



Vaasan yliopisto
UNIVERSITY OF VAASA

OSUVA Open
Science

This is a self-archived – parallel published version of this article in the publication archive of the University of Vaasa. It might differ from the original.

Training the Vocal Expression of Emotions in Singing: Effects of Including Acoustic Research-Based Elements in the Regular Singing Training of Acting Students

Author(s): Hakanpää, Tua; Waaramaa, Teija; Laukkanen, Anne-Maria

Title: Training the Vocal Expression of Emotions in Singing: Effects of Including Acoustic Research-Based Elements in the Regular Singing Training of Acting Students

Year: 2021

Version: Accepted manuscript

Copyright ©2021 Elsevier. This manuscript version is made available under the Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International (CC BY–NC–ND 4.0) license, <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Hakanpää, T., Waaramaa, T. & Laukkanen, A-M. (2021). Training the Vocal Expression of Emotions in Singing: Effects of Including Acoustic Research-Based Elements in the Regular Singing Training of Acting Students. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2020.12.032>

Training the vocal expression of emotions in singing: Effects of including acoustic research-based elements in the regular singing training of acting students

Tua Hakanpää, Teija Waaramaa*, Anne-Maria Laukkanen

Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University, Tampere, Finland

* Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University, Tampere, Finland

Communication Sciences, University of Vaasa, Vaasa, Finland

Address for correspondence

Tua Hakanpää, MA

Speech and Voice Research Laboratory

Åkerlundinkatu 5

33014 Tampere University

Tua.Hakanpaa@tuni.fi

+358405681172

'Declarations of interest: none'

Abstract

Objectives: This study examines the effects of including acoustic research-based elements of the vocal expression of emotions in the singing lessons of acting students during a seven-week teaching period. This information may be useful in improving the training of interpretation in singing.

Study design: Experimental comparative study

Methods: Six acting students participated in seven weeks of extra training concerning voice quality in the expression of emotions in singing. Song samples were recorded before and after the training. A control group of six acting students were recorded twice within a seven-week period, during which they participated in ordinary training. All participants sang on the vowel [a:] and on a longer phrase expressing anger, sadness, joy, tenderness, and neutral states. The vowel and phrase samples were evaluated by 34 listeners for the perceived emotion. Additionally, the vowel samples were analyzed for formant frequencies (F1–F4), sound pressure level (SPL), spectral structure (Alpha ratio = SPL 1500–5000 Hz - SPL 50–1500 Hz), harmonic-to-noise ratio (HNR), and perturbation (Jitter, Shimmer).

Results: The number of correctly perceived expressions improved in the test group's vowel samples, while no significant change was observed in the control group. The overall recognition was higher for the phrases than for the vowel samples. Of the acoustic parameters, F1 and SPL significantly differentiated emotions in both groups, and HNR specifically differentiated emotions in the test group. The Alpha ratio was found to statistically significantly differentiate emotion expression after training.

Conclusions: The expression of emotion in the singing voice improved after seven weeks of voice quality training. The F1, SPL, Alpha ratio, and HNR differentiated emotional expression. The variation in acoustic parameters became wider after training. Similar changes were not observed after seven weeks of ordinary voice training.

Key Words: voice quality, perceived emotion, acoustic analyses

1. Introduction

An important value of music lies in its capacity to express emotions. Many of the acoustic attributes that musicians use to express emotion are also important in vocal expression. It has been suggested that musical and vocal expressions share a common expressive code.^{1,2} This seems to offer an advantage for vocalists in the conveyance of emotions through their music. However, there are still many questions concerning emotional expression in the singing voice, especially regarding its training.

Plenty of techniques have been developed to train interpretation and emotional expression: some are based on the reflective system of information processing, consisting of rigorous self-observation (such as the Stanislavski method³ or Psychodrama⁴), and others try to engage the more subconscious associative route (such as TRE® or NLP). These methods and their variations are used in voice studios all over the world. The age-old master-apprentice tradition allows for an experimental approach towards these methods, which leads many teachers to use the “good parts” of different methods and rework them into exercises that suit their individual teaching styles. This is both good and bad: it is good in a sense that voice pedagogy keeps reinventing itself, but bad in the sense that the original idea sometimes gets lost in the metamorphoses and can result in a pseudo-therapeutic experiment laboratory where things can go awry fast.

The singing voice is an instrument where the self and the voice have a complex inter-relation. A singer’s vocal identity is composed of musical and self-identity, but it is also regulated by public and interpersonal transactions that influence perceptions of (or reactions to) the physical instrument and emotional self.⁵ The training of emotional expression in the singing studio may be difficult due to a) a lack of cognitive resources, such as episodic memory (the person hasn’t experienced these emotions), or due to identity-related beliefs about emotions, or b) motivational issues, such as social status and how one wants to present oneself to others or issues concerning the repression of emotions.⁶ This leaves the singing instructor in a difficult position: how should one address the expression of emotion in the singing voice without getting overly involved in the student’s emotional life? Furthermore, there are genre-typical esthetic demands that require genre-specific vocal techniques also when expressing emotions. Music pieces themselves regulate many acoustic variables normally used in the conveyance of emotions, like pitch, tempo, and also loudness to some extent. All of this complicates the emotional expression in singing and naturally also its training.

One of the most influential changes in teaching singing in the 21st century has been the increasing emphasis on voice science. This has led many teachers to re-examine how they teach vocal technique.⁷ Voice science has helped teachers use anatomy, physiology, and the principles of skill acquisition to improve vocal training.⁸ There are also a vast amount of studies concerning the acoustic characteristics used in the expression of emotions in singing and the differences between song genres in this respect.⁹⁻¹⁸ This leads to the question of whether this information can be exploited in training vocal expression in singing.

The field of emotional expression is naturally very complex. In addition to the basic emotions (happiness, sadness, fear, anger, disgust, and surprise) whose expressions have been found to be relatively universal,¹⁹⁻²² there are plenty of more subtle emotions whose expressions are culturally shaped.^{23,24} Emotions also have degrees (strong – weak) and nuance (e.g., cold or hot anger, depressive, submissive sadness vs. grief), etc. One way to simplify the topic for research or practical purposes is to classify emotions according to the activity (arousal) level involved and the

valence (negative – neutral – positive). Negative valence relates to something unpleasant, potentially threatening, while positive valence relates to something that is interpreted as pleasant and is potentially good for survival. According to this kind of classification, joy and anger have a higher activity level than sadness and tenderness, for example, and joy and tenderness represent a positive valence, while anger and sadness carry a negative valence. We concentrate here on the expression of joy, anger, sadness, and tenderness because of their opposite placement on the valence-activation scale and also because they are categories of emotion that are frequently encountered in the song literature. Table 1 summarizes some results from previous studies.

Table 1: Acoustic parameters found to code emotional expressions for joy, tenderness, sadness, and anger in singing. References ^{9,12,15–18} focus on the operatic voice, while references ^{10,11,25} concern the non-classical voice, non-specific voice technique, or both classical and non-classical voices.

	<i>Acoustic characteristic</i>		
<i>Perceptual characteristic</i>	Joy	Tenderness	Sadness
Pitch	Higher F0 floor, mean, and ceiling. ¹⁰ High on F0 variation, low on F0 rise and fall slopes. ¹⁶	Low F0 floor, mean, and ceiling. ¹⁰ Low on F0 variation and F0 rise and fall slopes. ¹⁶	Higher F0 floor, mean, and ceiling. ¹⁰ High on F0 variation, low on F0 rise and fall slopes. ¹⁶
Loudness	High loudness (AE ¹), low loudness (AE) variation, and moderate on loudness (AE) rise and fall slopes. ¹⁶ High equivalent sound level, low Hammarberg index, and low level difference between partials 1 and 2 (H1/H2). ¹⁸ Higher mean sound level and more short-term variability of sound level. ¹² High SPL. ¹¹	Low loudness (AE), high loudness (AE) variation and rise and fall slopes, low dynamics. ¹⁶ Low equivalent sound level, high Hammarberg index, and high level difference between partials 1 and 2 (H1/H2). ¹⁸ Low mean sound level and less short term variability of sound level. ^{12,17} Low SPL. ¹¹ Low vocal energy. ¹⁵	Low loudness (AE), high loudness (AE) variation, and moderate rise and fall slopes. ¹⁶ Low equivalent sound level, high Hammarberg index, and high level difference between partials 1 and 2 (H1/H2). ¹⁸ Low mean sound level and less short-term variability of sound level. ^{12,17} Low SPL. ¹¹ Low level of dynamics. ¹⁵ Low intensity. ⁹
Timbre	Low formant bandwidth, low formant amplitude, high formant frequency, and moderate low-energy frequency variation. ¹⁶ Low proportion energy <.5 kHz, low proportion energy <1 kHz, High spectral flatness, and high spectral centroid. ¹⁸ High Alpha ratio. ^{11,18} Shallow spectral slope ⁹ and narrow bandwidth. ¹¹	High formant bandwidth, moderate formant amplitude, small formant frequency, tendency for high low-frequency energy, and small low-energy frequency variation. ¹⁶ High proportion energy <.5 kHz, high proportion energy <1 kHz, low spectral flatness, low spectral slope, and low spectral centroid. ¹⁸ Low Alpha ratio. ^{11,18} Broad bandwidth. ¹¹	High formant bandwidth, low formant amplitude, small formant frequency, small low-energy frequency variation. ¹⁶ High proportion energy <.5 kHz, high proportion energy <1 kHz, low spectral flatness, low spectral slope, and low spectral centroid. ¹⁸ Low Alpha ratio. ^{11,18} Broad bandwidth. ¹¹
Tempo	Fast on tempo. ^{9,12,15–17,25}	Slowest on tempo. ^{9,12,15–17,25}	Slow on tempo. ^{9,12,15–17,25}
Irregularity of sound	Low on perturbation variation, high on perturbation level. ¹⁶ Less aperiodic fluctuation of F0, more irregular variation of the period amplitude. ¹¹ More jitter, less HNR. ¹⁰	High on perturbation variation, low on perturbation level, and little waveform irregularity. ^{16,17} More aperiodic fluctuation of F0, more irregular variation of the period amplitude. ¹¹ Less jitter, more HNR. ¹⁰	High on perturbation variation, low on perturbation level. ^{16,17} More aperiodic fluctuation of F0, more irregular variation of the period amplitude. ¹¹ More jitter, less HNR. ¹⁰

We know from previous research that the vocal characteristics change when expressing different emotions in the singing voice. Depending on the style of singing, we can deliberately enhance various characteristics to a greater or lesser degree to aid emotion recognition from the singing voice.^{9,10,25,26,11–18} Vocal expressions of high-activity emotions compared to expressions of low-activity emotions are typically characterized by a faster tempo, greater loudness and quick changes in amplitude, a lower level difference between the lowest partials H1 and H2, a flatter spectrum, a large extent of vibrato, local departures from pitch contour at tone onsets, and a higher degree of perturbation and noise (more jitter, lower HNR).^{10,13,14,17,27} Valence is coded in a more subtle way, through low or intermediate parameter values and differences in formant frequencies. According to Scherer et al.,¹⁶ the highest formant frequency mean (i.e., mean of the formant frequencies measured) was found in joy and lowest in tenderness. In turn, Waaramaa et al. found on average somewhat lower formant frequencies to characterize negative valence.²⁸ The samples were monopitched vowels produced by acting students in a speaking or speech-like singing voice. Furthermore, according to Waaramaa et al.,²⁹ samples with a synthetically raised F3 were more often perceived as positive in valence than those with the original F3 lowered or totally removed. These differences may be related to differences between singing and speaking or differences in the

¹ Loudness(ae) is a psychoacoustic measure that is designed to give a value to acoustically estimated loudness. According to Scherer et al. (2017), this correlates better with the vocal affect dimensions than with the raw signal energy.

strength or nuance of emotion expressed. For instance, sadness expressed in a whining voice may be expected to have higher formant frequencies than expressions of depressive sadness.

The supreme parameter used in the coding of emotions is loudness – or its principal acoustic correlate SPL. Many other parameters (fundamental frequency and also spectral slope) accompany variation in SPL, which is regulated by subglottic pressure (P_{sub}) and vocal fold adduction, and also by vocal tract acoustics. Varying vocal fold adduction along the axis from breathy to pressed induces a change from a steep to a gentle spectral slope (flattening of the spectrum). Loose adduction causes a large amplitude difference between the lowest partials (H1–H2, i.e., H1 dominates the spectrum), while the opposite is seen when the adduction is tight (a pressed, strained voice). Likewise, strong adduction results in stronger relative spectral energy above 500 Hz (or 1 kHz). Vocal tract resonances affect SPL. For example, raising F1 closer to F2 (by dropping the lower jaw and/or opening the mouth wider) increases it. ^{30–35}

Formant positioning affects also the timbral qualities of the voice. Raising the formant frequencies (e.g., by smiling or using a more frontal articulation) makes the voice timbre brighter. This increases perceived loudness, as it increases the sound energy in the higher frequency range (between 2 and 4 kHz), where the human ear is more sensitive. Lowering the formant frequencies (e.g., by yawning or protruding the lips or vocalizing with a retracted tongue or small mouth opening) darkens the timbre. Adduction and vocal tract acoustics interact. For example, when the adduction is loose, it lowers the amplitude of the formants and broadens their bandwidths. ^{8,32}

Perturbation (jitter and shimmer) may be introduced by an imbalance between P_{sub} and adduction, which causes irregular vocal fold vibration. The perceptual correlate is a rough, more or less hoarse quality. Turbulence noise may be increased by leaving a gap in the glottis and using a sufficient P_{sub} . The perceived voice contains a hissing component. Both perturbation and turbulence noise may contribute in a decrease in HNR.

This study aims to investigate whether an acoustic research-based parameter modulation technique could be helpful in training the vocal expression of emotions in singing. A 7 x 45 min. training routine was constructed, based on research results on the acoustic characteristics of emotional expressions and their perceptual and physiological correlates. This training was tested on acting students in addition to their ordinary training. The research questions were: 1. Does the specific training improve the recognition of emotions from the singing voice? 2. Do the acoustic differences between emotional expressions increase after the particular training? The hypotheses were: 1. The recognition of emotions will increase in the test group and will not change in the control group. 2. Acoustic differences between emotional expressions will increase in the test group. As markers of the latter, we hypothesize that (a) the number of significantly differentiating parameters will increase and (b) the range of parameters will increase.

2. Methods

2.1. Participants and recording

The participants of this study were six Finnish acting students (three males, three females, mean age 25 years, SD 4 years) with a minimum of two years of singing lessons and on average six years of singing experience (median 2.25 years). The control group consisted of six gender- and age-

matched Finnish acting students also with a minimum of two years of singing lessons. On average, they had two years of singing experience (median 1 year). All test subjects were native speakers of Finnish.

All subjects were instructed to perform an eight-bar musical excerpt composed especially for the test situation expressing the emotions of joy, tenderness, sadness, and anger plus a neutral state. They all sang using the syllables pa[pa:], da[da:], and fa[fa:], which in themselves in Finnish carry no meaning or emotional content as such. The excerpt was composed using the pentatonic scale in order to avoid sounding too major or too minor. The same test was issued before and after the teaching intervention.

Figure 1: The musical excerpt sung.



The modality of the song was f-pentatonic, and the pulse was 115 bpm (beats per minute) for all test subjects and every emotion portrayal. The emotion samples and neutrals were performed in a randomized order and repeated three times. The participants were asked to identify the take they liked the best and that take was selected for further analyses.

Recordings of the test group were made in the well-damped recording studio of Tampere University Speech and Voice Research Laboratory using a Brüel & Kjær Mediator 2238 sound level meter and 4188 microphone. The distance between the microphone and the test subjects' lips was 40 cm. Samples were recorded with an external soundcard (Focusrite Scarlett 2-i-4) and Sound Forge Pro 11.0 digital audio editing software using a 44.1 kHz sampling rate and a 16-Bit amplitude quantization. The sound recordings were calibrated for SPL measurements using a sine wave generator with a known SPL.

All control samples were recorded at recording studio 365 of the University of the Arts, Helsinki, using the same microphone and recording distance as for the test subjects. The RME Babyface Pro external sound card and Cubase 10 digital recording and audio editing software were used (44.1 kHz, 16-Bit). The sound recordings were calibrated for SPL measurements in the same way as for the test group.

In order to make the experiment as lifelike as possible, the subjects used a backing track with a neutral accompaniment that was played to them via a SONY MDR V-700 headset through Zoom H-4 in the test condition and a Sennheiser HD 25-SP II 60 Ohm headset in the control condition.

All samples were saved as .wav files for further analyses with Praat.³⁶

2.2. Listening test

A listening test was conducted in which 246 voice samples were replayed to 32 listeners. The participants completed a multiple-choice questionnaire on which emotion they perceived as being expressed.

The listening task was a web-based test with the randomized [ɑ:] vowel and phrase samples and six control samples (120 + 120 + 6). (The control samples were repeated samples of emotion portrayals selected at random.) The test was accessible through a browser by logging in with one's own password. Participants completed the test using their own equipment. The test was accessible from desktop, laptop, tablet computers and cellphones. The participants were instructed to use headphones to ensure the best possible sound quality. The voice samples were played in a randomized order and it was possible to play the samples as many times as needed. The questionnaire was in Finnish, and the listeners were Finnish speakers. The listening test took approximately 40 minutes to complete.

The number of listeners who completed the test was 32 (27 females, 5 males, no reported hearing defects). The total number of answers in the listening test was 8,160. There were 1632 answers from each emotion category and 1632 answers for the neutral portrayals.

2.3. Voice samples in acoustic analyses

The vowel [ɑ:] was extracted from the last bar in each sample for further analyses. The pitch was f1 (F4, 349.23 Hz) for females, and f (F3, 174.61 Hz) for males. The nominal duration of the extracted vowel (including the preceding consonant) was 3.13 s according to the notation and tempo of the song. A phrase was extracted consisting of two bars from the beginning of the melody. The [ɑ:] vowels and the phrases were extracted from the sung excerpts using Reaper audio editing software. The vowel samples were cut right after the preceding consonant (the Finnish language does not involve aspiration after voiceless plosives). The duration of the sample vowels varied between 1.2 s and 4.04 s depending on how the test subject had interpreted the time value of the notation. The tail end of the vowel was left as the singer interpreted it (nominal note duration 3.13 s), as previous studies have indicated that micromanaging the durations of written notes is one way of expressing emotions in the singing voice.¹³

2.4. Acoustic parameters under investigation

Twelve acoustic parameters were automatically extracted from the voice samples. All analyses were made using Praat software (Version 6.0.19).³⁶ The vowel samples (N = 120) were analyzed for the lowest formant frequencies from F1 to F4. SPL was measured with reference to the calibration signal recorded. The Alpha ratio, which reflects the mean strength of the higher spectrum partials as compared to the lower ones, was measured using the formula $\text{SPL } 1500\text{--}5000 \text{ Hz} - \text{SPL } 50\text{--}1500 \text{ Hz}$.^{37,38} The cut-off frequency was set to 1500 Hz instead of the more traditional 1000 Hz in order to better suit the analysis of the singing voice. The Harmonics-to-Noise Ratio (HNR) was also measured. Two measures of jitter and shimmer were used. For jitter, we measured the relative average perturbation and five-point period perturbation quotient, and for shimmer the three-point and the five-point amplitude perturbation quotient. The relative average perturbation (RAP) is the average absolute difference between an interval [glottal period] and the average of it and its two neighbors, divided by the average time between two consecutive points, and the five-point period perturbation quotient (ppq5) is the average absolute difference between an interval and the average of it and its four closest neighbors, divided by the average time between two consecutive points. The three-point amplitude perturbation quotient (Shimmer apq3) is the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbors, divided by the average amplitude, and the five-point amplitude perturbation quotient (apq5) is the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbors, divided by the average amplitude.

2.5. Statistical analyses

The results of the listening test were coded numerically for statistical analyses. Both intended and perceived emotions were given numbers (1= joy, 2 = tenderness, 3 = neutral, 4 = sadness, 5 = anger). Samples sung within the “test” condition were marked with 1, and those sung in the control group were marked with 2. Furthermore, the before condition was marked as 1 and the after condition as 2. The number of the correct (intended = perceived) answers for emotion are given as percentages and frequencies.

Results of the listening test were analyzed using four different statistical tests.

The first statistical test used was a binomial test (one proportion z-test) to evaluate the probability that the observed percentage of the correctly recognized emotions could have resulted from random guessing. The listening test contained five different emotions, which meant that the expected percentage of correctly recognized emotions in case of random guessing would be 20%. The statistically significant difference between guessing and correct recognition can be shown if the p-value of the test is <0.05 .

The second statistical test was the Unrelated samples t-test, which was used to compare the number of correct answers given for the test group samples and the number of correct answers given for the control group samples. The null hypothesis was that the two populations from which the two samples have been drawn have equal means. Separate t-tests were run for the first recording (before) and the second recording (after). Recognition between the test group samples and control samples was interpreted to differ statistically significantly if the p-value of the test was <0.05 .

The third statistical test used was Pearson’s Chi-squared test of homogeneity to evaluate the probability that two groups of results have the same percentage of correctly recognized emotions. The percentage of correctly recognized emotions is statistically significant in two groups of results if the p-value of the test is <0.05 . Pearson’s Chi-squared test was used to compare the correct recognition within the same population under different conditions, i.e., to determine if there was any difference in the recognition in the test group (and separately for the control group) between the two recordings.

The fourth statistical test used was Cronbach’s alpha to evaluate the internal consistency of listener evaluations. Values >0.7 indicate acceptable internal consistency.

To evaluate whether the parameter values extracted with Praat³⁶ differed across emotions for each parameter, we computed the Friedman test (a non-parametric alternative to the one-way repeated measures ANOVA) with SPSS (v.17; SPSS Inc., Chicago, IL). We ran the Friedman test separately for the test group and control group and the before and after conditions. Bonferroni corrections were used for multiple comparisons.

2.6. Training procedure

2.6.1. General structure

Test subjects participated in a workshop consisting of seven individual lessons, each lasting 45 minutes in duration. The aim was to introduce the basic acoustic characteristics typically observed

in the expressions of four particular emotional states and the perceptual correlates of these characteristics.^{11,39} After introducing the characteristics, we wanted to rehearse the voluntary variation of these voice characteristics so that they resulted in clearly recognizable emotional expression.

At the beginning of the teaching intervention, the rich tradition of emotion expression coaching in actor voice training through mind imagery, self-reflection, and interaction exercises was acknowledged and discussed. This was done because we felt it was important to point out that the method we were using was somewhat rigid and it was not our intention to imply that it would be a method the acting students should use exclusively when expressing emotions with the singing voice. After this, a more mechanical approach was agreed upon for the duration of the workshop. Exercises used for emotion expression included basic tension and release techniques, movement with singing, and work with different breathing patterns and standard drills for varying loudness, articulation, and timbre. Such drills can be found in many books and YouTube tutorials related to the art of singing.⁴⁰⁻⁴³

2.6.2. The parameter modulation technique

The participants were first offered the polar opposite scale of valence and activity of emotions and placed the target emotions there. We then introduced a system of acoustic parameter manipulation for expressing different emotions. As the parameter manipulation as such is quite a mechanical way of expressing emotion, we emphasized at all times that this exercise regime was just a tool for exploring the possibilities of voice quality in emotion expression, not a definitive way of arriving at stellar expression. For the purposes of this study, we asked the students to try out the following voice quality manipulations:

- Anger: loud volume with pressed phonation, very clear articulation and no vibrato.
- Sadness: soft voice with a few volume outbursts, more breathy phonation, unclear articulation, and a lot of vocal perturbation and noise.
- Tenderness: moderate loudness and projection, slightly breathy phonation, but clear articulation, no perturbation.
- Joy: loud and well projecting voice, phonation balance (neither breathy nor pressed), inclusion of vibrato acceptable (See Table 2).

Table 2: Acoustic parameters that were modulated during exercises using either more (+) or less (-) of that parameter.

<i>Emotion expression</i>	Perceived acoustic element			
	Volume - loudness	Phonation/ Sound Balance	Resonance/ Articulation	Perturbation/noise ²
Anger	++	+++	++	+
Sadness	-- (+-)	- (-+)	--	++
Tenderness	-	-	-	-
Joy	+	++	+	--

² noise: increase in jitter/shimmer, decrease in HNR

<i>Measurement</i>	SPL (& Alpha ratio)	Alpha ratio, HNR	F1–F4	Jitter, Shimmer, HNR
Acoustic Correlates				

The starting point of the voice modulations was the participants’ habitual neutral voice. In order to use the parameter modulation technique safely, students should be aware of what their individual optimal (well-balanced, effortless) voice use is like. The extent to which the parameter manipulation can be executed (i.e., how wide deviations from the optimum can be introduced) needs to be scaled individually and also for the esthetics of the singing style in use. In this study, the acting students were singing in non-classical styles, which allowed for the maximum amount of parameter manipulation.

As each exercise needs to be fitted individually to the students’ conceptual understanding of the voice and to their individual way of using it, a specific description of the exercises used cannot be given in the scope of this article. Instead, we will give a general description of how the parameter modulation was taught.

For volume control, we used exercises exploring the loudness range of each individual student from the softest possible to the loudest. We discussed each participant’s habitual loudness use, comfort loudness, air flow regulation, vocal fold adduction, and the influence of the oral cavity and mouth opening on the perceived loudness.

For phonation, we used phonation balance exercises fitted to the individual need of the student (soft attack and general “hypofunction” for the “hyperfunctional” student and *vice versa*). Vocal fold movement between the barely abducted and barely adducted is said to produce a resonant voice,⁴⁴ and the goal of these exercises was to establish this zone for the students so that they can safely depart from it. We also drilled polar opposite exercises ranging from very breathy voice through optimal sound balance to pressed phonation. The goal of these exercises was to clearly demonstrate the perceptual (both acoustic and tactile/sensory) differences between the different modes of phonation.

For resonance and articulation, we used exercises that shape the vocal tract in various ways. The articulatory exercises addressed different possibilities of the physiological positioning of the tongue, jaw, velum, and lips. The tongue position has a role in shaping the vowels and the general sound quality. The advancement of the root of the tongue results in the “fronting” of the sound, while retracting it makes the sound darker.^{32,45} We used exercises for moving the tongue forward and backward (genioglossus), flattening the tongue (hyoglossus, chondroglossus), pulling the tongue back and up while depressing the soft palate (palatoglossus) and working with the intrinsic muscles of the tongue. The point of these exercises was to acknowledge the major role that the tongue has in shaping the oral cavity and the resulting sound. The students were encouraged to look for sounds that would (in their opinion) fit the acoustic descriptions given to the target emotions.

Exercises of the jaw movement were presented on an open-closed continuum. The students were given different exercises addressing the relaxed, open and closed jaw positions and they were instructed to experiment with different jaw openings as well as the fixed jaw position (with a bite block). The aim of these exercises was to demonstrate the full range of jaw movement as well as the possibility to hold the jaw in place and still sound intelligible. Again, the students were encouraged to explore and pick different sounds to be used in emotion expression.

The velum exercises we used were either velum up (levator veli palatini, musculus uvulae) or velum down (palatoglossus, palatopharyngeus), and the way we approached them was through mind imagery instructions such as “smelling the flower” or “like you are just about to cry.”

The lips have a role in lengthening and shortening the vocal tract, which effects the formant frequencies and makes the voice color sound darker or shriller.^{32,46} This effect can be achieved using exercises extending the lips outwards (pouting) and retracting them sideways as in a smile. The lips (orbicularis oris) are continuous with the buccinator muscle, and ultimately with the superior pharyngeal constrictor,⁴⁷ and in this way any movement of the lips will have an effect on the shape of the oral cavity and its possibility of reinforcing formants. For this study, we used phonation with protruded and retracted lips, as well as with different lip openings and with restricted lip movement.

Another method we used was facial expressions. Facial expressions have been found to alter the voice quality, and finding a preferred sound through a facial expression is a known technique in singing instruction that lends itself readily to acoustic emotion expression.⁴⁸ By simply asking the student to make a sad, happy, angry, or tender face while phonating, we were able to generate different voice qualities through muscle movement.

Jitter and shimmer measures indirectly assess laryngeal function by quantifying acoustic correlates of irregular vocal fold vibration. Jitter measures f_0 perturbation and shimmer measures SPL perturbation, caused by vibratory variations from one vocal fold cycle to the next. Jitter is affected mainly because of the lack of control of vocal fold vibration and shimmer because of the reduction in SPL related tension in the vocal folds.⁴⁹ Although sound perturbations refer to abnormal stability of the period length, amplitude and waveform of the vocal fold cycle, the sound is somewhat tolerant of the wave's asymmetry, and a little bit of perturbation occurs in all natural sounds. Small irregularities in the acoustic wave are considered as normal variation associated with physiological body function and voice production.^{50–52} The occurrence of jitter or shimmer in voiced sound can be perceptually described as a hoarse, husky, or rough voice. As it is not exactly a desirable effect in a professional voice, we used extreme vibratos, both frequency (undulating between several semitones) and amplitude (crying-like volume changes) modulation, and even breaks in the voice to simulate perturbation in a voice-friendly way. For this exercise set, we did not use distorted sounds such as growls or screams. The extreme vibratos and tremolos were chosen for this exercise regime as they offer a perceptually concrete and singer-friendly way of practicing the unwanted undulation of sound. Adding a vocal tract-induced noise component to a steady vocal fold vibration cycle – such as we do in dist-sounds⁵³ – often takes a considerable amount of practice to be done safely, and as we had time restrictions, we felt that this was the better option.

3. Results

3.1. Recognition of emotions

The emotion appraisals (answers) given in the listening test indicate that it is possible to recognize emotion from the singing voice. The overall recognition was 47.7% (x2 z-value 62.57, $p = 0.000$). The recognition of phrases at 52.7% (x2 z-value 52.28, $p = 0.000$) was slightly better than the recognition of short vowel samples at 42.6% (x2 z-value 36.24, $p = 0.000$).

3.1.1 Recognition from vowel samples

For the purposes of this study, we are mainly interested in the short vowel samples, as they are seen as a carrier of information about the voice quality and as such are reflective of the usefulness of the practice regime used in the teaching intervention. Our hypothesis was that by teaching specific voice use (different voice qualities,) we could improve the recognition of emotion from the singing voice.

We first ran unrelated t-tests to determine if the recognition differed between the test group samples and the control samples. We established that there were no outliers in the data, as assessed by an inspection of the boxplot. The distribution of correct answers given in the listening test was normally distributed in all other conditions except for the test group before condition, as assessed by the Shapiro-Wilk test ($p > .05$). In the test group before condition, the data distribution was not normally distributed (Statistic 0.927, df 30, $p = 0.041$). There was homogeneity of variance for the correct answers in the listening test for the test group and the control group, as assessed by Levene's test for the equality of variances ($p = 0.166$ in the before condition; $p = 0.201$ in the after condition). There were 30 voice samples evaluated by 34 listeners in the test group and 30 voice samples evaluated by 34 listeners in the control group. There were more correct recognitions of intended emotions in the test group in both the before ($M = 15$, $SD = 10$) and the after ($M = 17$, $SD = 10$) conditions. For the control group, the correct recognition of intended emotions was $M = 14$ ($SD = 8$) in the before condition and $M = 11$ ($SD = 8$) in the after condition (Table 3). The mean difference in the correct answers given in response to the heard samples was 1.50 (95% CI, -3.28 to 6.28) higher in the test group in the before condition in comparison to the control group and 5.97 (95% CI, 1.31 to 10.62) higher in the test group in the after condition in comparison to the control group. In the before condition, there was no statistically significant difference in the mean recognition of emotion between the test group and control group ($t(58) = 0.628$, $p = 0.532$). There was a statistically significant difference in mean correct recognition of emotion between the test group samples and the control group samples after the teaching intervention ($t(58) = 2.565$, $p = 0.013$).

The results indicate that for the test group samples, the recognition of emotion increased in all emotion portrayals in the after condition. The recognition of neutral samples decreased in the after condition. For the control group samples, the recognition of emotion decreased for the after condition in all other emotion portrayals except joy. The recognition of neutral also increased (Table 3).

Table 3: Number of correctly recognized vowel samples.

Expressed feeling	Correctly recognized samples BEFORE	Correctly recognized samples AFTER	Change in recognition
<i>Test Group</i>			
Joy	33	46	13
Tenderness	63	77	14
Neutral	191	122	-69
Sadness	112	123	11
Anger	86	141	55
Correctly recognized samples all together	485	509	

Control Group

Joy	32	38	6
Tenderness	93	43	-50
Neutral	78	89	11
Sadness	99	89	-10
Anger	110	78	-32
Correctly recognized samples all together	412	337	

The internal consistency of the answers was tested with Cronbach’s alpha, and the results showed a mean consistency of 0.93. Anger yielded the most consistent answers, while neutral yielded the least consistent answers (Table 4).

We ran Pearson’s Chi-squared test to see if there was a statistically significant difference between the answers given for samples recorded before and after the seven-week training period. For the sake of comparison, we also ran the test for the control group samples recorded at the seven-week interval. Pearson’s Chi-squared test showed a significant difference in the answers given for the neutral and anger portrayals in the test group and tenderness and anger portrayals in the control group. The recognition of anger increased by 28.4% in the test group from before to after training. The recognition of neutral decreased by 20.6%. In the control group, the situation was reversed: the recognition of emotion decreased from before to after in tenderness (24.4%) and anger (15.7%) (Table 4).

Table 4: Correctly recognized short vowel samples in the listening test, the internal consistency of answers, and the statistical significance of the percentual difference in answers given from the samples recorded before and after the exercise regime.

		Test group			Control group		
		<i>before</i>	<i>after</i>	<i>Pearson's Chi-squared</i>	<i>before</i>	<i>after</i>	<i>Pearson's Chi-squared</i>
Joy	% of recognition	16.2%	22.5%	H0:%1 = %2	15.7%	18.6%	H0 = %1 = %2
	z-value	1.37	0.91	2.7	-1.54	-0.49	0.6
	p-value	0.17	0.363	0.103	0.123	0.624	0.431
	Cronbach's alpha	0.76	0.86		0.86	0.87	
Tenderness	% of recognition	30.9%	38.2%	H0:%1 = %2	45.6%	21.2%	H1 = %1<>%2
	z-value	3.89	6.51	2.4	9.14	0.42	27.2
	p-value	0.00	0	0.118	0	0.0674	0

	Cronbach's alpha	0.87	0.92		0.57	0.54	
Neutral	% of recognition	81.4%	60.8%	H1:%1<>%2	38.2%	43.6%	H0 = %1 = %2
	z-value	23.64	14.56	23.1	6.51	8.44	1.2
	p-value	12.46	0	0	0.000	0	0.268
	Cronbach's alpha	0.10	0.9		0.52	0.75	
Sadness	% of recognition	54.9%	60.8%	H0:%1 = %2	48.5%	43.6%	H0 = %1 = %2
	z-value	12.46	14.56	1.4	10.19	8.44	1
	p-value	0.00	0	0.229	0	0	0.321
	Cronbach's alpha	0.74	0.88		0.93	0.96	
Anger	% of recognition	42.2%	70.6%	H1:%1<>%2	53.9%	38.2%	H1 = %1<>%2
	z-value	7.91	18.06	33.5	12.11	6.51	10.1
	p-value	0.00	0	0	0	0	0.001
	Cronbach's alpha	0.97	0.96		0.97	0.96	

3.1.2. Recognition from phrases

Recognition from phrases seemed to be easier than recognition from the vowel samples in this study.

There were 30 voice samples evaluated by 34 listeners in the test group and 30 voice samples evaluated by 34 listeners in the control group. There were no outliers in the data, as assessed by an inspection of the boxplot. The distribution of correct answers given in the listening test was normally distributed, as assessed by the Shapiro-Wilk test ($p > .05$), except in the test group before condition, where the data was not normally distributed (Statistic 0.917, $df = 30$, $p = 0.023$). The homogeneity of variances was observed, as assessed by Levene's test for equality of variances ($p = .689$ in the before condition and $p = .218$ in the after condition). In the before condition, the phrases were better recognized from the control group samples ($M = 18$, $SD = 10$) than from the test group samples ($M = 17$, $SD = 9$). In the after condition, the situation was reversed. Phrases were better recognized from the test group samples ($M = 20$, $SD = 8$) than from the control group samples ($M = 17$, $SD = 9$) (Table 5). There were no statistically significant differences in recognition. The mean difference in the correct answers given in response to the heard samples was -0.67 (95% CI, -5.52 to 4.18) in the test group in the before condition in comparison to the control group, and 3.23 (95% CI, -1.25 to 7.71) in the test group in the after condition in comparison to the control group. In the before condition, there was no statistically significant difference in the mean recognition of emotion between the test group and control group ($t(58) = -0.275$, $p = 0.784$). There was no statistically significant difference in the mean correct recognition of emotion between the test group samples and the control group samples after the teaching intervention ($t(58) = 1.445$, $p = 0.154$).

The results indicate that for the test group samples, the recognition of emotion increased in all emotion portrayals in the after condition. The recognition of neutral samples decreased in the after condition. For the control group samples, the recognition of emotion increased for the after

condition in tenderness and neutral and decreased in joy and sadness. The recognition of anger was similar in both conditions (Table 5).

Table 5: Correctly recognized phrases in the listening test.

Expressed feeling	Correctly recognized samples BEFORE	Correctly recognized samples AFTER	Change in recognition
<i>Test Group</i>			
Joy	71	92	21
Tenderness	114	115	1
Neutral	100	97	-3
Sadness	112	157	45
Anger	112	130	18
Correctly recognized samples all together	509	591	
<i>Control Group</i>			
Joy	87	80	-7
Tenderness	121	125	4
Neutral	73	90	17
Sadness	142	101	-41
Anger	109	109	0
Correctly recognized samples all together	532	505	

The internal consistency of the answers was tested with Cronbach's alpha, and it showed a mean consistency of 0.87. Anger yielded the most consistent answers, while neutral yielded the least consistent answers (Table 6).

We ran Pearson's Chi-squared test to see if there was a statistically significant difference between the answers given for samples recorded before and after the seven-week training period. For the sake of comparison, we also ran the test for the control group samples recorded at the seven-week interval. Pearson's Chi-squared test showed a significant difference for answers given for the neutral and anger portrayals in the test group and tenderness and anger portrayals in the control group. The recognition of anger increased by 10.1% from the test group samples from before to after training. The recognition of joy increased by 11.8% and the recognition of sadness increased 22.5%. In the control group, the recognition of tenderness increased by 2% but decreased in joy by 3.4% and in sadness by 20.1%, while the recognition of anger remained the same (Table 6).

Table 6: Correctly recognized phrase samples in the listening test, the internal consistency of answers, and the statistical significance of the percentual difference in answers given from the samples recorded before and after the exercise regime.

Recognition from phrase		Test group		Pearson's Chi-squared	Control group		Pearson's Chi-squared
		before	after		before	after	
Joy	% of recognition	34.8%	46.6%	H1:%1<>%2	42.6%	39.2%	H0:%1 = %2
	z-value	5.29	9.49	5.9	8.09	6.86	0.5
	p-value	0.00	0	0.016	0	0	0.481
	Cronbach's alpha	0.83	0.91		0.91	0.87	
Tenderness	% of recognition	56.4%	57.1%	H0:%1 = %2	59.3%	61.3%	H0:%1 = %2
	z-value	12.99	13.23	0	14.04	14.74	0.2
	p-value	0.00	0	0.875	0	0	0.686
	Cronbach's alpha	0.91	0.67		0.24	0.94	
Neutral	% of recognition	49.5%	49%	H0:%1 = %2	35.8%	44.1%	H0:%1 = %2
	z-value	10.54	10.36	0	5.64	8.61	3
	p-value	0.00	0	0.921	0	0	0.086
	Cronbach's alpha	0.93	0.85		0.9	0.81	
Sadness	% of recognition	55.4%	77.9%	H1:%1<>%2	69.6%	49.5%	H1:%1<>%2
	z-value	12.64	20.69	23.3	17.71	10.54	17.1
	p-value	0.00	0	0	0	0	0
	Cronbach's alpha	0.96	0.93		0.98	0.95	
Anger	% of recognition	54.9%	65 %	H1:%1<>%2	53.4%	53.4%	H0:%1 = %2
	z-value	12.46	16.04	4.3	11.94	11.94	0
	p-value	0.00	0	0.037	0	0	1
	Cronbach's alpha	0.94	0.92		0.98	0.98	

3.2. Acoustic results

A Friedman test was run to determine if there were differences in the usage of different sound parameters in different emotion expressions. Pairwise comparisons were performed (SPSS, 2019) with a Bonferroni correction for multiple comparisons. The acoustic parameters F1 and SPL differed significantly between the expressions of emotions in both the before and after conditions

(Table 7, Table 8). In addition, there was a statistically significant difference of the HNR between emotions in the test group samples. In the samples recorded before the teaching intervention, F3, jitter, and shimmer distinguished emotions in the test group and F4 in the control group, but the effect was not repeated in the samples recorded after the teaching intervention/waiting period. Instead, the Alpha ratio was found to differ statistically significantly between the emotions in the after condition for the test group.

Post hoc analysis of the before condition revealed statistically significant differences in F1 between sadness (Mdn = 620 Hz) and joy (Mdn = 801 Hz) ($p = .010$) and between sadness and anger (Mdn = 807 Hz) ($p = .001$) in the test group. In the control group, differences were found between sadness (Mdn = 684 Hz) and joy (Mdn = 811 Hz) ($p = .019$) and between sadness and anger (Mdn = 887 Hz) ($p = .010$). In the after condition, post hoc analysis showed statistically significant differences between tenderness (Mdn = 658 Hz) and anger (Mdn = 852 Hz) ($p = .019$), between sadness (Mdn = 658 Hz) and anger ($p = .019$), and between neutral (Mdn = 658 Hz) and anger ($p = .019$) in the test group samples.

For SPL, post hoc analyses for the before condition revealed statistically significant differences between tenderness (Mdn = 76 dB) and anger (Mdn = 86 dB) ($p = .035$), between tenderness and joy (Mdn = 86 dB) ($p = .019$), and between sadness (Mdn = 75 dB) and joy ($p = .035$) in the test group samples. In the control group samples, differences were found between tenderness (Mdn = 71 dB) and anger (91 dB) ($p = .019$), between tenderness and joy (Mdn = 84 dB) ($p = .000$), and between sadness (Mdn = 73 dB) and anger ($p = .035$).

In the after condition, statistically significant differences in SPL were found between sadness (Mdn = 73 dB) and joy (Mdn = 82 dB) ($p = .019$) and between sadness and anger (Mdn = 87 dB) ($p = .001$) in the test group. Statistically significant differences for the control group were found between tenderness (Mdn = 67 dB) and anger (Mdn = 80 dB) ($p = .019$), between sadness (Mdn = 66 dB) and anger ($p = .035$), and between neutral (Mdn = 67 dB) and anger ($p = .019$).

In addition to these, in the test group after condition, post hoc tests revealed statistically significant differences in the Alpha ratio between neutral (Mdn = -27 dB) and joy (Mdn = -20 dB) ($p = .035$) and between tenderness (Mdn = -27 dB) and joy ($p = .035$), and in HNR between sadness (Mdn = 17 dB) and joy (Mdn = 21 dB) ($p = .019$).

Table 7: The Friedman test for correlated samples recorded before the teaching intervention.

BEFORE

The Friedman test for correlated samples

	Test group			Control group		
	df	x2	sig.	df	x2	sig.
F1	4	21.067	.000	4	13.733	.008
F2	4	8.400	.078	4		8 .092
F3	4	9.600	.048	4	7.600	.
F4	4	7.600	.107	4	11.467	.022

Alpha ratio	4	6.267	.108	4	6.800	.147
SPL	4	17.467	.002	4	21.733	.000
HNR	4	10.400	.034	4	7.092	.131
Jitter rap	4	10.475	.033	4	3.322	.505
Jitter ppq	4	11.898	.018	4	1.544	.819
Shimmer apq3	4	11.067	.026	4	3.467	.483
Shimmer apq5	4	11.200	.024	4	5.517	.238

Table 8: The Friedman test for correlated samples recorded after the teaching intervention.

AFTER

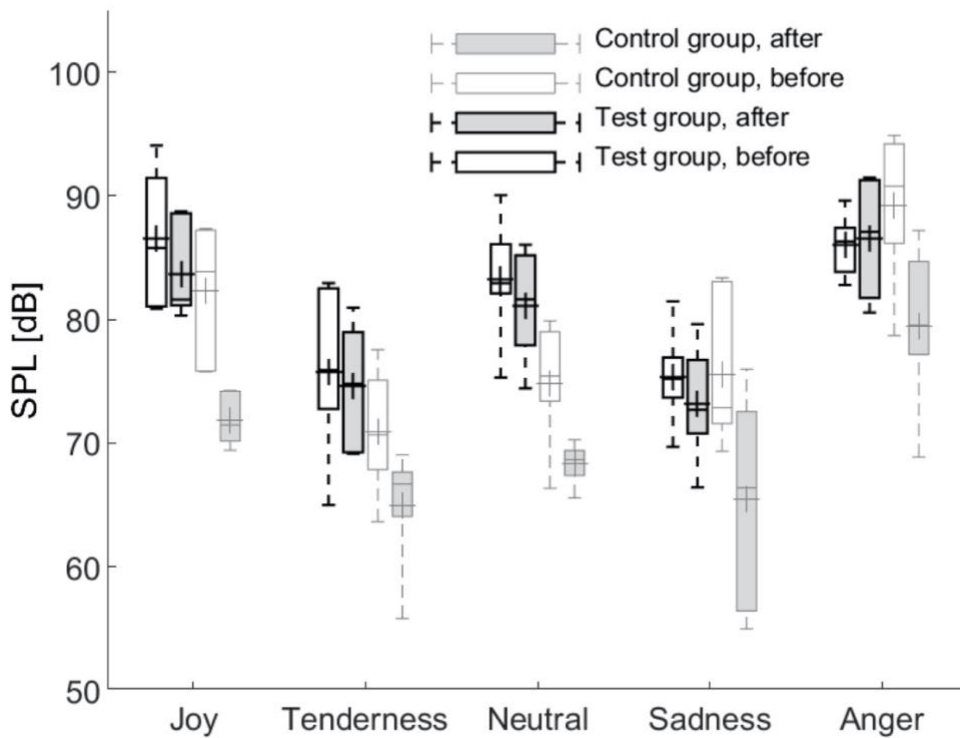
The Friedman test for correlated samples

	Test group			Control group		
	df	x2	sig.	df	x2	sig.
F1	4		.003	4	15.018	.005
F2	4	3.930	.416	4	5.754	.218
F3	4	1.600	.809	4	3.733	.443
F4	4	4.772	.312	4	1.263	.868
Alpha ratio	4	17.263	.002	4	4.772	.312
SPL	4	21.193	.000	4	17.825	.001
HNR	4	10.386	.034	4	.421	.981
Jitter rap	4	5.895	.207	4	.587	.964
Jitter ppq	4	7.604	.107	4	2.400	.663
Shimmer apq3	4	3.228	.520	4	2.847	.584
Shimmer apq5	4	8.140	.087	4	3.856	.426

3.2.1 SPL:

For the test group, SPL increased on the continuum of tenderness → sadness → neutral → joy → anger. Low intensity emotions (sadness, tenderness) were sung with a lower volume than the emotions with higher activity. The effect of the exercise routine can be seen in the wider variety in SPL control in the after condition as opposed to the before condition in the test group. In comparison to the control group, the test group showed a more consistent SPL between the recordings. In the control group, singers sang slightly louder the first time and softer the last time (Figure 1). SPL differed significantly between emotional expressions for both groups.

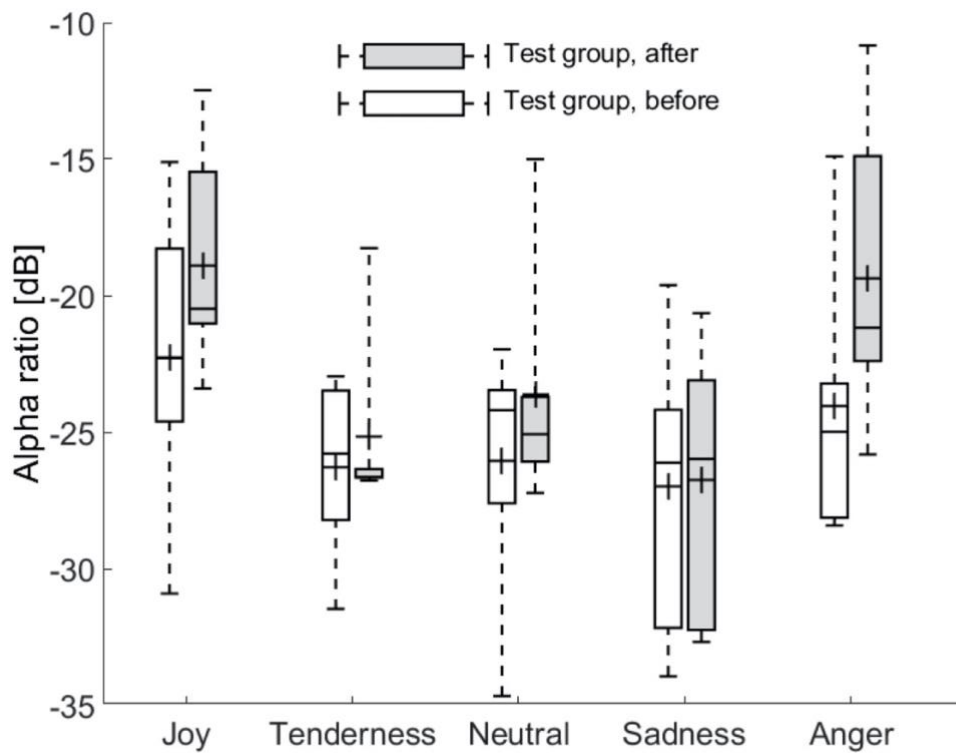
Figure 2: Test group SPL control before and after the training intervention and control group volume control in two separate recordings.



3.2.2. Alpha ratio:

All of the samples for high-activity emotions (joy and anger) were characterized by a larger Alpha ratio than for the low-activity emotions (sadness and tenderness). The effects of the training routine can be seen again in the wider variety of Alpha ratio usage in the test group from before to after samples. The difference in the test group after samples in comparison to the before samples indicates that the test group started to vary their sound balance more after the teaching intervention (Figure 2).

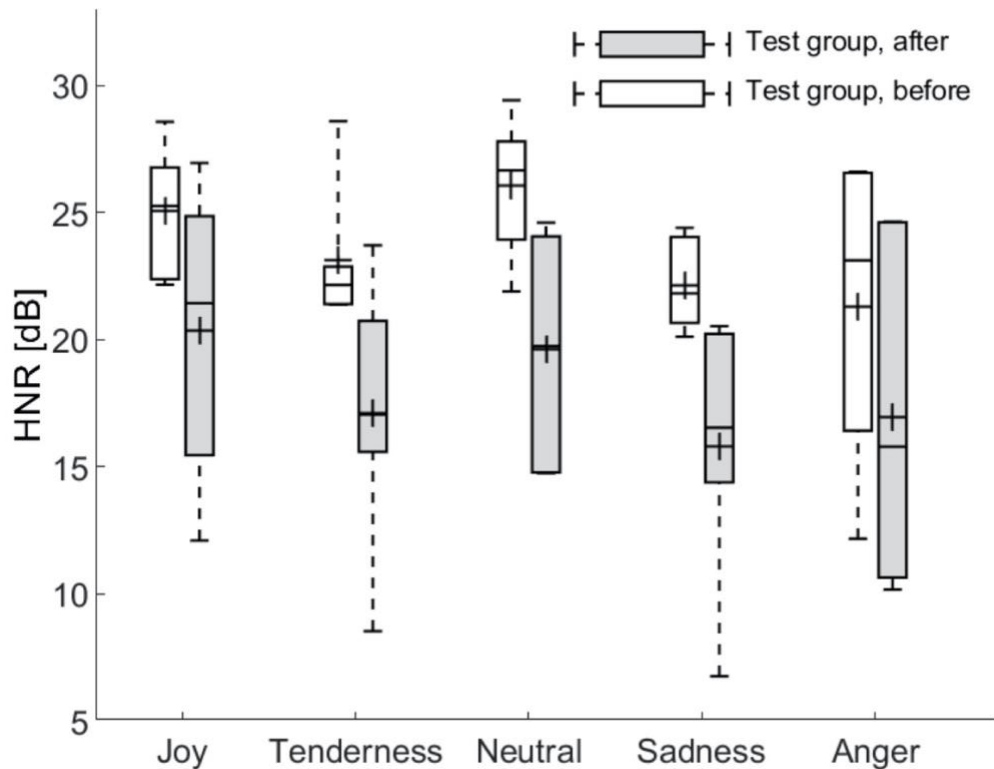
Figure 3: Measured Alpha ratio in the test group samples before and after the training intervention.



3.2.3. HNR:

In this study, HNR did show a statistically significant difference between the emotions in the test group in both the before and after conditions. The HNR decreased in the second recording for the test group, suggesting the use of more noise components in the signal. The most harmonic content was found in joy and the least in sadness. For the control group, HNR increased for the second recording, suggesting a more sonorous sound. The effects of the training routine were again observed in the increased variety of the use of the HNR component in the test group (Figure 3).

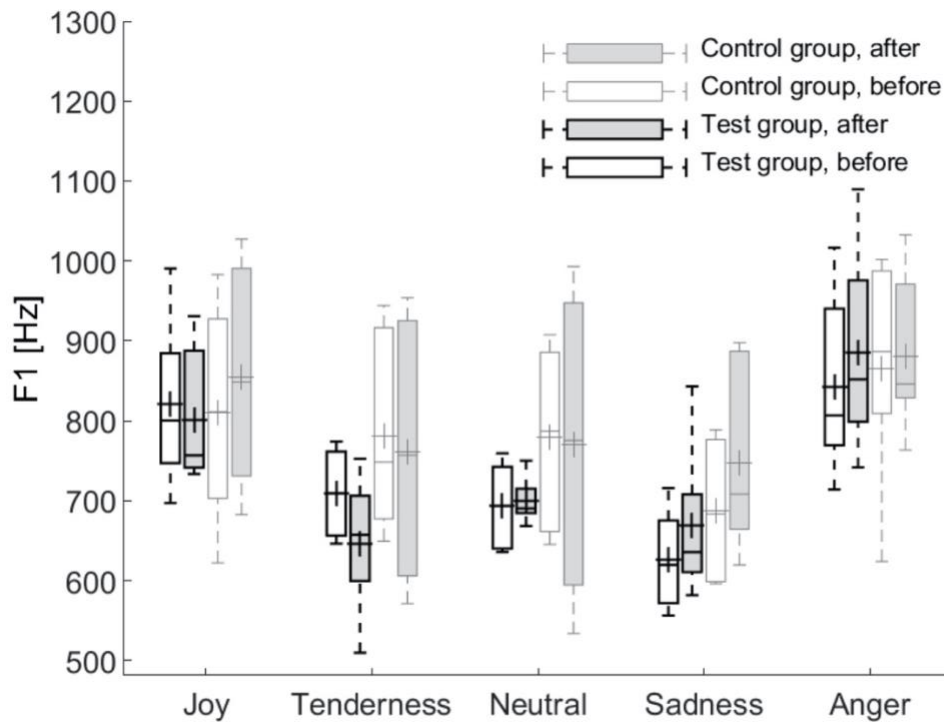
Figure 4: Measured HNR in the test group samples before and after training.



3.2.4. Formants:

The overall formant structure of the samples revealed a few distinctive patterns in regard to emotion expression. In sadness, F1 was lower, and F2, F3, and F4 were higher in comparison to other emotions and neutral, suggesting a more diffuse formant pattern. In anger, the opposite was true: F1 was higher and F2, F3, and F4 were lower in both groups. A similar but less pronounced formant pattern could be found in joy. In tenderness, the first formant was positioned slightly higher than in sadness, but it was still relatively low; other formants were positioned fairly high. In neutral expression, the first formant was neither high nor low, while the second and the third formants were in a relatively low position. The formant structure was most compact in anger and then it scattered in a sequence of anger → joy → neutral → tenderness → sadness. F1 was found to be statistically significant in differentiating the emotional expressions in both groups (Figure 4).

Figure 5: Range of F1 positioning in recorded samples.



4. Discussion

This study investigated the effects of a specific training strategy, the “parameter manipulation technique,” for emotional expression in the singing voice. The technique is based on accumulated research findings on the acoustic characteristics of vocal emotion expression. The aim was to see whether the specific training improves the recognition of emotions from the singing voice and whether the acoustic differences between emotional expressions increase after the particular training. We hypothesized that the recognition of emotions would increase in the test group and not change in the control group, and that the number of significantly differentiating parameters and the range of the parameters would increase after training. Whilst one needs to take caution in extrapolating data that have been produced by a small number of people repeating specific deliberate tasks while expressing requested emotions, the results seem to support the hypotheses at least partially. Our results suggest that training with the parameter modulation technique increased the correct recognition of emotional expression from the short vowel and phrase samples.

The number of significantly distinguishing parameters did not increase in our investigation, but the range of how various parameters were used became broader.

For the vowels, we found a statistically significant difference in mean correct recognition between the test group samples and the control group samples after the teaching intervention ($p = 0.013$). The recognition was 4.5% units better from the test group samples after the intervention. Our results show that for the test group samples, recognition of emotion increased in all emotion portrayals in the after condition. The recognition of neutral samples decreased in the after condition. It is fairly common to get a lot of “neutral” answers with the type of forced choice questionnaire that we used

for the listening test. If the listeners are not sure what they heard, they are more likely to select “neutral.”⁵⁴ The decrease in neutral answers in the test group after condition samples can be therefore interpreted as an increase in the expressivity of the singing voice. For the control group, vowel sample recognition of emotion decreased for the after condition in all other emotion portrayals except joy. The recognition of neutral on the other hand increased. This can be interpreted as a difficulty in arriving at a specific emotion appraisal on the listener side. This lends support to our hypothesis that the systemic parameter manipulation of the singing voice can help in building expressivity in contrast to a situation where the students receive just regular voice tuition without rehearsing the use of voice quality in emotional expression.

For phrases, there was no statistically significant difference found in the recognition of emotions between the test and control groups (even though for the test group, the recognition was significantly better in the after condition as compared to the before condition). Emotions were already recognized better from phrases than vowels at the beginning. This can be interpreted as resulting from the fact that phrases contain more acoustic variables and thus more redundancy, which makes it easier to convey emotions through them. Furthermore, the training technique used targeted skills to vary loudness and voice quality, not various other aspects relevant at the level of the phrase. As such, the result that the recognition of emotions improved for vowels and not for phrases supports the suggestion that the result is due to the training intervention and not other intervening factors (like various elements of the ordinary actors’ training obtained during the seven-week period).

For the test group vowel samples in the before condition, the Friedman test found statistical significance in differentiating emotion expression for F1, SPL, HNR, F3, jitter(s), and shimmer(s). In the samples recorded after the teaching intervention only, F1, SPL, HNR, and Alpha ratio were found to be statistically significant in distinguishing the emotion expressions. For the control group, F1, SPL, and F4 showed statistical significance in differentiating the emotions in the samples collected in the former recording, but only F1 and SPL remained statistically significant in the samples collected in the latter recording. This means that the results did not give support to our hypothesis that the number of differentiating acoustic parameters would increase. The result could be interpreted to show that rehearsing stereotypic use of voice parameters narrows down the choice of parameters. On the other hand, it may reflect a more focused and conscious use of certain key parameters and the reduction of parameters that may not be related to expression as such or that may not have any relevance to emotion recognition.

The exercise routine we used to train the parameters of emotional expression in the sound signal included exercises for volume control, phonation balance, articulation, and extreme vibrato, but excluded resonant voice training and distortions of sound signal. Resonant voice with a singer’s ring at around 3000 Hz has been identified as a typical marker of a quality voice that can carry over a large orchestra.^{32,55–57} The resonance features can be explored in a context of emotional expression by altering the carrying power of the voice from a very resonant, sonorous voice to a non-resonant, muffled sound. Articulation and vocal tract settings affect the resonance frequencies of the vocal tract.¹⁵ Clear articulation, which can be viewed as the voluntary maximization of the intelligibility of one’s singing for the benefit of the listener, is another mark of a good singer. There are, however, several different configurations of the vocal tract that can lead to perceptually similar sounding vowels and consonants, depending on the pitch used and individual singing style.⁵⁸ As the resonance and articulation have plenty of overlapping anatomical and physiological elements and as time was scarce with our seven-week training regime, we felt that it was not efficient in a teaching situation to try to compartmentalize these two functions in the context of emotional expression through sound qualities. This is especially the case since jaw and lip movements have been widely

studied and linked to both an increase or decrease in clarity of articulation and the resonance of sound, which in turn have been linked to the perception of loudness.^{44,59-68} The five bodily systems needed for singing are the nervous, breathing, phonation, resonance, and articulatory systems.⁴⁷ Already for the perceptually seemingly easy task of volume control, the students need to control these systems simultaneously. It was our pedagogical challenge to adjust the exercises of the “parameter modulation” model to correspond with the students’ own way of using and coordinating these systems. Our solution to this problem was to use the most simple instruction of moving clearly definable anatomical organs on the vertical, horizontal, and anterior/posterior axis, and, when this was not possible, to use vocal sounds that could be easily segregated, demonstrated, and mimicked using the auditory feedback loop.^{47,69} In this way, we felt that the possibilities of misunderstandings due to a lack of conceptual knowledge or terminology could be avoided and the idea of the “parameter modulation” technique could be conveyed impartially to all participants.

There is so much individual variation in singing voices and the way that we use them that there will never be a ready-made mold that will fit all singers. However, we can say that based on previous research and the results of this investigation, SPL/loudness plays a large role in emotion expression, phonation balance is one of the defining factors, and formant frequencies is another.¹⁵⁻¹⁸ When we move on to longer units of sound, such as phrases, verses, or songs, there will be many more factors that will contribute to the expressivity of the singing voice.

This study is limited by the small number of participants, and therefore the potential effect of individual factors cannot be excluded. On the other hand, the recognition rate was very similar for both the test and the control group samples before training, which indicates that the groups expressed emotions in a similarly recognizable way. The fact that the (singing) participants were acting students also may affect the results, since they can be expected to have a better baseline ability to convey emotion through their voices. However, the acting students were chosen as subjects as this kind of training fits their curriculum, and they are supposed to be more motivated and also more capable of exploring their vocal resources. The samples were recorded in two different locations, but with similar equipment and similar (studio) surroundings. The listener-participants used their own devices in the listening test, which creates possible fluctuations in audio quality from one participant to another. However, the flexible use of listening equipment allowed us to have significantly more listener-participants than the use of fixed listening equipment.

5. Conclusions

The parameter modulation exercise routine helped in expressing emotions vocally, as indicated by the increase in the correct recognition of emotions in the listening tests.

The seven-week training program helped the acting students to broaden their voice use and tactically implement the usage of F1 positioning, SPL, HNR, and Alpha ratio in the expression of emotion in comparison to the control group, who did not receive instructions in parameter modulation.

Training different voice qualities may help in expressing emotions in the singing voice.

6. References:

1. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull.* 2003;129(5):770-814. doi:10.1037/0033-2909.129.5.770
2. Scherer KR. Expression of emotion in voice and music. *J Voice.* 1995;9(3):235-248. doi:10.1016/S0892-1997(05)80231-0
3. Stanislavski K, Männistö M. *Näyttelijän Työ.*; 2015.
4. Karp M, Holmes P, Tavon KB, Moreno ZT, Sprague K. *The Handbook of Psychodrama.*; 2005. doi:10.4324/9780203977767
5. Sweet B, Parker EC. Female Vocal Identity Development: A Phenomenology. *J Res Music Educ.* 2019;67(1):62-82.
6. Quirin M, Kazén M, Kuhl J. When Nonsense Sounds Happy or Helpless: The Implicit Positive and Negative Affect Test (IPANAT). *J Pers Soc Psychol.* 2009. doi:10.1037/a0016063
7. Mesiä S. *Developing Expertise of Popular Music and Jazz Vocal Pedagogy Through Professional Conversations: A Collaborative Project among Teachers in Higher Music Education in the Nordic Countries.* The Sibelius Academy of the University of the Arts Helsinki *Studia Musica* 77; 2019.
8. Titze IR, Verdolini Abbot K. *Vocology the Science and Practice of Voice Habilitation.* NCVS; 2012.
9. Jansens S, Bloothoof G, de Krom G. Perception And Acoustics Of Emotions In Singing. *Proc Fifth Eur Conf Speech Commun Technol.* 1997;0:0-3. <http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=9747D0A838F2790BD0161DCF94739C2E?doi=10.1.1.56.8871>.
10. Livingstone SR, Choi DH, Russo FA. The influence of vocal training and acting experience on measures of voice quality and emotional genuineness. *Front Psychol.* 2014;5(MAR). doi:10.3389/fpsyg.2014.00156
11. Hakanpää T, Waaramaa T, Laukkanen A-M. Comparing Contemporary Commercial and Classical styles - Emotion expression in singing. *J Voice.* 2019. doi:https://doi.org/10.1016/j.jvoice.2019.10.002

12. Sundberg J, Iwarsson J, Hagegård H. A Singer's Expression of Emotions in Sung Performance. In: Fujimura O, Hirano M, eds. *Vocal Fold Physiology: Voice Quality Control*. Singular Pub Group; 1995:357.
13. Sundberg J. Emotive Transforms: acoustic patterning of speech Its Linguistic and Physiological Bases. *Phonetica*. 2000;(57):95-112.
14. Eyben F, Salomão GL, Sundberg J, Scherer KR, Schuller BW. Emotion in the singing voice—a deeperlook at acoustic features in the light ofautomatic classification. *EURASIP J Audio, Speech, Music Process*. 2015;2015(1):19. doi:10.1186/s13636-015-0057-6
15. Scherer KR, Trznadel S, Fantini B, Sundberg J. Recognizing emotions in the singing voice. *Psychomusicology Music Mind, Brain*. 2017. doi:10.1037/pmu0000193
16. Scherer KR, Sundberg J, Fantini B, Trznadel S, Eyben F. The expression of emotion in the singing voice: Acoustic patterns in vocal performance. *J Acoust Soc Am*. 2017;142(4):1805-1815. doi:10.1121/1.5002886
17. Scherer KR, Sundberg J, Tamarit L, Salom??o GL. Comparing the acoustic expression of emotion in the speaking and the singing voice. *Comput Speech Lang*. 2015;29(1):218-235. doi:10.1016/j.csl.2013.10.002
18. Sundberg J, Salomão GL, Scherer KR. Analyzing Emotion Expression in Singing via Flow Glottograms, Long-Term-Average Spectra, and Expert Listener Evaluation. *J Voice*. 2019. doi:10.1016/j.jvoice.2019.08.007
19. Darwin C, Ekman P. The expression of the emotions in man and animals (3rd ed.). *Expr Emot man Anim 3rd ed*. 1872. doi:10.1037/10001-000
20. Izard CE. Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. *Perspect Psychol Sci*. 2007;2(3):260-280. doi:10.1111/j.1745-6916.2007.00044.x
21. Izard CE. Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol Rev*. 1992;99(3):561-565. doi:10.1037//0033-295X.99.3.561
22. Purves, D., Cabeza, R., Huettel, S., LaBar, K., Platt, M. & Woldorff M 2. *Principles of Cognitive Neuroscience, Second Edition.*; 2013.
23. Bericat E. The sociology of emotions: Four decades of progress. *Curr Sociol*. 2016. doi:10.1177/0011392115588355

24. Scherer KR. The dynamic architecture of emotion: Evidence for the component process model. *Cogn Emot*. 2009. doi:10.1080/02699930902928969
25. G.M. K, V.P. M. Acoustical correlates of emotional content on vocalized speech. *Sov Phys Acoust*. 1976;22:208-2011.
26. Siegwart H, Scherer KR. Acoustic concomitants of emotional expression in operatic singing: The case of lucia in *Ardi gli incensi*. *J Voice*. 1995;9(3):249-260. doi:10.1016/S0892-1997(05)80232-2
27. Sundberg J. Expressivity in singing. A review of some recent investigations. *Logop Phoniatr Vocology*. 1998;23(3):121-127. doi:10.1080/140154398434130
28. Waaramaa T, Laukkanen AM, Alku P, Väyrynen E. Monopitched expression of emotions in different vowels. *Folia Phoniatr Logop*. 2008;60(5):249-255. doi:10.1159/000151762
29. Waaramaa T, Alku P, Laukkanen A-M. The role of F3 in the vocal expression of emotions. *Logop Phoniatr Vocology*. 2006;31(4):153-156. doi:10.1080/14015430500456739
30. Fant G. *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations*. 2nd ed. Berlin/Boston: De Gruyter, Inc.; 1971.
31. Pulkki V, Karjalainen M. *Communication Acoustics an Introduction to Speech, Audio and Psychoacoustics*. John Wiley & Sons, Ltd; 2015.
32. Sundberg J. *The science of the singing voice*. Dekalb North IllUniversity Press. 1987.
33. Titze IR. *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice-Hall; 1994.
34. Titze IR, Baken RJ, Bozeman KW, et al. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *J Acoust Soc Am*. 2015. doi:10.1121/1.4919349
35. Welch G, Thurman L, Theimer A, Grefsheim E, Feit P. how your vocal tract contributes to basic voice qualities. In: *Bodymind and Voice*. ; 2000:449-469.
36. Boersma P, Weenink D. Praat. 2014.
37. Frøkjær-Jensen B, Prytz S. Registration of voice quality. *Bruel-Kjaer Technol Rev*. 1976;3:3-17.
38. Lã FMB, Sundberg J. Pregnancy and the singing voice: Reports from a case study. *J Voice*. 2012. doi:10.1016/j.jvoice.2010.10.010

39. Hakanpää T, Waaramaa T, Laukkanen A-M. Emotion Recognition From Singing Voices Using Contemporary Commercial Music and Classical Styles. *J Voice*. 2018. doi:10.1016/j.jvoice.2018.01.012
40. Behrman A, Haskell J. *Exercises for Voice Therapy*. Second Edi. Plural publishing; 2013.
41. Linklater K. *Freeing the Natural Voice*. Drama Publishers; 2006.
42. Brown OL. *Discover Your Voice*. New York: Delmar cenage learning; 2007.
43. Miller Richard. *The Structure of Singing System and Art in Vocal Technique.*; 1986.
44. Myers BR, Finnegan EM. The effects of articulation on the perceived loudness of the projected voice. *J Voice*. 2015. doi:10.1016/j.jvoice.2014.07.022
45. Edmondson JA, Esling JH. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*. 2006. doi:10.1017/S095267570600087X
46. Tartter VC. Happy talk: Perceptual and acoustic effects of smiling on speech. *Percept Psychophys*. 1980;27(1):24-27. doi:10.3758/BF03199901
47. Seikel J, King D, Drumright D. *Anatomy and Physiology for Speech, Language, and Hearing.*; 2009. <http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract%5Cnhttp://books.google.com/books?hl=en&lr=&id=LFBOhaD1JHwC&oi=fnd&pg=PR7&dq=Anatomy+%26+Physiology+for+Speech,+Language,+and+Hearing&ots=9Gv5WCUMRI&sig=gPuraE2s72r2cJoHaeObmWkLGAs>.
48. Aura M, Geneid A, Bjørkøy K, Rantanen M, Laukkanen AM. The Nasal Musculature as a Control Panel for Singing—Why Classical Singers Use a Special Facial Expression? *J Voice*. 2019. doi:10.1016/j.jvoice.2017.12.016
49. Farrús M, Hernando J. Using Jitter and Shimmer in speaker verification. *IET Signal Process*. 2009. doi:10.1049/iet-spr.2008.0147
50. Brockmann M, Drinnan MJ, Storck C, Carding PN. Reliable jitter and shimmer measurements in voice clinics: The relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *J Voice*. 2011. doi:10.1016/j.jvoice.2009.07.002
51. Orlikoff RF, Baken RJ. The effect of the heartbeat on vocal fundamental frequency

- perturbation. *J Speech Hear Res.* 1989. doi:10.1044/jshr.3203.576
52. Titze IR. A model for neurologic sources of aperiodicity in vocal fold vibration. *J Speech Hear Res.* 1991. doi:10.1044/jshr.3403.460
 53. Borch DZ, Sundberg J, Lindestad P a, Thalén M. Vocal fold vibration and voice source aperiodicity in “dist” tones: a study of a timbral ornament in rock singing. *Logoped Phoniatr Vocol.* 2004;29(6):147-153. doi:10.1080/14015430410016073
 54. Marcel Z, Eerola T. Self-Report Measures and Models. In: Juslin PN, Sloboda JA, eds. *Handbook of Music and Emotion Theory, Research, Applications.* Oxford university press; 2012:187-221.
 55. Sundberg J. Level and center frequency of the singer’s formant. *J Voice.* 2001. doi:10.1016/S0892-1997(01)00019-4
 56. Lee S-H, Kwon H-J, Choi H-J, Lee N-H, Lee S-J, Jin S-M. The Singer’s Formant and Speaker’s Ring Resonance: A Long-Term Average Spectrum Analysis. *Clin Exp Otorhinolaryngol.* 2008. doi:10.3342/ceo.2008.1.2.92
 57. Barrichelo VMO, Heuer RJ, Dean CM, Sataloff RT. Comparison of singer’s formant, speaker’s ring, and LTA spectrum among classical singers and untrained normal speakers. *J Voice.* 2001. doi:10.1016/S0892-1997(01)00036-4
 58. Rua Ventura SM, Freitas DRS, Ramos IMAP, Tavares JMRS. Morphologic differences in the vocal tract resonance cavities of voice professionals: An MRI-based study. *J Voice.* 2013. doi:10.1016/j.jvoice.2012.11.010
 59. Sundberg J. Articulatory interpretation of the “singing formant.” *J Acoust Soc Am.* 1974. doi:10.1121/1.1914609
 60. Vurma A. Amplitude Effects of Vocal Tract Resonance Adjustments When Singing Louder. *J Voice.* 2020. doi:10.1016/j.jvoice.2020.05.020
 61. Titze IR, Story BH. Acoustic interactions of the voice source with the lower vocal tract. *J Acoust Soc Am.* 1997. doi:10.1121/1.418246
 62. Acker BF. Vocal tract adjustments for the projected voice. *J Voice.* 1987. doi:10.1016/S0892-1997(87)80028-0
 63. Sundberg J. What’s so special about singers? *J Voice.* 1990. doi:10.1016/S0892-

1997(05)80135-3

64. Garnier M, Henrich N, Smith J, Wolfe J. Vocal tract adjustments in the high soprano range. *J Acoust Soc Am*. 2010. doi:10.1121/1.3419907
65. Fant G. *Acoustic Theory of Speech Production*. The Hague: Mouton; 1960.
66. Schulman R. Articulatory dynamics of loud and normal speech. *J Acoust Soc Am*. 1989. doi:10.1121/1.397737
67. Dromey C, Ramig LO. Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. *J Speech, Lang Hear Res*. 1998. doi:10.1044/jslhr.4105.1003
68. Wohlert AB, Hammen VL. Lip Muscle Activity Related to Speech Rate and Loudness. *J Speech, Lang Hear Res*. 2000. doi:10.1044/jslhr.4305.1229
69. Tourville JA, Guenther FH. The DIVA model: A neural theory of speech acquisition and production. *Lang Cogn Process*. 2011. doi:10.1080/01690960903498424