chosen approach also shows some limitations. The mean VT model is based on an overall number of six subjects, which is rather small. This was the reason why we only used MRI datasets obtained from professional tenor singers, which show reduced inter-subjective variations. These singers are able to sing a vowel constantly and in a reproducible way regarding pitch and register. Furthermore, the VT replicas consist of acoustically hard materials, which is assumed to be not the case in vivo. Thus, it will be the content of future work to first evaluate the acoustic reflection characteristics of VT tissue and in a second step to reproduce these characteristics with synthetic rubber materials. Nevertheless, the presented results exhibit good agreements with the physiologically expected and measured formant frequencies showing the validity of this study.

**Author Contributions:** Conceptualization, S.K. and A.L.; methodology, J.P., A.L. and S.F.; software, A.L.; validation, J.P., A.L. and S.K.; formal analysis, J.P. and A.L.; investigation, J.P., A.L. and S.K.; resources, M.E., B.R. and S.B.; data curation, J.P.; writing—original draft preparation, J.P.; writing—review and editing, A.L. and S.K.; visualization, J.P.; supervision, S.B. and M.D.; project administration, S.K.; funding acquisition, S.K.

**Funding:** This research was funded by Else Kröner-Fresenius Stiftung under grant agreement number 2016 A78. The authors also gratefully acknowledge the support of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative.

**Acknowledgments:** We thank Sebastian Trunk, Chair of Chemical Engineering, Friedrich-Alexander-University Erlangen-Nürnberg, Germany for his help with the production of the 3D-printed vocal tracts.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Titze, I. *Principles of Voice Production*; Prentice Hall: Englewood Cliffs, NJ, USA, 1994.
2. Titze, I.R. Nonlinear source–filter coupling in phonation: Theory. *J. Acoust. Soc. Am.* **2008**, *123*, 1902–1915. [[CrossRef](#)] [[PubMed](#)]
3. Kniesburges, S.; Thomson, S.L.; Barney, A.; Triep, M.; Sidlof, P.; Horacek, J.; Brucker, C.; Becker, S. In vitro experimental investigation of voice production. *Curr. Bioinform.* **2011**, *6*, 305–322. [[CrossRef](#)] [[PubMed](#)]
4. Alipour, F.; Brucker, C.; Cook, D.D.; Gommel, A.; Kaltenbacher, M.; Mattheus, W.; Mongeau, L.; Nauman, E.; Schwarze, R.; Tokuda, I.; et al. Mathematical models and numerical schemes for the simulation of human phonation. *Curr. Bioinform.* **2011**, *6*, 323–343. [[CrossRef](#)]
5. Mittal, R.; Erath, B.D.; Plesniak, M.W. Fluid dynamics of human phonation and speech. *Annu. Rev. Fluid Mech.* **2013**, *45*, 437–467. [[CrossRef](#)]

6. Döllinger, M.; Kaltenbacher, M. Preface: Recent Advances in Understanding the Human Phonatory Process. *Acta Acust. United Acust.* **2016**, *102*, 195–208. [[CrossRef](#)]
7. Zhang, Z. Mechanics of human voice production and control. *J. Acoust. Soc. Am.* **2016**, *140*, 2614–2635. [[CrossRef](#)] [[PubMed](#)]
8. Stevens, K.N. *Acoustic Phonetics*; MIT Press: Cambridge, MA, USA, 2000; Volume 30.
9. Sundberg, J. Formant structure and articulation of spoken and sung vowels. *Folia Phoniatr. Logop.* **1970**, *22*, 28–48. [[CrossRef](#)]
10. Wendler, J. *Lehrbuch der Phoniatrie und Pädaudiologie*; Georg Thieme Verlag: Stuttgart, Germany, 2005.
11. Fant, G. *Acoustic Theory of Speech Production*; Walter de Gruyter: Berlin, Germany, 1970; Number 2.
12. Mermelstein, P. Articulatory model for the study of speech production. *J. Acoust. Soc. Am.* **1973**, *53*, 1070–1082. [[CrossRef](#)] [[PubMed](#)]
13. Story, B.H.; Titze, I.R.; Hoffman, E.A. Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Am.* **1996**, *100*, 537–554. [[CrossRef](#)]
14. Kitamura, T.; Takemoto, H.; Adachi, S.; Honda, K. Transfer functions of solid vocal-tract models constructed from ATR MRI database of Japanese vowel production. *Acoust. Sci. Technol.* **2009**, *30*, 288–296. [[CrossRef](#)]
15. Aalto, D.; Aaltonen, O.; Happonen, R.P.; Jääsaari, P.; Kivelä, A.; Kuorrtti, J.; Luukinen, J.M.; Malinen, J.; Murtola, T.; Parkkola, R.; et al. Large scale data acquisition of simultaneous MRI and speech. *Appl. Acoust.* **2014**, *83*, 64–75. [[CrossRef](#)]
16. Echternach, M.; Sundberg, J.; Arndt, S.; Markl, M.; Schumacher, M.; Richter, B. Vocal tract in female registers—A dynamic real-time MRI study. *J. Voice* **2010**, *24*, 133–139. [[CrossRef](#)] [[PubMed](#)]
17. Echternach, M.; Sundberg, J.; Markl, M.; Richter, B. Professional opera tenors' vocal tract configurations in registers. *Folia Phoniatr. Logop.* **2010**, *62*, 278–287. [[CrossRef](#)] [[PubMed](#)]
18. Arnela, M.; Dabbaghchian, S.; Blandin, R.; Guasch, O.; Engwall, O.; Van Hirtum, A.; Pelorson, X. Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds. *J. Acoust. Soc. Am.* **2016**, *140*, 1707–1718. [[CrossRef](#)] [[PubMed](#)]
19. Blandin, R.; Arnela, M.; Laboissière, R.; Pelorson, X.; Guasch, O.; Hirtum, A.V.; Laval, X. Effects of higher order propagation modes in vocal tract like geometries. *J. Acoust. Soc. Am.* **2015**, *137*, 832–843. [[CrossRef](#)] [[PubMed](#)]
20. Takemoto, H.; Mokhtari, P.; Kitamura, T. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. *J. Acoust. Soc. Am.* **2010**, *128*, 3724–3738. [[CrossRef](#)] [[PubMed](#)]
21. Arnela, M.; Guasch, O.; Alías, F. Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations. *J. Acoust. Soc. Am.* **2013**, *134*, 2946–2954. [[CrossRef](#)] [[PubMed](#)]
22. Arnela, M.; Blandin, R.; Dabbaghchian, S.; Guasch, O.; Alías, F.; Pelorson, X.; Van Hirtum, A.; Engwall, O. Influence of lips on the production of vowels based on finite element simulations and experiments. *J. Acoust. Soc. Am.* **2016**, *139*, 2852–2859. [[CrossRef](#)]
23. Traser, L.; Burdumy, M.; Richter, B.; Vicari, M.; Echternach, M. The effect of supine and upright position on vocal tract configurations during singing—A comparative study in professional tenors. *J. Voice* **2013**, *27*, 141–148. [[CrossRef](#)]
24. Traser, L.; Burdumy, M.; Richter, B.; Vicari, M.; Echternach, M. Weight-bearing MR imaging as an option in the study of gravitational effects on the vocal tract of untrained subjects in singing phonation. *PLoS ONE* **2014**, *9*, e112405. [[CrossRef](#)]
25. Echternach, M.; Markl, M.; Richter, B. Dynamic real-time magnetic resonance imaging for the analysis of voice physiology. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2012**, *20*, 450–457. [[CrossRef](#)] [[PubMed](#)]
26. Sundberg, J. *The Science of the Singing Voice*; Northern Illinois Press: DeKalb, IL, USA, 1987.
27. Lucero, J.C.; Lourenço, K.G.; Hermant, N.; Van Hirtum, A.; Pelorson, X. Effect of source–tract acoustical coupling on the oscillation onset of the vocal folds. *J. Acoust. Soc. Am.* **2012**, *132*, 403–411. [[CrossRef](#)] [[PubMed](#)]
28. Titze, I.; Riede, T.; Popolo, P. Nonlinear source–filter coupling in phonation: Vocal exercises. *J. Acoust. Soc. Am.* **2008**, *123*, 1902–1915. [[CrossRef](#)] [[PubMed](#)]
29. Birk, V.; Sutor, A.; Döllinger, M.; Bohr, C.; Kniesburges, S. Acoustic impact of ventricular folds on phonation studied in ex vivo human larynx models. *Acta Acust. United Acust.* **2016**, *102*, 244–256. [[CrossRef](#)]

30. Zheng, X.; Bielamowicz, S.; Luo, H.; Mittal, R. A computational study of the effect of false vocal folds on glottal flow and vocal fold vibration during phonation. *Ann. Biomed. Eng.* **2009**, *37*, 625–642. [[CrossRef](#)] [[PubMed](#)]
31. Alipour, F.; Jaiswal, S.; Finnegan, E. Aerodynamic and acoustic effects of false vocal folds and epiglottis in excised larynx models. *Ann. Otol. Rhinol. Laryngol.* **2007**, *116*, 135–144. [[CrossRef](#)]
32. Kniesburges, S.; Birk, V.; Lodermeier, A.; Schützenberger, A.; Bohr, C.; Becker, S. Effect of the ventricular folds in a synthetic larynx model. *J. Biomech.* **2017**, *55*, 128–133. [[CrossRef](#)]
33. Horáček, J.; Uruba, V.; Radolf, V.; Veselý, J.; Bula, V. Airflow visualization in a model of human glottis near the self-oscillating vocal folds model. *Appl. Comput. Mech.* **2011**, *5*, 21–28.
34. Echternach, M.; Sundberg, J.; Baumann, T.; Markl, M.; Richter, B. Vocal tract area functions and formant frequencies in opera tenors' modal and falsetto registers. *J. Acoust. Soc. Am.* **2011**, *129*, 3955–3963. [[CrossRef](#)]
35. Kniesburges, S.; Hesselmann, C.; Becker, S.; Schlücker, E.; Döllinger, M. Influence of vortical flow structures on the glottal jet location in the supraglottal region. *J. Voice* **2013**, *27*, 531–544. [[CrossRef](#)]
36. 3D Slicer. Available online: <http://www.slicer.org> (accessed on 07 August 2018).
37. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [[CrossRef](#)] [[PubMed](#)]
38. Traser, L.; Birkholz, P.; Flügge, T.V.; Kamberger, R.; Burdumy, M.; Richter, B.; Korvink, J.G.; Echternach, M. Relevance of the implementation of teeth in three-dimensional vocal tract models. *J. Speech Lang. Hear. Res.* **2017**, *60*, 2379–2393. [[CrossRef](#)] [[PubMed](#)]
39. Lodermeier, A.; Becker, S.; Döllinger, M.; Kniesburges, S. Phase-locked flow field analysis in a synthetic human larynx model. *Exp. Fluids* **2015**, *56*, 77. [[CrossRef](#)]
40. Kniesburges, S.; Lodermeier, A.; Becker, S.; Traxdorf, M.; Döllinger, M. The mechanisms of subharmonic tone generation in a synthetic larynx model. *J. Acoust. Soc. Am.* **2016**, *139*, 3182–3192. [[CrossRef](#)] [[PubMed](#)]
41. Lodermeier, A.; Tautz, M.; Becker, S.; Döllinger, M.; Birk, V.; Kniesburges, S. Aeroacoustic analysis of the human phonation process based on a hybrid acoustic PIV approach. *Exp. Fluids* **2018**, *59*, 13. [[CrossRef](#)]
42. Scherer, R.C.; Shinwari, D.; De Witt, K.J.; Zhang, C.; Kucinski, B.R.; Afjeh, A.A. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *J. Acoust. Soc. Am.* **2001**, *109*, 1616–1630. [[CrossRef](#)] [[PubMed](#)]
43. Thomson, S.L.; Mongeau, L.; Frankel, S.H. Aerodynamic transfer of energy to the vocal folds. *J. Acoust. Soc. Am.* **2005**, *118*, 1689–1700. [[CrossRef](#)] [[PubMed](#)]
44. Rupitsch, S.J.; Ilg, J.; Sutor, A.; Lerch, R.; Döllinger, M. Simulation based estimation of dynamic mechanical properties for viscoelastic materials used for vocal fold models. *J. Sound Vib.* **2011**, *330*, 4447–4459. [[CrossRef](#)]
45. Durst, F.; Heim, U.; Ünsal, B.; Kullik, G. Mass flow rate control system for time-dependent laminar and turbulent flow investigations. *Meas. Sci. Technol.* **2003**, *14*, 893. [[CrossRef](#)]
46. Howe, M.; McGowan, R. Voicing produced by a constant velocity lung source. *J. Acoust. Soc. Am.* **2013**, *133*, 2340–2349. [[CrossRef](#)]
47. Alku, P.; Pohjalainen, H.; Airaksinen, M. Aalto Aparat—A freely available tool for glottal inverse filtering and voice source parameterization. In Proceedings of the Subsida: Tools and Resources for Speech Sciences, Malaga, Spain, 21–23 June 2017.
48. Pohjalainen, H.; Airaksinen, M.; Airas, M.; Alku, P. Aalto Aparat—Manual v2.0. Available online: <http://research.spa.aalto.fi/projects/aparat/AaltoAparatManual.pdf> (accessed on 30 August 2019).
49. Airas, M. TKK Aparat: An environment for voice inverse filtering and parameterization. *Logop. Phoniater. Vocol.* **2008**, *33*, 49–64. [[CrossRef](#)] [[PubMed](#)]
50. Vainio, M.; Airas, M.; Järvikivi, J.; Alku, P. Laryngeal voice quality in the expression of focus. In Proceedings of the Eleventh Annual Conference of the International Speech Communication Association, Chiba, Japan, 26–30 September 2010.
51. Kohler, M.; Vellasco, M.M.; Cataldo, E. Analysis and classification of voice pathologies using glottal signal parameters. *J. Voice* **2016**, *30*, 549–556.
52. Matsuzaki, H.; Motoki, K.; Miki, N. A study of the simplification of the three-dimensional vocaltract model using finite element method. In Proceedings of the 18th International Congress on Acoustics (ICA), Kyoto, Japan, 4–9 April 2004; Volume 1.



53. Liljencrants, J.; Fant, G. Computer program for VT-resonance frequency calculations. *STL-QPSR* **1975**, *16*, 15–21.
54. Sundberg, J.; Lindblom, B.; Liljencrants, J. Formant frequency estimates for abruptly changing area functions: A comparison between calculations and measurements. *J. Acoust. Soc. Am.* **1992**, *91*, 3478–3482. [[CrossRef](#)] [[PubMed](#)]
55. Fleischer, M.; Pinkert, S.; Mattheus, W.; Mainka, A.; Mürbe, D. Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall. *Biomech. Model. Mechanobiol.* **2015**, *14*, 719–733. [[CrossRef](#)] [[PubMed](#)]
56. Peterson, G.E.; Barney, H.L. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* **1952**, *24*, 175–184. [[CrossRef](#)]
57. Sundberg, J. The acoustics of the singing voice. *Sci. Am.* **1977**, *236*, 82–91. [[CrossRef](#)]
58. Flanagan, J.L. *Speech Analysis Synthesis and Perception*; Springer: Berlin, Germany, 1972; Volume 2.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).