How Anatomy Influences the Sound of a Voice

Marik Roos, Veronika Weber, Christoph Reuter
Institut für Musikwissenschaft, Universität Wien, 1090 Wien, E-Mail: marik.roos@univie.ac.at

Background

Each vocal tract is unique and so is the sound of a voice. The perception not only of personality traits and emotional states in speech, but also of the speaker‘s gender can be predicted more accurately by timbre rather than fundamental frequency [10]. Also, voices that are either perceived as male or female show significant differences in timbre features such as formant structures and Mel Frequency Cepstral Coefficients, especially when it comes to the mfcc2, 4, 5, 7, 9 and 13 [9].

The aim of this study is to gain a better understanding of the underlying connections between the vocal anatomy and the sound it produces by analyzing the timbre changes as results of different methods of surgical alteration of the vocal tract.

Patients of ‚Voice Feminization Surgery‘ provide exceptionally suitable examples of surgically altered vocal anatomy, especially since there are different methods of surgery to compare with one another. In this study, five of the most prominent methods with overlapping procedures were compared to link timbre changes specifically to a certain anatomical alteration.

In *Cricothyroid Approximation* (CTA), the thyroid and cricoid cartilage are stitched together to generate a permanent contraction and thus release of the cricothyroid muscle.

Figure 1: Illustration of Cricothyroid Approximation [2]

In *Feminization Laryngoplasty* (FemLar), parts of the vocal fold as well as anterior thyroid cartilage are being removed to shorten the vocal folds and narrowing the larynx.

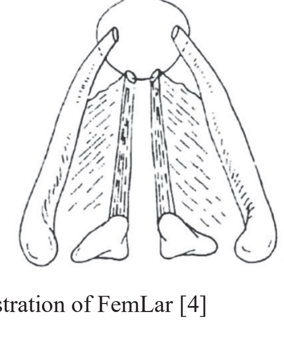

Figure 2: Illustration of FemLar [4]

In FemLar plus Thyrohyoid Elevation (FLT), the thyroid cartilage is stitched to the hyoid bone, thus shortening the larynx in addition to the basic FemLar procedure.

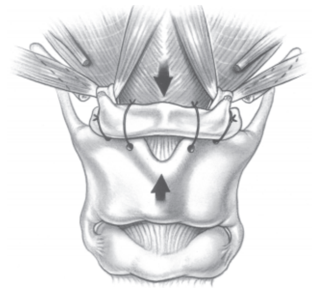

Figure 3: Illustration of FLT [1]

The main method of *Vocal Folds Shortening and Retro-displacement of Anterior Commissure* (VFSRAC) is to stitch the vocal folds together at approximately 1/3 of length and stabilize them additionally by attachment to the posterior larynx.

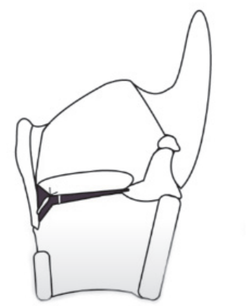

Figure 4: Illustration of VFSRAC [3]

In *Vocal Fold Webbing* (VFW), the vocal folds are being stitched together at about 1/3 of length.

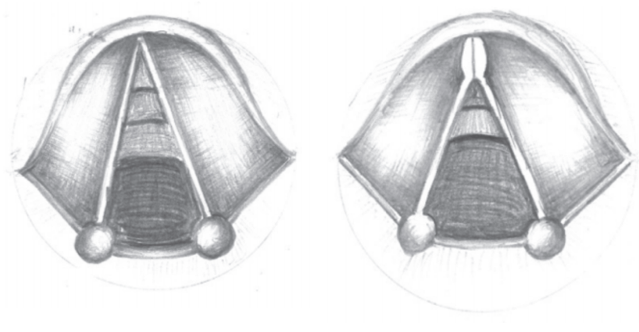

Figure 5: Illustration of Vocal Fold Webbing [8]

As the methods feature similar alterations, the timbral analyses are hypothesized to show similarities accordingly. CTA and FLT share the elevation of the larynx, VFSRAC and VFW the webbing of vocal folds, and FLT and FemLar share the removal of cricoid cartilage and vocal fold tissue.

Methods

N=129 patients who underwent one of the five methods of Voice Feminization Surgery introduced above provided 258 recordings in total, each set of recordings containing one pre- and one post-op example of the same sentence. From each recording, different and cross-culturally comparable vowels were extracted and analyzed using timbre feature extractor scripts made by Isabella Czedik-Eysenberg combining MIRToolBox [5] and Mining Suite [6] as well as the Python library „Parselmouth“ [7]. Paired as well as independent samples *t*-tests and regression along with factor analyses were conducted between singular vowels and timbre feature mean values of each voice to estimate significant differences between pre- and post-op timbres as well as different methods of altering the vocal tract. Decreased tonal or band energy were not interpreted due to possible effects of the individual healing process on a voice‘s physical strength and roughness features, since the post-op examples varied between from being recorded 3 weeks up to 3 months after surgery.

Results

Each of the five procedures showed the overall intended effect of a significant increase in pitch. However, the changes in timbre differed significantly for each method.

For CTA, no significant changes in formant structure as well as spectral centroid could be observed. The increase of pitch was significant (t=2.906, p=.009, d=1.3, df=18).

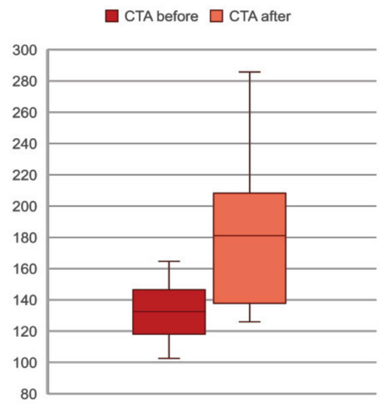

Figure 6: Increase of average pitch before and after CTA (t=2.906, p=.009, d=1.3, df=18).

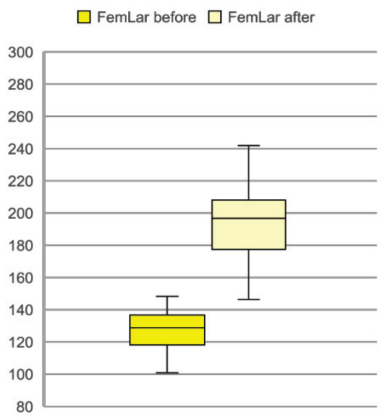

Figure 7: Increase of average pitch before and after FemLar (t=9.601, p<.001, d=3.506, df=28).

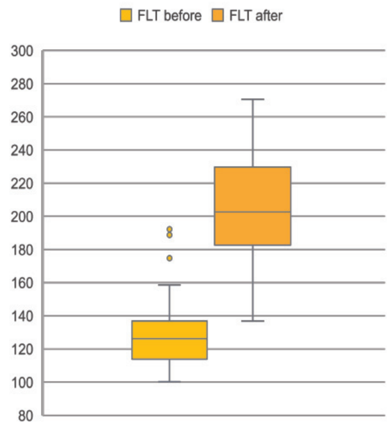

Figure 8: Increase of average pitch before and after FLT (t=14.422, p<.001, d=2.856, df=100).

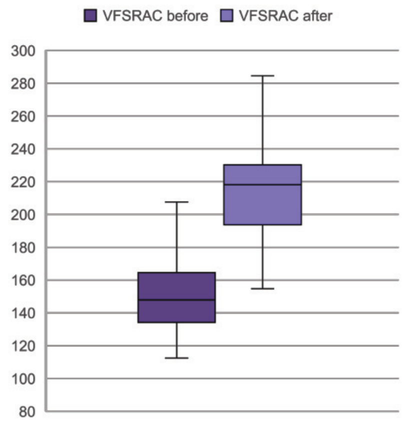

Figure 9: Increase of average pitch before and after VFSRAC (t=2.906, p<.001, d=2.551, df=68).

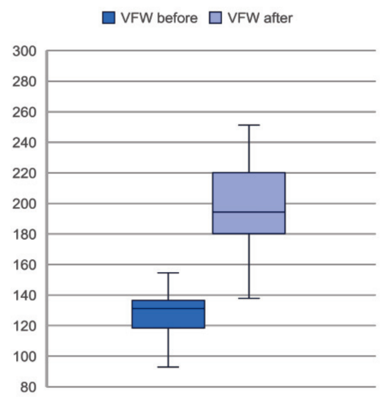

Figure 10: Increase of average pitch before and after VFW (t=9.616, p<.001, d=3.205, df=34).

Significant changes in mfccs could be observed for every procedure, mfcc2, 4, 5, 6, 7, 9 and 11 having predominantly increased while mfcc1, 3, 8 and 12 having decreased, though not consistently for each method. Interestingly, similar mfcc changes could be observed in similar methods, for example mfcc 3 decreasing significantly for both methods using the FemLar-technique, mfcc 7 increasing for both methods that involve elevation of the larynx, mfcc 1, 6 and 9 showing the same changes in both procedures using vocal fold webbing.

Table 1: Significantly higher (green) and lower (orange) post-op mfcc (1 to 13) changes with medium to large effect sizes, sorted by surgical method

	CTA [df=9]	FemLar [df=14]	FLT [df=50]	VFSRAC [df=34]	VFW [df=17]
1			*d*=.421 *p*=.004	*d*=.399 *p*=.024	*d*=.626 *p*=.017
2			*d*=.548 *p*<.001		
3		*d*=.687 *p*=.019	*d*=.791 *p*<.001		
4		*d*=1.291 *p*<.001	*d*=.843 *p*<.001		
5				*d*=.712 *p*<.001	
6				*d*=.333 n. s.	*d*=.468 n. s.
7	*d*=.842 *p*=.011		*d*=.550 *p*<.001		
8			*d*=.378 *p*=.009		
9	*d*=.548 n. s.	*d*=.456 n. s.	*d*=.942 *p*<.001	*d*=.536 *p*=.003	*d*=.546 *p*=.033
10					
11			*d*=.500 *p*<.001		
12	*d*=.948 *p*=.015				
13	*d*=.602 n. s.				

The average Spectral Centroid was increased after each procedure, but significantly only in FLT (t=3.092, p=.003, d=.433, df=50), VFSRAC (t=3.715, p<.001, d=.628, df=34), and VFW (t=3.033, p=.008, d=.594, df=17).

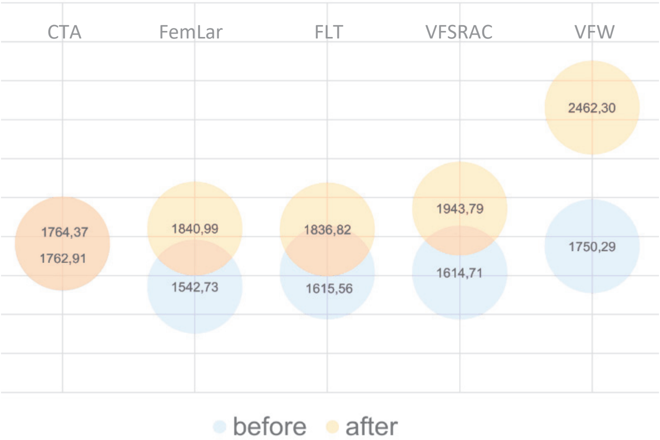

Figure 11: Changes of Spectral Centroid [Hz] before and after surgery.

Formant changes could mainly be observed as results of FLT and VFW. In FLT, the first three formants were on average lowered after the procedure (f_1: t=2.695, p=.01, d=.377; f_2: t=2.013, p=.05, d=.282; f_3: t=3.342, p=.002, d=.468; df=50), in VFW a significant effect could only be observed for the first formant (t=2.521, p=.022, d=.594, df=17).

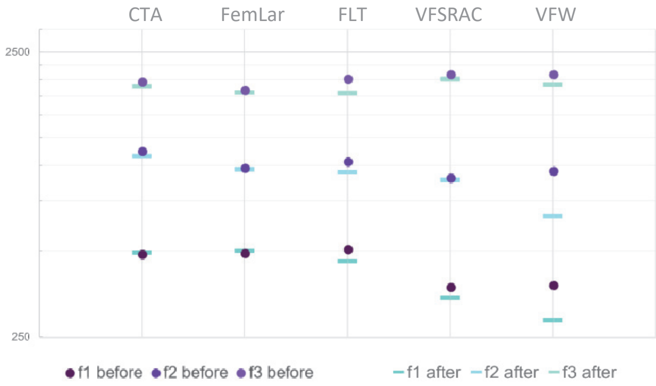

Figure 12: Changes of the first three formants [Hz] before and after surgery.

Discussion

The data indicate that the shortening of the vocal tract by elevation of the larynx (CTA, FLT) predicts less increase of spectral centroid and higher mfcc7. This alteration of the sound source in relation to the vocal tract could explain an envelope shift thus leading to the lowered formants. However, this could not explain the significantly lower formant structures for VFW, where a similar effect could be observed in spite of a greatly increased Spectral Centroid mean. A possible explanation for the altogether too low formants might be a hoarseness due to the healing process in some patients since the recordings were obtained at different time intervals after the surgery. Since the participants were asked to share only the kind of information they were comfortable sharing, not every recording came with information about the speaker's height, age, or point in time after surgery.

The increase of mfcc4 (which in some previous research has been attributed to rough or inharmonic features [10]) could also be related to the natural healing process respectively hoarseness and should therefore not be interpreted too hastily.

The increase of mfcc5 for VFSRAC could be explained by a more stable suspension of the vocal folds since mfcc5 has been attributed to perceived contentment in previous research [10]. The shortening of vocal folds by suturing in VFSRAC as well as VFW may influence mfcc6 while shortening by reducing the size of the larynx as achieved by laryngoplasty procedures (FemLar and FLT) seems to predict changes in mfcc3 and 4.

The reduction of (any) resonating tissue could be responsible for the increase of mfcc9 in all procedures, but this effect was certainly largest in procedures where the most oscillatory tissue (e. g. mucosal or elastic cartilage tissue of the actual vocal folds) is removed in the process of webbing.

Since mfcc2, 4, 5, 7, 9 and 13 have previously been attributed to a gender specific timbre [9], changes in those parameters might seem to be beneficial to the purpose of voice feminization. Since the greatest independent (from height, age etc.) biological difference between voices which are either perceived as male or female is the amount of cartilage and mucosal tissue, mfcc9 seems to be the most important factor to alternate, although the combination of various mfccs might provide the best results. Also, VFW and FemLar provide the biggest effects regarding average pitch as well as spectral centroid, but share only a medium effect increase of mfcc9 regarding timbre. This emphasises previous findings regarding timbre being independent from pitch when it comes to gender (or general voice) recognition [10].

It is to be noted that there were many individual differences among the patients, which could be explained by age, height, native language, stage of healing or amount of speech therapy previously attended. It could, for instance, be the case that elevating of the larynx (CTA and FLT) might not be necessary, especially for patients of average female height, thus cancelling out the effect of unnecessarily shifted formants in some patients. Interestingly, the greatest differences in spectral centroid could be observed in those procedures without elevating methods.

Overall, more data about individual differences in patients is needed to explain variances in the outcomes. The overall surgical results were fascinatingly successful and generated a better understanding on which part of the anatomy influences which timbre feature, since the similarities in the surgical methods were directly reflected by the significant changes in timbre features, especially mfccs.

References

[1] Friedman, M., Lim, J., & Tanyeri, H. (1998). Thyrohyoid Suspension for correction of obstructive sleep apnea. *Operative Techniques in Otolaryngology – Head and Neck Surgeries, 9*(3), 138–141.

[2] Kanagalingam, J., Georgalas, C., Wood, G., Ahluwalia, S., Sandhu, G., & Cheesman, A. (2005). Cricothyroid Approximation and Subluxation in 21 Male-to-Female Transsexuals. *The Laryngoscope*, 115, 611–618.

[3] Kim, H. (2017). A New Conceptual Approach for Voice Feminization: 12 Years of Experience. *The Laryngoscope*, 127, 1102–1108.

[4] Kunachak, S., Prakunhungsit, S., & Sujjalak, K. (2000). Thyroid cartilage and vocal fold reduction: a new phonosurgical method for male-to-female transsexuals. *Ann Otol Rhinol Laryngol*, 109(11), 1082–1086.

[5] Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction from Audio. *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*, 237–244.

[6] Lartillot, O. (2019). Miningsuite: A Comprehensive Matlab Framework for Signal, Audio and Music Analysis, Articulating Audio and Symbolic Approaches. *SMC 2019 Proceedings of the 16th Sound & Music Computing Conference*. 2019, 489-489.

[7] Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1-15.

[8] Lennie, Y., Leareng, K., & Evered, L. (2020). Perioperative considerations for transgender women undergoing routine surgery: a narrative review. *British Journal of Anaesthesia*, 124(6), 702–711.

[9] Reuter, C., Wöllner, C., Czedik-Eysenberg, I., & Hohagen, J. (2016). *Tönend bewegte Normen: Akustische und kinetische Morphs bei Sängerinnen und Sängern*. Poster presented at "Akustik und Musikalische Hörwahrnehmung": 32. Jahrestagung 2016 der Deutschen Gesellschaft für Musikpsychologie (DGM), Wien, Österreich.

[10] Roos, M., Reuter, C. (2016). *Ich höre, wer Du bist! Über die Wahrnehmung geschlechtsspezifischer Persönlichkeitsmerkmale in der menschlichen Stimme*. Poster presented at "Akustik und Musikalische Hörwahrnehmung": 32. Jahrestagung der Deutschen Gesellschaft für Musikpsychologie (DGM), Wien, Österreich.