

This is the accepted manuscript of the article: Saldías, M., Guzman, M., Miranda, G., Laukkanen, AM., Stoney, J., Alku, P. & Sundberg, J. (2021) The Vocal Tract in Loud Twang-Like Singing While Producing High and Low Pitches, which has been published in *Journal of Voice*, 35 (5): 807.e1-807.e23 (Available online 15 April 2020). <https://doi.org/10.1016/j.jvoice.2020.02.005>.

© <2021>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## The vocal tract in loud twang-like singing while producing high and low pitches

Marcelo Saldías, Anne-Maria Laukkanen, Marco Guzmán, Gonzalo Miranda, Justin Stoney, Paavo Alku, Johan Sundberg.

**Abstract:** Twang-like vocal qualities have been related to a megaphone-like shape of the vocal tract (epilaryngeal tube and pharyngeal narrowing, and a wider mouth opening), low-frequency spectral changes, and tighter/increased vocal fold adduction. Previous studies have focused mainly on loud and high-pitched singing, comfortable low-pitched spoken vowels or are based on modeling and simulation. There is no data available related to twang-like voices in loud, low-pitched singing. **Purpose:** This study investigates the possible contribution of the lower and upper vocal tract configurations during loud twang-like singing on high and low pitches in a real subject. **Methods:** One male contemporary commercial music singer produced a sustained vowel [a:] in his habitual speaking pitch (B2) and loudness. The same vowel was also produced in a loud twang-like singing voice on high (G4) and low pitches (B2). Computerized tomography, acoustic analysis, inverse filtering, and audio-perceptual assessments were performed. **Results:** Both loud twang-like voices showed a megaphone-like shape of the vocal tract, being more notable on the low pitch. Also, low-frequency spectral changes, a peak of sound energy around 3 kHz and increased vocal fold adduction were found. Results agreed with audio-perceptual evaluation. **Conclusions:** Loud twang-like phonation seems to be mainly related to low-frequency spectral changes (under 2 kHz) and a more compact formant structure. Twang-like qualities seem to require different degrees of twang-related vocal tract adjustments while phonating in different pitches. A wider mouth opening, pharyngeal constriction, and epilaryngeal tube narrowing may be helpful strategies for maximum power transfer and improved vocal economy in loud CCM singing and potentially in loud speech. Further studies should focus on vocal efficiency and vocal economy measurements using modeling and simulation, based on real-singers' data.

**Key Words:** Nonclassical singing – Vocal tract imaging – Voice source – Formant frequencies – Physiology of singing.

## **Short Title:** Vocal tract in loud twang-like phonation

### **Introduction**

Twang is a voice quality that has been perceptually related to a "bright," "ringy," "brassy," "metallic," or simply "twangy" voice production [1-5]. These voice qualities are widely used by some Contemporary Commercial Music (CCM) singers (e.g., pop, rock, country, and musical theatre) [3,5,6]. Nevertheless, some authors have considered that twang is also used by classical operatic singers [6,7]. Additionally, twang has been commonly related to loud and high-pitched singing [3] and loud phonation [2,4,5].

In voice sciences, the production of twang has been anatomically associated to lower vocal tract adjustments (Fig. 1.1): epilaryngeal tube narrowing and pharyngeal constriction [1,8,9]. Titze [1] states that pharyngeal narrowing differentiates twang from the production of operatic voice quality, where the pharynx is kept wider. Therefore, twang is expected to be a vocal quality used in CCM singing and not necessarily in classical singing. Classical singers need to project their voices over the sound of a symphony orchestra without electronic amplification, and this is accomplished with the aid of 'ring' by male singers and by females when singing in lower pitch range (below 500 Hz) [10]. Vocal "ring" in Western Classical singing style is the result of the amplification of high overtones in the range of 2-3 kHz due to a clustering of higher frequency formants, i.e., the third, fourth and fifth formants. This formant cluster is known as the singer's formant [10]. The singer's formant and, thus, vocal ring, is caused by a narrowed epilaryngeal tube and a wider pharynx [1]. In the classical singing style, the "ring" is combined with a dark and warm vocal quality due to the widened pharynx, and in general, a lengthened vocal tract due to a lowered position of the larynx [1,10]. This vocal tract configuration is used in the technique known as "covered singing" [11].

On the other hand, CCM singers typically use electronic amplification and thus do not need to be heard over an orchestra by using the operatic type of "ring" [1,12]. However, they can also produce a loud and "resonant" voice quality, as in the case

of twang, which is characterized by a brighter and whiter timbre than the operatic quality. Since both the epilaryngeal tube and the pharynx are expected to be narrowed in twang, the resulting “resonant” quality should not be related to “ring”, i.e., no formant cluster is expected around 2-3 kHz [1]. Nevertheless, some authors have found a spectral peak of energy in the singer’s formant region in CCM singers when singing on a high pitch and greater loudness (e.g., belting) [8,13,14]. This peak of energy has been related to the increased amplitude of the fourth formant (the actor’s or the speaker’s formant) [15]. A similar spectral peak has been found in country music singers when singing and speaking [16], and in actors with a good voice quality [15].

Sundberg and Thalén [3] found that the first and second formants ( $F_1$  and  $F_2$ ) increased in frequency whereas the third and fifth formants ( $F_3$  and  $F_5$ ) decreased when a female CCM singer sang in high pitches with twang, compared to a neutral voice quality. The authors concluded that the formant distribution observed would be the most relevant aspect related to the perception of “twanginess” and sound pressure level (SPL) increase. Similar changes in  $F_1$  and  $F_2$  have been previously found in a modeling study by Story et al. [9] in several spoken vowels produced with a twangy quality ([u:], [i:], [æ], and [a:]) in a comfortable, habitual speaking pitch and loudness. Therefore, the acoustical features of twang seem to result not only from a specific lower vocal tract configuration but of an entire vocal tract setting: a shortened vocal tract due to a higher laryngeal position and lip retraction, epilaryngeal and pharyngeal narrowing, wider lip and jaw opening, and a more constricted oral cavity due to an increased tongue height (Fig. 1.2) [3,5,9]. This overall anatomical description has been related to a “megaphone-like shape” of the vocal tract [17].

Voice source characteristics have been also found to be relevant in twang [3,18]. In a synthesis study, Titze et al. [18] found that a decrease in the open quotient estimated from the flow signal produced the most significant impact on the perception of twang, compared to the epilaryngeal tube and pharynx narrowing, and the shortening of the vocal tract length, respectively. Similarly, Sundberg and Thalén [3] observed through inverse filtering of the flow signal that twang was characterized

by a higher closed quotient (the ratio between the closed phase, i.e. zero flow duration of the flow signal period and the duration of the fundamental period), a lower AC flow amplitude (peak-to-peak flow amplitude), a lower value of  $H_1-H_2$  (the level difference between the source fundamental and the second source spectrum partial, in decibels) and a lower value of the normalized amplitude quotient (NAQ). Based in these results the authors suggested that twang was characterized by a greater adduction of the vocal folds compared to a neutral voice quality.

According to Sundberg and Thalén [3] a terminological confusion has emerged as it comes to the use of the word 'twang'. LoVetri [19] has suggested not to use the term "twang" (especially concerning belting) since it is associated with the timbre of country music, specifically in Nashville (in the USA). According to the Estill Voice model [5], twang is commonly used in Appalachian and Country music styles, and "it has also found its way" into other types of music (e.g., Gospel and R&B). In order to avoid confusion and respect the cultural background of twang, we will use the term "twang-like voice quality" from now on to refer to the voice that is characterized as bright, brassy or metallic, not necessarily representing the singers' voice quality in the country singing style.

To summarize the previous findings, twang-like voices seem to be related to both upper and lower vocal tract specific postures and greater vocal fold adduction (Fig. 1.2). So far, twang-like voice quality has been studied mainly in loud and high-pitched singing [3,8] and in comfortable low-pitched spoken vowels [9]. Also, most of the data available have been inferred from modeling and simulation studies [9,18]. Therefore, there is no data related to twang-like voices in loud, low-pitched singing. The present study aims to investigate the possible contribution of the lower and upper vocal tract configuration on loud twang-like singing on both high and low pitches in one CCM singer. Since twang-like voices seem to be related to multiple physiologic, physical and perceptual aspects, this study exploits vocal tract imaging, spectral and inverse filtering analysis, and audio-perceptual assessment to answer four questions:

1. Are there differences in the vocal tract and laryngeal setting between loud twang-like voice qualities on high and low pitches?
2. If there are differences, what spectral effects do they cause?
3. Are there differences in voice source, reflecting differences in glottal adduction?
4. What auditory-perceptual effects do the characteristics of the vocal tract and voice source cause?

Although we expect some anatomical similarities between loud twang-like productions on high and low pitches (a constricted epilaryngeal tube and pharynx, and a wide mouth opening and large oral cavity concordant with a megaphone-like shape of the vocal tract), prominent differences in the vocal tract setting are also expected during the high-pitched production, including a higher laryngeal position and wider mouth opening, due to pitch elevation [17,20]. Moreover, based on earlier results of high-pitched singing [20-22], we expect to find a more forward tilted thyroid cartilage and frontally upward tilted cricoid cartilage in the high-pitched twang-like voice quality compared to the low-pitched twang-like voice quality.

Acoustically, high  $F_1$  and  $F_2$ , and relatively low  $F_3$  and  $F_5$  are expected in both twang-like voices, resulting in a strong energy concentration at 1-2 kHz and high SPL [3].  $F_1$  and  $F_2$  are expected to be higher in the high pitch due to the commonly used resonance strategy in high pitches (e.g. greater jaw opening and higher laryngeal position) [13,17]. Also, a spectral peak of energy around 3kHz (increased acoustic energy in the singer's formant region) is expected in both twang-like voices due to epilaryngeal tube narrowing [8,13,14]. Furthermore, voice source characteristics reflecting relatively high glottal adduction are expected in both twang-like productions (e.g. lower values in NAQ and  $H_1$ - $H_2$ ) [3]. Finally, auditory perception of brightness and twanginess in the voice are expected to be found for twang-like voices in both pitches [18].

## Methods

### *Ethical considerations*

The present study was reviewed and approved by the University of Chile Hospital Review Board. Since the commonly used endoscopic assessment visualizes the pharynx and larynx from above, Computerized Tomography (CT) was used in the present study to obtain both high-resolution images of the sagittal and transversal planes and three-dimensional reconstruction of the vocal tract [23]. As CT procedures are related to health risks due to ionizing radiation [24] data were obtained from only one volunteer, a co-author of the present study who participated in a previous CT study [25]. The subject was a male semi-professional singer (32 years old), with a healthy voice condition and with more than three years of CCM vocal training. To confirm the normal vocal condition of the singer, a nasoendolaryngoscopic assessment (Olympus ENF-P4; Olympus, Center Valley, PA, USA) was performed by an experienced laryngologist.

### *Acquisition of CT images and audio samples*

Computerized tomography of the vocal tract was carried out using a SOMATON Sensation 64 (Siemens Healthcare, Erlangen, Germany) CT machine (100 kV of voltage, 0.4 seconds of time of rotation, and 1.2 mm of slice thickness) by a radiologist at the Department of Imaging and Radiology at the University of Chile Hospital. During image acquisition, the singer was in supine position inside the CT machine with his head stabilized by a frame, while performing three different vocal tasks: 1) a sustained vowel [a:] in a comfortable habitual speaking pitch and loudness, using a speaking voice quality (B<sub>2</sub>, 123.67 Hz); 2) a sustained vowel [a:] at a high pitch (G<sub>4</sub>, 394.29 Hz) and loudness, using a twang-like singing voice quality; and 3) a sustained vowel [a:] in the comfortable habitual speaking pitch (B<sub>2</sub>, 122.28 Hz) using a similar loudness and singing voice quality as in the high-pitched twang-like production. The pitch in the first task was comparable to what was used in the study by Story et al. [9]. In order to compare the speaking voice and the low-pitch twang-like singing voice, the same pitch was asked in both tasks.

Both high and low-pitched twang-like singing voice productions were required to be performed on vowel [a:] with an increased loudness, as similar as possible in both samples. During sampling, pitch was controlled by an electronic keyboard. Audio feedback was provided to the singer through headphones. In order to minimize the amount of radiation, vocal tasks were performed only once, for at least 15 seconds each. The singer rehearsed each task before CT image acquisition. In addition, one of the experimenters verified that the singer's voice productions matched the vocal tasks asked.

The twang-like voice quality on the high pitch seemed to represent 'belting'. According to McGlashan et al. [14], belting may be divided into at least two subtypes, which according to the pedagogical method 'Complete Vocal Technique' [6] are called 'Overdrive' and 'Edge'. Therefore, three blinded judges trained in Complete Vocal Technique (CVT™) assessed the audio samples. These judges agreed that the high-pitched twang-like voice sample represented Overdrive mode with reduced density (a belted-like sound with a shout-like character, that is "less than 50% filled out" or "with significantly less weight in the sound", resembling a mixed voice) [14,20].

Due to the increased noise produced by the CT machine at the moment of the image acquisition, audio samples were recorded previously using a professional condenser-omnidirectional microphone (Rode, Long Beach, CA; model NT2-A) at a distance of 30 cm from the singer's mouth, in an acoustically treated room in the Voice Research Laboratory at the University of Chile. The singer maintained an upright position during the recording of the same three tasks that were used for the CT procedure. Each task was recorded three times in a row. A USB audio interface (Focusrite Scarlett 8i-6; Focusrite Audio Engineering, High Wycombe, UK) and the software Pro Tools 9.0 (Avid Corporation, Burbank, CA) were used to digitally record the audio samples in WAV format (44.1 kHz of sampling rate; 16 bits). The input level was controlled to be the same for all three voice tasks.

## *CT image analysis*

An experienced head and neck radiologist and two speech-language pathologists specialized in vocology analyzed the CT images using OsiriX software (version 5.0.2, 64 bit, Pixmeo SARL, Bernex, Switzerland). We considered similar measures to those used previously by Saldías et al. [25] in a study of belting. Therefore, anatomic sagittal distances, sagittal cross-sectional areas, and epilaryngeal tube outlet and pharyngeal inlet areas were calculated (Fig.2 and Fig. 3). Additionally, since high-pitched belted-like voice qualities have been attributed to particular positioning of the laryngeal cartilages [20-22], the angles of inclination of the thyroid and cricoid cartilages were measured.

Since it is known that different singing styles may use different glottal length in the same pitches e.g. due to different amount of involvement of extralaryngeal muscles in the pitch control [26], glottal length was also measured for each voice sample. Midsagittal and transversal images were used to obtain the inclination of the laryngeal cartilages and the glottal length, respectively (Fig. 4). Finally, volumetric measures of the epilaryngeal tube were calculated. A total of seventeen anatomic measures were included in the analysis.

### a. Midsagittal measures

- *Anatomic distances (mm):* 1) VL - vertical length of the vocal tract (measured from the lowermost edge of the anterior arch of the first vertebra to the vocal folds), 2) HL - horizontal length of the vocal tract (measured from the lowermost edge of the anterior arch of the first vertebra to the narrowest point between the lips), 3) LO - lip opening (measured from the upper edge of the lower lip to the lower edge of the upper lip), 4) JO - jaw opening (measured from the lowermost edge of the jawbone to the anterior end of the hard palate), 5) TDH - tongue dorsum height (measured from the lowermost edge of the jaw bone to the uppermost point of the tongue dorsum), 6) OW - oropharynx width (measured from the lowermost edge of the second vertebra to the most posterior part of the tongue contour, following a line to the anterior uppermost edge of the jawbone), 7) VE - velum elevation (measured from the upper edge



of the hard palate end to the anterior lowest point of the uvula), and 8) HW - hypopharynx width (measured from the lowest point of the pharynx to the internal edge of the epiglottis, tangent to the inferior surface of the jawbone).

- *Sagittal cross-sectional areas (cm<sup>2</sup>):* 1) A<sub>1</sub> - oral cavity (measured from the lips to the velum), 2) A<sub>2</sub> - the pharyngeal region (measured from the A<sub>1</sub> ending to the line between the lower edge of opisthion in the occipital bone and the lowermost edge of the jawbone contour), and 3) A<sub>3</sub> - the epilaryngeal region (measured from the A<sub>2</sub> ending to the vocal folds).

The sagittal cross-sectional areas were defined broadly for a rough description, as has been done in earlier studies [25,27]. To be more exact, in defining A<sub>3</sub> the sagittal area of the epilaryngeal tube should be considered from the uppermost edge of the epiglottis cartilage to the uppermost edge of the arytenoid cartilages (Fig. 1.1). However, both epiglottic and arytenoid cartilages are movable structures, and their position changes differently depending on the pitch and voice quality. Hence, the limits of A<sub>3</sub> were determined based on an external and fixed structure (the occipital bone) for a better comparison between voice qualities. A more accurate measurement of A<sub>3</sub> is not supposed to alter the general trends observed in the present study. In addition, the reference points used to measure A<sub>3</sub> in the sagittal plane were different than those used for both measurements of the epilaryngeal tube outlet area and for calculations of the epilaryngeal tube volume. Therefore, transversal and volumetric measures are not influenced by the anatomical delimitations of A<sub>3</sub>.

b. Epilaryngeal tube outlet (A<sub>e</sub>) and pharyngeal inlet (A<sub>p</sub>) areas (cm<sup>2</sup>)

Both the A<sub>e</sub> and A<sub>p</sub> areas were measured from a reconstructed plane of the epilaryngeal tube outlet (Fig. 3). The reconstructed plane consisted of the estimated midline of the epilaryngeal tube. A<sub>e</sub> and A<sub>p</sub> were calculated perpendicular to the reconstructed plane.

- *Epilaryngeal tube outlet area ( $A_e$ )*: the outlet of the epilaryngeal tube was considered as the region just below the collar of the epiglottis, where the ring of the epilaryngeal folds is still fully visible.
- *Pharyngeal inlet area ( $A_p$ )*: the inlet of the lower pharynx was considered as the region just above the collar of the epiglottis, where the epilaryngeal ring is no longer visible.

Additionally, the ratio between the inlet of the lower pharynx and the outlet of the epilaryngeal tube ( $A_p/A_e$ ) was calculated for each task.

### c. Laryngeal setting

- *Inclination of laryngeal cartilages*: Two helplines were constructed to calculate the angles of inclination of the thyroid and cricoid cartilages. A horizontal helpline (A) (Fig. 4) was traced from the anterior nasal spine to the most posterior edge of the *clivus* (the anterior limit of the *foramen magnum*). A vertical helpline (B) was traced perpendicular to the helpline A, from the *clivus* to the level of the lowermost edge of the sixth vertebra. The two laryngeal angles calculated were defined as 1) TCI - thyroid cartilage inclination (the angle between the helpline B and a line from the anterior laryngeal commissure to the posterior laryngeal commissure, following the most superior edge of the vocal folds, following the most superior edge of the vocal folds; the angle was measured below the TCI line), and 2) CCI – cricoid cartilage inclination (the angle between the helpline B and a line from the lowermost edge of the anterior portion of the cricoid cartilage and the lowermost edge of the posterior portion of the cricoid cartilage; the angle was measured below the CCI line).
- *Glottal length (mm)*: the distance between the middle point of the anterior laryngeal commissure to the middle point of the posterior laryngeal

commissure, when the vocal processes of the arytenoid cartilages are in contact (Fig.4).

- d. Volumetric measures: the total volume of the epilaryngeal tube was calculated based on its three-dimensional reconstruction from the CT images obtained for each vocal task. The anatomical limits considered for the epilaryngeal tube volumetric measurement were the same as described in Fig. 1.1 [1,10].

Each anatomical measure was obtained twice by the same experimenter in order to increase measurement reliability. Differences between the first and the second measurement were calculated to obtain a reference limit of error. Thus, differences between samples (speaking voice quality, twang-like voice quality on the high pitch, and twang-like voice quality on the low pitch) were considered as such only if the difference was larger than the reference limit of error. Moreover, a difference obtained was considered as a potential measurement error when it was equal or smaller than the limit of error.

#### *Acoustic analysis*

The acoustic analysis was performed using PRAAT software (version 6.0.43, 64-bit edition for MacOS 10.7 or later). SPL, fundamental frequency ( $f_0$ ), formant frequencies, and long-term-average spectrum (LTAS) were obtained from the most stable and representative sample previously recorded for each vocal task. The same samples were used later in the inverse filtering and audio-perceptual analysis. Thus, the acoustic, audio-perceptual, and inverse filtering analyses included three samples in total.

LTAS analysis was computed using a filter bandwidth of 25 Hz and Hann-windowing. LTAS analysis included the following measures:

- a. The difference in the sound energy (in dB) between the first formant region ( $L_1$ ) and the fundamental frequency range ( $L_0$ ) [15]. Although vocal tasks were related to a sustained vowel, slight variations in the fundamental frequency and the

frequency of the first formant are expected. Therefore, it may not be possible to measure the amplitude of the first harmonic and that of the first formant. Thus, an estimate for the level difference of the fundamental and the first formant (i.e.,  $L_1 - L_0$ ) was calculated [15] instead of only the peak values. The difference between the SPL in the fundamental frequency range and that in the first formant region was calculated automatically by Praat. For the low-pitched samples  $L_0$  was measured within the frequency range of 50-300 Hz, whereas  $L_1$  was measured from a spectral peak within the frequency range of 300-800 Hz. The high-pitched sample in twang-like voice quality was performed with a higher  $f_0$  (394.29 Hz). Also, the first formant frequency rose (to 800 Hz, approximately). Thus,  $L_0$  was calculated within the frequency range of 50-400 Hz, and  $L_1$  within the frequency range of 400-900 Hz. Figure 5 represents the different bandwidths considered for the two pitches.

- b. The sound energy difference between the highest strongest peak in the frequency range of 2-4 kHz and the highest strongest peak in the frequency range of 0-2 kHz. This parameter of the spectrum analysis is known as the singing power ratio (SPR). Although SPR estimates the sound energy in the singer's formant region, it has been described as a tool for quantitative evaluation of the singing voice quality rather than for tracking the presence or absence of the singer's formant [28].

As a complementary visual analysis, the LTAS of each audio signal was plotted. To obtain a better visual comparison of the differences in sound energy distribution between the voice qualities, the peak amplitude of each LTAS was normalized, i.e. set to zero. The frequency range considered in the plotting process was between 50 and 8000 hertz.

In addition, the first five formant frequencies ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ , and  $F_5$ ) and the frequency differences between them ( $F_1 - F_2$ ,  $F_3 - F_4$ ,  $F_4 - F_5$ , and  $F_3 - F_5$ ) were estimated through FFT spectrum and a broad-band spectrogram (256 Hz bandwidth), following the procedures previously described in a study by Guzmán et al. [27]. The approximation

of formant frequencies was obtained from FFT spectrum analysis. The stronger spectral peaks or the frequency in the middle of two adjacent, equally strong peaks, were considered near to the formant frequencies. As a comparison, spectrograms were used. There, formant frequency estimation was considered in the middle part of the strongest energy bands.

### *Inverse filtering analysis*

Since a previous study by Sundberg and Thalén [3] showed specific features of the voice source during twang-like voice production, inverse filtering analysis was included in the present study. Previously recorded speech pressure signals (the most stable and representative for each voice quality) were inverse filtered by one of the experimenters, who has long experience in glottal inverse filtering analysis. Inverse filtering was done using quasi closed phase (QCP) analysis [29] implemented in the Aalto Aparat tool [30]. QCP is based on the estimation of the vocal tract transfer function from glottal closed phases over several glottal cycles by utilizing weighted linear prediction with a specific attenuated main excitation (AME) weight function [31]. QCP was selected as the glottal inverse filtering method in the current study because it was shown in Airaksinen et al. [29] to yield the highest estimation accuracy compared to four other known inverse filtering methods. QCP analysis was conducted in 100-ms frames using the sampling frequency of 8 kHz. The obtained time-domain estimates of the glottal flow were parameterized with the following parameters:

#### a. Time-domain parameter:

- *Closing quotient (CIQ)*: ratio between the duration of the glottal closing phase (i.e. the time-difference between the instants of the maximum and minimum flow during the decreasing phase of the glottal pulse) and the duration of the fundamental period.

b. Amplitude-based parameters:

- *Amplitude quotient (AQ)*: the ratio between the AC flow amplitude and the amplitude of the negative peak of the differentiated flow (also known as maximum flow declination rate, MFDR).
- *Normalized amplitude quotient (NAQ)*: the ratio between AQ and the duration of the fundamental period.

c. Frequency-domain parameters:

- *The level difference between the first and second harmonic ( $H_1-H_2$ )*: a measure for the tilt of the glottal flow spectrum obtained from the level difference (in decibels) of the amplitudes of the source fundamental ( $H_1$ ) and the second spectrum partial ( $H_2$ ).
- *Harmonic richness factor (HRF)*: a measure for the tilt of the glottal flow spectrum obtained from the ratio between the sum of the amplitude levels of harmonics above the fundamental and the amplitude level of the fundamental (in decibels).
- *Parabolic spectral parameter (PSP)*: the pitch-synchronously computed spectrum obtained by fitting a second-order polynomial to the flow spectrum on a logarithmic scale computed over a single glottal cycle.

*Auditory-perceptual analysis*

Five voice teachers (different from the first group of CCM singing teachers that assessed the accuracy of the vocal qualities performed by our subject) served as blinded judges for the auditory-perceptual analysis of the degree of brightness and twanginess. Each of the five voice teachers had more than five years of experience in CCM music and Broadway-style singing. The analysis was carried out using two four-point Likert-type scales:

- a. The scale for the degree of brightness of the voice considered: “0” as “no brightness at all”, “1” was “slight brightness”, “2” was “moderate brightness”, and “3” was “very bright”.

- b. The scale for the degree of twanginess of the voice considered: “0” as “no twanginess at all”, “1” was “slight twanginess”, “2” was “moderate twanginess”, and “3” was “very twangy”.

Although the judges were instructed to play the samples in their personal devices, high-quality over-ear headphones were requested during the auditory-perceptual assessment. Judges could play the samples as many times as needed to be sure of each answer. The mean values of the raters’ answers for each sample were calculated. Since only three voice samples were included (one for each voice quality), a simple audio-perceptual analysis task was considered. Thus, no inter- and intra-rater reliability analysis was performed.

## **Results**

### *Sagittal anatomic distance measures*

Anatomic distances (mm) calculated from the midsagittal images of the vocal tract for neutral speaking voice and both twang-like voice productions (on the high and low pitches) are presented in Table 1 and Figure 6.1. Changes compared to the speaking voice were observed in lip and jaw opening which increased in both twang-like voice qualities, on the high and low pitch. Changes were also found for oropharynx width, which increased during the twang-like voice quality on the high pitch and decreased during the twang-like voice quality on the low pitch compared to the speaking voice. As can be seen in Figure 6.1, the most evident change occurred during the twang-like voice quality on the low pitch, where the oropharynx narrowed due to a more backward and slightly raised tongue position. Hypopharyngeal sagittal distance also increased during the twang-like voice quality on the high pitch and showed a decrease during the twang-like voice quality on the low pitch. The velum position rose during the twang-like voice quality on the high pitch, while on low pitch it stayed similar to the position seen in the speaking voice quality, except that a slight closure of the velopharyngeal opening was present. Tongue dorsum height decreased during the high-pitched twang-like voice quality, while it seemed to increase during the low-pitched twang-like voice production.

However, the difference observed between the low-pitched twang-like voice quality and the speaking voice production stayed within the limit of error. The vertical length of the vocal tract decreased during both twang-like voice qualities, on the high and low pitch, which is related to a higher larynx position. Finally, the horizontal length of the vocal tract decreased during both twang-like voice qualities, though to a greater degree on the low pitch.

### *Laryngeal setting*

The results for the inclination of laryngeal cartilages are presented in Table 2 and Figure 6.2. The thyroid cartilage angle showed a decrease during the twang-like voice quality on the high pitch, which is related to a greater forward tilt of the thyroid cartilage compared to the other two voice productions. Oppositely, the thyroid cartilage angle increased in the twang-like voice quality on the low pitch. The cricoid cartilage angle increased during both twang-like voice qualities, on the high and low pitch, showing a backward rotation during both twang-like voice productions. Glottal length (mm) results are presented in Table 2 and Figure 7. Glottal length increased during the twang-like voice quality on the high pitch, as can be expected due to the pitch increase. Although both speaking and twang-like voice qualities on the low pitch were performed using similar  $f_0$  (123.67 Hz and 122.28 Hz, respectively), glottal length decreased during the twang-like voice quality.

### *CT area measures*

Cross-sectional areas ( $\text{cm}^2$ ), measured from the midsagittal images of the vocal tract obtained from the CT scannings, are presented in Table 3 and Figure 6.1. Oral cavity sagittal area ( $A_1$ ) increased during both twang-like voice qualities, on the high and low pitch. The pharyngeal region ( $A_2$ ) showed an increase during the twang-like voice quality on the high pitch compared to the speaking voice and decreased during the twang-like voice quality on the low pitch. The epilaryngeal region ( $A_3$ ) became larger during the twang-like voice quality on the high pitch, decreasing during the twang-like voice quality on the low pitch.



The epilaryngeal outlet area and the pharyngeal inlet area ( $\text{cm}^2$ ) are presented in Table 3 and Figure 8.  $A_p$  area decreased in both twang-like voices compared to the speaking voice. As can be seen in Figure 8,  $A_p$  showed a discreet decrease during the high pitch. The most evident change occurred during the twang-like voice quality on the low pitch. The  $A_e$  area became smaller during the twang-like voice quality on the low pitch, while a slight increment was observed during the twang-like voice quality on the high pitch.  $A_p/A_e$  ratio changed differently, decreasing for high-pitched twang-like voice quality and increasing for the low-pitched twang-like voice quality when compared to the speaking voice.

### *Volumetric measures*

Results for the volumetric measures of the epilaryngeal tube are displayed in Table 4 and Figure 9. During both twang-like voice qualities, on both the high and low pitches, the ET volume decreased (Table 4). This reduction was much more prominent during the twang-like voice quality on the low pitch than on the high pitch. Although volume values decreased in both twang-like voice qualities, it seems that the epilaryngeal tube volume was modified in an anterior-posterior and medial direction during the twang-like voice quality on the low pitch (Fig. 9). On the high pitch, instead, the lower portion of the ET (the ventricular region) was medially narrowed, while the middle and upper portions widened sagittally and medially.

### *Acoustic analysis results*

Table 5 shows the results of LTAS variables for speaking voice and twang-like voice quality on high and low pitches. Figure 10.1 shows LTAS comparisons between the three acoustic samples. It should be noted that the measurements in Table 5 concern average sound energy from particular frequency bands, not just peak values as seen in spectra in Figure 10.1.

An important increase was seen for  $L_1-L_0$ , which was much clearer during the twang-like voice quality on the high pitch (14.27 dB) than during the twang-like voice quality in the low pitch (5.14 dB). These increases were mainly related to the higher energy

found in  $L_1$  during both twang-like voices on the high and low pitches compared to the speaking voice. Singing power ratio also increased during both twang-like voice qualities compared to the speaking voice. This increase was much more prominent on the low pitch (14.30 dB) than on the high pitch (7.46 dB).

SPL and formant frequency results are presented in Table 6. Formant distribution is shown in Figure 10.2. SPL was higher in both twang-like voice qualities. The highest SPL was observed for the high-pitched twang-like voice quality (88 dB), followed by the low-pitched twang-like voice quality (74 dB) and the speaking voice (67 dB). Frequency of  $F_1$  increased for both twang-like productions compared to the speaking voice. The increment was greater for the twang-like quality on the high pitch than on the low pitch.  $F_2$  and  $F_4$  also increased for both high-pitched and low-pitched twang-like voice qualities. Similar to  $F_1$ , frequencies of  $F_2$  and  $F_4$  increased more for the high-pitched twang-like production than for the low-pitched twang-like version.  $F_3$  and  $F_5$  decreased for the low-pitched twang-like voice quality. Oppositely,  $F_3$  and  $F_5$  increased for the high-pitched twang-like voice quality.

Related to formant frequency differences (Table 6),  $F_1$  and  $F_2$  were closer in the twang-like voice quality on the low pitch and more separated on the high pitch, reflecting the predominant increase of  $F_1$  on the low pitch and that of  $F_2$  on the high pitch.  $F_2$  and  $F_3$  were closer together during the low-pitched twang-like voice quality because  $F_3$  decreased. In contrast,  $F_2$  and  $F_3$  moved further apart during the high-pitched twang-like voice quality because  $F_3$  increased.  $F_3$  and  $F_4$  also moved further apart during both twang-like voices compared to the speaking voice. However, a greater frequency difference was observed during the low-pitched twang-like voice quality due to a decrease in  $F_3$ .  $F_4$  and  $F_5$  got closer during both twang-like productions. This frequency difference was smaller during the low-pitched twang-like voice quality due to a higher  $F_4$  and a lower  $F_5$ . Finally, the frequency difference between  $F_3$  and  $F_5$  was smaller for both twang-like voices. The low-pitched twang-like sample showed the smallest difference between  $F_3$  and  $F_5$ . However, no formant clustering was observed in any of the samples.

### *Inverse filtering analysis results*

Measures obtained from inverse filtering of the three voice samples are presented in Table 7. The time-domain and amplitude-based parameters obtained (CIQ, AQ, and NAQ) showed a decrease in both twang-like voice quality samples compared to the speaking voice. Although NAQ decreased similarly in both samples, a remarkable difference was observed in AQ, which showed a greater decrease on the high pitch than on the low pitch. Differently, CIQ was somewhat lower on the low pitch than the high pitch. The frequency-domain parameters,  $H_1$ - $H_2$  and PSP, also showed a decrease in both twang-like voice productions, both being lower on the high pitch (11.2 dB and 0.09 dB, respectively) than the low pitch (7.2 dB and 0.08 dB, respectively). The harmonic richness factor (HRF) increased in both twang-like voice qualities. HRF was higher on the high pitch (10.6 dB) than the low pitch (4.7 dB).

Time-domain waveforms of the glottal flow and voice source spectrum are respectively presented in Figure 11.1 and 11.2. for each voice sample. As can be seen in Figure 11.1, the cessation speed of the airflow was faster in both twang-like voice qualities, showing a more skewed closing phase of the flow pulse. The voice source spectrums (Figure 11.2) show the diminished difference between the amplitudes of  $H_1$  and  $H_2$ , due mainly to an increase in  $H_2$ . Also, it is possible to observe a less steep spectral tilt in both twang-like qualities, i.e. in twang-like qualities the higher harmonics are stronger compared to the lower ones.

### *Auditory-perceptual results*

Results from the auditory-perceptual analysis, performed by five blinded listeners, showed that both samples with twang-like voice qualities obtained higher values for the degree of brightness and twanginess in the voice compared to the speaking voice. Specifically, the twang-like voice quality on the low pitch obtained an average of 2.8 for the degree of brightness and 2.4 for the degree of twanginess. The twang-like voice quality on the high pitch was rated with an average of 1.8 for the degree of brightness and 2.0 for the degree of twanginess. Finally, speaking voice obtained

an average of 1.2 for both the degree of brightness and twanginess.

## **Discussion**

The present study was designed to investigate the possible contribution of the lower and upper vocal tract configuration on loud twang-like singing voices on high and low pitches in one male CCM singer. Results showed interesting similarities and differences in the vocal tract configuration and acoustic, inverse filtering, and perceptual evaluation when the three samples included in the present study were compared.

### *The megaphone-like shaped vocal tract in loud twang-like singing*

The loud twang-like voices in our study showed a megaphone-like shape of the vocal tract [17]. As seen in figures 6.1 and 9, and in tables 1 and 3, lip and jaw opening, and the oral cavity area were wider during both twang-like voices compared to the speaking voice. Also, both twang-like voices showed a shortened vocal tract due to a higher laryngeal position (Table 1). Differences between twang-like voices were found in the lower vocal tract configuration, in which both the pharynx and epilaryngeal tube (ET) were narrower in the low pitch compared to the high pitch (figures 6.1, 8, and 9, and tables 1 and 3). Pharyngeal sagittal measures and the  $A_e$  area suggest that the ET and pharynx were wider during the high-pitched twang-like voice compared to both low-pitched voices. However, the  $A_p$  area was smaller than for the speaking voice, which may have resulted from a raised larynx [32] and a greater jaw and lip opening [33]. The volumetric measures (Table 4) confirmed that the ET was also narrowed in both twang-like voices. Nevertheless, the reduction of the ET volume was more prominent in the low-pitched twang-like voice, showing a constriction of both the vestibular and ventricular regions (Figure 9). Therefore, although both twang-like voices seem to be related to a megaphone-like shape of the vocal tract, this configuration was more discrete in the high-pitch twang-like voice compared to the lower-pitched version, and was characterized mainly by an increased jaw and lip opening and a narrowing of the ET rather limited to the lower (ventricular) portion.

### Laryngeal setting during high and low-pitched twang-like singing

The inclination of the laryngeal cartilages in the high-pitched twang-like voice quality agrees with Overdrive mode with reduced density, which is characterized by a tilted thyroid cartilage (Table 2 and Figure 6.2) [20]. As for the angle of inclination of the cricoid cartilage, it increased, so the back of the cricoid cartilage showed a more down and backward rotation. Findings are concordant with the longer glottis (Table 2 and Figure 7) which, in turn, reflects the increased vocal fold length expected in the higher pitch. According to Honda et al. [21] the high-pitched voice production can be achieved by thyroid cartilage tilting, which may be facilitated by the horizontal (forward) movement of the hyoid bone due to the genioglossus and geniohyoid muscle action. The external frame function theory [26] supports this assumption. Thus, vowel articulation (a more retracted or forward tongue) may play an important role in the extralaryngeal pitch control. It is worth mentioning that the anterior movement of the hyoid bone may also promote a wider ET and pharynx [26], which is concordant with Overdrive mode with reduced density anatomical features [20]. High-pitched sounds have also been related to a backward cricoid cartilage rotation due to the activation of the cricothyroid muscles [22,37].

Regarding glottal length, it was shorter during the low-pitched twang-like voice compared to the speaking voice (Table 2 and Figure 7). Since  $f_0$  was controlled to be the same during both low-pitched samples (Table 6), the length of the vocal folds (the membranous glottis) was expected to remain the same. However, the low-pitched twang-like voice was asked to be performed loudly (similar to the twang-like voice on the high pitch). It is well known that  $f_0$  increases with loudness, due to higher subglottic pressure (causing increased stiffness of the vocal folds related to the increased amplitude of vibration) [35]. Thus, shortening and thickening of the vocal folds during the loud twang-like voice on the low pitch may have been a necessary strategy to avoid  $f_0$  rise, keeping the laryngeal framework stabilized. Also, since thyroarytenoid muscle contraction during louder and lower pitches tends to raise  $f_0$  because the effective stiffness of the vibrational portion of the vocal folds increases, an external pitch control mechanism should be necessary to shorten and thicken the

vocal folds in order to keep  $f_0$  constant [26,35]. In radiographic observations, Sonninen et al. [26] found a smaller distance between the cricoid cartilage and the spine during lower pitches at loud phonation when “open singing mode” was used (as in belting), compared to “covered singing mode” (as in Western classical singing). Also, both a small distance between the thyroid and arytenoid cartilages (shorter vocal folds) and an increased distance between the cricoid and arytenoid cartilages were found. Authors suggested that these findings are concordant with the cricopharyngeal muscle contraction. Cricopharyngeal muscle activation may pull the cricoid cartilage backward (toward the spine), thickening the vocal folds on lower pitches during loud phonation in a non-classical mode of singing. Our CT data are concordant with extralaryngeal pitch control mechanisms, possibly used to thicken the vocal folds without increasing thyroarytenoid muscle contraction (figures 6.2 and 7).

### *Spectral features of loud twang-like singing*

#### *Formant distribution and vocal tract configuration*

The shortened and megaphone-like shaped vocal tract during both twang-like voices agree with the formant distribution found. As seen in Table 6 and Figure 10.2, all formant frequencies were the highest during the high-pitched twang-like voice. Similarly,  $F_1$ ,  $F_2$ , and  $F_4$  increased in frequency during the low-pitched twang-like quality. These results are concordant with the shortened vocal tract found (mainly due to a higher laryngeal position). Although vocal tract shortening is expected to increase all formant frequencies [32,35,37],  $F_3$  and  $F_5$  decreased during the low-pitched twang-like voice. Hence, other changes different from the laryngeal position may also be related to the formant distribution found.

The  $F_1$  increased in both twang-like voices, which agrees with the greater lip and jaw opening [36-38]. Lip opening was wider during the high-pitch twang-like voice, which may be necessary because  $F_1$  needs to be raised higher in a higher pitch [17]. Pharyngeal narrowing has been also related to an increase in  $F_1$  [32,35]. A greater jaw opening moves the jaw backward, thus diminishing the pharynx [38]. Our subject

showed an increased jaw opening most during the twang-like voice quality on the low pitch (Figure 6.1), which in turn may have narrowed the hypopharynx near the glottis, raising  $F_1$ . Similar observations have been made by Sundberg et al. in different substyles of belting [13]. Lindblom and Sundberg [38] have also proposed that the position of the tongue body in the pharynx may increase  $F_1$ , which is in agreement with our subject whose tongue body was retracted toward the pharyngeal region in the low-pitched twang-like voice (figures 6.1 and 9).

The increase in  $F_2$  during the low-pitched twang-like voice may have been related mainly to tongue position [36-38]. According to Lindblom and Sundberg [38], a greater constriction of the oral cavity produced by the tongue raising toward the palatal region rises  $F_2$ . In the study by Story et al. [9], the results from a male subject (native American English speaker) showed an increase in  $F_2$  for all the vowels performed using a twangy quality. Authors described twangy vowels as more “fronted” and “spread-lipped. According to Hillenbrand et al. [39], the vowel [æ] in American English is characterized by a lower  $F_1$  and a higher  $F_2$  compared to the vowel [a:] from the same phonetic system. Based on the previous data, it is possible to infer that twang-like vowels may result from mixing a wider jaw opening (as in vowel [a:]) with a more forward tongue dorsum (as in vowel [æ]). Thus, vowel articulation seems to be a relevant aspect related to twang. Nevertheless, the increase of  $F_2$  in our study, during the low-pitched twang-like voice, was discrete (Table 6 and Figure 10.2) which is concordant with the slight increase in the tongue height found (Table 1 and Figure 6.1). These findings may be explained by the fact that the singer produced a sustained vowel [a:] under the Chilean Spanish phonetic system, in which the [a:] vowel is characterized by a high  $F_1$  and a low  $F_2$  [40]. Moreover, the vowel [æ] does not exist in the Chilean Spanish phonetic system. Considering the formant frequencies obtained in the present study for  $F_1$  and  $F_2$  during the low-pitched twang-like voice, it seems that the vowel quality was quite well preserved to be [a:] instead of [æ], which in turn limited the increase in  $F_2$ .

The increase in  $F_3$  during the high-pitched twang-like voice seems to be related to the combination of a more fronted tongue position and to the greater posterior area

of the oral cavity [33], which also agrees with the increase in  $A_2$ . In the low-pitched twang-like voice, the decrease in  $F_3$  would be related to a space produced in the front cavity between the tongue tip and the lower incisors [3,33]. This front space in the oral cavity has been associated with a more retracted tongue body. Our CT and acoustic data (figures 6.1 and 9, and Table 6) agree with the previous findings.

The  $F_4$  increased in both twang-like voices. According to Takemoto et al. [41] the laryngeal cavity (i.e., the epilaryngeal tube) generates the fourth formant, which in turn is sensitive to the ET shape. A reduction of the vestibular area of the ET lowers  $F_4$ , whereas a reduction of the ventricular area, rises  $F_4$ . Since  $F_4$  increased during the low-pitched twang-like voice, a major reduction may have been produced at the ventricular region. In the high-pitched twang-like voice  $F_4$  showed the greatest increase (Table 6). Since ET narrowing was limited to the ventricular region (Figure 9) it is possible to expect a greater increase of the  $F_4$ .

As for  $F_5$ , Titze and Story [42] observed that when the ET cross-sectional area was computationally narrowed, all formant frequencies were “attracted” to the 2.5-3.0 kHz region and, therefore,  $F_1$  and  $F_2$  rose, whereas  $F_4$  and  $F_5$  lowered. Thus,  $F_5$  during the low-pitched twang-like voice seemed to be lowered due to ET narrowing (Table 6 and Figures 6.1, 8, and 9). In the high-pitched twang-like voice the ET narrowing was discrete, and the shortening of the vocal tract was the greatest. Therefore, the increase in  $F_5$  agrees with the vocal tract configuration found.

#### *The singer's format region*

Both twang-like voices showed an increased SPR (Table 5) due to the increased energy concentration around 3 kHz, indicating a less steep spectral tilt (Figure 10.1). Our results are concordant with previous findings [8,13,14]. It should be noted that although the distance between  $F_3$  and  $F_5$  decreased in both twang-like voices, no formant clustering was found (Table 6 and figures 10.1, and 10.2). This may be explained by the small pharynx to epilarynx ratio ( $A_p/A_e$ ). Therefore, the energy concentration around 3 kHz seems to be mainly related to the increased frequency and amplitude of  $F_4$  (figures 10.1 and 10.2) [15,41]. In addition, the low-pitched



twang-like voice showed the greater SPR increase (Table 5). A possible explanation to this would be that the high-pitched twang-like voice quality in our study was categorized as Overdrive mode with reduced density, following the CVT terminology. According to Thuesen et al. [20], Overdrive with reduced density is characterized by less spectral energy around 3kHz and therefore, by a steeper spectral tilt compared with a full metallic and full density Overdrive mode. Anatomically, the reduced density version of Overdrive has been related to a wider ET and pharynx [20]. Our acoustic and anatomic data are concordant with this description.

#### *L<sub>1</sub>-L<sub>0</sub> from LTAS analysis*

L<sub>1</sub>-L<sub>0</sub> increased in both twang-like voices. The higher L<sub>1</sub>-L<sub>0</sub> was found during the high pitch (Table 5), which also showed the greater SPL (Table 6). Based on Kitzing [43], the increased L<sub>1</sub>-L<sub>0</sub> suggest that twang-like voices may be related to greater glottal adduction, similar to pressed voices. According to Gauffin and Sunberg [44,45], pressed phonation is related to a decreased amplitude of the fundamental, reflecting the limited amplitude of vibration of the vocal folds. Although L<sub>1</sub>-L<sub>0</sub> increased, the twang-like voices in our study may not be necessarily related to a more pressed phonation. Both L<sub>0</sub> and L<sub>1</sub> increased during the twang-like voices compared to the speaking voice, with L<sub>1</sub> showing a greater increase (Table 5). Thus, L<sub>1</sub>-L<sub>0</sub> results may be mainly related to changes in L<sub>1</sub>. Since L<sub>0</sub> also increased, the amplitude of vibration of the vocal folds would not be as limited as in pressed phonation. On the other hand, it has been shown that SPL depends on the amplitude of the harmonics closest to F<sub>1</sub> [44,45] and the higher harmonics gain more in amplitude than the lower ones as SPL rises. Therefore, L<sub>1</sub>-L<sub>0</sub> is expected to increase with a higher SPL [46]. Our both twang-like voices were asked to be performed with an increased loudness (Table 6), which agrees with the relation between SPL and L<sub>1</sub> previously described. Regarding the high-pitched twang-like voice, increases in L<sub>1</sub>-L<sub>0</sub> have been previously related with higher SPL and greater jaw opening, the latter one related to formant tuning [46]. Hence, the greater increase in L<sub>1</sub>-L<sub>0</sub> seems to agree with the presence of a resonance strategy (Figure 10.1 and Figure 10.2), in which the fundamental was not the predominant partial but the second harmonic which was boosted by F<sub>1</sub> that

was raised to match it by lowering the jaw and raising the larynx. Spectral characteristics previously found in belting support this assumption [17,20].

### *Voice source features of loud twang-like singing*

Both twang-like voices showed a decreased  $H_1-H_2$  (Table 7), suggesting increased glottal adduction [3,35,47]. Similar trends were observed previously by Sundberg and Thalén in twang [3]. Our PSP and HRF data are concordant with  $H_1-H_2$  results. Also, PSP and HRF indicate a smaller spectral decay (Figure 11.2) [48]. CIQ, AQ, and NAQ decreased in both twang-like samples (Table 7). AQ showed the greatest decrease in the high pitch, whereas NAQ showed almost the same decrease in both twang-like samples. NAQ results agree with those obtained previously by Sundberg and Thalen concerning twang [3]. Sundberg et al. [49] have suggested that NAQ reflects the degree of pressed phonation. According to the authors, classical singing would be closer to breathy phonation (NAQ near to 0.16), whereas blues singing style utilizes vocal qualities closer to pressed phonation (NAQ near to 0.10). Nevertheless, blues showed a higher NAQ than pressed voice. Based on Sundberg et al. findings [49], our NAQ results may indicate that although NAQ was low for both twang-like voices, the voice was not necessarily pressed. In relation to NAQ and AQ results, Björkner et al. [50] proposed that AQ would reflect more accurately the phonation type within the same voice phonating at different fundamental frequencies than NAQ, which would be better when comparisons are made between genders or between different subjects. Therefore, the high-pitched twang-like voice may have been produced with more adducted vocal folds than the low-pitched version.  $H_1-H_2$ , PSP and HRF results are concordant with this assumption. As for CIQ, it is expected to decrease with increased glottal adduction [50]. In the present study, CIQ decreased in both twang-like voices, being somewhat lower on the low pitch than on the high pitch. CIQ data do not agree with the rest of the inverse filtering data obtained. According to the literature, a possible explanation for this discrepancy would be that it is difficult to determine precisely the beginning and end of the closing phase in order to obtain CIQ. Thus, NAQ and AQ would be more robust than CIQ [48,50].

### Auditory perception of brightness and twanginess

The strongest brightness and twanginess were perceived in the low-pitched twang-like voice quality. Both the highest SPR (and therefore, a narrower ET) and the formant distribution (Table 5 and figures 10.1, and 10.2) may have contributed to the increased perception of brightness [1]. According to Titze et al. [18] the increased perception of twanginess would be related to ET and pharyngeal narrowing, and vocal tract shortening. However, the major effect on twanginess perception would be related to a lower open quotient and therefore, to increased glottal adduction [18]. Our inverse filtering data and auditory-perceptual results are concordant with previous findings on perceived twanginess. Although the high-pitched twang-like voice quality was perceptually assessed as brighter and twangier than the speaking voice sample, the degree of brightness and twanginess perceived was lower than that of the low-pitched twang-like voice quality. The auditory-perceptual analysis is concordant with the lower increase in SPR and the discrete ET and pharyngeal narrowing, which in turn agrees with the Overdrive mode with reduced density [20].

### Limitations of the study

It is important to bear in mind that our results were obtained from one male subject. Also, no inter- and intra-rater reliability analysis was performed in the auditory-perceptual analysis task. Therefore, the generality of the results should be considered cautiously. In addition, data were obtained only for the vowel [a:]. Thus, it is not possible to generalize the results to other vowels. According to Sadolin [6], some vowels (e.g., [u:]) would be problematic to be produced in a twangy quality because the tongue position would make ET narrowing more difficult. Hence, some vowels may need other articulatory adjustments of the vocal tract in order to achieve twang-like voice quality. Another limitation of the study is that vocal tasks were performed in a noisy environment during CT image acquisition. Although headphones provided audio feedback to the singer, the possible increase of SPL due to environmental noise cannot be ignored entirely. Since environmental noise was controlled to be maximally reduced during audio sample acquisition, changes

found in CT data may have been of greater magnitude than the acoustic, inverse filtering and audio-perceptual changes, especially for the low-pitched twang-like voice quality. On the other hand, CT acquisition was performed while the singer was in a supine position, which is known to affect laryngeal position (larynx rises) [51]. Nevertheless, a raised larynx was observed in both loud twang-like voices compared to the speaking voice. Thus, it is possible to infer that loud phonation in twang-like voices may be effectively related to a higher larynx. Finally, it should be noted that the adduction of the vocal folds was assessed indirectly from acoustic and inverse filtering data. Transversal CT images of the glottal length (Figure 7) were considered to estimate the length of the vocal folds only. Since CT images in Figure 7 represent averages over different time instants during vocal fold vibration for each voice quality, direct measurements of the degree of adduction were not possible to make.

#### *A brief consideration of vocal efficiency and vocal economy*

Although our spectral and inverse filtering data showed that twang-like voices are closer to the pressed extreme of the phonation type continuum, our CT data suggest that differences between twang-like and speech-like phonation observed are unlikely to depend on increased glottal adduction only. The CT analysis confirmed the presence of a megaphone-like shaped vocal tract, which is considered to be a high-impedance system (an increased inertive reactance, or inertance, over a wide frequency range) [17,52]. An inertive vocal tract changes the glottal flow waveform increasing the skewing of the flow pulse at the closing phase (the cessation speed of the airflow becomes faster) [53,54] (Figure 11.1), which in turn increases the maximum flow declination rate (MFDR) [52], strengthening the higher harmonics and increasing the SPL [45,53,54]. Thus, changes observed in the present study related to acoustics and inverse filtering measures are more likely to be related to modifications in the acoustic impedance of the vocal tract than merely to the greater glottal adduction. Although both twang-like voices seem to be related to an increased inertance of the vocal tract, the role and the magnitude of the changes of the supraglottal inertance, as well as the strategy used to modify the acoustic impedance, may be different between pitches.

Considering that high frequencies are naturally louder than low frequencies (i.e., both glottal source power and the radiated SPL increases with higher fundamental frequencies) [35], most likely high pitch productions do not necessarily need as prominent anatomical changes of the vocal tract as low pitch productions. According to earlier modelling results concerning source-filter interaction, an optimal voice production will occur when a harmonic from the source matches a peak of inertance in the vocal tract [55]. The maximum inertance is located just below a vocal tract resonance (as in the case of  $F_1$ ) [17]. According to Titze and Worley [17] the particular resonance strategy used by male belters (a greater jaw opening and higher larynx) would raise  $F_1$  so that the second harmonic would match the peak of inertance, which would strengthen the voice source. Therefore, it is possible to infer that the resonance strategy used by our singer during the high pitch was enough to achieve the task, boosting the second harmonic due to an increased inertance. However, if a twangier and brighter sound is aimed at (e.g., full metallic and full density Overdrive mode), a greater ET and pharynx narrowing would be needed in order to increase the acoustic impedance of the vocal tract to achieve maximum power transfer [56].

Oppositely, since lower-pitched sounds are more difficult to be produced loudly than high-pitched sounds, increased voice strain is expected when loud voice production is required. However, the notable changes in the vocal tract configuration during the low-pitched twang-like voice quality may have been a necessary strategy to achieve loud phonation and maximum power transfer, avoiding voice strain [56]. According to Titze and Worley [17], impedance interaction between the filter and the source would result not only in a higher energy power transfer but in strengthening the source, which in turn may increase vocal efficiency. However, higher glottal adduction is expected when a megaphone like-shape of the vocal tract is used [56], which in turn may generate potential damage on vocal folds tissues due to increased impact stress [57]. About this issue, nonlinear source-filter interaction models have suggested that the presence of an increased supraglottic impedance in this type of vocal tract setting may protect vocal folds tissue from damage, as since when the reactance above the vocal folds increases due to greater backpressure (supraglottal

and intraglottal pressures), vocal fold collision diminishes [52]. Thus, a brighter and louder voice may be achieved without excessive effort and with a reduced risk of damage because the impact stress (impact force per unit area) [35,57] is diminished.

Most data available related to this topic have been inferred from modeling and simulation studies. Future studies should focus on gathering more data related to the “price of decibels” in loud twang-like voices by MRI and CT based modelling and calculation of vocal tract impedance and simulation to study how much gain (in dB) would be obtained from the vocal tract configuration and how much would be obtained from the voice source. Also, synthesis experiments combined with a listening test could be also used in the future to confirm the perceptual effects of vocal tract modifications found in the twang-like voice qualities studied.

## **Conclusion**

The characteristics associated with loud twang-like voices seem to depend on several levels: epilaryngeal tube and pharynx narrowing, jaw opening, tongue position, and vocal fold adduction. The adequate combination of these levels would result in a brighter twang-like sound, in both high and low fundamental frequencies. Different pitches seem to require different degrees of twang-related vocal tract adjustments. It should be emphasized that the main features of twang-like voices may seem to be related to low-frequency spectral changes (under 2 kHz). Thus, vowel articulation (lip and jaw opening, oral cavity constriction, tongue body position, and tongue dorsum height) would be relevant aspects to consider when utilizing twang-like phonation. Auditory-perceptual tests of spectrum-manipulated samples would be needed in order to confirm this assumption. Moreover, although glottal adduction seems to be greater in twang-like voices, the acoustic impedance (inertance) related to a megaphone-like shape of the vocal tract (specifically related to the ET narrowing and jaw opening) should be considered as an important feature that may promote vocal economy in this type of loud voice production. Future studies should focus on real-subject based modeling and simulation to estimate the “price of decibels” in loud twang-like voices.

## **Acknowledgments**

The authors thank all the voice teachers who participated as blinded judges of the sound samples. Also, the authors thank Paula Cancino, who illustrated figures 1.1 and 1.2 of the present study.

## **Declaration of interests**

The authors report no declarations of interest.

## **References**

1. Titze IR. Acoustic interpretation of resonant voice. *J Voice*. 2001;15(4):519-528.
2. Lombard LE, Steinhauer KM. A novel treatment for hypophonic voice: Twang therapy. *J Voice*. 2007;21(3):294-299.
3. Sundberg J, Thalén M. What is “Twang”? *J Voice*. 2010;24(6):654-660.
4. Estill J, Klimek MM, Obert K, Steinhauer K. Estill voice training level one: Figures for voice control. Pittsburgh: Estill Voice Training Systems International; 2005.
5. Estill J, Klimek MM, Obert K, Steinhauer K. Estill voice training level two: Figure combinations for six voice qualities. Pittsburgh: Estill Voice Training Systems International; 2005.
6. Sadolin C. Complete vocal technique. Copenhagen: Shout Publishing; 2014.
7. Kayes G. Singing and the actor. 2<sup>nd</sup> Ed. New York: A&C Black Publishers; 2004.
8. Yanagisawa E, Estill J, Kmucha S, Leder S. The contribution of aryepiglottic constriction to “ringing” voice quality: A videolaryngoscopic study with acoustic analysis. *J Voice*. 1989;3:342-350.
9. Story BH, Titze IR, Hoffman EA. The relationship of vocal tract shape to three voice qualities. *J. Acoust. Soc. Am*. 2001;109(4):1651-1667.
10. Sundberg S. Articulatory interpretation of the “singing formant”. *J. Acoust. Soc. Am*. 1974;55(4):838-844.
11. Hertegård S, Gauffin J, Sundberg J. Open and covered singing as studied by means of fiberoptics, inverse filtering, and spectral analysis. *J Voice*. 1990;4(3):220-230.

12. McCoy S. *Your voice: An inside view*. 2<sup>nd</sup> Ed. Delaware: Inside View Press; 2012.
13. Sundberg J, Thalén M, Popeil L. Substyles of belting: Phonatory and resonatory characteristics. *J Voice*. 2012;26(1):44-50.
14. McGlashan J, Thuesen MA, Sadolin C. Overdrive and edge as refiners of “belting”? An empirical study qualifying and categorizing “belting” based on audio perception, laryngostroboscopic imaging, acoustics, LTAS, and EGG. *J Voice*. 2016 in press.
15. Leino T. Long-term average spectrum in screening of voice quality in speech: Untrained male university students. *J Voice*. 2009;23(6):671-676.
16. Cleveland TF, Sundberg J, Stone RE. Long-term-average spectrum characteristics of country singers during speaking and singing. *J Voice*. 2001;15(1):54-60.
17. Titze IR, Worley AS. Modeling source-filter interaction in belting and high-pitched operatic male singers. *J. Acoust. Soc. Am*. 2009;126(3):1530-1540.
18. Titze IR, Bergan CC, Hunter EJ, Story B. Source and filter adjustments affecting the perception of the vocal qualities twang and yawn. *Logoped Phoniatr Vocol*. 2003;28(4):147-55.
19. LoVetri, J. (2012). The confusion about belting: A personal observation. *Journal of the New York Singing Teachers' Association*, Sept-Oct, 4-7.
20. Thuesen MA, McGlashan J, Sadolin C. Investigating laryngeal “tilt” in same-pitch phonation: Preliminary findings of vocal mode metal and density parameters as alternatives to cricothyroid-thyroid “mix”. *J Voice*. 2018. Article in press. doi: 10.1016/j.jvoice.2018.02.023.
21. Honda K, Hirai H, Masaki S, Shimada Y. Role of vertical larynx movement and cervical lordosis in F0 control. *Lang Speech*. 1999;42(4):401-411.
22. Unteregger F, Honegger F, Potthast S, Zwicky S, Schiwowa J, Storck C. 3D analysis of the movements of the laryngeal cartilages during singing. *Laryngoscope*. 2017;127(7):1639-1643.



23. Tom K, Titze IR, Hoffman EA, Story BH. Three-dimensional vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness. *J. Acoust. Soc. Am.* 2001;109(2):742-747.
24. Amis ES, Jr, Butler PF, Applegate KE, Birnbaum SB, Brateman LF, Heyezi JM, Mettler FA, Morin RL, Pentecost MJ, Smith GG, Strauss KJ, Zeman RK. American College of Radiology white paper on radiation dose in medicine. *J Am Coll Radiol.* 2007 May;4(5):272-284.
25. Saldías M, Guzmán M, Miranda G, Laukkanen AM. A computerized tomography study of vocal tract setting in hyperfunctional dysphonia and in belting. *J Voice.* 2019;33(4):412-419.
26. Sonninen A, Hurme P, Laukkanen A-M. The external frame function in the control of pitch, register, and singing mode: Radiographic observations of a female singer. *J Voice.* 1999;13(3):319-340.
27. Guzman M, Laukkanen AM, Krupa P, Horáček J, Švec JG, Geneid A. Vocal tract and glottal function during and after vocal exercising with resonance tube and straw. *J Voice.* 2013;27(4):523.e19-523.e34.
28. Omori K, Kacker A, Carroll LM, Riley WD, Blaugrund SM. Singing power ratio: Quantitative evaluation of singing voice quality. *J Voice.* 1996;10(3):228-235.
29. Airaksinen M, Raitio T, Story B, Alku P. Quasi closed phase glottal inverse filtering analysis with weighted linear prediction. *IEEE/ACM Trans. Audio, Speech, Language Process.* 2014;22(3):596-607.
30. Alku P, Pohjalainen H, Airaksinen M. Aalto Aparat - A freely available tool for glottal inverse filtering and voice source parameterization. *Proc. Subsidia: Tools and Resources for Speech Sciences, Malaga, Spain, June 21-23, 2017.*
31. Alku P, Pohjalainen J, Vainio M, Laukkanen AM, Story B. Formant frequency estimation of high-pitched vowels using weighted linear prediction. *J. Acoust. Soc. Am.* 2013;134(2):1295-1313.
32. Sundberg J, Nordström PE. Raised and lowered larynx: The effect on vowel formant frequencies. *STL-QPSR.* 1976;17(2-3):035-039.
33. Sundberg J. Formant structure and articulation of spoken and sung vowels. *Folia Phoniatr Logop.* 1970;22:28-48.

34. Storck C, Gehrler R, Fischer C, Wolfensberger M, Honegger F, Friedrich G, Gugatschka M. The role of the cricothyroid joint anatomy in cricothyroid approximation surgery. *J Voice*. 2011;25(5):632-637.
35. Titze IR. Principles of voice production. Iowa City: National Center of Voice and Speech; 2000.
36. Fant G. Acoustic theory of speech production. The Hague: Mouton; 1960.
37. Kent RD. Vocal tract acoustics. *J Voice*. 1993;7(2):97-117.
38. Lindblom B, Sundberg J. Acoustical consequences of lip, tongue, jaw, and larynx movement. *J. Acoust. Soc. Am.* 1971;50(4):1166-1179.
39. Hillendrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 1995;97(5):3099-3111.
40. Díaz S, Cisternas P, López I. Acoustic characteristics of the Chilean Spanish vowels produced by subjects residents in the city of Santiago. *Revista Chilena de Fonoaudiología*. 2015;14:92-102.
41. Takemoto H, Adachi S, Kitamura T, Mokhtari P, Honda K. Acoustic roles of the laryngeal cavity in vocal tract resonance. *J. Acoust. Soc. Am.* 2006;120(4):2228-2238.
42. Titze IR, Story BH. Acoustic interactions of the voice source with the lower vocal tract. *J. Acoust. Soc. Am.* 1997;101(4):2234-2243.
43. Kitzing P. LTAS criteria pertinent to the measurement of voice quality. *J. Phon.* 1986;14:477-482.
44. Gauffin J, Sundbegr J. Data on the glottal voice source behavior in vowel production. *STL-QPSR*. 1980;21(2-3):61-70.
45. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *J Speech Lang Hear Res*. 1989;32:556-565.
46. Master S, De Biase N, Chiari BM, Laukkanen AM. Acoustic and perceptual analysis of Brazilian male actors' and nonactors' voices: Long-term average spectrum and the "actor's formant". *J Voice*. 2008;22(2):146-154.
47. Gordon M, Ladefoged P. Phonation types: A cross-linguistic overview. *J. Phon.* 2001;29(4):383-406.

48. Alku P, Strik H, Vilkman E. Parabolic spectral parameter: A new method for quantification of the glottal flow. *Speech Commun.* 1997;22:67-79.
49. Sundberg J, Thalén M, Alku O, Vilkman E. Estimating perceived phonatory pressedness in singing from flow glottograms. *J Voice.* 2004;18(1):56-62.
50. Björkner E, Sundberg J, Alku P. Subglottal pressure and normalized amplitude quotient variation in classically trained singers. *Logoped Phoniatr Vocol.* 2006;31(4):157-165.
51. Traser L, Burdumy M, Ritcher B, et al. 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.
52. Titze IR. Voice training and therapy with a semi-occluded vocal tract: Rationale and scientific underpinnings. *J Speech Lang Hear Res.* 2006;49:448-459.
53. Rothenberg M. Acoustic interaction between the glottal source and the vocal tract. In: Stevens KN, Hirano M (eds.). *Vocal Fold Physiology.* University of Tokyo Press, Tokyo. 1981: 305-328.
54. Rothenberg M. Source-tract acoustic interaction and voice quality. In: Lawrence VL, ed. *Transcripts of the 12<sup>th</sup> Symposium Care of Professional Voice, Part I.* New York, NY: The Voice Foundation; 1983:25-31.
55. Story BH, Laukkanen A-M, Titze IR. Acoustic impedance of an artificially lengthened and constricted vocal tract. *J Voice.* 2000;14:455-469.
56. Titze IR. Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to low dimensional phonation model. *J. Acoust. Soc. Am.* 2002;111(1):367-376.
57. Horáček J, Laukkanen AM, Šidlof P, Murphy P, Švec JG. Comparison of acceleration and impact stress as possible loading factors in phonation: A computer modeling study. *Folia Phoniatr Logop.* 2009;61:137-145.

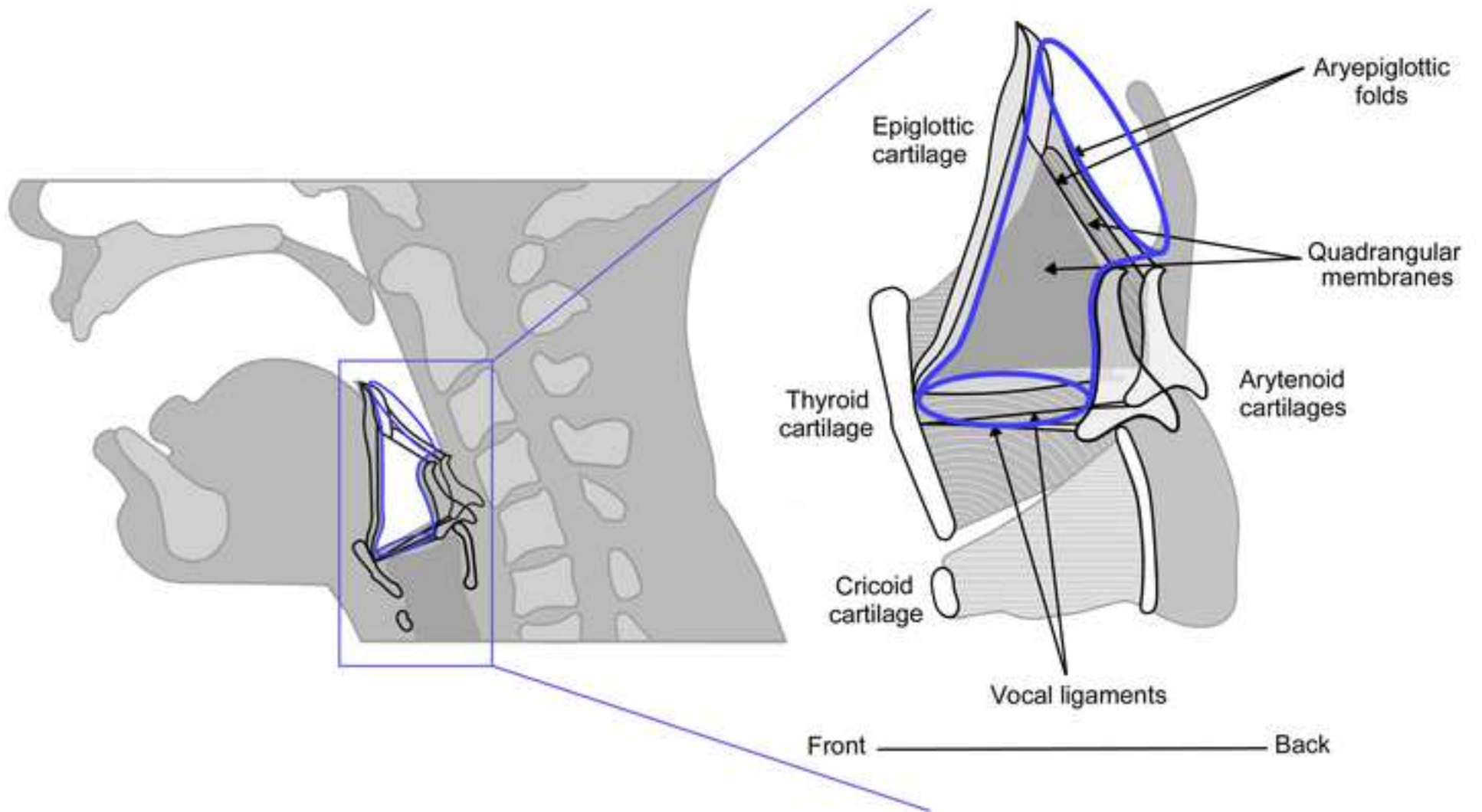


Figure 1.1

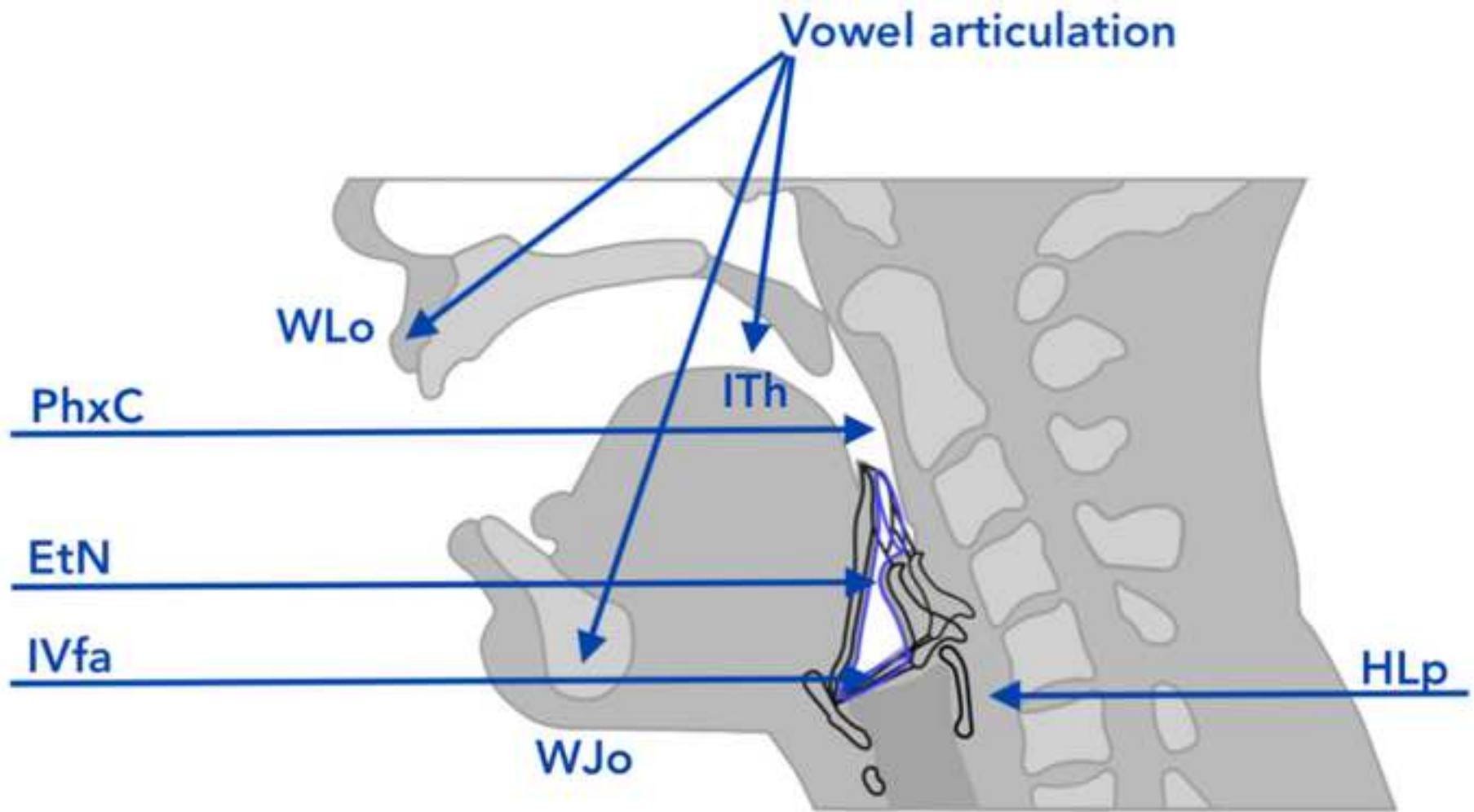


Figure 1.2



Figure 2

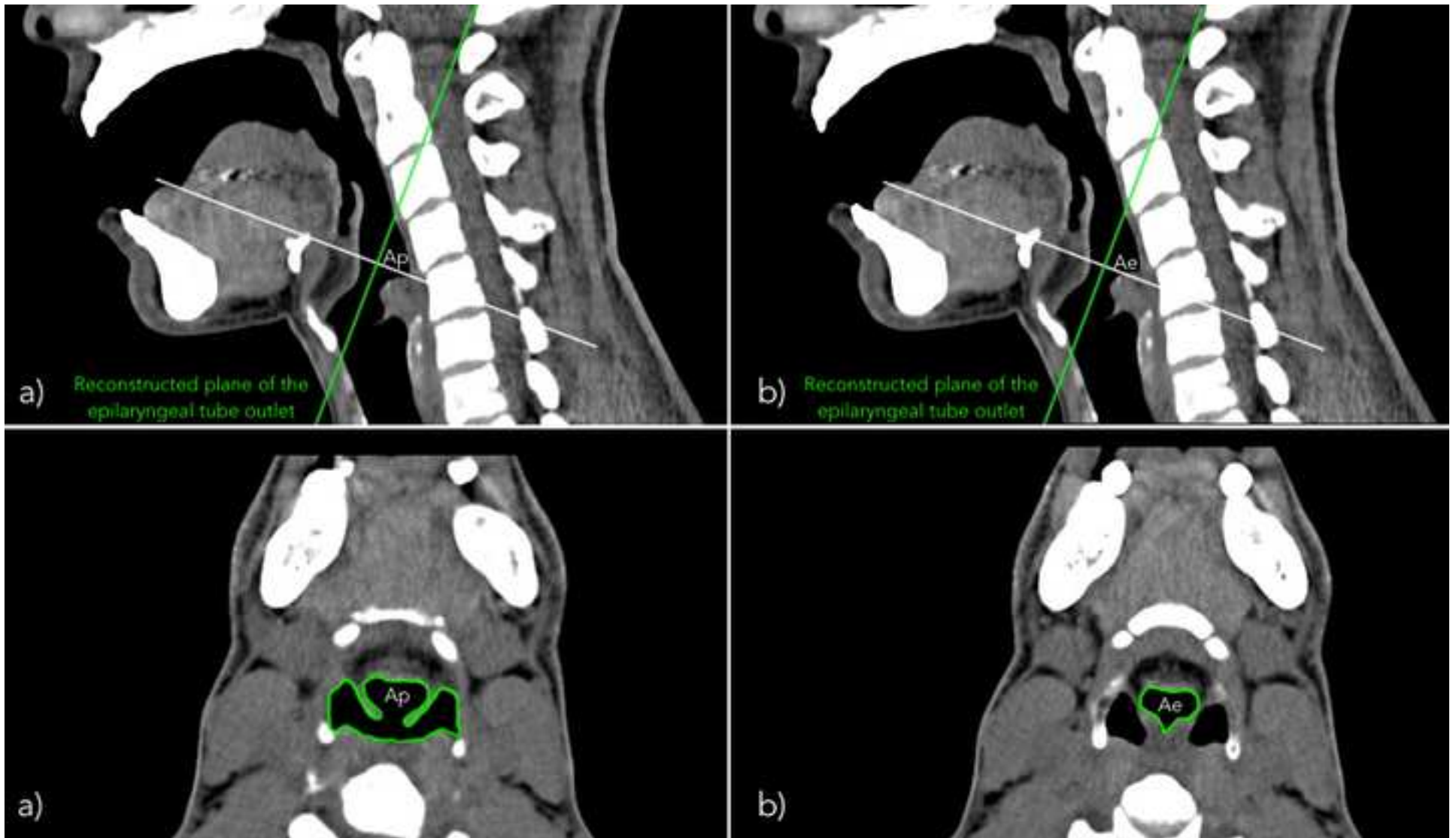


Figure 3

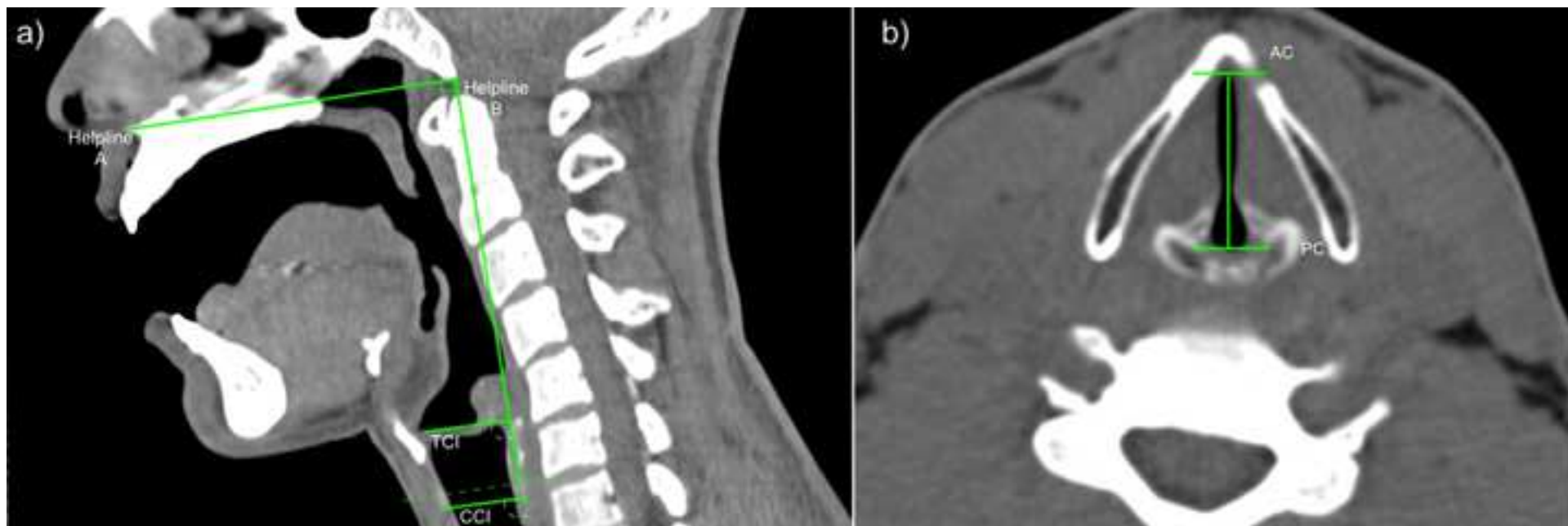


Figure 4



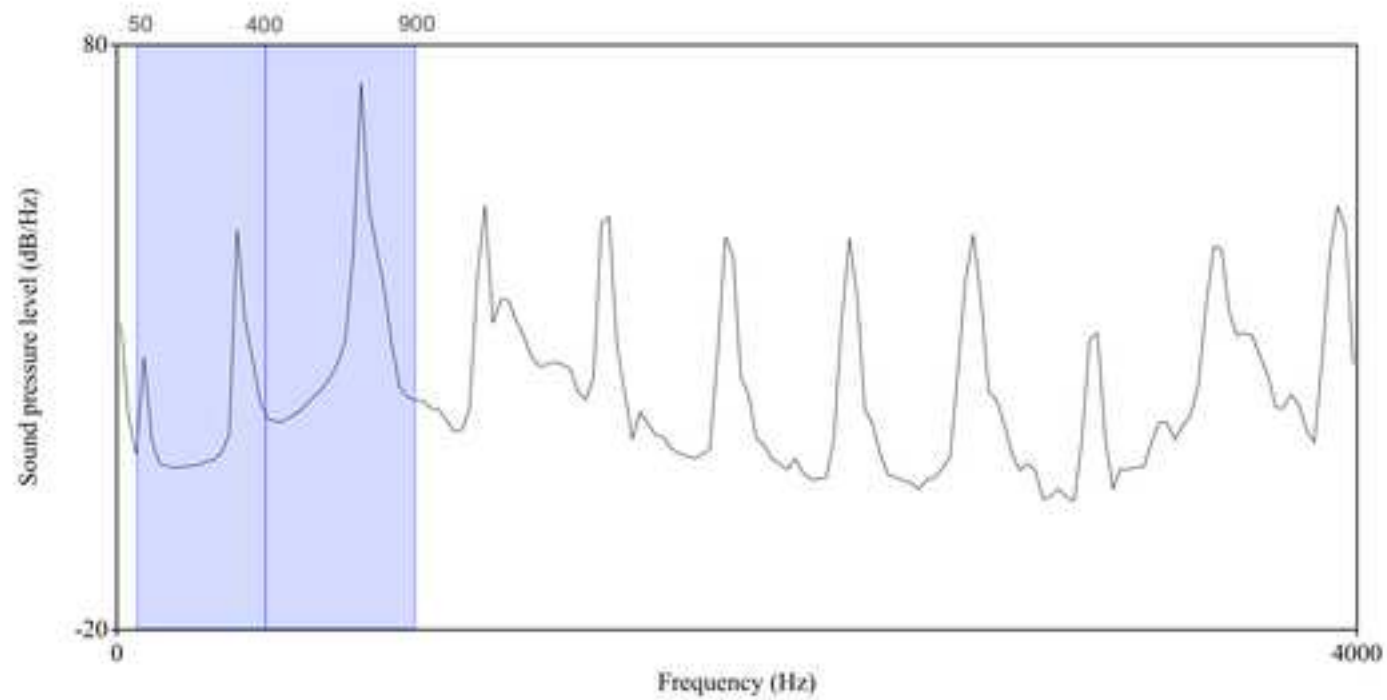
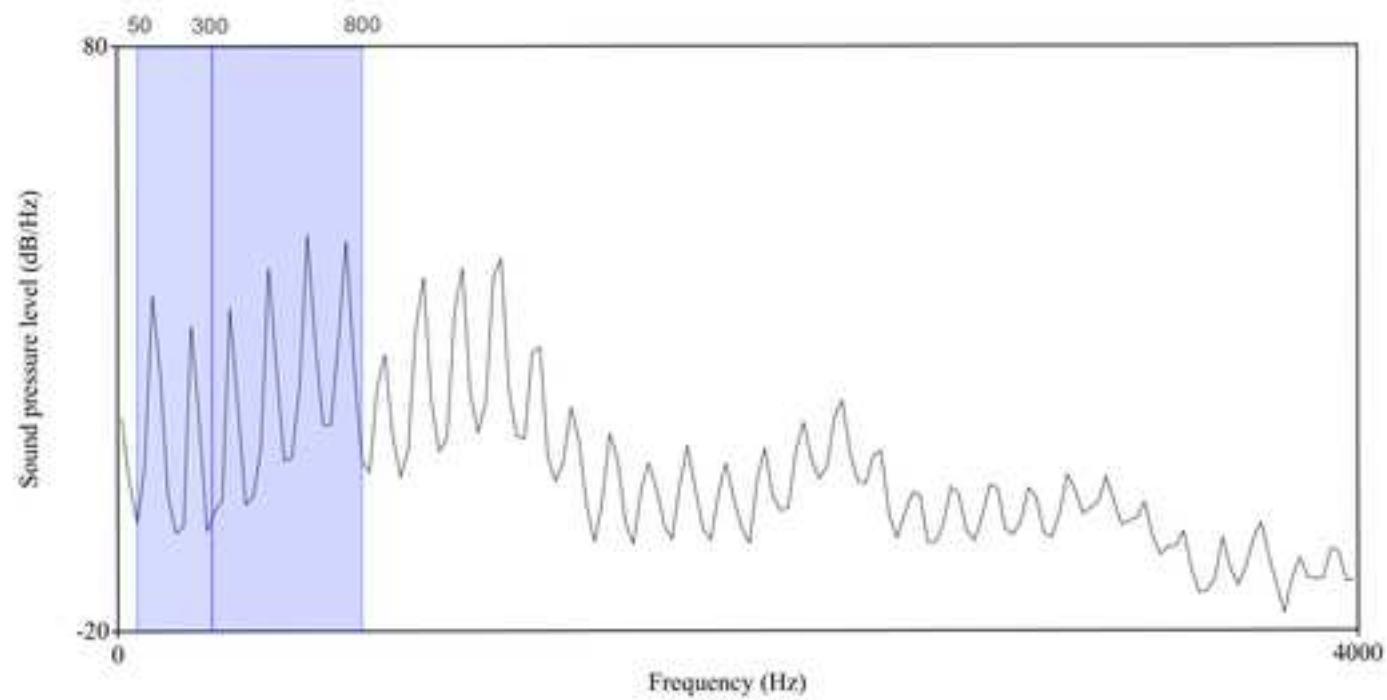


Figure 5

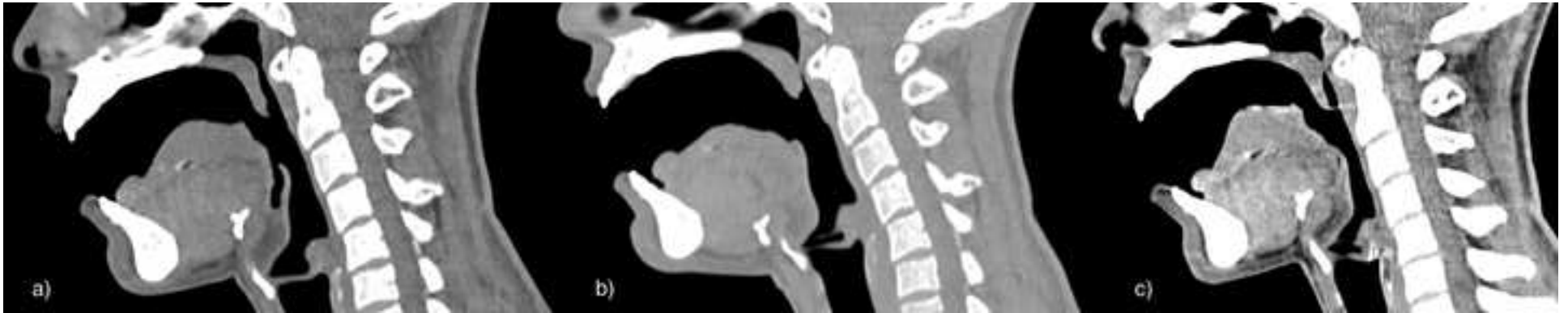


Figure 6.1

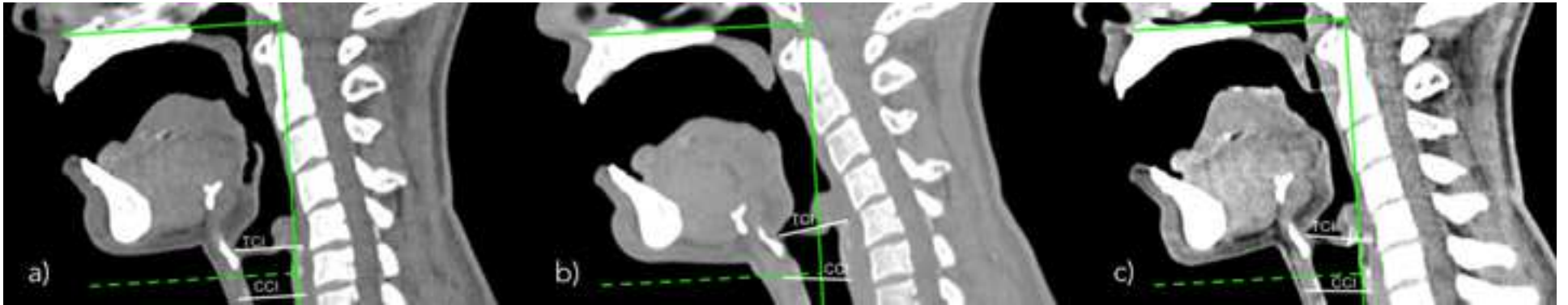


Figure 6.2



Figure 7

Ap



Ae



Figure 8

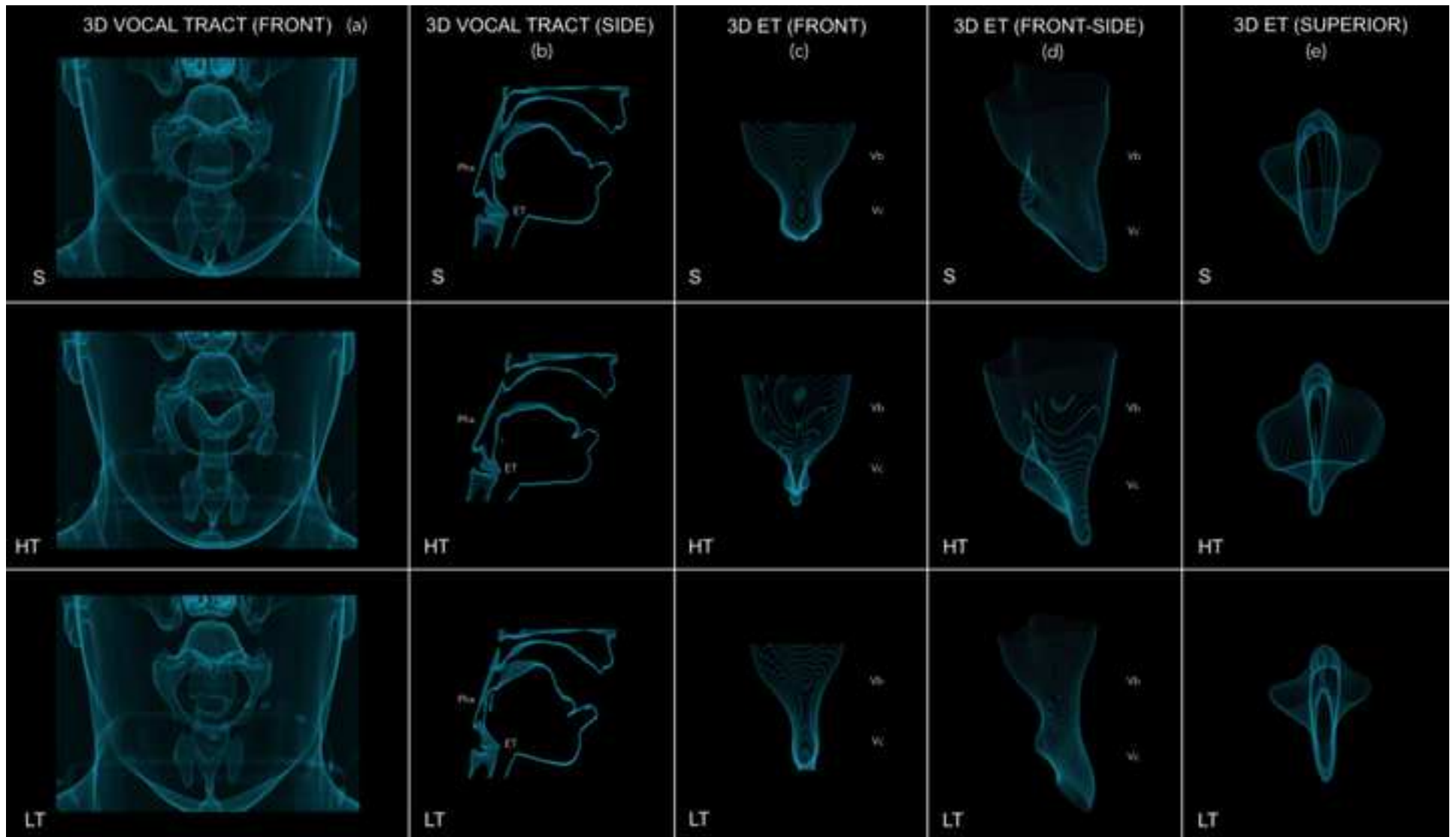


Figure 9

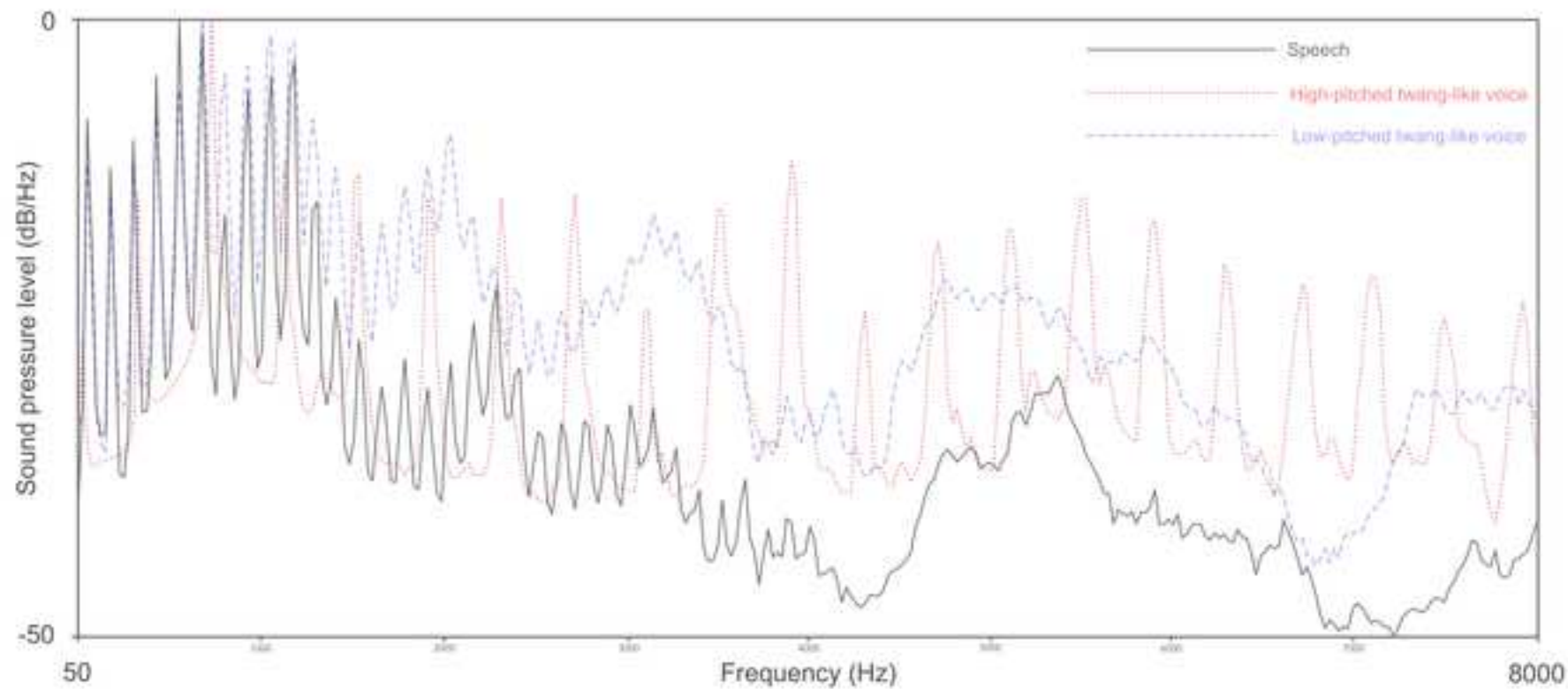


Figure 10.1

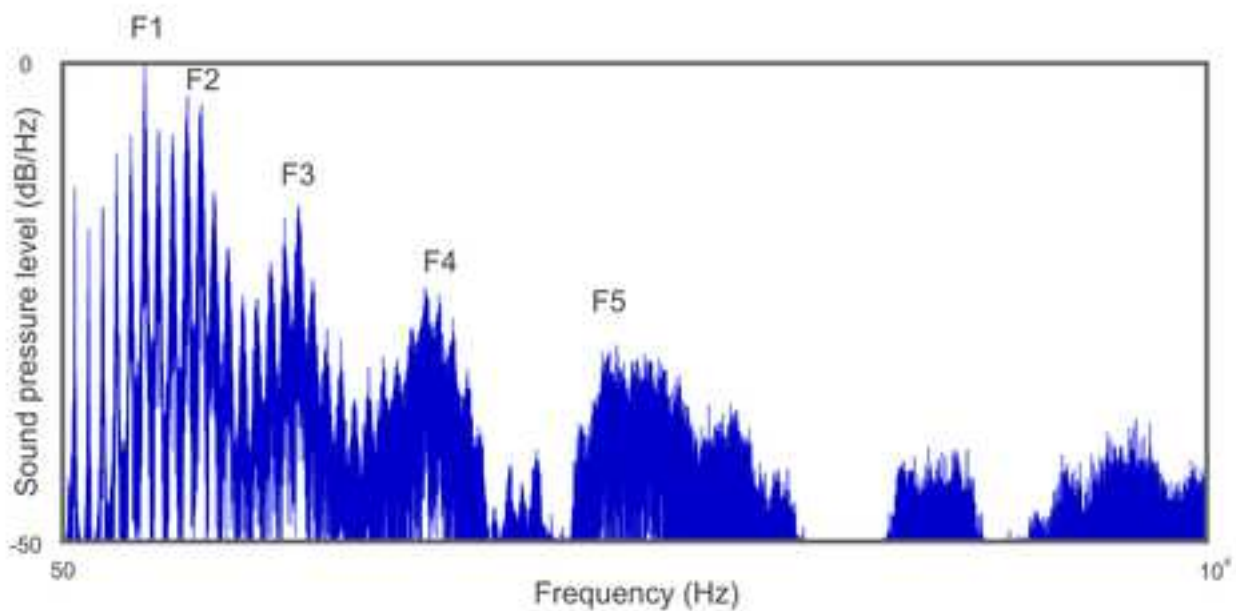
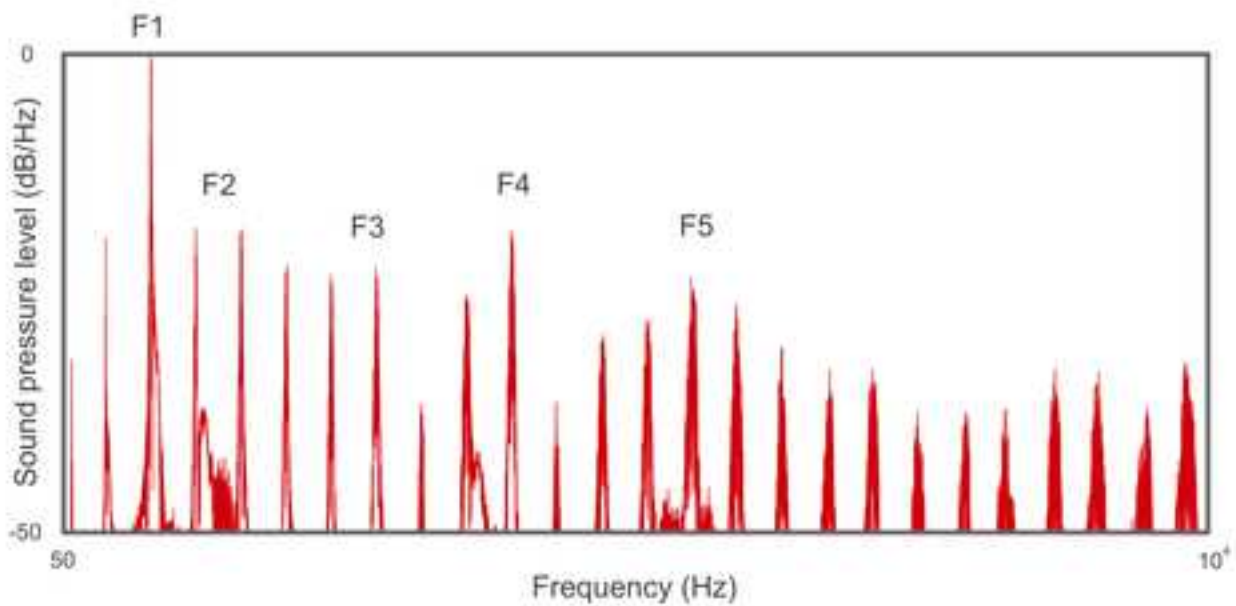
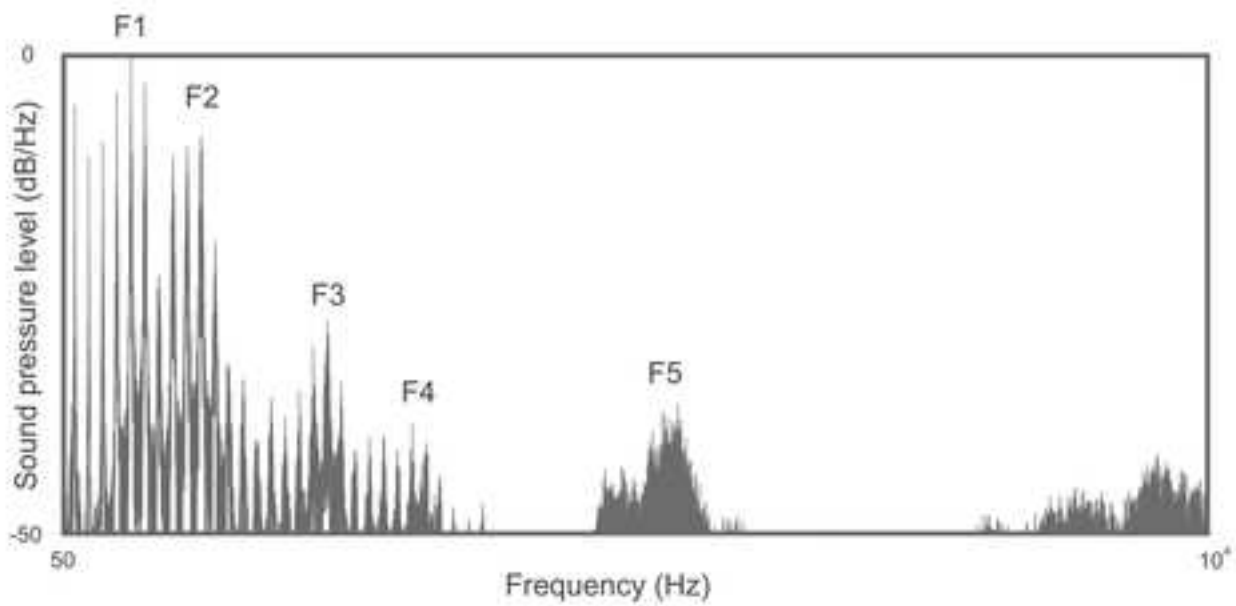


Figure 10.2



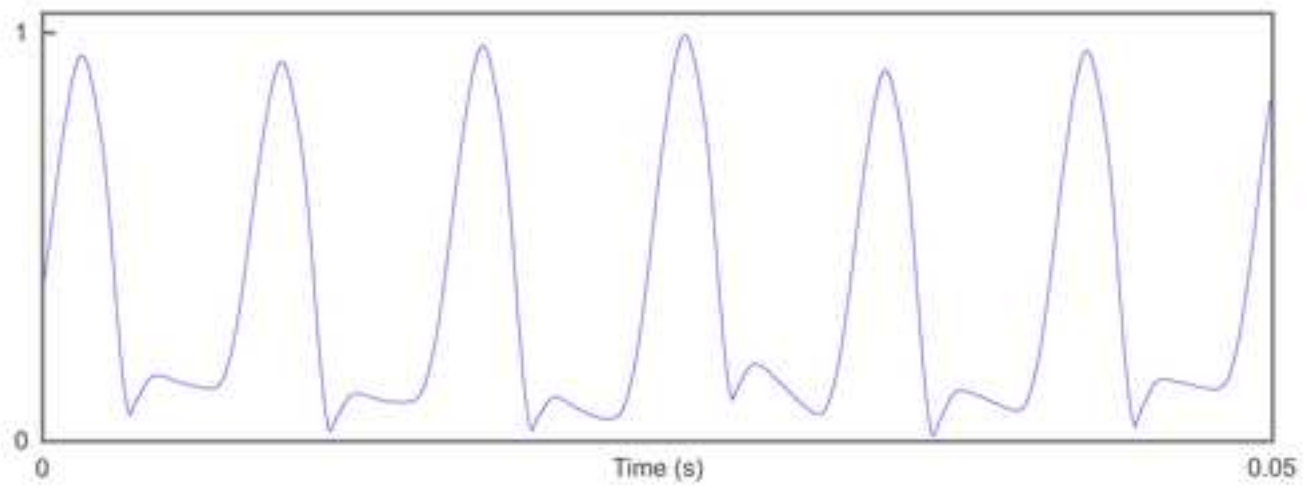
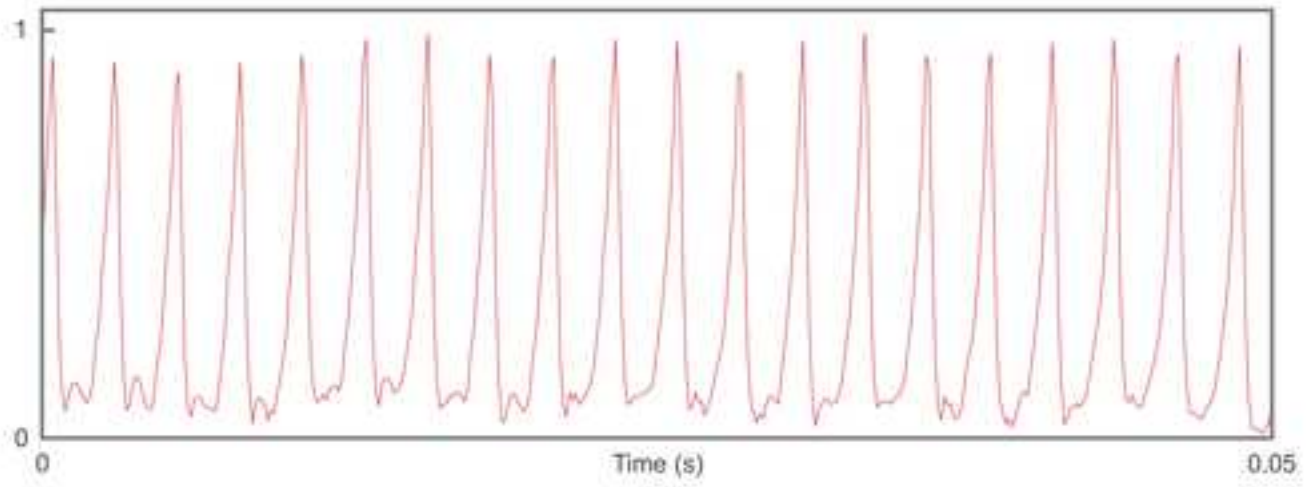
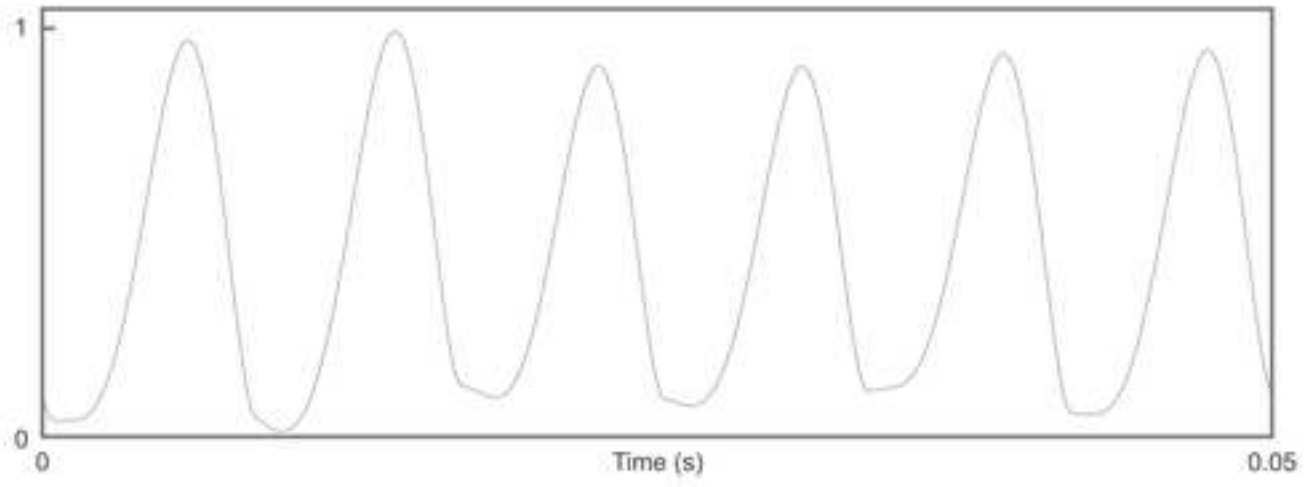


Figure 11.1

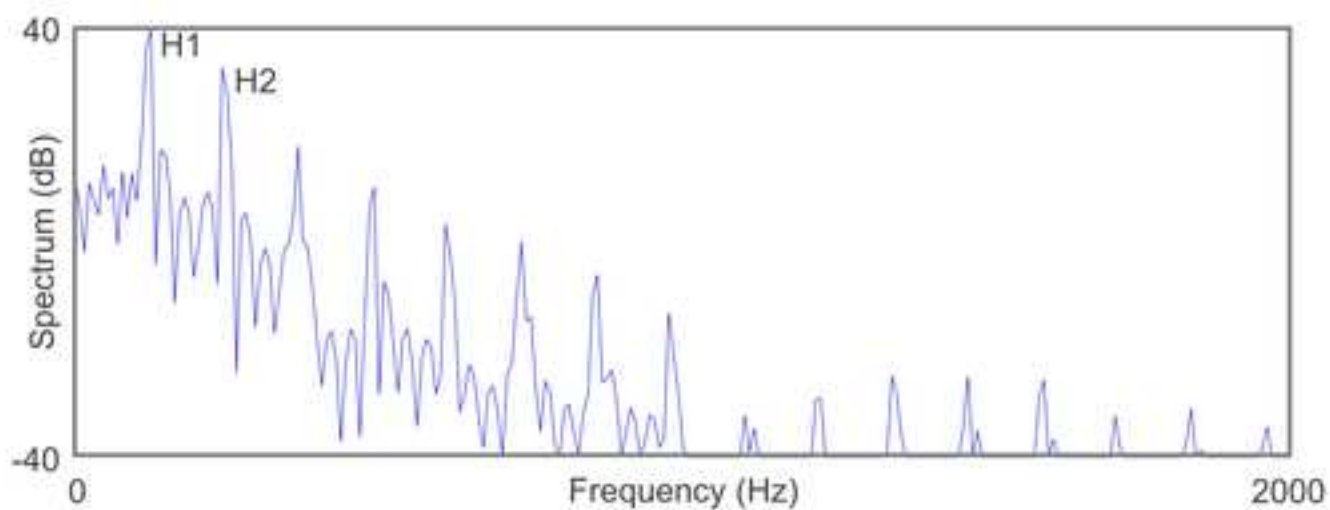
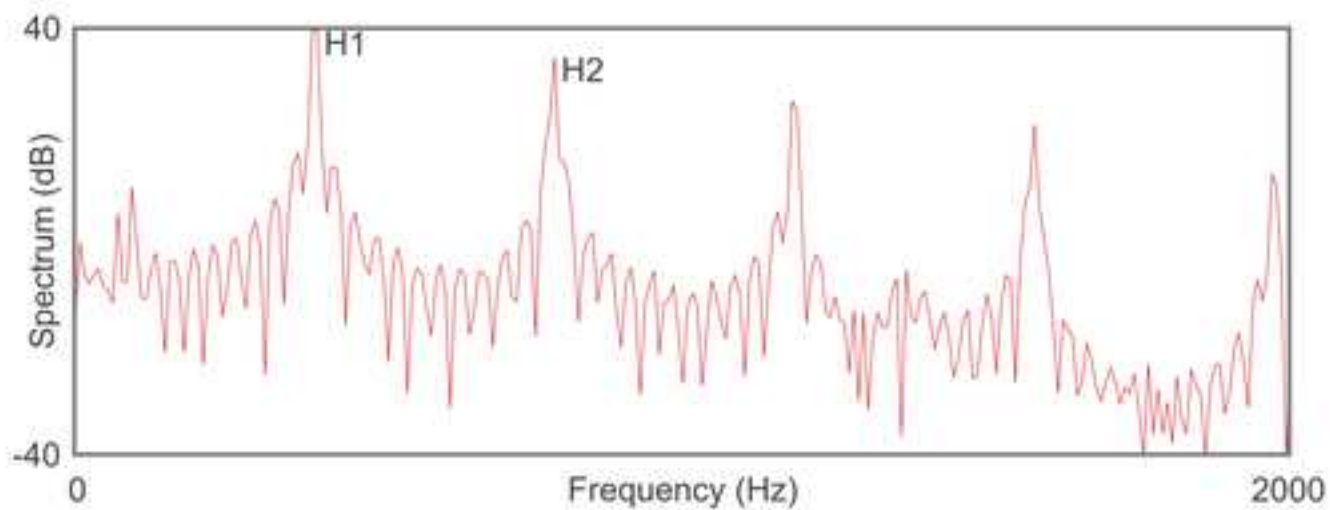
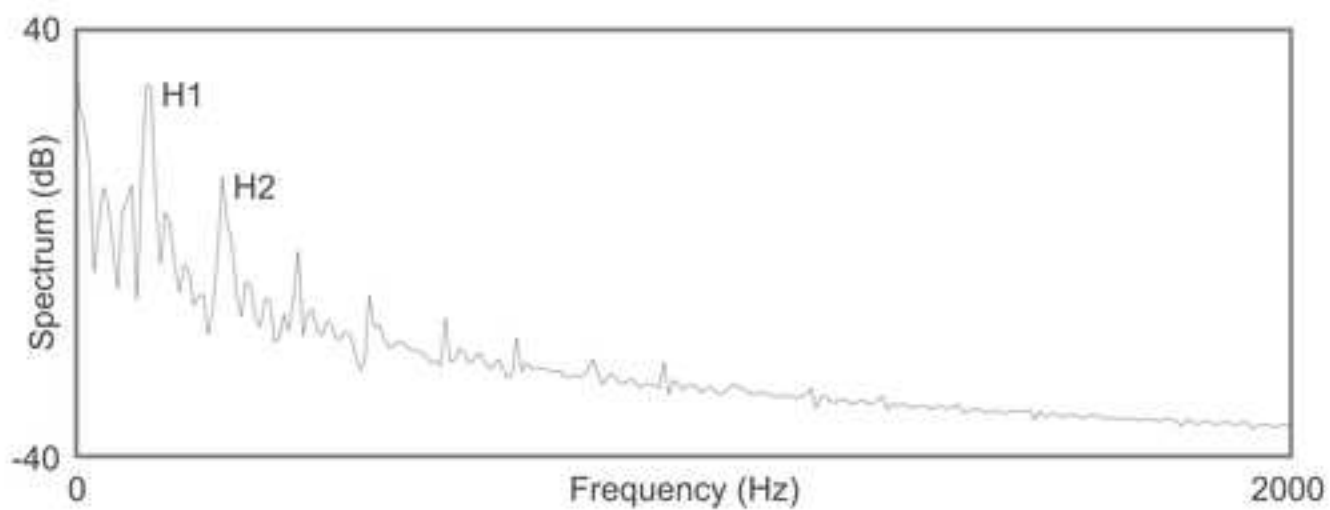


Figure 11.2

Table 1: Anatomic distances (mm) calculated from the midsagittal images of the vocal tract obtained from the CT measurements performed during the speaking voice and both twang-like voice qualities on the high and low pitches.

Distances (mm)	Speaking voice (SV)			High pitch twang-like voice quality (HT)						Low pitch twang-like voice quality (LT)					
	1	2	Mean	1	2	Mean	Error	<b>SV - HT Difference</b>	%	1	2	Mean	Error	<b>SV - LT Difference</b>	%
VL	91.70	91.50	91.60	85.20	84.50	84.85	0.7	<b>6.75</b>	7.36	85.30	86.50	85.90	1.2	<b>5.7</b>	6.22
HL	107.30	107.80	107.55	106.40	106.60	106.50	0.2	<b>1.05</b>	0.97	104.40	103.60	104.00	0.8	<b>3.55</b>	3.30
LO	30.80	28.00	29.40	39.00	38.60	38.80	0.4	<b>-9.4</b>	31.97	37.80	36.20	37.00	1.6	<b>-7.6</b>	25.85
JO	105.00	102.30	103.65	110.70	106.90	108.80	3.8	<b>-5.15</b>	4.96	111.40	107.40	109.40	4.0	<b>-5.75</b>	5.54
TDH	72.80	74.00	73.40	62.50	62.70	62.60	0.2	<b>10.8</b>	14.71	74.70	77.60	76.15	2.9	<b>-2.75*</b>	3.74
OW	14.20	13.90	14.05	16.50	15.50	16.00	1.0	<b>-1.95</b>	13.87	6.56	6.78	6.67	0.22	<b>7.38</b>	52.52
VE	31.30	31.30	31.30	26.80	25.90	26.35	0.9	<b>4.95</b>	15.81	31.10	30.20	30.65	0.9	<b>0.65*</b>	2.07
HW	17.70	19.40	18.55	21.20	21.30	21.25	0.1	<b>-2.7</b>	14.55	17.0	16.4	16.7	0.6	<b>1.85</b>	9.97

Notes: Differences between samples were considered as such only if the difference was larger than the reference limit of error. Differences within the limit of error are marked (\*). Abbreviations: VL, vertical length of the vocal tract; HL, horizontal length of the vocal tract; LO, lip opening; JO, jaw opening; TDH, tongue dorsum height; OW, oropharyngeal width; VE, velum elevation; HW, hypopharyngeal width.

Table 2: Laryngeal cartilages inclination degrees and glottal length (mm) calculated from the midsagittal and transversal images of the vocal tract obtained from the CT measurements performed during the speaking voice and both twang-like voice qualities on the high and low pitches.

Measures	Speaking voice (SV)			High pitch twang-like voice quality (HT)						Low pitch twang-like voice quality (LT)					
	1	2	Mean	1	2	Mean	Error	<b>SV - HT Difference</b>	%	1	2	Mean	Error	<b>SV - LT Difference</b>	%
Thyroid Angle	89.10	89.64	89.37	80.045	79.79	79.91	0.25	<b>9.46</b>	10.58	98.62	98.83	98.72	0.21	<b>- 9.35</b>	10.46
Cricoid angle	89.67	89.075	89.37	93.07	91.90	92.48	1.17	<b>- 3.11</b>	3.47	92.16	93.20	92.68	1.04	<b>- 3.31</b>	3.70
Glottal length (mm)	28.20	28.60	28.40	31.30	31.40	31.35	0.1	<b>- 2.95</b>	10.38	25.50	26.20	25.85	0.7	<b>2.55</b>	8.97

Notes: Differences between samples were considered as such only if the difference was larger than the reference limit of error.

Table 3: Cross-sectional areas (cm<sup>2</sup>) calculated from the midsagittal and transversal images of the vocal tract obtained from the CT measurements performed during the speaking voice and both twang-like voice qualities on the high and low pitches.

Areas (cm <sup>2</sup> )	Speaking voice (SV)			High pitch twang-like voice quality (HT)						Low pitch twang-like voice quality (LT)					
	1	2	Mean	1	2	Mean	Error	<b>SV - HT Difference</b>	%	1	2	Mean	Error	<b>SV - LT Difference</b>	%
A1	23.70	24.70	24.20	30.50	31.20	30.85	0.7	<b>- 6.65</b>	27.47	27.40	28.60	28.00	1.2	<b>- 3.80</b>	15.70
A2	6.20	6.00	6.10	6.80	7.10	6.95	0.3	<b>- 0.85</b>	13.93	5.50	5.30	5.40	0.2	<b>0.70</b>	11.47
A3	8.70	8.70	8.70	9.50	9.60	9.55	0.1	<b>- 0.85</b>	9.77	6.30	5.50	5.75	0.8	<b>2.95</b>	33.90
Ap	5.62	5.60	5.61	5.13	5.15	5.14	0.02	<b>0.47</b>	8.37	3.57	3.56	3.56	0.01	<b>2.05</b>	36.54
Ae	1.70	1.70	1.70	1.77	1.78	1.77	0.01	<b>-0.07</b>	4.11	0.80	0.80	0.80	0.0	<b>0.90</b>	52.94
Ap/Ae	3.30	3.29	3.29	2.89	2.89	2.89	0.0	<b>0.40</b>	12.15	4.46	4.45	4.45	0.01	<b>- 1.16</b>	35.25

Notes: Differences between samples were considered as such only if the difference was larger than the reference limit of error. Abbreviations: A1, oral cavity; A2, pharyngeal region; A3, epilaryngeal region; Ap, area of the inlet of the lower pharynx; Ae, area of outlet of the epilaryngeal tube; Ap/Ae, pharynx to epilarynx ratio.

Table 4: Volumetric measures of epilaryngeal tube (mm<sup>3</sup>) calculated from the CT images during the speaking voice and both twang-like voice qualities on the high and low pitches.

Voice quality	Speaking voice (SV)	High pitch twang-like voice quality (HT)	<b>SV - HT Difference</b>	%	Low pitch twang-like voice quality (LT)	<b>SV - LT Difference</b>	%
ET Volume (mm <sup>3</sup> )	1.919	1.511	<b>0.408</b>	21.26	0.9528	<b>0.9662</b>	50.34

Table 5: Results of the acoustic analysis for LTAS variables obtained from the speaking voice and both twang-like voice qualities on the high and low pitches.

Voice quality	L0	L1	L1-L0 (dB)	SPR (dB)
Speaking Voice (SV)	28.79	38.19	9.36	-28.49
Twang-like quality at high pitch (HT)	36.98	60.56	23.58	-21.03
<b>Difference between SV and HT</b>	<b>- 8.19</b>	<b>- 22.37</b>	<b>- 14.22</b>	<b>7.46</b>
Twang-like quality at low pitch (LT)	32.49	47.15	14.50	-14.19
<b>Difference between SV and LT</b>	<b>- 3.7</b>	<b>- 8.96</b>	<b>5.14</b>	<b>14.30</b>

Table 6: Results of the fundamental frequencies ( $F_0$ ) and formant frequency values obtained from the speaking voice and both twang-like voice qualities on the high and low pitches.

Voice quality	Fundamental frequency (Hz)	SPL (dB)	Formant frequencies (Hz)					Formant frequencies differences (Hz)				
			F1	F2	F3	F4	F5	F2-F1	F3-F2	F4-F3	F5-F4	F5-F3
Speaking Voice (SV)	123.67 (B2)	67	660	1200	2300	3100	5330	540	1100	800	2230	3030
Twang-like quality at high pitch (HT)	394.29 (G4)	88	820	1580	2780	3800	5490	760	1200	1020	1690	2710
<b>Difference between SV and HT</b>	<b>- 270.62</b>	<b>-21</b>	<b>-160</b>	<b>-380</b>	<b>-480</b>	<b>-700</b>	<b>-160</b>	<b>-220</b>	<b>-100</b>	<b>-220</b>	<b>540</b>	<b>320</b>
Twang-like quality at low pitch (LT)	122.28 (B2)	74	750	1220	2100	3400	4780	470	880	1300	1380	2680
<b>Difference between SV and LT</b>	<b>1.39</b>	<b>-7</b>	<b>-90</b>	<b>-20</b>	<b>200</b>	<b>-300</b>	<b>550</b>	<b>70</b>	<b>220</b>	<b>-500</b>	<b>850</b>	<b>320</b>



Table 7: Results of the inverse filtering parameters obtained from the speaking voice and both twang-like voice qualities on the high and low pitches.

Inverse filtering data	Speaking Voice (SV)	Twang-like quality at high pitch (HT)	<b>Difference between SV and HT</b>	Twang-like quality at low pitch (LT)	<b>Difference between SV and LT</b>
NAQ	0.213	0.137	<b>0.076</b>	0.138	<b>0.075</b>
AQ	1.760	0.346	<b>1.414</b>	1.129	<b>0.631</b>
CIQ	0.425	0.297	<b>0.128</b>	0.241	<b>0.184</b>
H1-H2 (dB)	15.014	3.837	<b>11.177</b>	7.823	<b>7.191</b>
PSP (dB)	0.197	0.105	<b>0.092</b>	0.117	<b>0.08</b>
HRF (dB)	-7.308	3.305	<b>10.613</b>	-2.584	<b>4.724</b>
T (ms)	8.264	2.532	<b>5.732</b>	8.197	<b>0.067</b>

## Table legends:

**Table 1:** Anatomic distances (mm) calculated from the midsagittal images of the vocal tract obtained from the CT measurements performed during the speaking voice and both twang-like voice qualities on the high and low pitches.

**Table 2:** Laryngeal cartilages inclination degrees and glottal length (mm) calculated from the midsagittal and transversal images of the vocal tract obtained from the CT measurements performed during the speaking voice and both twang-like voice qualities on the high and low pitches.

**Table 3:** Sagittal cross-sectional areas and epilaryngeal tube outlet and pharyngeal inlet areas (cm<sup>2</sup>) obtained from the CT images of the speaking voice and both twang-like voice qualities on the high and low pitches.

**Table 4:** Volumetric measures of epilaryngeal tube (mm<sup>3</sup>) calculated from the CT images during the speaking voice and both twang-like voice qualities on the high and low pitches.

**Table 5:** Results of the acoustic analysis for LTAS variables obtained from the speaking voice and both twang-like voice qualities on the high and low pitches.

**Table 6:** Results of the fundamental frequencies (F0) and formant frequency values obtained from the speaking voice and both twang-like voice qualities on the high and low pitches.

**Table 7:** Results of the inverse filtering parameters obtained from the speaking voice and both twang-like voice qualities on the high and low pitches.

## Figure legends:

**FIGURE 1.1. Lower vocal tract structures related to twang.** Epilaryngeal tube (ET) position within the vocal tract (to the left), and schematic anatomical representation of the ET (to the right). ET is defined as the space between the vocal folds and the aryepiglottic folds [10]. Aryepiglottic folds connect the lateral portions of the epiglottis to the arytenoid cartilages. Two membranes (the quadrangular membranes) form the lateral walls of the ET, while the epiglottis cartilage and the arytenoid cartilages form the anterior and posterior walls, respectively [1]. The

small ET (2.5-3.5 cm in length and 0.2-1.0 cm<sup>2</sup> average cross-sectional area) opens into a larger tube, the pharynx [10].

**FIGURE 1.2. Schematic representation of the anatomic aspects that would contribute to twang-like voice production.** Lower vocal tract adjustments represented by PhxC (pharynx constriction), EtN (epilaryngeal tube narrowing), and HLp (higher laryngeal position). Upper vocal tract adjustments related to vowel articulation represented by WLo (wider lip opening), WJo (wider jaw opening), and ITh (increased tongue height). Voice source represented by IVfa (increased vocal fold adduction).

**FIGURE 2. Measurement of the sagittal distances and areas.** (a) To the left, measurement of sagittal distances: 1. vertical position of the larynx (VL), 2. horizontal length of the vocal tract (HL), 3. lip opening (LO), 4. jaw opening (JO), 5. tongue dorsum height (TDH), 6. oro-pharyngeal width (OW), 7. velum elevation (VE), and 8. hypo-pharyngeal width (HW). (b) To the right, measurement of sagittal cross-sectional areas: A<sub>1</sub>. oral, A<sub>2</sub>. pharyngeal, and A<sub>3</sub>. epilaryngeal.

**FIGURE 3. Measurement of the epilaryngeal tube outlet and pharyngeal inlet areas.** (a) To the left: the reconstructed plane of the epilaryngeal tube outlet (above) and the pharyngeal inlet area (A<sub>p</sub>) (below). (b) To the right: the reconstructed plane of the epilaryngeal tube outlet (above) and the epilaryngeal outlet area (A<sub>e</sub>) (below).

**FIGURE 4. Measurement of the inclination of the laryngeal cartilages and glottal length.** (a) To the left: thyroid cartilage inclination (TCI) and cricoid cartilage inclination (CCI). (b) To the right: glottal length, from the anterior commissure (AC) to the posterior commissure (PC).

**FIGURE 5. L<sub>1</sub>-L<sub>0</sub> bandwidths.** (a) Bandwidth considered for both low-pitched samples (speaking and twang-like voice quality). The frequency range considered for L<sub>0</sub> was 50 to 300 Hz. The frequency range considered for L<sub>1</sub> was 300 to 800 Hz ; (b) Bandwidth considered for the high-pitched twang-like voice quality. The frequency range considered for L<sub>0</sub> was 50 to 400 Hz. The frequency range considered for L<sub>1</sub> was 400 to 900 Hz .

**FIGURE 6.1. Sagittal images.** (a) To the left: speaking voice quality; (b) Center: twang-like voice quality on the high pitch; (c) To the right: twang-like voice quality on the low pitch.

**FIGURE 6.2. Inclination of laryngeal cartilages.** (a) To the left: speaking voice quality; (b) Center: twang-like voice quality on the high pitch; (c) To the right: twang-like voice quality on the low pitch. Thyroid cartilage inclination = TCI; cricoid cartilage inclination = CCI.

**FIGURE 7. Glottal length.** (a) To the left: speaking voice quality; (b) Center: twang-like voice quality on the high pitch; (c) To the right: twang-like voice quality on the low pitch.

**FIGURE 8. Epilaryngeal tube outlet and pharyngeal inlet areas.** (a) To the left: speaking voice quality; (b) Center: twang-like voice quality on the high pitch; (c) To the right: twang-like voice quality on the low pitch. The upper images represent the  $A_p$  areas and the images below the  $A_e$  area for each voice quality.

**FIGURE 9. Three-dimensional (3D) representation of the vocal tract and the epilaryngeal tube.** 3D representation of the vocal tract was performed for each voice quality in different planes: (a) the vocal tract in the frontal plane; (b) the vocal tract in the sagittal plane. Phx: pharynx. ET: epilaryngeal tube; (c) the epilaryngeal tube in the frontal plane. Vb: vestibular region. Vc: ventricular region; (d) the epilaryngeal tube in a front-side view. Vb: vestibular region. Vc: ventricular region; (e) the epilaryngeal tube in the transversal plane in a superior view (from above). Upper row: speaking voice quality (S); Middle row: high-pitched twang-like voice quality (HT); Lower row: low-pitched twang-like voice quality (LT).

**FIGURE 10.1. Normalized long-term average spectrum.** Comparison of the differences in sound energy between the speaking voice (solid line), the high-pitched twang-like voice quality (dotted line), and the low-pitched twang-like voice quality (dashed line). The peak amplitude of each LTAS was normalized, set to zero. The frequency range considered in the plotting process was between 50 and 8000 hertz.

**FIGURE 10.2. Formant distribution based on FFT spectrum analysis.** The location of the first five formants ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ , and  $F_5$ ) is indicated for each voice sample: the speaking voice (top), the twang-like voice quality on the high pitch (middle), and the twang-like voice quality on the low pitch (bottom).

**FIGURE 11.1. Time-domain waveforms of the glottal flow.** The speaking voice (top), the twang-like voice quality on the high pitch (middle), and the twang-like voice quality on the low pitch (bottom).

**FIGURE 11.2. Voice source spectrums.** The speaking voice (top), the twang-like voice quality on the high pitch (middle), and the twang-like voice quality on the low pitch (bottom).