Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2008;60:249–255
DOI: 10.1159/000151762

Published online: September 2, 2008

Monopitched Expression of Emotions in Different Vowels

Teija Waaramaa[a] Anne-Maria Laukkanen[a] Paavo Alku[b] Eero Väyrynen[c]

[a]Department of Speech Communication and Voice Research, University of Tampere, Tampere,
[b]Department of Signal Processing and Acoustics, Helsinki University of Technology, Espoo, and
[c]MediaTeam, University of Oulu, Oulu, Finland

Key Words
Voice quality · Inverse filtering · Voice source · Formants · Perception of emotions

Abstract
Fundamental frequency (F_0) and intensity are known to be important variables in the communication of emotions in speech. In singing, however, pitch is predetermined and yet the voice should convey emotions. Hence, other vocal parameters are needed to express emotions. This study investigated the role of voice source characteristics and formant frequencies in the communication of emotions in monopitched vowel samples [a:], [i:] and [u:]. Student actors (5 males, 8 females) produced the emotional samples simulating joy, tenderness, sadness, anger and a neutral emotional state. Equivalent sound level (L_{eq}), alpha ratio [SPL (1–5 kHz) – SPL (50 Hz–1 kHz)] and formant frequencies F1–F4 were measured. The [a:] samples were inverse filtered and the estimated glottal flows were parameterized with the normalized amplitude quotient [$NAQ = f_{AC}/(d_{peak}T)$]. Interrelations of acoustic variables were studied by ANCOVA, considering the valence and psychophysiological activity of the expressions. Forty participants listened to the randomized samples (n = 210) for identification of the emotions. The capacity of monopitched vowels for conveying emotions differed. L_{eq} and NAQ differentiated activity levels. NAQ also varied independently of L_{eq}. In [a:], filter (formant frequencies F1–F4) was related to valence. The interplay between voice source and F1–F4 warrants a synthesis study.

Introduction

Fundamental frequency (F_0), the main correlate of pitch, its variations and sound pressure level (SPL), which is mainly heard as loudness, are well known to be among the most important variables in emotion expressions [1, 2]. Both of them tend to increase in accordance with psychophysiological activity level, being higher in emotions with high arousal (e.g. joy and anger) and lower in emotions with less arousal (e.g. tenderness and depressive sadness) [1, 3]. In the literature on emotions it is widely agreed that there are four basic emotions (joy, anger, fear and sadness) which are universal, not culture-related. These basic or primary emotions represent both high arousal activity level (joy, anger and fear) and low arousal activity level (sadness). Additionally, there are also so-called secondary or social emotions, which are culturally related and therefore more difficult to define (such as longing, boredom or satisfaction).

Voice quality has been considered a crucial variable in differentiating emotions which are communicated

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1021–7762/08/0605–0249$24.50/0

Accessible online at:
www.karger.com/fpl

Teija Waaramaa, MA
Department of Speech Communication and Voice Research
University of Tampere
FI–33014 Tampere (Finland)
Tel. +358 3 3551 7104, Fax +358 3 3551 6063, E-Mail teija.waaramaa@uta.fi

through a subtle coloring of the voice [2, 4, 5]. Thus, voice quality is a paralinguistic means of signaling differences in meaning in speech [6]. Voice quality and F_0 may vary individually in conveying vocal emotional content [7].

Voice quality is defined by Laver [8] in a broad sense as the individual coloring of the speaker's voice, which is determined both by the voice source (a result of vocal fold vibration) and vocal tract characteristics [9]. Thus, voice quality results from both phonatory and articulatory characteristics. In order to study the voice source, inverse filtering is needed to separate the resonances from the signal. The voice source varies together with F_0, SPL (or equivalent sound level, L_{eq}) and also with tempo [10–14], and therefore, the individual role of voice quality in emotional expression has not been simple to study. However, it is reasonable to assume that the voice source may also vary independently, not merely related to variation in F_0 and SPL [15, 16]. Moreover, F_0 and SPL seem to interact, SPL typically rising together with F_0. F_0 is also one of the means to increase SPL [17].

In singing, the variations in pitch are always predetermined, and the relative loudness and duration are also more or less predirected. Yet the vocalist's voice should be emotionally expressive. Thus, singers, compared to speakers, need to use different strategies to convey emotional expressions to the audience. This acoustic conveyance of emotions has been investigated among other things by analyzing the vocalist's interpretation *of the music* and the expression of the emotions *of the character* portrayed, and also by investigating the *vocalist's own psychophysiological state* while performing [18]. From the perceptual viewpoint, the listeners' individual abilities to perceive emotional quality and the acoustic characteristics of the voice signal used in the perception process are of importance [18]. Vocalists' and also actors' expressions and performance therefore need to be tightly controlled. One means of improving this control is to practice with monopitched sounds. Hence, the control concentrates on the voice quality itself, not only on the prosodic features.

As every vowel has its own vocal tract setting, it is plausible that the expression-related acoustic changes are different when they occur in the context of a different vowel. The same articulatory movement, for example moving the tongue forward in the oral cavity, causes different acoustic changes in different vowels [19].

The present study investigated voice quality parameters (voice source and formant frequencies) in monopitched emotional expressions. It may be assumed that, when the variation of F_0 is eliminated, the role of voice quality in the expression of emotions would become clearer. Different vowels were studied since it was hypothesized that there might be (1) differences between vowels in conveying emotional content and (2) differences in the possible expression-related changes in the formant structure of different vowels.

Materials and Methods

Subjects and Recordings

The material for the present study was recorded in the University of Tampere in a well-damped studio using a digital recorder Tascam DA-20 and a Brüel & Kjær 4165 microphone, placed at distance of 40 cm from the subject's lips. Thirteen graduating professional actors (5 males and 8 females) with normal voices and without any known pathologies of the larynx or hearing served as subjects. They produced in random order three monopitched prolonged steady vowels, [a:], [i:] and [u:], separated from each other, expressing randomly four emotional states: anger, joy, sadness, tenderness, and a neutral emotional state at a comfortable speaking pitch. These emotional states were chosen since they represent both high and low activity level and positive and negative emotional valences. Intensity and duration varied freely in the expressions, but vowel pitch was standardized. The material contained 195 vowel samples (13 actors × 5 emotional states × 3 vowels).

Perceptual Analysis

The samples were replayed to 40 listeners, consisting of university teachers and students (20 males, 20 females, mean age 38 years in females and 39 in males). An equal number of subjects from both genders were used to investigate possible gender differences in the perception of emotional expressions. The computer program Judge (developed by Svante Granqvist, KTH, Stockholm) was used to evaluate the samples. The participants listened to the samples using Sennheiser HD 530 II headphones. The Judge program replayed the samples in a different randomized order for every listener. All of the recorded samples (n = 195) were used in the listening test. Fifteen of them were repeated in order to study intrarater reliability. The listeners' task was to state for each of the 210 samples which emotion they perceived in the sample. A visual analog scale (0–1,000 units) was used. One end, 0, was labeled 'neutral' (no emotion), and the other end (1,000 units) was labeled according to each emotional state that was simulated in the study. The listeners were allowed to listen to the samples as many times as they felt they needed. However, the participants were recommended to listen to each sample only once, if possible, because it was of interest to study the very first and thus the basic reaction to or perception of the signal heard and hence to avoid any speculations which were prone to arise from the samples (average 2,336 ms in males and 2,472 ms in females). A confusion matrix of the listeners' answers was calculated in order to investigate the percentage of similar answers between the participants. Intrarater reliability was studied by calculating the percentage of similar answers given by the listeners to the repeated samples (n = 15). The differences between the genders in the results of the listening test were also studied.

Table 1. Correct recognition of the emotions expressed in different vowels

Emotion	[a:]	[i:]	[u:]
Neutral	46%	39%	40%
Sadness	42%	57%	58%
Joy	41%	43%	28%
Anger	73%	64%	67%
Tenderness	61%	48%	35%

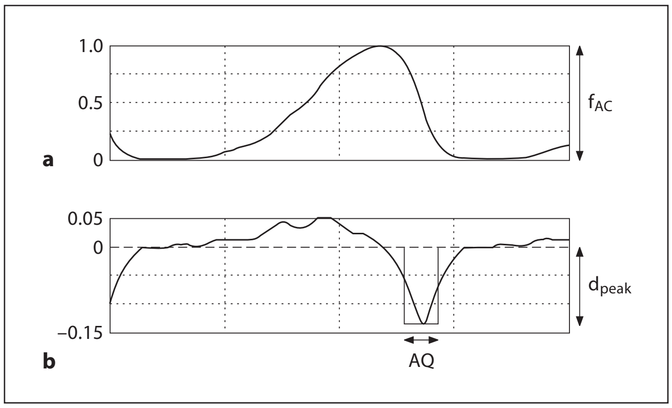

Fig. 1. Inverse filtered signal. **a** Glottal flow. f_{AC} = AC flow amplitude. **b** First derivative of glottal flow. d_{peak} = Negative peak amplitude of the derivative. AQ (amplitude quotient): f_{AC}/d_{peak}. Time on the horizontal axis, flow on an arbitrary scale on the vertical axis.

Acoustic Analysis

The 195 samples were analyzed for F_0, L_{eq}, alpha ratio and formant frequencies F1–F4 with a signal analysis system named Intelligent Speech Analyser (ISA), developed by Raimo Toivonen, MScEng. F_0 was measured to ensure that the expressions were monopitched since the actual Hz value is typically connected with the psychophysiological activity level and hence may vary along with intensity. The alpha ratio reflects voice quality by showing the sound level difference between the range above and below 1 kHz [20]. It was calculated here by subtracting the L_{eq} in the range 50 Hz–1 kHz from the L_{eq} in the range 1–5 kHz. The alpha ratio naturally depends on both the voice source and filter characteristics. Formant frequencies were measured on spectrograms and the FFT average spectra were taken from the middle portion of each vowel sample.

In order to study the voice source and formant frequencies separately, the voice signal was inverse filtered using the Iterative Adaptive Inverse Filtering method [21], which uses the acoustic speech pressure signal as the input, thereby enabling the study of natural speech without the inconvenience and restrictions inherent in using a flow mask over the face. Since inverse filtering techniques require vowels with high F1 in order to be accurate, only [a:] vowels were inverse filtered in the present study. The resulting voice source was parameterized by calculating the normalized amplitude quotient [$NAQ = f_{AC}/(d_{peak}T)$] [22]. NAQ measures the relative time of the glottal closing phase from two amplitude domain values, peak-to-peak AC flow (f_{AC}) and the amplitude of the negative peak (d_{peak}) of the first derivative of the flow. T is the fundamental period length (fig. 1). NAQ has been shown to reflect phonation type, being low in hyperfunctional (pressed) voice and high in hypofunctional (breathy) voice [22]. NAQ also correlates with SPL [22].

Statistical Analysis

Analysis of covariance (ANCOVA, SPSS-15, Chicago, Ill., USA) was used to study the interrelations of acoustic variables in emotional expressions. Two main characteristics of expressions were considered: valence and psychophysiological activity level. Valence (the affective value of an emotion on an axis positive – neutral – negative) and psychophysiological activity level (on an axis high – medium – low) were assigned arbitrary numbers by the authors to enable statistical analysis. The positive emotions joy and tenderness were assigned a positive value 1, a neutral emotional state 0, and the negative emotions anger and sadness –1. For the psychophysiological activity level joy and anger were given a positive value 1, neutrality 0, and tenderness and sadness a negative value –1. Thus, there were fewer choices (n = 3) in the study of valence and the psychophysiological activity level than in the actual emotions (n = 5).

Dependent variables were: (1) Filter (= weighted sum variable of formant frequencies F1–F4 constructed with principal component analysis), (2) NAQ (measured only in vowel [a:]), (3) L_{eq}, and (4) alpha ratio. The effect of gender was included in the models. L_{eq} was set as a covariate in order to study the dependence of other characteristics on it. Bonferroni correction was used. Of the emotions expressed in [a:], the data for 1 female subject had to be excluded, those for 4 females in [u:], and all samples expressing tenderness in vowel [u:] produced by female subjects were completely excluded due to analysis problems.

Results

Listening Test

Emotions were recognized with 50% accuracy in the listening test (table 1). The percentage for intrarater reliability was 59%. The best recognized emotion was anger in all vowels with 68% accuracy, while joy was the least recognized emotion with 37% accuracy. Sadness was the most often chosen emotion for an answer (24% of all answers), while joy was the most seldom chosen emotion (15%). There were differences in the perception between the vowels. Vowel [a:] conveyed best tenderness, vowels [i:] and [u:] sadness. Anger was well conveyed by all the vowels studied, however, vowel [a:] was a somewhat better conveyor of anger than the other vowels. Emotions conveyed by vowel [u:] were quite poorly recognized, especially positive emotions.

Table 2. Averages of L_{eq}, alpha ratio and formant frequencies F1–F4 in vowels [a:], [i:] and [u:], and NAQ averages for vowel [a:] for males and females separately

	[a:]					[i:]					[u:]				
	neutral	sadness	joy	anger	tender-ness	neutral	sadness	joy	anger	tender-ness	neutral	sadness	joy	anger	tender-ness
Males															
L_{eq}	65	58	67	72	57	62	56	64	68	56	62	59	64	68	57
Alpha ratio	–10	–10	–8	–7	–9	–14	–18	–13	–6	–18	–27	–28	–25	–23	–29
NAQ	0.11	0.16	0.1	0.08	0.18										
F1	611	603	646	650	659	306	284	461	301	323	306	314	327	319	327
F2	1,068	1,060	1,124	1,073	1,107	2,024	2,033	2,119	2,050	2,080	646	629	685	672	633
F3	2,575	2,627	2,700	2,713	2,657	2,588	2,713	2,735	2,903	2,683	2,433	2,459	2,403	2,494	2,567
F4	3,372	3,381	3,407	3,372	3,492	3,329	3,372	3,390	3,643	3,669	2,993	3,191	3,118	3,415	3,518
Females															
L_{eq}	62	55	63	72	55	60	56	56	63	66	60	56	56	70	70
Alpha ratio	–4	–6	–4	–1	–7	–15	–23	–15	–19	–12	–22	–30	–21	–26	–26
NAQ	0.11	0.15	0.14	0.1	0.17										
F1	678	692	719	778	668	377	380	382	428	425	377	393	396	458	458
F2	1,190	1,222	1,295	1,225	1,284	2,538	2,455	2,557	2,199	2,592	651	668	657	751	773
F3	2,891	3,004	3,182	2,931	3,082	3,163	3,192	3,112	3,152	3,308	2,739	2,455	2,830	2,552	2,764
F4	3,954	3,892	4,180	3,941	4,162	4,078	4,121	4,110	3,898	4,038	3,708	3,898	4,212	3,583	3,583

Valence was perceived with 70.5% and activity with 76.5% accuracy of the answers given. This result may suggest fewer difficulties in the perception of valence and psychophysiological activity level compared to the recognition of actual emotions. Valence was perceived correctly in 76% of [a:] and [i:] vowels and in 60% of vowel [u:]. Psychophysiological activity level was recognized best in vowel [i:] with 86% correct answers. The corresponding percentage for [a:] was 73% and for [u:] 71%.

Some minor nonsignificant gender differences were seen in the answers: males perceived the emotions with 48% accuracy, females with 52% accuracy on average for all emotions expressed. No gender differences in perception were found regarding emotions expressed by the same or opposite gender.

Acoustic Analysis

The averages of the parameters measured are given in table 2 separately for both genders.

Effects of Filter

In all three vowels filter (F1–F4) was related to gender, which was to be expected. In vowel [a:] filter was also related to valence ($F_{1,51} = 6.18$, $p = 0.016$), differentiating between positive and negative emotions (Bonferroni adjusted $p = 0.016$). The interaction effect between valence and gender was not significant. Valence was thus related to formant frequencies in a similar way in both genders. In positive emotions, the average frequencies of the formants tended to be somewhat higher than in negative emotions.

In vowels [i:] and [u:], unlike for [a:], filter was not significantly related to valence. As expected, formant patterns in emotional expressions differed between vowels. The perceptual effect of the formant changes was most likely related to an interplay between voice source and frequency relations between adjacent formants.

Effects of Voice Source

In all three vowels, L_{eq} was associated with activity, which was to be expected. Significant differences between low and medium (Bonferroni test, $p = 0.001$) and low and high ($p < 0.001$) activity levels were observed. In [i:] and [u:], only L_{eq} and activity level were significantly related. L_{eq} was highest in anger and lowest in tenderness and sadness in both genders. There was a high negative correlation between L_{eq} and NAQ in both males and females. NAQ was smallest in anger in both genders and greatest in tenderness. In [a:], NAQ was related to the psychophysiological activity level ($F_{1,57} = 27.9$, $p < 0.001$). Differences in NAQ also remained significant when L_{eq} was set as a covariate, suggesting differences in phonation type independent of voice loudness (fig. 2).

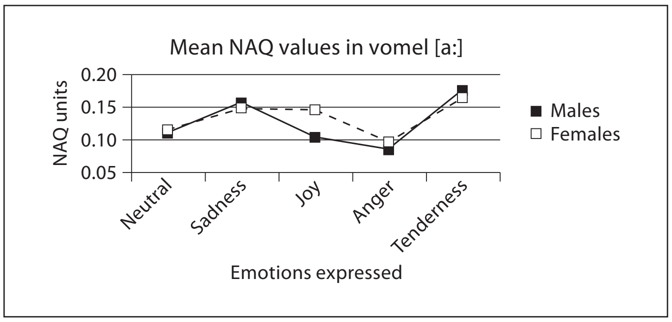

Fig. 2. Mean NAQ values in monopitched emotional expressions for males and females in vowel [a:].

Unlike L_{eq} and NAQ, L_{eq} and alpha ratio did not show any correlation. This is most likely due to the fact that alpha ratio – spectral energy distribution – is also affected by resonances.

Discussion

The perception of emotion samples revealed that the listeners were more likely to hear the negative coloring of voices than the positive coloring. This tendency may have been inherited in the course of evolution: on the one hand, humans have had to be aware of a possible threat and be sensitively attuned to negative signals especially. On the other hand, positive stimuli did not require any necessary reaction for survival. The listening test also revealed that emotion perception in different vowels varies. The open back vowel [a:] conveyed better anger, tenderness and neutrality than the other two vowels, most likely due to its evenly spread (nondiffuse) formant structure, which gives more freedom for expression. Alpha ratio was significantly higher in vowel [a:] than in [i:] and [u:], which was to be expected. This may explain the better recognition of emotions in [a:]. In all emotions studied, alpha ratio values were significantly higher in [a:] than in the other vowels. This may be due to louder formant amplitudes and greater amount of spectral energy, which may have had some perceptual relevance. Anger was conveyed remarkably well by all vowels studied, implying that it may not be related to filter functions but rather to the voice source, and may thus not be vowel-dependent. Joy was slightly better recognized in the front vowel [i:] than in [a:], but distinctively better in [i:] than in [u:]. This result for joy may be connected to the formant structure of [i:] where F2–F4 are all relatively high in frequency, which, in turn, may give a bright sound to the vowel when signaling positive emotions.

In the writing of song lyrics, the words tend to be chosen such that the vowels support the mood of the message of the song, e.g. /a/ and /e/ may be used when brightness is needed in the expressions, and /u/ and /y/ when darker voice timbre is desired. The resonances create the acoustic structures of the vowels, and so the formant frequencies differ between them. Thus, their ability to modify perceptually relevant resonances is different.

In earlier investigations, formant frequencies (F2–F4) have been found to be higher in positive valence than in negative valence [16, 23]. The higher frequencies may be due to the smiling position of the lips, which shortens the vocal tract. This effect should affect the highest formants of [i:] especially. However, sadness was signaled well by both [i:] and [u:] vowels. Their diffuse formant structures and smaller amount of spectral energy in the higher formant frequency area (alpha ratio was significantly lower than in [a:]) may account for this result. Since there was no significant difference between valence and formant frequencies in [i:] and [u:], the use of the formant amplitudes by varying L_{eq} may play a role in darkening the timbre of the voice for signaling negative emotions in vowels [i:] and [u:]. The perceptual value of a formant, its loudness, is influenced by the amplitude of the formant and the sensitivity of hearing at the frequency range of the formant. Formant amplitude, in turn, is affected by tilting of the voice source spectrum, by formant tuning (how close a match there is between a voice source partial and a formant) and by the distance of formants from each other. Formant amplitude is higher when the voice source spectrum has stronger overtones (louder voice or more pressed phonation type), and when the formant frequency decreases and hence the formant coincides with a lower and thus a stronger overtone. Furthermore, distance between formant frequencies has an effect on formant amplitudes; if two formants are close to each other, their amplitudes become 6 dB stronger and the area between them 12 dB [9].

F_0 and SPL interact with the voice source characteristics (glottal flow velocity waveform) [24, 25]. In the present study, loudness was allowed to vary, since a strict control of SPL would have affected the phonation type too much. Phonation type along the axis hypofunctional-hyperfunctional (pressed) is reflected in NAQ and spectral energy distribution, i.e. in this case in alpha ratio. Smooth, almost sinusoidal glottal flow velocity waveform (and thus high NAQ) characterizes a soft, breathy, hypofunctional phonation type, while a steeper waveform is seen

in a pressed, hyperfunctional phonation type (and therefore NAQ would be low) [9]. A smooth voice source then has a steeper spectrum slope (resulting in lower alpha ratio). If psychophysiological activity level is high, phonation type is hyperfunctional and consequently, if it is low, phonation type is hypofunctional. In emotional expressions with high psychophysiological activity level, the spectrum of the voice signal tends to be flatter and the glottal volume velocity waveform sharper than in the emotional expressions with low psychophysiological activity level [23, see also ref. 17]. In the latter, the spectrum is more tilting and the waveform is smoother, sometimes almost sinusoidal.

The grade of hypo-/hyperfunctionality of the phonation type is reflected in alpha ratio and NAQ. In the present study, voice quality was more hyperfunctional in joy and anger, and more hypofunctional in tenderness and sadness, alpha ratio being lower in tenderness and sadness and NAQ being smaller in joy and anger. Thus, the spectral slope was flatter in a more hyperfunctional and steeper in a more hypofunctional phonation type. The regulation of the vocal apparatus can be presumed to be holistic, especially in emotional expressions. In emotional expressions (here in vowel [a:]) NAQ seemed to have an independent effect, not combined with L_{eq}. This result was in line with earlier findings [3, 15]. However, no gender differences were found in NAQ, which contrasts with earlier findings [26]. Gender differences in NAQ may reflect differences in F_0 *control* mechanisms (not differences in F_0 per se, since NAQ by definition is normalized according to F_0).

Conclusions

(1) Monopitched vowels [a:], [i:] and [u:] differed in their capacity to convey emotions.

(2) In [a:], filter (formant frequencies F1–F4) was related to valence.

(3) Voice source characteristics (reflected in NAQ) appeared to have a role in expression, not merely related to L_{eq}.

(4) In vowels [i:] and [u:] L_{eq} was the only statistically significant variable in emotional expressions. Thus, either voice source and filter characteristics are used differently in different vowels or due to differences in the vocal tract setting in different vowels the same phonatory or articulatory characteristics have different acoustic consequences. It should be noted, however, that the data sample was smaller for [u:] in females.

(5) The perceptual effects of the interplay between voice source and formant frequencies in different vowels warrant synthetic study.

Acknowledgments

This study was supported by the Academy of Finland (grants No. 200807, No. 200859 and No. 200997). The authors would like to thank Hanna-Mari Pasanen, MSc, of the Unit for Science, Technology and Innovation Studies (TaSTI), University of Tampere, for statistical analysis.

References

1 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.
2 Murray IR, Arnott JL: Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. J Acoust Soc Am 1993;93:1097–1108.
3 Laukkanen AM, Vilkman E, Alku P, Oksanen H: On the perception of emotions in speech: the role of voice quality. Scand J Log Phon Voc 1997;22:157–168.
4 Scherer KR: Vocal communication of emotion: a review of research paradigms. Speech Commun 2003;40:227–256.
5 Gobl C, Ní Chasaide A: The role of voice quality in communicating emotion, mood and attitude. Speech Commun 2003;40:189–212.
6 Campbell N, Mokhtari P: Voice quality: the 4th prosodic dimension. Proc 15th Int Congr Phonet Sci, Barcelona, August 2003, pp 2417–2420.
7 Ladd DR, Silverman KEA, Tolkmitt F, Bergmann G, Scherer KR: Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. J Acoust Soc Am 1985;78:435–444.
8 Laver J: The Phonetic Description of Voice Quality. Cambridge, Cambridge University Press, 1980.
9 Fant G: Acoustic Theory of Speech Production. With Calculations Based on X-Ray Studies of Russian Articulations, ed 2. The Hague, Mouton, 1970.
10 Granström B, Nord L: Ways of exploring speaker characteristics and speaking styles. Proc 12th Int Congr Phonet Sci, Aix-en-Provence, August 1991, pp 278–281.
11 Laukkanen AM: On Speaking Voice Exercises: a Study on the Acoustic and Physiological Effects of Speaking Voice Exercises Applying Manipulation of the Acoustic-Aerodynamic State of the Supraglottic Space and Artificially Modified Auditory Feedback; doct diss Medical School University of Tampere, Tampere, 1995.
12 Alku P, Vilkman E, Laukkanen AM: Estimation of amplitude features of the glottal flow by inverse filtering speech pressure signals. Speech Commun 1998;24:123–132.

13 Sonesson B: On the anatomy and vibratory pattern of the human vocal folds: with special reference to a photo-electrical method for studying the vibratory movements. Acta Otolaryngol Suppl 1960;156:1–80.
14 Sundberg J, Andersson M, Hultqvist C: Effects of subglottal pressure variation on professional baritone singers' voice sources. J Acoust Soc Am 1999;105:1965–1971.
15 Laukkanen AM, Vilkman E, Alku P, Oksanen H: Physical variations related to stress and emotional state: a preliminary study. J Phonet 1996;24:313–335.
16 Waaramaa T, Laukkanen AM, Alku P, Björkner E, Leino T: Perception of emotions in mono-pitched vowels; in Rantala L (ed): Puheopin laitos. Report 5. Department of Speech Communication and Voice Research, University of Tampere, 2007.
17 Alku P, Vintturi J, Vilkman E: The effect of fundamental frequency per se on vocal intensity in soft, normal and loud phonation. Proc 25th World Congr Int Assoc Logop and Phoniatr, Montréal, 2001.
18 Scherer KR: Expression of emotion in voice and music. J Voice 1995;9:235–248.
19 Fant G: Speech Sounds and Features. Cambridge, Massachusetts Institute of Technology, Colonial Press, 1973.
20 Frøkjær-Jensen B, Prytz S: Registration of voice quality. Brüel Kjær Tech Rev 1973;3:3–17.
21 Alku P: Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Commun 1992;11:109–118.
22 Alku P, Bäckström T, Vilkman E: Normalized amplitude quotient for parametrization of the glottal flow. J Acoust Soc Am 2002;112:701–710.
23 Waaramaa T, Alku P, Laukkanen AM: The role of F3 in the vocal expression of emotions. Logoped Phoniatr Vocol 2006;31:153–156.
24 Gauffin J, Sundberg J: Spectral correlates of glottal voice source waveform characteristics. J Speech Hear Res 1989;32:556–565.
25 Sundberg J, Gauffin J: Waveform and spectrum of the glottal voice source. STL-QPSR 1978;2–3:35–50.
26 Airas M, Alku P: Emotions in vowel segments of continuous speech: analysis of the glottal flow using the normalised amplitude quotient. Phonetica 2006;63:26–46.