Influence on Spectral Energy Distribution of Emotional Expression

*Marco Guzman, †Soledad Correa, ‡Daniel Muñoz, and §Ross Mayerhoff, *‡*Santiago,* †*Valparaiso, Chile and* §*Detroit, Michigan*

Summary: Purpose. The aim of this study was to determine the influence of emotional expression in spectral energy distribution in professional theater actors.

Study Design. The study design is a quasi-experimental study.

Method. Thirty-seven actors, native Spanish speakers, were included. All subjects had at least 3 years of professional experience as a theater actor and no history of vocal pathology for the last 5 years. Participants were recorded during a read-aloud task of a 230-word passage, expressing six different emotions (happiness, sadness, fear, anger, tenderness, and eroticism) and without emotion (neutral state). Acoustical analysis with long-term average spectrum included three variables: the energy level difference between the F_1 and fundamental frequency (F_0) regions, ratio between 1–5 kHz and 5–8 kHz, and alpha ratio.

Results. All the different emotions differ significantly from the neutral state for alpha ratio and 1–5/5–8 kHz ratio. Only significant differences between "joy," "anger," and "eroticism" were found for L1–L0 ratio. Statistically significant differences between genders for the three acoustical variables were also found.

Conclusion. The expression of emotion impacts the spectral energy distribution. On the one hand emotional states characterized by a breathy voice quality such as tenderness, sadness, and eroticism present a low harmonic energy above 1 kHz, high glottal noise energy, and more energy on F_0 than overtones. On the other hand, emotional states such as joy, anger, and fear are characterized by high harmonic energy greater than 1 kHz (less steep spectral slope declination), low glottal noise energy, and more energy on the F_1 than F_0 region.

Key Words: Emotions–Actor–Spectral energy–LTAS–Voice quality–Timbre.

INTRODUCTION

Human speech transmits multiple layers of information.[1] In addition to the linguistic messages, the speech acoustic signal also carries information about the identity, age, geographic origin, attitude, and emotional state of the speaker. The present study is focused on how emotions are encoded in the speech acoustic signal.

Component process theory[2,3] conceptualizes emotion as an episode of temporary synchronization of all major subsystems of organismic functioning represented by five components (cognition, physiological regulation, motivation, motor expression, and monitoring-feeling) in response to the evaluation or appraisal of an external or internal stimulus event as relevant to the central concerns of the organism.

The important role of vocal cues in the expression of emotion, both felt and feigned, and the powerful effects of vocal affect expression on interpersonal interaction and social influence have been recognized ever since antiquity.[4] Charles Darwin, in his pioneering monograph on the expression of emotion in animals and humans, underlined the primary significance of the voice as a carrier of affective signals.[4]

Emotions have been described in a three-dimensional space where arousal (activation), valence (pleasure), and control (power) represent each dimension.[5] Commonly analyzed acoustic parameters for such a description of emotion in speech have been fundamental frequency (F_0) (level, range, and contour),[6–9] duration at phoneme or syllable level,[6,9–11] interword silence duration,[9,11] voiced/unvoiced duration ratio in utterance level,[7–10] energy related to the waveform envelope (or amplitude, perceived as intensity of the voice),[7,11] location of the first three formant frequencies (related to the perception of articulation),[9] and the distribution of the energy in the frequency spectrum (particularly the relative energy in the high- vs the low-frequency region, affecting the perception of voice quality).[7,9]

Most studies related to emotions in speech have focused on the role of F_0,[12,13] sound pressure level (SPL), speech rate, segment duration, and overall prosody.[14] The role of voice quality (or timbre) in conveying emotions has been studied to a lesser extent. Voice quality or timbre can be defined as a combination of voice source characteristics (an airflow pulsation resulting from vocal fold vibration) and vocal tract (formant frequencies). Related to this definition, Laukkanen et al,[15] in a study performed with acted emotions, suggested that voice source characteristics (relative open time of the glottis and speed quotient) seemed to communicate the psychophysiological activity level related to an emotional state, whereas formant frequencies seemed to be used to code valence of the emotions, for example, whether the emotion is positive or negative. The higher formants, F_3 and F_4, seemed to have greater values in positive emotions and lower in negative emotions. In a more recent study where the vowel (a:) was extracted from simulated emotions and inverse filter was applied, it was reported that the role of F_3 alone is not crucial in determining the perceived valence of emotion. Results reflected difficulties in the

Accepted for publication August 20, 2012.

From the *School of Communication Sciences, University of Chile, Santiago, Chile; †School of Communication Sciences, University of Valparaiso, Valparaiso, Chile; ‡Faculty of Medicine, University of Chile, Santiago, Chile; and the §Department of Otolaryngology, Wayne State University, Detroit, Michigan.

Address correspondence and reprint requests to Marco Guzman, School of Communication Sciences, University of Chile, Avenida Independencia 1029, Santiago, Chile. E-mail: guzmanvoz@gmail.com

Journal of Voice, Vol. 27, No. 1, pp. 129.e1-129.e10
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2012.08.008

perception of short synthesized samples.[16] This study was also carried out with actors; hence, the acted emotions were assessed.

Scherer[3] presented a model for future research on vocal affect expression. In his model, he hypothesized that individual emotions would differ from each other on a number of different acoustic parameters. Since the study by Scherer, several researchers have continued to suggest that the acoustic properties of speech rate, voice intensity, voice timbre, and F_0 are among the most powerful cues in terms of their effects on listeners' ratings of emotional expressions.[17–21]

The fact that listener-judges are able to reliably recognize different emotions on the basis of vocal cues alone implies that the vocal expression of emotions is differentially patterned. There is considerable evidence that emotion produces changes in respiration,[22–25] phonation, and articulation, which in turn partly determine the parameters of the acoustic signal.[9] Furthermore, much evidence points to the existence of phylogenetic continuity in the acoustic patterns of vocal affect expression.[26]

It has been shown that listeners have a greater-than-chance ability to label emotion only listening to the audio sample, which shows that there are clearly audio properties that are linked to specific emotions, which humans can detect consciously or unconsciously.[27] Several studies have shown that anger and happiness/joy are generally acoustically characterized by high mean F_0, wider pitch range, high speech rate, increases in high-frequency energy, and usually increases in the rate of articulation.[3,6,28] Sadness is characterized by decrease in mean F_0, slightly narrow pitch range, and slower speaking rate.[6] Kienast and Sendlmeier,[29] in a study in which utterances were produced by male and female German actors enacting different emotional states, analyzed spectral and segmental changes caused by the emotions in speech. Their study showed that anger has the highest accuracy of articulation compared with other emotions that they analyzed (happiness, fear, boredom, and sadness). They also analyzed the spectral balance of fricative sounds. Their analysis revealed that two different groups can be observed, one containing fear, anger, and happiness (increased spectral balance compare with that in neutral state) and the other containing boredom and sadness (decreased spectral balance compare with that in neutral state).

Juslin and Laukka[30] conducted a meta-analysis of 104 vocal emotion studies. Among the more notable acoustic findings in this analysis was that the emotions generally referred to as positive, such as happiness and tenderness, tend to show more regularity in F_0, rate, and intensity than do negative emotions, such as anger or sadness.

Moreover, studies that have performed emotion classification experiments with the aim to automatically detect emotions in speech have shown that some emotions are often confused with each other. For example, acoustic classifiers often confuse sadness with boredom or neutrality and happiness is often confused with anger. In contrast, sadness is almost never confused with anger.[7,31,32]

Considering the data collection process, there are two approaches that have been used to study acoustic and perceptual differences in emotions. Some research have been performed with natural emotional expression; however, most of the researchers have used actors to elicit a specific response. According to Banse and Scherer,[7] for ethical and practical reasons, it is not feasible to conduct a study with naive participants by using experimental induction of "real" emotions (in such a way that vocal expression will result). The authors also state that even if one succeeded by clever experimentation to induce a few affective states, it is most likely that their intensities would be rather low and unlikely to yield representative vocal expression. Based on these reasons, in our study, participants were professional actors who acted different emotional states.

The aim of this study was to determine the influence of emotional expression in spectral energy distribution in professional theater actors.

METHOD

Participants

Thirty-seven professional theater actors, native Spanish speakers (16 women and 21 men), were included in this study. None of the subjects had any known pathology of the larynx. The average age of the subjects was 36 years, with a range of 25–43 years. All subjects had at least 3 years of professional experience as a theater actor and no history of vocal pathology for the last 5 years. None of them reported a history of voice therapy before conducting this study. All participants had a similar performance educational background, normal or corrected-to-normal vision, and no reported hearing impairment.

For the purposes of this study, it was felt that the most appropriate approach was to use actors to produce utterances for analysis, even though the emotions themselves remain artificial. The policy of using recordings of actors for emotional voice analysis has been found to be representative of real emotions by a number of scientists.[3,7,14,33] The reasons by Banse[7] for using acted emotions were detailed in the Introduction. Furthermore, the eventual loss of realism in the emotional expression is largely offset by the benefits of both being able to script the dialog and closely control the conditions of the recording process.

Recordings

Participants were recorded during a read-aloud task of a 230-word passage, expressing six different emotions. The duration of each recording was approximately 90 seconds. A Presonus Pre-amplifier plus an analog/digital converter, model Bluetube DP, and a Rode condenser microphone, model NT2A (Steinberg Media Technologies GmbH, Hamburg, Germany), were used to capture the voice samples. This microphone was selected on the basis that the manufacturer's specifications include a flat frequency response from 20 to 20 000 Hz. The microphone was positioned 10 cm with an angle of 45° from the mouth of the participants who remained standing. The recording took place in an acoustically treated room, and samples were recorded digitally at a sampling rate of 44 kHz and 16 bit. The voice signals were captured and recorded using the software *Wavelab*, Version 4.0c (PreSonus Audio Electronics, Baton Rouge, LA) installed on a Dell laptop Inspiron 1420. The audio signal was calibrated using a 220 Hz tone at 80 dB produced with a sound generator for further sound level measurements. The SPL of this reference sound was measured with the sound level meter (American Recorder Technologies, Simi Valley, CA, model SPL-8810), also positioned at a distance of 10 cm from the generator.

To avoid the effects of differences in phonemic structure on the acoustic variables, standardized language material was used. Each subject was asked to read and interpret the text "Monologue of Amadeo"[34] expressing six different basic emotions: happiness, sadness, fear, anger, tenderness, and eroticism. Samples without emotion (neutral state) were also recorded. Although tenderness and eroticism have not been widely assessed in previous studies, they were included in this study because of their common use in acting practice and real life. The recorded utterances were chosen to be semantically empty phrases, which can be equally valid when spoken in any of the emotions to be analyzed. They are also referred to as "emotionless" or "emotionally empty" phrases because they can legitimately take on different emotions depending on the context. Participants were instructed to put themselves into the respective emotional state with the help of self-induction techniques, and they were not given any detailed instructions concerning the emotional expressions. The sequence of interpretation of each emotion was the same for all actors. This resulted in a total number of 259 samples (37 subjects × 7 emotional states). Additionally, 10% of samples were randomly repeated in this sequence to determine whether judges were consistent in their perceptions (intrarater reliability analysis). Samples were edited with the software *GoldWave*, Version V5.57 (Goldwave, St. John's, NL, Canada), and acoustical analysis with long-term average spectrum (LTAS) was performed. No tests were available to assess the possible carryover effect of emotional expression throughout the sequence, that is, the difference between the early states versus late states. Participants were not asked to control the vocal intensity because it could interfere with the expression of emotion during interpretation. Nevertheless, sound level was measured as mentioned above for further sound level analysis.

Listening test

To determine how well the data represent each emotional state, we conducted human evaluation tests with five native Spanish speakers (four men and one woman; mean age of 45.5 years with a range of 39–47 years). This group of blinded judges consisted of professional theater actors with more than 10 years of experience teaching theater. The order of recordings was randomized. Samples from each emotion category were played to the listeners, and they were asked to rate the emotional expression in utterances using a 10-point scale (1, very poor and 10, very good), that is, judges were given the actual emotion before rating and then they rated the perceived quality of the sample. Decoders could replay each emotion as many times as they wanted before making their determination and moving on to the next recording. The evaluation was performed in a well-dampened studio using a laptop computer (Dell Inspiron 1470) and a high-quality loudspeaker (Audioengine 2, Sao Paulo, Brazil). The listeners were located at approximately 2 m from the loudspeaker. The samples were played at a normal conversational loudness throughout the test.

Because investigator partiality may add errors of its own during the process of perceptual assessment of voice quality, a listening test to assess the degree of breathiness during expression of emotions was carried out. In this test, five blinded judges (two men and three women; mean age of 39.2 years with a range of 37–43 years) were asked to rate the degree of breathiness in utterances using a 10-point scale (1, very pressed voice and 10, very breathy voice). The five judges were speech-language pathologists with at least 11 years of experience in the assessment and rehabilitation of voice disorders. This listening test was performed using the same conditions and methods as the listening test used to evaluate the emotional state previously described.

Acoustical analysis

Acoustical analysis with LTAS was performed. In the LTAS window, the acoustical variables in this study were the (1) energy level difference between the F_1 and F_0 regions (L1–L0), that is, the energy level difference between 300–800 Hz and 50–300 Hz, which provides information on the mode of phonation; (2) energy level difference between 1–5 kHz and 5–8 kHz, which may provide information about noise in the glottal source (breathy voice quality); and (3) alpha ratio, that is, the energy level difference between 50 and 1000 Hz and 1–5 kHz, which provides information on the spectral slope declination. For all the acoustic variables, the energy of each spectral segment was calculated automatically by averaging intensity.

The LTAS spectra for each subject were obtained automatically by *Praat*, Version 5.2 developed by Paul Boersma and David Weenink[35] from the Institute of Phonetic Sciences of the University of Amsterdam, The Netherlands. For each sample, Hanning window and a bandwidth of 100 Hz were used. To perform the LTAS, unvoiced sounds and pauses were automatically eliminated from the samples by *Praat* software using the pitch corrected version with standard settings. The advantage of eliminating the voiceless sounds and pauses is that they can affect the average of voiced segments and mask the information from the voice source, especially in the band between 5 and 8 kHz.[36,37] The amplitude values of the spectral peaks were normalized to control for loudness variations between subjects. This process was accomplished automatically by assigning the intensity of the strongest partial, a value of 0, and each subsequent partial, a proportional value, compared with this peak intensity. For the purposes of analyzing the data, equivalent sound level (Leq) was also measured for every emotional state in each recorded samples. Leq was used as the measure of vocal loudness because it gives an average over a long time window, whereas SPL is computed over a short-time window.

Descriptive statistics were calculated for the variables, including mean and standard deviation. The analysis was performed using *Stata 12* (StataCorp. 2011; StataCorp LP, College Station, TX).

Reliability analysis and subjects selection

Friedman nonparametric two-way analysis of variance and Kendall coefficient of concordance were used. Kendall coefficient of concordance ranges from 0 to 1, with 0 meaning no agreement across raters (inter-judges agreement). The null hypothesis is that there is no agreement between judges. In addition, Friedman nonparametric two-way analysis of variance nested in each subject was used for test agreement across raters by subject. Friedman value indicates intrarater reliability. Then, if judges agree (statistically significant interrater agreement), subjects were sorted according to their average score obtained from all judges for all emotions. Those subjects whose score is below the 25th percentile and whose analysis of variance is not statistically significant will be removed from the analysis. Last, an additional reliability analysis across actors' emotions by the degree of breathiness using the same statistical models for obtaining interrater agreement was performed.

Acoustic parameters analysis

Three multiple linear regression models were conducted for each acoustic parameter, considering emotions, gender, and its interaction, explicative for the models. With this, we evaluated the influence of these explicative variables in acoustic parameters and its statistically significant differences. In addition, Pearson linear correlation coefficient (r) for evaluating the correlation between the acoustic parameters was calculated. Last, Kruskal-Wallis nonparametric analysis of variance for loudness level evaluation comparing emotions with the neutral state was performed. An alpha of 0.05 was used for the statistical procedures.

The experiments were conducted with the understanding and the written consent of each participant. This study was approved by the research ethics committee of the School of Communication Disorders of the University of Valparaiso, Chile.

RESULTS

Reliability analysis and subjects selection

Interrater and intrarater reliability in understanding actors' emotional expression is shown in Table 1. All judges understand joy as joy, anger as anger, and so forth. Friedman values indicate that there was also intrarater reliability. We removed subjects (actors) who did not reach statistical significance in the Friedman model and whose score was below the 25th percentile (Table 2). According to this, six actors were eliminated before the final statistical analysis. This indicates that the eliminated actors had poor expression of emotions and that there was agreement among the raters indicating this.

Results by emotion

Figure 1 and Table 3 show the differences by emotion. All the different emotions differ significantly from the neutral state ($P < 0.001$) for alpha ratio and the difference between 1–5 kHz and 5–8 kHz. Only significant differences between "joy," "anger," and "eroticism" were found for L1–L0, when compared with the neutral state ($P < 0.001$). For the other emotions, this difference did not reach statistical significance ($P = 0.349$).

Result by gender

Differences by gender are shown in Figure 2. Statistically significant differences between men and women ($P < 0.001$) according to the linear regression model for alpha ratio were found (male -18.44 ± 3.25 and female -17.46 ± 3.29). There are significant differences between men and women ($P < 0.001$) for L1–L0 (male: -4.21 ± 7.07 and female: 0.96 ± 5.64). There are also significant differences between men and women ($P < 0.001$) for 1–5/5–8 kHz (male: -8.47 ± 4.10 and female: -9.18 ± 4.46). Therefore, the data presented indicate that there are statistically significant differences for all LTAS parameters analyzed by gender.

Interaction between the two variables (emotion and gender)

Figure 3 shows the interaction between the two variables. There are statistically significant differences ($P < 0.001$) in the interaction between emotion and gender, that is, the expression of certain emotion by either a man or woman is different, for alpha ratio and 1–5/5–8 kHz ratio. On the other hand, there are no significant differences for L1–L0 depending on gender, that is, the expression of a certain emotion by either a man or woman was acoustically equal ($P = 0.312$).

Correlation analysis

The correlation between acoustic parameters was as follows: between alpha ratio and L1–L0, $r = 0.49$ ($P < 0.0001$); alpha and 1–5/5–8 kHz, $r = -0.32$ ($P < 0.0001$); and L1–L0 and

TABLE 1.
Interrater and Intrarater Reliability Analysis Across Emotions

	Judges' Scores (Mean ± SD)					Interrater	
Emotions	1	2	3	4	5	Kendall	*P* Value
Joy	6.94 ± 2.22	7.05 ± 2.33	6.18 ± 2.18	7.35 ± 2.38	6.72 ± 1.96	0.80	<0.0001
Anger	7.72 ± 1.96	7.81 ± 1.98	6.89 ± 2.22	7.86 ± 2.07	7.37 ± 2.13	0.75	<0.0001
Eroticism	6.91 ± 1.96	6.91 ± 2.01	6.13 ± 2.21	7.18 ± 1.98	6.83 ± 2.03	0.74	<0.0001
Fear	6.59 ± 1.81	6.78 ± 1.82	5.72 ± 2.03	6.51 ± 1.98	6.37 ± 1.87	0.70	<0.0001
Neutral	7.37 ± 1.56	7.72 ± 1.40	6.86 ± 1.75	7.29 ± 1.63	7.32 ± 1.68	0.59	<0.0001
Tenderness	6.02 ± 1.70	6.37 ± 1.91	4.91 ± 2.16	6.18 ± 1.88	6.18 ± 1.80	0.69	<0.0001
Sadness	6.70 ± 1.66	6.37 ± 2.04	5.67 ± 1.91	6.48 ± 2.08	6.11 ± 1.72	0.77	<0.0001
Intrarater	Friedman = 2.7; *P* < 0.0001	Friedman = 3.2; *P* < 0.0001	Friedman = 3.1; *P* < 0.0001	Friedman = 4.2; *P* < 0.0001	Friedman = 2.9; *P* < 0.0001		

TABLE 2.
Subjects Removed From the Analysis

Subject ID	Average Judges' Score	Friedman's *P* Value
13	3.51	0.161
22	2.91	0.113
24	4.77	0.125
26	5.42	0.196
32	4.68	0.309
37	5.05	0.183

1–5/5–8 kHz, $r = -0.32$ ($P < 0.0001$). This information is presented in Figure 4.

Sound level

Because SPL is an important factor influencing the spectral slope and spectral energy levels in the LTAS analysis, Leq measures were included in this study. The mean Leq and standard deviation measured during expression of emotions are summarized in Table 4. The mean Leq measured for joy was 76.91 ± 3.68 dB, anger was 81.28 ± 3.46 dB, eroticism was 66.92 ± 5.51 dB, fear was 70.69 ± 6.12 dB, tenderness was 68.33 ± 3.96 dB, sadness was 66.85 ± 4.53 dB, and the neutral state was 69.57 ± 4.40 dB.

Perceptual assessment of the degree of breathiness

Reliability analysis, mean, and standard deviation for the perceptual assessment of the degree of breathiness during expression of emotions are presented in Table 5. Data show consistency between the judges' evaluation of the degree of breathiness across emotions. The highest degree of breathiness was found in eroticism (8.25 ± 1.56), the lowest value was demonstrated by anger (3.18 ± 1.67). The mean value measured for joy was 4.006 ± 1.05, fear was 7.05 ± 1.25, tenderness was 6.23 ± 0.83, sadness was 6.71 ± 0.94, and neutral states was 5.39 ± 0.80.

DISCUSSION

Spectral tilt

In the present study, results reveal that all the different emotions differed significantly from the neutral state ($P < 0.001$) for alpha ratio. All the evaluated emotional states obtained a higher value of this parameter, that is, a less negative number. This information suggests that the spectral slope declination was less steep for all emotions in comparison with the neutral state. Joy, anger, and fear showed the lowest slope declinations, which represent a more accentuated increased energy in the higher harmonics of the spectrum. In other words, there is less difference between the energy of the lower harmonics and energy of the higher harmonics for these three emotions. Furthermore, anger, joy, and fear were the emotions in which the actors demonstrated the highest Leq, which physiologically implies more subglottic pressure and possibly more glottal adduction. Moreover, anger and joy were the emotional states rated with the most pressed phonation by blinded judges. Related to this, Gauffin and Sundberg[38] suggested that a hyperfunctional adjustment of the vocal folds produces more intense harmonics by increasing the closed phase of the vibration cycle. This increased velocity of the closed phase results in a modification in the glottal flow wave shape affecting the spectral slope declination. On the other hand, during phonation with less glottal adduction, the harmonics that lost intensity are replaced by noise. Furthermore, previous studies have demonstrated that SPL is not linearly correlated to the spectral envelope; an increase in SPL does not correspond to the same increase in decibels at all frequencies of the spectrum.[39] When increasing SPL, the gain in decibels in the region of high frequencies is greater than in the region of low frequencies.[40–43] For this reason, alpha ratio (spectral slope declination) probably changed because of the sound level variation during the expression of the emotional states in this study.

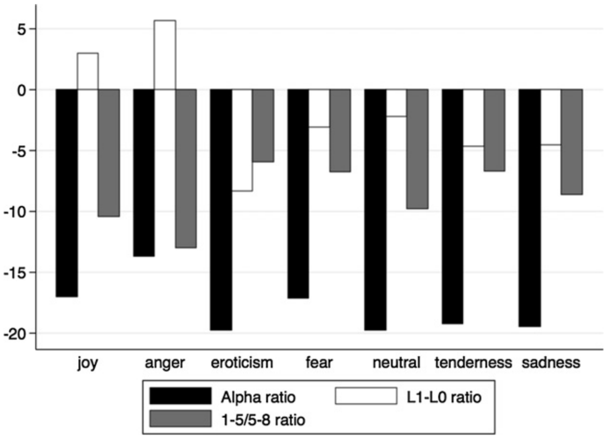

FIGURE 1. Acoustic parameters (mean) by emotions.

In a study comparing actresses' and nonactresses' voices through LTAS analysis to study the existence of the actor's formant cluster, the authors found increased alpha ratio for both groups in loud voice production.[44] This indicates that the louder the voice is, the less steep the slope will be (more energy in harmonics above 1 kHz). The same results were shown in a previous work designed to characterize vocal projection strategies in actors. The alpha ratio was directly related to the increased loudness perception.[45]

On the other hand, eroticism, sadness, and tenderness produced the lowest values of Leq in this study. Additionally, these three emotional states were rated with breathier voice quality than joy and anger. They showed the greatest slope declinations between emotions. This represents less energy in the higher harmonics of the spectrum; hence, there is more difference between the harmonic energy below 1 kHz and above 1 kHz. Related to this finding, Hammarberg et al[46] pointed out that one of the main features of the LTAS for breathy voice production is a low-energy concentration in the region of 0.4–4 kHz, corresponding to the main formants, with a steep slope of the curve to 5 kHz. Therefore, the increased spectral tilt demonstrated in eroticism, sadness, and tenderness might be explained by the low Leq values and this in turn was probably produced by a decreased glottal resistance because those emotions were perceptually rated as breathy.

Supporting our results, Banse and Scherer[7] stated that joy, anger, and fear generally seem to be characterized by an

TABLE 3.
Acoustic Parameters Analysis by Emotion (Mean ± SD)

	Emotions						
Parameters	Joy	Anger	Eroticism	Fear	Neutral	Tenderness	Sadness
Alpha ratio	−17.02 ± 3.55	−13.74 ± 2.54	−19.76 ± 2.06	−17.15 ± 2.81	−19.79 ± 2.64	−19.27 ± 2.55	−19.47 ± 1.57
L1–L0	2.98 ± 4.03	5.63 ± 5.62	−8.33 ± 4.96	−3.12 ± 7.23	−2.21 ± 4.89	−4.68 ± 5.87	−4.57 ± 5.01
1–5/5–8	−10.47 ± 4.54	−13.04 ± 3.80	−5.98 ± 3.21	−6.75 ± 3.00	−9.81 ± 3.09	−6.72 ± 3.83	−8.63 ± 3.55

increase in the mean F_0 and mean energy, whereas in sadness, a decrease in the mean F_0 and mean energy is usually found. Despite F_0 not being assessed in our study, one could expect higher values in joy, anger, and fear because vocal intensity and voice F_0 are normally interdependent. Also supporting out outcomes, Scherer[47] and Murray and Arnott[48] point out that joy, anger, and fear seem to be associated with increases in high-frequency energy, whereas sadness shows decreased high-frequency energy.

Glottal source noise

The difference between 1–5 kHz and 5–8 Hz provides information about noise in the glottal source. In the present study, all the different emotions differed significantly from the neutral state ($P < 0.001$) for 1–5/5–8 kHz. Joy and anger obtained the lowest values of 1–5/5–8 kHz ratio (more negative number). This suggests that these two emotions have less spectral energy in the 5–8 kHz region, which should be related to a decreased quantity of noise in the glottal source (less breathy voice quality). This is concordant to the results obtained in the listening test. Joy and anger showed the lowest values on the degree of breathiness. On the other hand, eroticism and tenderness obtained the highest values on the difference between 1–5 kHz and 5–8 kHz (less negative number). These values reflect a higher spectral energy on the 5–8 kHz region than joy and anger, suggesting the presence of higher glottal noise energy. Related to this, eroticism and tenderness were perceptually rated breathier than joy and anger. According to Yanagihara,[49] a high level of energy between 5 and 8 kHz is related to the components of noise emission. Hurme and Sonninem[50] also found that the spectrum of voice disorders because of paralysis of the vocal folds, the high concentration of energy is greater than 1 kHz. Concordant to these studies, Hammarberg et al[46] reported that in breathy voice quality, with incomplete glottic closure of the vocal folds, a high concentration of energy in the region of 5–8 kHz is found. Despite the fact that all participants in the present study had normal voices, all of them produced a breathy voice quality in both eroticism and tenderness according to the results from the listening test. This probably caused a high concentration of energy on the spectral region between 5 and 8 kHz, thus a high value on the difference between 1–5 kHz and 5–8 kHz. Of course this breathy voice quality was produced as part of the vocal interpretation process of emotions.

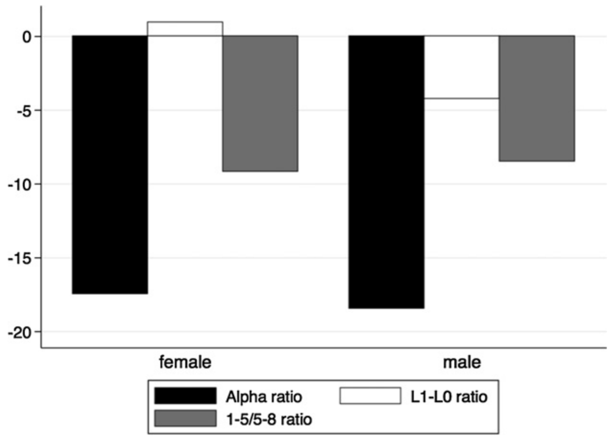

FIGURE 2. Acoustic parameters (mean) by gender.

Observing alpha ratio in eroticism, interestingly we find that this emotion has the lowest value (excluding neutral state) and thus a decreased harmonic energy on 1–5 kHz region. In addition, recall that eroticism demonstrated perceptually the highest value on the degree of breathiness. Therefore, it is possible to state that in this case, a breathier voice was associated with less harmonic energy in the 1–5 kHz region and more noise energy in the 5–8 kHz region. In other words, there was an inverse relationship between alpha ratio and the difference between 1–5 kHz and 5–8 kHz. This information is consistent with the correlation analysis in which we found a negative correlation ($r = -0.32$) between alpha ratio and 1–5/5–8 kHz ratio ($P < 0.0001$).

Mode of phonation

As it was stated before, L1–L0 (energy level difference between the first formant and F_0) provides information on the mode of phonation. In this study, significant differences only between joy, anger, and eroticism were found for L1–L0, compared with the neutral state ($P < 0.001$). Both joy and anger showed more energy on the first formant (F_1) than F_0, whereas eroticism revealed more energy on F_0 than F_1. As a result, negative values were obtained for eroticism and positive values for joy and anger. In this regard, Gauffin and Sundberg[38] suggest that in the spectrum of voices that have high glottal adduction, L0 (energy of F_0) is weak, and in the spectrum of voices with low glottal adduction, L0 is stronger. These variables were also correlated with the perception of a strained voice and breathy voice. Kitzing[51] noted that a strong L0 and low L1 (energy of F_1) were present in the spectrum of breathy voices, whereas a weak L0 and strong L1 in strained voice, indicating hyperadduction and hypoadducion of the vocal folds, respectively. Similar results were reported by Master et al[45] and Sundberg et al.[52]

In the present study, eroticism was perceptually characterized by raters as the breathiest emotion, whereas joy and

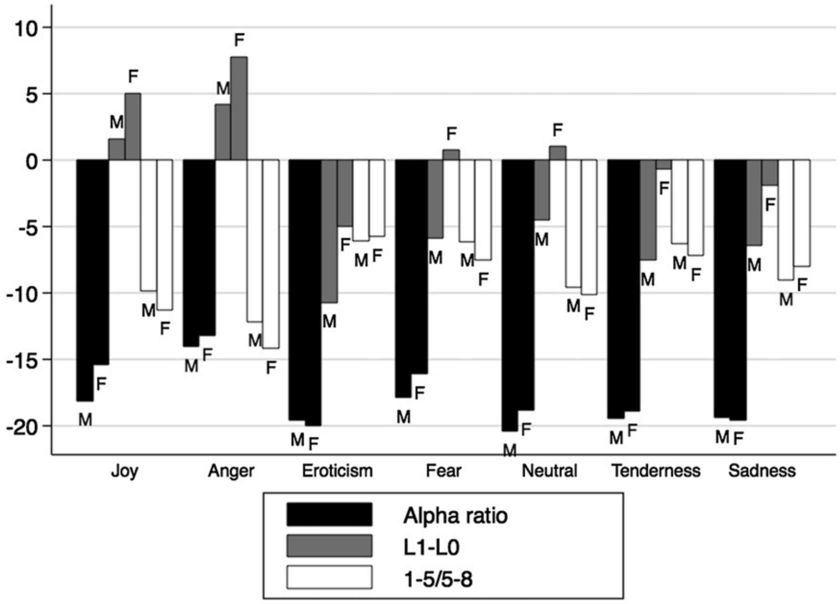

FIGURE 3. Acoustic parameters (mean) by gender and emotion.

anger were perceptually rated as the most pressed emotional states. Moreover, objective analysis of sound level showed that joy and anger obtained the highest values of Leq. From both the acoustic analysis and perceptual information together, it could be inferred that eroticism was produced with the lowest degree of glottal adduction among all the emotional states.

Correlation analysis

Correlation analysis revealed a positive correlation ($r = 0.49$) between alpha ratio and L1–L0 and a negative correlation ($r = -0.32$) between L1–L0 and 1–5/5–8 kHz differences. Moreover, a negative correlation ($r = -0.32$) between alpha ratio and 1–5/5–8 kHz ratio ($P < 0.0001$) was found as noted before. These findings seem to be logical from the physiologic point of view. The alpha ratio provides information on the spectral slope declination, 1–5/5–8 kHz difference provides information about noise in the glottal source, and L1–L0 provides information on the mode of phonation. The three features (spectral slope declination, glottal noise, and mode of phonation) are parts of the same physiologic concept, glottal resistance. As the glottal resistance varies, these three spectral parameters should change in a related way according to our outcomes. In a recent work by Master et al,[44] results showed a positive correlation between the alpha ratio and L1–L0 difference. Authors suggested that this information reflects the relation between phonation mode and amplitude of the harmonics in the high-frequency region.

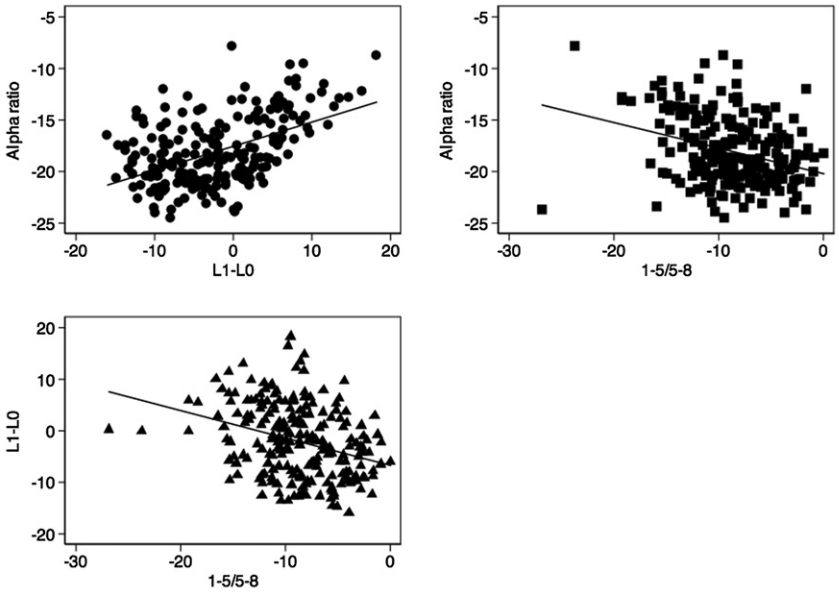

FIGURE 4. Linear correlation scatter plots between acoustic parameters.

TABLE 4.
Analysis of Equivalent Intensity Level (Leq) During the Expression of Emotions

Emotions	Mean ± SD	Neutral	*P* Value
Joy	76.91 ± 3.68	69.57 ± 4.40	<0.001
Anger	81.28 ± 3.46	69.57 ± 4.40	<0.001
Eroticism	66.92 ± 5.51	69.57 ± 4.40	0.327
Fear	70.69 ± 6.12	69.57 ± 4.40	0.154
Tenderness	68.33 ± 3.96	69.57 ± 4.40	0.108
Sadness	66.85 ± 4.53	69.57 ± 4.40	0.445

Conceptually, glottal resistance is the quotient of the subglottic pressure divided by the airflow rate.[53] This measurement is intended to reflect the overall resistance of the glottis and, by extension, serve as an estimate of the valving characteristics, whether too tight (hyperfunctional), too loose (hypofunctional), or normal. For high glottal resistance, contact quotient is greater than in a voice with low glottal resistance. Therefore, based on cited works,[49,50] a breathier voice quality (mode of phonation), more glottal noise energy (above 5 kHz), and less spectral harmonic energy (between 1–5 kHz) should be found in voice production with low glottal resistance, such as eroticism and tenderness. Furthermore, the amplitude of higher harmonics are particularly sensitive to the phase velocity of closing phase of the vibration cycle, that is, the speed at which the airflow decreases at the end of open phase. The faster the speed of closure is, the greater the subglottic pressure and the most intense high harmonics of the spectrum will be. This increased energy in the harmonics above 1 kHz should be indirectly reflecting less glottal noise and less energy at the F_0 in comparison with the energy of the first formant, hence, a more resonant and brilliant voice quality should be perceived.

Gender differences

Statistically significant differences between men and women ($P < 0.001$) according to the linear regression model for all acoustic variables (alpha ratio, L1–L0, and 1–5/5–8 kHz difference) were found. For alpha ratio, female voice obtained a higher mean value than male voice. This indicates a less steep spectral slope declination for women than men. Regarding L1–L0, female voices showed less negative values than male voices. For 1–5/5–8 kHz, female voices showed a lower value than male voices, which should be related to a more decreased quantity of noise in the glottal source (less breathy voice quality) in women than men. These results suggest that men might have had a more breathy voice quality and less energy above 1 kHz than women. These outcomes are surprising because several previous studies have reported opposite differences between female and male acoustic spectral features. Klatt and Klatt[54] and Hillenbrand et al[55] reported that women were judged to be breathier than men. H1 amplitude measures were also generally greater for women than men. Similarly, Mendoza et al[37] found greater levels of aspiration noise, located in the spectral regions corresponding to the third formant, causing the female voices to have a more "breathy" quality than the male voices. In this regard, Van Borsel et al[56] concluded that breathiness indeed may contribute to the perception of femininity.

Related to the mode of phonation, the results from our study suggest a lower and stronger F_0 than F_1 in men, which could reflect less glottal adduction and hence a more flow mode of phonation. In terms of aerodynamic measurements, flow phonation has been defined as "that type of phonation that has the highest possible glottogram amplitude that can be combined with a complete glottal closure."[38] Inspection of inverse-filtered waveforms for flow phonation in fact reveals a slightly positive minimum flow offset, implying barely abducted vocal folds.[57] Nevertheless, some caution is necessary when drawing conclusions with respect to the voice differences between women and men in the present study. To analyze gender differences, we took all the samples together including six emotions and neutral state. Therefore, the mean of these samples do not represent the normal mode of phonation used in previous researches in which female voices were more breathy than male voices, moreover is interesting.

Interaction between gender and emotions

Additionally, the analysis of interaction between the two variables (emotion and gender) showed that there are statistically significant differences ($P < 0.001$) in the interaction between emotion and gender, that is, the expression of certain emotion by either a man or woman is different with regard to alpha ratio and 1–5/5–8 kHz difference. This indicates that women used more glottal adduction than men during expression of emotions,

TABLE 5.
Judges' Reliability by Degree of Breathiness During the Expression of Emotions

	Judges' Scores (Mean ± SD)					Interrater	
Emotions	1	2	3	4	5	Kendall	*P* value
Joy	4.03 ± 1.22	3.93 ± 1.33	3.90 ± 1.18	4.01 ± 1.38	4.16 ± 1.96	0.91	<0.0001
Anger	3.25 ± 0.96	3.09 ± 0.98	3.12 ± 1.22	3.25 ± 1.07	3.16 ± 1.13	0.85	<0.0001
Eroticism	8.35 ± 1.16	8.22 ± 1.01	8.06 ± 1.21	8.54 ± 0.98	8.06 ± 1.03	0.84	<0.0001
Fear	7.19 ± 0.81	7.19 ± 0.82	7.03 ± 1.03	6.96 ± 1.98	6.90 ± 1.87	0.90	<0.0001
Neutral	5.51 ± 1.56	5.38 ± 1.40	5.35 ± 1.75	5.38 ± 1.63	5.32 ± 1.68	0.89	<0.0001
Tenderness	6.33 ± 0.70	6.03 ± 0.91	6.20 ± 1.16	6.36 ± 0.88	6.20 ± 0.80	0.89	<0.0001
Sadness	6.90 ± 1.66	6.48 ± 2.04	6.54 ± 1.91	6.70 ± 2.08	6.93 ± 1.72	0.87	<0.0001

such as joy, anger, and fear, as part of the interpretation. Moreover, male actors used a more breathy voice quality, thus more glottal noise, than women in emotions such as eroticism, sadness, and tenderness as part of the interpretation. This suggested that in this study, the difference in female and male voices could be attributed to the interpretation of emotions rather than the normal vocal physiology.

The fact that women used more glottal adduction than men during expression of emotions such as joy, anger, and fear is a very interesting result that could be discussed in relation to the higher prevalence of voice disorders in women.[58,59] In our study, this vocal expression was part of acted emotional expression; studying whether this behavior is present in normal phonation would be an interesting topic for future research.

CONCLUSION

The expression of emotion impacts the spectral energy distribution. Emotional states characterized by a breathy voice quality, such as tenderness, sadness, and eroticism, present a low harmonic energy above 1 kHz, high glottal noise energy, and more energy on F_0 than overtones. On the other hand, emotional states such as joy, anger, and fear are characterized by high harmonic energy above 1 kHz (less steep spectral slope declination), low glottal noise energy, and more energy in the F_1 than F_0 region. It is important to recall that only acted emotions were assessed in this study and not naturally occurring emotions because of the reasons presented above. Although recordings of actors for emotional voice analysis have been found to be representative of real emotions by a number of scientists, further research with real emotions might be addressed to corroborate the findings of the present work.

REFERENCES

1. Xu Y. Speech melody as articulatorily implemented communicative functions. *Speech Commun*. 2005;46:220–251.
2. Scherer KR. On the nature and function of emotion: a component process approach. In: Scherer KR, Ekman P, eds. *Approaches to Emotion*. Hillsdale, NJ: Erlbaum Press; 1984:293–318.
3. Scherer KR. Vocal affect expression: a review and a model for future research. *Psychol Bull*. 1986;99:143–165.
4. Scherer KR. Interpersonal expectations, social influence, and emotion transfer. In: Blanck PD, ed. *Interpersonal Expectations: Theory, Research, and Application*. Cambridge, England: Cambridge University Press; 1993: 316–336.
5. Schlosberg H. Three dimensions of emotion. *Psychol Rev*. 1954;61:81–88.
6. Murray IR, Arnott JL. Toward to simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *J Acoust Soc Am*. 1993; 93:1097–1108.
7. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol*. 1996;70:614–636.
8. Nwe TL, Foo SW, De Silva LC. Speech emotion recognition using hidden Markov models. *Speech commun*. 2003;41:603–623.
9. Scherer KR. Vocal correlates of emotion. In: Manstcad A, Wagner H, eds. *Handbook of Psychophysiology: Emotion and Social Behavior*. London, UK: Wiley Press; 1989:165–197.
10. Douglas-Cowie E, Campbell N, Cowie R, Roach P. Emotional speech: towards a new generation of databases. *Speech Commun*. 2003;40:33–60.
11. Borden GJ, Harris KS. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Baltimore, MD: Williams & Wilkins Press; 1984.
12. Schröder M, Cowie R, Douglas-Cowie E, Westerdijk M, Gielen S. In: *Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis*. Vol. 1. Aalborg, Denmark: Eurospeech; 2001.
13. Douglas-Cowie E, Cowie R, Schröder M. The Description of Naturally Occurring Emotional Speech. 15th International Conference of Phonetic Sciences, Barcelona, Spain, 2003.
14. Scherer KR. Vocal communication of emotion: a review of research paradigms. *Speech Commun*. 2003;40:227–256.
15. Laukkanen A-M, Vilkman E, Alku P, Oksanen H. On the perception of emotions in speech: the role of voice quality. *Scand J Logoped Phoniatr Vocol*. 1997;22:157–168.
16. Waaramaa T, Alku P, Laukkanen AM. The role of F3 in the vocal expression of emotions. *Logoped Phoniatr Vocol*. 2006;31:153–156.
17. Juslin PN. Perceived emotion expression in synthesized performances of a short melody: capturing the listener's judgment policy. *Musicae Scientiae*. 1997;1:225–256.
18. Juslin PN. Cue utilization in communication of emotion in music performance: relating performance to perception. *J Exp Psychol Hum Percept Perform*. 2000;26:1797–1813.
19. Juslin PN, Madison G. The role of timing patterns in recognition of emotional expression from musical performance. *Music Percept*. 1999;17: 197–221.
20. Lieberman P, Michaels SB. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *J Acoust Soc Am*. 1962;34:922–927.
21. Scherer KR, Oshinsky JS. Cue utilization in emotion attribution from auditory stimuli. *Motiv Emot*. 1977;15:123–148.
22. Pettersen V, Bjørkøy K. Consequences from emotional stimulus on breathing for singing. *J Voice*. 2009;23:295–303.
23. Foulds-Elliott S, Thorp CW, Cala S, Davis P. Respiratory unction in operatic singing: effects of emotional connection. *Logoped Phoniatr Vocol*. 2000;25:151–168.
24. Miller R. *The Structure of Singing*. New York, NY: Shirmer Books; 1986.
25. Chapman J. *Singing and Teaching Singing: A Holistic Approach to Classical Voice*. San Diego, CA: Plural; 2006.
26. Scherer KR. On the symbolic functions of vocal affect expression. *J Lang Soc Psychol*. 1988;7:79–100.
27. Pittam J, Scherer KR. Vocal expression and communication of emotion. In: Lewis M, Haviland JM, eds. *Handbook of Emotions*. New York, NY: Guilford Press; 1993:185–197.
28. Davitz JR. Auditory correlates of vocal expression of emotional feeling. In: Davitz R, ed. *The Communication of Emotional Meaning*. New York, NY: MacGrav-Hill; 1964:101–112.
29. Kienast M, Sendlmeier W. *Acoustical Analysis of Spectral and Temporal Changes in Emotional Speech*. New Castle, Northern Ireland: ISCA Workshop on Speech and Emotion; 2000:145–151.
30. Juslin P, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull*. 2003; 129:770–814.
31. Ververidis D, Kotropolos C. Automatic speech classification to five emotional states based on gender information. Proceedings of Eusipco. 2004: 341–344.
32. Yildirim S, Bulut M, Lee CM et al. An acoustic study of emotions expressed in speech. Proceedings ICSLP. 2004:2193–2196.
33. Williams CE, Stevens KN. Emotions and speech: some acoustic correlates. *J Acoust Soc Am*. 1972;52:1238–1250.
34. Miralles A. *Monólogos para ejercicio*. Madrid, Spain: La Avispa Press; 1984:56–58.
35. Boersma P, Weenink D. Praat Manual: Doing Phonetics by Computer (V. 5. 0.23). [Computer program]. Available at: http://www.praat.org/. Accessed September 3, 2008.
36. Linville S, Rens J. Vocal tract resonance analysis of aging voice using the long term average spectra. *J Voice*. 2001;3:323–330.
37. Mendoza E, Valencia N, Muñoz J, Trujillo H. Differences in voice quality between men and women: use of the long-term average spectrum. *J Voice*. 1996;1:59–66.
38. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *J Speech Lang Hear Res*. 1989;32:556–565.

39. Nordemberg M, Sundberg J. Effect on LTAS of vocal loudness variation. TMH-Quarterly Progress and Status Report, Royal Institute of Technology. 2003;45:87–91.
40. Bloothooft G, Plomp R. The sound level of the singer's formant in professional singing. *J Acoust Soc Am*. 1986;79:2028–2033.
41. White P. A study of the effects of vocal loudness intensity variation on children's voices using long-term average spectrum analysis. *Logoped Phoniatr Vocol*. 1998;23:111–120.
42. White P, Sundberg J. Spectrum effects of subglottal pressure variation in professional baritones singers. TMH-Quarterly Progress and Status Report, Royal Institute of Technology. 2000;4:29–32.
43. Ternström S. Very loud speech oversimulated environmental noise tends to have a spectral peak in the F1 region. *J Acoust Soc Am*. 2003;13. 2296.
44. Master S, De Biase N, Madureira S. What about the "actor's formant" in actresses' voices? *J Voice*. 2012;26:e117–e122.
45. Master S, De Biase N, Chiari BM, Laukkanen AM. Acoustic and perceptual analyses of Brazilian male actors' and nonactors' voices: long-term average spectrum and the actor's formant. *J Voice*. 2008;22:146–154.
46. Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*. 1980;90: 441–451.
47. Scherer KR. Expression of emotion in voice and music. *J Voice*. 1995;9: 235–248.
48. Murray I, Arnott J. Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech. *Comput Speech Lang*. 2008;22: 107–129.
49. Yanagihara N. Significance of harmonic changes and noise components in hoarseness. *J Speech Lang Hear Res*. 1967;10:531–541.
50. Hurme P, Sonninem A. Normal and disordered voice qualities: listening tests and long-term spectrum analyses. In: Hurme P, ed. *Papers in Speech Research*. Finland: University of Jyväskylä: Publications of the Department of Communication Studies, number 1; 1985.
51. Kitzing P. LTAS criteria pertinent to the measurement of voice quality. *J Phonetics*. 1986;14:477–482.
52. Sundberg J, Titze I, Scherer R. Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *J Voice*. 1993;7:15–29.
53. Stemple JC. *Clinical Voice Pathology Theory and Management*. San Diego, CA: Singular Publishing Group; 2000.
54. Klatt D, Klatt L. Analysis, synthesis and perception of voice quality variations among female and male talkers. *J Acoust Soc Am*. 1990;87:820–857.
55. Hillenbrand J, Cleveland R, Erickson R. Acoustic correlates of breathy vocal quality. *J Speech Lang Hear Res*. 1994;37:769–778.
56. Van Borsel J, Janssens J, De Bodt M. Breathiness as a feminine voice characteristic: a perceptual approach. *J Voice*. 2009;23:291–294.
57. Verdolini K, Druker DG, Palmer PM, Samawi H. Laryngeal adduction in resonant voice. *J Voice*. 1998;12:315–327.
58. Nelson Roy N, Merrill R, Thibeault S, Parsa R, Gray S, Smith E. Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res*. 2004;47:281–293.
59. Russel A, Oates J, Greenwood KM. Prevalence of voice problems in teachers. *J Voice*. 1998;12:467–479.