Journal of Nonverbal Behavior

10.1007/s10919-008-0055-9

Original paper

In a Nervous Voice: Acoustic Analysis and Perception of Anxiety in Social Phobics' Speech

Petri Laukka[1] ✉, Clas Linnman[1], Fredrik Åhs[1], Anna Pissiota[1], Örjan Frans[1], Vanda Faria[1], Åsa Michelgård[2], Lieuwe Appel[3], Mats Fredrikson[1] and Tomas Furmark[1]

(1) Department of Psychology, Uppsala University, P.O. Box 1225, 751 42 Uppsala, Sweden

(2) Department of Neuroscience, Psychiatry, Uppsala University, Uppsala, Sweden

(3) Uppsala Imanet AB, Uppsala, Sweden

✉ **Petri Laukka**

Email: Petri.Laukka@psyk.uu.se

Received: 14 November 2007 **Accepted:** 27 June 2008 **Published online:** 18 July 2008

Abstract

This study investigated the effects of anxiety on nonverbal aspects of speech using data collected in the framework of a large study of social phobia treatment. The speech of social phobics (*N* = 71) was recorded during an anxiogenic public speaking task both before and after treatment. The speech samples were analyzed with respect to various acoustic parameters related to pitch, loudness, voice quality, and temporal aspects of speech. The samples were further content-masked by low-pass filtering (which obscures the linguistic content of the speech but preserves nonverbal affective cues) and subjected to listening tests. Results showed that a decrease in experienced state anxiety after treatment was accompanied by corresponding decreases in (a) several acoustic parameters (i.e., mean and maximum voice pitch, high-frequency components in the energy spectrum, and proportion of silent pauses), and (b) listeners' perceived level of nervousness. Both speakers' self-ratings of state anxiety and listeners' ratings of perceived nervousness were further correlated with similar acoustic parameters. The results complement earlier studies on vocal affect expression which have been conducted on posed, rather than authentic, emotional speech.

Keywords Anxiety - Fear - Vocal expression - Social phobia - Speech

---

Introduction

Emotional communication is of major importance for social competence and the regulation of social behavior. For example, it is possible for us to communicate vital information to others by expressing emotions and we can infer others' probable behaviors by recognizing their expressions. Emotion expressions may also regulate social behavior by inducing affective responses in the perceiver (e.g., Darwin 1872/1998; Ekman 2003; Russell et al. 2003). Among the nonverbal cues that people use for inferring others' emotional states, voice characteristics are frequently reported (Planalp et al. 1996; Scherer and Ceschi 2000). Thus, it is not surprising that nonverbal communication of affect through the voice (i.e., vocal expression) has been reported to have an impact on many social phenomena ranging from the quality of close relationships (Koerner and Fitzpatrick 2002) to patients' satisfaction with medical care (Haskard et al. 2008) and communication effectiveness (Biersack and Kempe 2005).

Interest in vocal expression has increased considerably in recent years (see Cowie et al. 2001), and recent reviews have shown that vocal emotion expressions of specific emotions (e.g., anger, fear, happiness, sadness) are generally recognized with accuracy above chance, also cross-culturally, and are associated with relatively distinct acoustic characteristics (Juslin and Laukka 2003; Laukka 2008). However, one problem with previous research on vocal expressions, and emotion expression in general, is that it has mostly been conducted on posed expressions. In a typical study, actors are asked to portray expressions and listeners are then asked to identify the portrayed expressions in recognition tests. This line of research has produced important findings, indeed most of what we know about vocal expressions is obtained with posed expressions, but the use of posed expressions also has certain limitations. For example, it has been noted that although posed expressions must be relatively similar to ‘real’ expressions so that communication can be successful (Davitz *1964*), they may also be exaggerated and more intense than authentic everyday expressions (Scherer *1986*). There is also a possibility that some aspects of the voice which convey affective information are not under voluntary control (Bachorowski and Owren *1995*). It has further been noted that the prototypical expressions of basic emotions, which have been the focus of a majority of studies of posed expressions, are rarely encountered in spontaneous everyday conversations (e.g., Cowie and Cornelius *2003*; Devillers et al. *2005*). The investigation of authentic expressions of affective states other than basic emotions is thus an important endeavor in order to increase the ecological validity of research on vocal expression.

Studying Authentic Vocal Expressions

Researchers have attempted to study authentic vocal expressions in a number of ways. For instance, various affect induction methods have been applied in order to study the effects of the manipulation on the voice (e.g., Aubergé et al. *2006*; Bachorowski and Owren *1995*; Barrett and Paus *2002*; Bonner *1943*; Johnstone et al. *2005*). Another line of research has investigated spontaneous emotional speech from real-life conversations (e.g., Devillers et al. *2005*; Eldred and Price *1958*; Greasley et al. *2000*; Forsell et al. *2007*; Lee and Narayanan *2005*; Litman and Forbes-Riley *2006*). These kinds of investigations are valuable, but also have limitations. For one thing, it is difficult to induce strong and well-differentiated emotional reactions in laboratory settings, which makes the study of intense emotional states difficult. Further, the study of real-life conversations is made complicated by the fact that one rarely has any control over what emotions, if any, the speakers actually were experiencing.

Another complication is that authentic emotion expressions are shaped to a certain degree by both *push* and *pull* effects (Scherer 1989; see also Ekman and Friesen 1969). Push effects involve various internal processes of the organism that are influenced by the emotional response (e.g., activation of the autonomic nervous system). Pull effects, on the other hand, involve external conditions such as social norms and cultural display rules. Pull effects may lead to strategically posed expressions (i.e., an emotional expression without a corresponding subjective emotional feeling), and also to the opposite phenomenon, namely masked expressions, where a subjective emotional feeling is present but the expression is hidden. Thus, even authentic expressions may often be partly posed. Ideally, one would like to control for the presence of both push and pull effects in studies of emotion expression, and thus be able to disentangle how the resultant nonverbal behavior is influenced by physiological emotional reactions on the one hand and consciously applied expression strategies on the other.

Contribution of the Current Study

What is evident from the above brief review is that no one method of studying vocal expressions is a panacea. The present study fills a gap in the literature by presenting an experimental study on vocal expression where relatively intense affect was induced in a laboratory setting. In short, the speech of patients with social phobia was recorded in an anxiety provoking situation (i.e., giving a public speech) both before and after pharmacological treatment, using data originally collected in the framework of a large ongoing project on treatment for social phobia (e.g., Furmark et al. 2005). The speech samples were also analyzed regarding a number of acoustic parameters known from previous research to be involved in the communication of affect (see Juslin and Laukka 2003; Scherer 1986, for reviews), and all speech samples were evaluated by listeners regarding perceived anxiety. Some patients (responders) responded to the treatment and reported significantly decreased anxiety during the public speaking task from pre- to post-treatment, whereas other patients (non-responders) did not report any decrease in anxiety and served as a control group. This design allowed us to investigate the effects of the anxiety manipulation on both the nonverbal aspects of speech (as indexed by the various acoustic parameters) and the listeners' perception of affect in the speech samples. In other words, the design allowed us to closely investigate the associations between experienced, expressed, and perceived affect.

Previous Studies on the Effects of Anxiety on Speech

Anxiety occurs when a person experiences a situation as personally threatening, either physically or psychologically, which triggers physiological responses and various coping strategies. It is not usually considered as a basic emotion, but rather a combination of one or more negative affects, including fear, uncertainty, distress, apprehension, and worry (e.g., Lang *1985*). Because anxiety most often involves fear, it is considered to belong to the fear family of emotions. However, anxiety typically occurs more frequently and for longer periods of time in everyday life than fear itself. Also, fear stimulates avoidance and escape, while anxiety appears when the threat is unavoidable (Lazarus *1991*). A further distinction is often made between state and trait anxiety, where *state anxiety* is considered an emotional response to a personally threatening situation and *trait anxiety* reflects the existence of stable individual differences in the tendency to respond with state anxiety in the anticipation of threatening situations (e.g., Spielberger et al. *1983*). It should be noted that in the present study we focus on the effects of state anxiety, rather than trait anxiety, on vocal expression.

A couple of previous studies have examined the effects of induced state anxiety on the nonverbal aspects of speech (see Siegman *1987*, for a review). A majority of these studies have focused on temporal aspects of speech, and the most frequent finding has been that anxiety is associated with dysfluencies of speech (e.g., increased pausing; Eldred and Price *1958*; Hofmann et al. *1997*; Kasl and Mahl *1965*; Lewin et al. *1996*; Mahl *1956*; Pope et al. *1970*). A few studies have also looked at effects of induced anxiety on a wider range of vocal measures, but evidence is in general scarce and conflicting results are common for measures other than dysfluency. For example, Hagenaars and van Minnen (*2005*) found that anxiety was associated with lowered pitch variability, whereas Fuller et al. (*1992*) found anxiety to be associated with increased jitter (i.e., pitch perturbations), but no other pitch measures (see also Smith 1977). The induction methods used in the above studies (e.g., manipulation of interview topic, autobiographical memories, or anticipation of exams) allow only induction of relatively mild anxiety that in many cases may not be separable from general arousal or stress (but see Lewin et al. 1996, for an example of stronger induced anxiety). In contrast, the present study investigated the effects of relatively strong anxiety and also considered acoustic parameters (i.e., *voice cues*) related to each of the four main dimensions of vocal expression: pitch, loudness, voice quality, and temporal aspects of speech (see Table 1).

Table 1 Description of the voice cues included in the current study, and hypothesized directions of effects of anxiety and fear based on previous research

Voice cue	**Description**	**Subjective experience**	**Effect of anxiety**[a]	**Effect of fear**[b]
Pitch cues				
F0 M	Mean fundamental frequency	Pitch	na	+
F0 SD	Fundamental frequency variability (standard deviation)	Pitch variability	na	−
F0 max	Maximum fundamental frequency	Pitch	na	+
Intensity cues				
Intensity M	Mean voice intensity	Loudness of speech	na	±
Voice quality cues				
HF 500	Relative proportion of spectral energy above, versus below,	Voice quality (much high-frequency energy corresponds to a sharp	na	±

Voice cue	Description	Subjective experience	Effect of anxiety[a]	Effect of fear[b]
	500 Hz	voice quality)		
Temporal cues				
Speech rate	Number of syllables divided by the total duration of the speech	Velocity of speech	na	+
% Silence	Ratio of silent parts and the total duration of speech (incl. silent parts)	Dysfluencies of speech	+	na

Note. [a] Based on results from Siegman (*1987*). [b] Based on results from Juslin and Laukka (*2003*)

(+) indicates an increase of respective voice cues; (−) indicates a decrease of respective voice cues; (±) indicates predictions in opposing directions; (na = not applicable) indicates that no predictions can be made based on previous research

As noted previously, anxiety is in many ways similar to fear, and it may be difficult to draw exact boundaries between the two affects. Therefore, we show a summary of relevant results from studies of expression of both fear and anxiety in Table 1. While the only reliable prediction that could be made from studies of induced anxiety (as reviewed in Siegman *1987*) was that an increase in anxiety should be associated with an increase in dysfluency, fear has been reliably associated with a wider range of voice cues. The results for fear in Table 1 are taken from a review of 104 studies of vocal expression (Juslin and Laukka *2003*), and are thus based on a much larger material than the comparable results for anxiety. The majority of studies on the vocal expression of fear have been conducted on posed expressions, though a couple of studies on authentic fear expressions do exist (e.g., Devillers et al. *2005*; Kuroda et al. *1976*; Williams and Stevens *1969*). However, Williams and Stevens (*1972*) compared posed and authentic fear expressions and reported that the most marked difference was that larger effects were found for the portrayals. For the above reasons we will compare the present results also with the literature on (mainly posed) vocal expression of fear.

A recent meta-analysis of the literature on recognition of anxiety (Harrigan et al. 2004) reported that state anxiety was better communicated through the voice than through other nonverbal channels. However, the studies included in the meta-analysis did not separate the effects of the verbal (what is said) and nonverbal (how it is said) aspects of vocal communication. To address this issue, in the present study we investigated the recognition of state anxiety through strictly nonverbal aspects of the voice. This was achieved by content-masking the speech stimuli through low-pass filtering (which obscures the linguistic content of the speech but preserves nonverbal affective cues) before using them in listening tests (see Scherer et al. 1972). Previous reports have shown that anxiety can be perceived from content-masked speech (e.g., McNally et al. 2001), but no previous study has attempted to relate listeners perceptions of anxiety from content-filtered speech with acoustic measurements of the voice.

Research Questions and Hypotheses

The following specific research questions were addressed in the present study:

(1) Does state anxiety have an effect on the nonverbal aspects of speech, as measured by a variety of acoustic parameters?

(2) Can experienced state anxiety be perceived from the nonverbal aspects of speech?

(3) What are the associations between experienced, expressed, and perceived state anxiety?

Based on previous research (see Table 1) we made the following specific predictions for the voice cues under study. A decrease in state anxiety should be associated with (a) a decrease in mean and maximum pitch and an increase in pitch variability, and (b) a decrease in speech rate and proportion of pauses. No specific predictions could be made for the voice quality or voice intensity cues because previous studies on expression of fear show opposing trends, perhaps because different studies have investigated fear of different emotion intensities (e.g., mild fear or panic fear, see Juslin and Laukka *2001*, *2003*).

Method

Collection of the Emotional Speech Stimuli

Participants

The data were originally collected in the framework of a large ongoing project concerning treatment for social phobia conducted at Uppsala University (e.g., Furmark et al. *2002*; Furmark et al. *2005*). For the purposes of the present study we selected 71 patients (from a total of 108 patients) who either clearly responded, or clearly did not respond, to a pharmacological anxiolytic treatment for social phobia. Twenty-four patients (12 male, 12 female, mean age = 32 years) originally participated in the study by Furmark et al. (*2005*), and 47 patients (14 male, 33 female, mean age = 36 years) were taken from an as yet unpublished study by the same group of researchers.

All participants met the DSM-IV criteria for social phobia (social anxiety disorder), as revealed by structured clinical interviews (Structured Clinical Interview for DSM-IV; First et al. *1998*) administered by a clinical psychologist, and exhibited marked public speaking anxiety. According to the DSM-IV, social phobia is defined as ‘a marked and persistent fear of one or more social or performance situations in which the person is exposed to unfamiliar people or possible scrutiny by others’ (American Psychiatric Association *1994*, p. 416). Participants were recruited through newspaper advertising and a written consent was obtained from each participant. We refer to Furmark et al. (*2005*) for a complete description of the screening procedure and inclusion and exclusion criteria.

Treatment Procedure

The patients received pharmacological treatment consisting of either SSRI (selective serotonin reuptake inhibitors) or non-SSRI drugs or a placebo treatment. They were randomly allocated to treatment or control (placebo) groups in a double-blind fashion (see Furmark et al. 2005 for details). Study drugs were ingested orally on a daily basis. The study period was 6 weeks for the patients from Furmark et al. (2005) and 8 weeks for the unpublished sample. All patients visited the clinic weekly for assessment of complications and to receive new supplies of medication. The studies were approved by the Uppsala University Medical Faculty Ethical Review Board and the Swedish Medical Products Agency.

Recording of Speech Samples

The speech of the patients was recorded immediately before the first treatment (baseline) and 2–4 hr after the final treatment (post-trial). The recordings were made while the patients took part in a PET assessment and lay in the scanner. The patients were instructed to prepare a speech about a vacation or travel experience about 20 min before the scanning began. When the scanning began, the patients were asked to start speaking and to continue for approximately 2 min until they received instructions to stop. The speech was performed before an audience of six to eight persons who silently observed the patient's performance. One of the observers also had a video camera pointed at the speaker and recorded the speech (see below), and the patients were further instructed to watch the audience in order to increase observational anxiety. We chose a public speaking task as our method of anxiety provocation because speaking in public is the most prevalent social fear both in individuals with social phobia and in the community at large (Furmark et al. 1999; Kessler et al. 1998). Also, we have previously shown that individuals with social phobia experience relatively high anxiety in this type of public speaking tasks, as indicated by both self-reports and measurements of cerebral blood flow (e.g., Tillfors et al. 2001).

For the 24 patients from Furmark et al. (*2005*), the speech was recorded from close distance with a Sony CCD-F450E portable video camera (Sony Corp., Tokyo) using the built-in microphone. For the 47 patients from the unpublished sample, the speech was recorded with a Panasonic NV-GS50 digital portable video camera (Matsushita Electric Industrial Co., Ltd, Osaka, Japan) using a Sony ECM-MS908C stationary external microphone (Sony Corp., Tokyo) located at a distance of 30 cm from the patient's mouth.

Assessment of Treatment

Response to treatment was determined by the Clinical Global Impression improvement item (CGI-I) (Zaider et al. *2003*) administered by an experienced psychiatrist at baseline and post-trial. Patients with a score of 1 or 2 (i.e., very much or much improved) on the CGI-I at follow-up were classified as responders, whereas patients with a score of 4 (no change) or higher were considered as non-responders. Patients with a score of 3 (minimally improved) were excluded from this study, because we wanted to concentrate on patients who clearly responded, or clearly did not respond, to the treatment. Forty-one speakers (24 female, 17 male, mean age = 36 years) were classified as responders and 30 speakers (21 female, 9 male, mean age = 33 years) as non-responders. The different treatments were approximately equally distributed across responders (10 SSRI, 25 non-SSRI, 6 placebo) and non-responders (6 SSRI, 14 non-SSRI, 10 placebo).

Self-reports of State Anxiety

The state scale of the Spielberger State-Trait Anxiety Inventory (STAI-S; Spielberger et al. *1983*) was administered directly after completion of the speech task at baseline and post-trial to estimate retrospectively how anxious the patients had felt during the public speaking task. The STAI-S scale consists of 20 self-descriptive items and measures how the subject feels at a particular moment (e.g., 'I feel nervous') rated on a scale from 1 to 4 (*not at all, somewhat, moderately so, very much so*). The essential qualities evaluated by the STAI-S scale are feelings of apprehension, tension, nervousness, and worry. The STAI-S scores can range from 20 to 80 and higher scores indicate higher levels of state anxiety. A mean STAI-S score of 35 is considered normative for adults (Spielberger et al. *1983*), and the literature suggests defining high anxiety at 1 standard deviation above the normative mean (i.e., STAI-S > 45).

Acoustic Measures

The speech stimuli used in this study consisted of the first 10 s of speech recorded at baseline and post-trial for each of the 71 patients. A wide range of studies has shown that 10 s is long enough to allow accurate social and interpersonal perception from expressive behavior, including nonverbal aspects of speech (Ambady and Rosenthal *1992*). Acoustic voice cues that can be measured from the speech signal can be broadly divided into those related to (a) fundamental frequency (F0), (b) voice intensity, (c) voice quality, and (d) temporal aspects of speech. We measured voice cues related to all of these aspects of the voice. However, because the recordings were not conducted with the primary aim of speech analysis, the sound quality was not optimal with a constant level of background noise (e.g., from the PET scanner). Therefore, it was not possible to measure voice cues that are susceptible to measurement error due to background noise (e.g., jitter, shimmer, formant frequencies; see Perry et al. *2000*). All measurements were conducted using the speech analysis software PRAAT (Boersma and Weenink *2007*). Below we detail how 7 voice cues were measured (F0 M, F0 max, F0 SD, Intensity M, HF500, Speech rate, and % Silence).

The fundamental frequency (*F0*) of the voice represents the rate with which the vocal folds open and close across the glottis during phonation, and is strongly related to our auditory impression of voice pitch. The mean (*F0 M*), maximum (*F0 max*), and standard deviation (*F0 SD*) of F0 were extracted using an autocorrelation algorithm. Detection errors (including octave jumps and detection of periodicity in unvoiced speech segments or background noise) were manually corrected; but no correction was applied when the algorithm failed to detect periodicity in voiced segments (e.g., because of irregular phonation).

Voice intensity reflects the effort required to produce the speech and is subjectively heard as the loudness of the voice. The mean voice intensity was measured as the mean level contour in dB (*Intensity M*). This index does not have an absolute meaning, but allows for comparisons within the same speaker. Mean voice intensity was only measured for the patients from the unpublished sample, where recording level and distance from mouth to microphone was kept constant across speakers and conditions.

One index of voice quality is provided by the relative proportion of total acoustic energy above versus below a certain cut-off frequency (Scherer et al. 1991). As the proportion of high frequency energy in the spectrum increases, the voice sounds sharper and less soft. We calculated the long-term average spectrum using a Cooley-Tukey integer algorithm with a frequency band of 0–5 kHz. The amount of high-frequency energy in the spectrum was measured as the relative proportion of energy found above versus below a cut-off frequency of 500 Hz (*HF500*; see Juslin and Laukka 2001; Laukka et al. 2005). Like voice intensity, HF500 was only measured for the patients from the unpublished sample and only for those parts of the samples that included phonation.

Speech rate refers to the velocity of the speech. A measure of speech rate was obtained by counting the number of syllables from a literal transcription of the speech, and dividing this number with the total duration of the speech produced (excluding silent periods and pauses). An additional temporal measure was obtained by calculating the ratio of silent parts and the total duration of the speech sample including the silent parts (*% Silence*). A period was defined as silent if no speech activity could be observed from either the amplitude envelope or spectrogram.

Listening Test

Participants

A total of 16 listeners aged between 20 and 33 years took part in the listening test (mean age = 24 years, 8 women). All participants were Swedish speaking and they received movie tickets in return for their confidential and voluntary participation.

Speech Stimuli

Before being entered into the listening test, the speech stimuli were content-masked by low-pass filtering. All frequencies above 500 Hz were removed using the Praat software (Hann-shaped filter). This procedure eliminates phonetic information and renders the speech unintelligible and sounding muffled. Nevertheless, affective information transmitted by F0, voice intensity and temporal aspects of speech are largely preserved (e.g., Scherer et al. 1972). Also some information related to voice quality seems to be spared. It has been reported that laryngeal voice quality can be perceived from low-pass filtered speech (van Bezooijen and Boves 1986). The filtering procedure served double purposes in the present study: first, it helped to preserve the integrity of the patients because they could not be identified by the listeners, and second, it forced the listeners to focus on the nonverbal content, as opposed to the linguistic content, of the speech.

Design and Procedure

All listeners judged the nervousness (‘in your opinion, how nervous does the speaker sound?’) of all speech stimuli (N = 142) on a scale from 0 (*not nervous at all*) to 10 (*very nervous*). In order to avoid misunderstandings, ‘nervousness’ was carefully explained to the listeners as feelings of apprehension, fear, nervousness, tension, and worry. These words were chosen because they are identical to the essential qualities evaluated by the STAI-S scale that was used to assess the speakers’ level of experienced anxiety (Spielberger et al. 1983). Listening experiments were conducted individually using custom computer software in a room with dampened acoustics, and the participants listened to the stimuli through Beyerdynamic DT 770 Pro headphones (Beyerdynamic GmbH & Co.KG, Heilbronn, Germany). The presentation order of the stimuli was randomized for each listener. There were no time constraints and the listeners were allowed to listen to each stimulus as many times as necessary to make a judgment.

Results

Validation of the Anxiety Manipulation

To begin with, we wanted to ascertain that the treatment had an effect on patients' self-reports of anxiety in our sample. Therefore we conducted a mixed model analysis of variance (ANOVA) on the patients' STAI-S scores, with group (two levels: responder and non-responder) as a between subjects factor and treatment (two levels: baseline and post-trial) as a within-subjects factor. The analyses revealed that significant main effects of both group ($F_{1, 69} = 6.84$, partial $\eta^2 = 0.09$, $p < 0.05$) and treatment ($F_{1, 69} = 104.86$, partial $\eta^2 = 0.60$, $p < 0.001$) were qualified by a significant interaction effect ($F_{1, 69} = 34.15$, partial $\eta^2 = 0.33$, $p < 0.001$). The patients' STAI-S scores as a function of group and treatment are shown in Fig. 1. As can be seen, the interaction effect indicated that STAI-S scores decreased significantly more from baseline to post-trial for responders than for non-responders. *Post hoc* multiple comparisons (Fisher's *LSD*) further revealed that the group difference between responders and non-responders was significant at post-trial but not at baseline ($p < 0.001$), though the decrease from baseline to post-trial was significant for both responders ($p < 0.0001$) and non-responders ($p < 0.01$). The patients' self-reports of anxiety thus confirmed that the responder group did experience significantly less anxiety after, but not before, treatment than did the non-responder group.

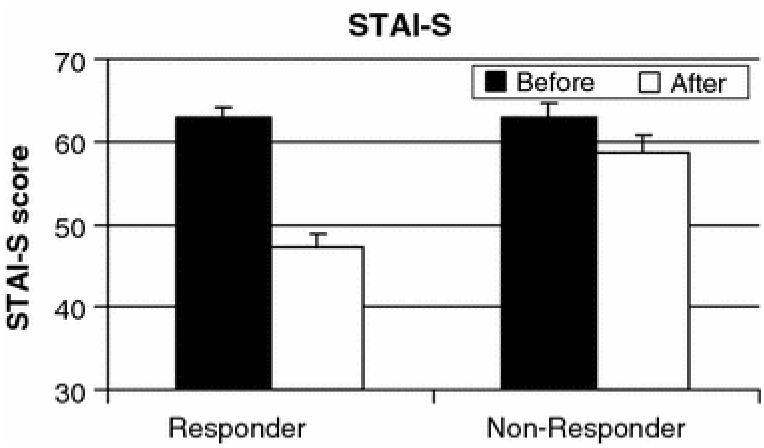

Fig. 1 Experienced anxiety (STAI-S) as a function of treatment for both responders and non-responders. Error bars represent the standard error

The STAI-S scores also confirmed that the anxiety manipulation resulted in relatively intense anxiety. As shown in Fig. 1, the mean STAI-S scores were above the suggested threshold for high anxiety (STAI-S > 45; see Spielberger et al. *1983*) for both responders and non-responders at both measurement occasions. The responders' mean STAI-S score at post-trial was not significantly above the threshold for high anxiety as evidenced by the 95% confidence limits of the mean (STAI-S = 47.17, ±3.27), but indicates that also patients who responded to the treatment reported moderately high anxiety during the public speaking task.

Effects of Anxiety on Acoustic Measures

Next we investigated if the treatment had an effect on the nonverbal aspects of the patients' speech as indexed by the voice cues. Separate ANOVAs (mixed model) were conducted for each voice cue with group (2 levels: responder and non-responder) as a between subjects factor and treatment (2 levels: baseline and post-trial) as a repeated measure. The results of the ANOVAs are shown in Table 2. Because we wanted to assess if the nonverbal behavior of responders and non-responders would show different patterns before and after treatment, we were mainly interested in studying the interaction between group and treatment. As shown in Table 2, significant interaction effects coupled with at least one significant main effect were found for F0 M, F0 max, and % Silence, whereas for HF500 only the interaction effect was significant. The mean values of these voice cues, across all patients, are shown in Fig. 2 as a function of group and treatment. Because men and women differ with regard to voice pitch, and because the responder and non-responder groups included different numbers of men and women, the raw values of the pitch cues (F0 M, F0 SD, and F0 max) were *z*-transformed separately for men and women, respectively, before inclusion in the analyses. However, in Fig. 2 we show the raw values in order to facilitate the interpretation of the data.

Table 2 Summary of results from analyses of variance of voice cues

Voice cue	*df*	Group (G)		Treatment (T)		G × T	
		F	Partial η^2	*F*	Partial η^2	*F*	Partial η^2
F0 M	1, 69	0.57 ns	0.01	30.78***	0.31	7.74**	0.10
F0 SD	1, 69	1.03 ns	0.01	0.30 ns	0.00	0.95 ns	0.01
F0 max	1, 69	2.47 ns	0.04	10.02**	0.13	6.34*	0.08
Intensity M	1, 45	2.39 ns	0.05	6.20*	0.12	2.20 ns	0.05
HF 500	1, 45	0.12 ns	0.00	1.21 ns	0.03	6.92*	0.13
Speech rate	1, 69	0.19 ns	0.00	0.40 ns	0.01	0.34 ns	0.00
% Silence	1, 69	4.28*	0.06	4.13*	0.06	18.70***	0.21

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

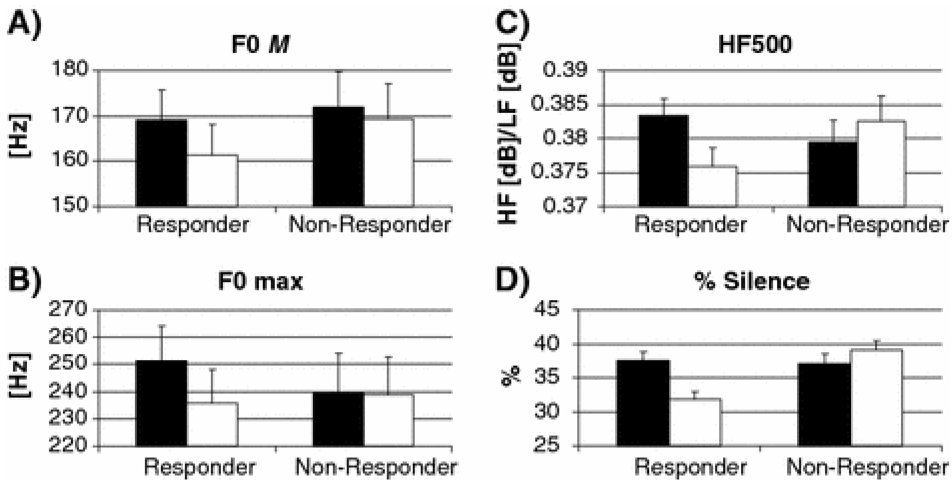

Fig. 2 Acoustic measures that showed changes as a function of group and treatment. (**a**) Mean fundamental frequency (F0 M), (**b**) maximum fundamental frequency (F0 max), (**c**) proportion of spectral energy above versus below 500 Hz (HF500), (**d**) proportion of silent pauses (% Silence); solid bars = baseline; open bars = post-trial. Error bars represent the standard error

---

As can be seen in Fig. 2, the interaction effects indicated that responders decreased significantly more than non-responders from baseline to post-trial for F0 M and F0 max, whereas for HF500 and % Silence responders decreased while non-responders showed a small increase from baseline to post-trial. *Post hoc* multiple comparisons (Fisher's *LSD*) revealed that the changes from baseline to post-trial were significant for responders, but not for non-responders, for F0 M, F0 max, and % Silence (p's < 0.0001), as well as for HF500 ($p < 0.01$). To summarize, the results indicate that a change in experienced state anxiety was associated with changes in several nonverbal aspects of speech for patients who responded to the treatment. The changes were in the hypothesized direction for F0 M, F0 max and % Silence, but our predictions regarding F0 SD and Speech rate could not be confirmed by the data.

Effects of Anxiety on Perception of Nervousness

First we wished to establish if the listeners were able to rate the nervousness of the speech samples consistently. The internal consistency of ratings among the judges, as measured by Cronbach's alpha, was high for the nervousness scale (alpha = 0.91) indicating that the listeners' nervousness ratings were internally consistent.

The effect of anxiety on the listeners' perception of nervousness in the patients' voices was then investigated by conducting a mixed model ANOVA on the mean nervousness ratings across all listeners. Group (2 levels: responder and non-responder) served as a between subjects factor and treatment (2 levels: baseline and post-trial) as a within subjects factor. The ANOVA revealed a significant main effect of treatment ($F_{1, 69} = 5.43$, partial $\eta^2 = 0.07$, $p < 0.05$) and a significant Group X Treatment interaction ($F_{1, 69} = 4.17$, partial $\eta^2 = 0.06$, $p < 0.05$). The listeners' mean ratings of nervousness are shown in Fig. 3 as a function of group and treatment. As shown, the listeners rated the responders' speech to be less nervous after treatment than before treatment, while the speech of non-responders was judged to be approximately equally nervous at both baseline and post-trial. *Post hoc* tests (Fisher's *LSD*) revealed that the decrease from baseline to post-trial was significant for responders ($p < 0.01$), but not for non-responders. Also, the group difference between responders and non-responders was significant at post-trial ($p < 0.05$), but not at baseline. These results indicate that a change in experienced anxiety was accompanied by perceptible changes in nonverbal expressive behavior for patients who responded to the treatment.

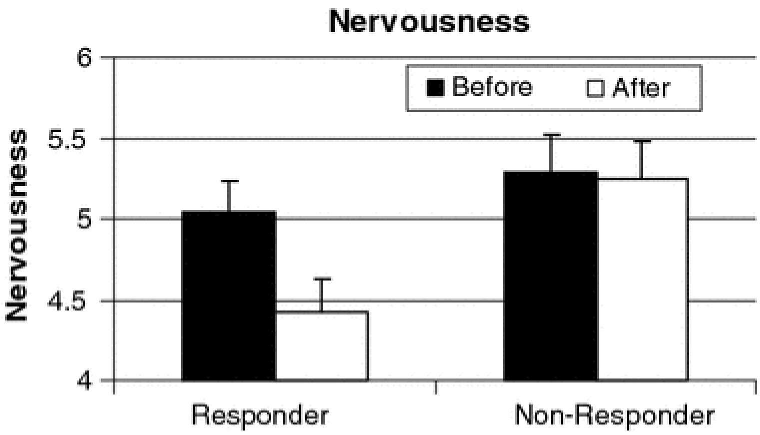

Fig. 3 Listeners ratings of nervousness (0 = min, 10 = max) as a function of treatment for responders and non-responders. Error bars represent the standard error

---

Correlations Between Acoustic Measures and Self-reports of Anxiety and Listeners' Perception of Nervousness

In the final stage of the analyses, we wanted to relate the acoustic measurements with the patients' self-reports of anxiety and the listeners' ratings of nervousness. First, we computed the correlations (Pearson *r*) between voice cues and patients' STAI-S scores and listeners' mean nervousness ratings, see Table 3. The correlations were computed for the post-trial values only, because the variance at baseline was deemed too constrained for the analyses to be valuable. Also, because men's and women's voices differ with regard to pitch level, the raw values of the pitch cues (F0 M, F0 SD, and F0 max) were separately normalized for men and women using the *z*-transformation before inclusion in the analyses.

Table 3 Correlations (Pearson *r*) between voice cues and (a) speakers' self-reports of anxiety (STAI-S) and (b) listeners' ratings of nervousness at post-trial

Voice cue	*N*	STAI-S	Nervousness
F0 M	71	−0.09 ns	0.33**
F0 SD	71	−0.24*	−0.12 ns
F0 max	71	−0.18 ns	0.13 ns
Intensity M	47	−0.20 ns	−0.59***
HF 500	47	0.21 ns	0.15 ns
Speech rate	71	0.08 ns	0.05 ns
% Silence	71	0.36**	0.34**

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As shown in Table 3, both STAI-S scores and nervousness ratings were positively correlated with % Silence. Nervousness ratings were further positively correlated with F0 M and negatively correlated with Intensity M, and STAI-S scores were negatively correlated with F0 SD. These results suggest that listeners utilized pitch and intensity as well as temporal cues in judging the nervousness of the speakers. However, though we previously showed that changes in experienced state anxiety were associated with changes in several voice cues for responders, these changes were largely not reflected in the correlations between STAI-S scores and voice cues. One possible explanation of this finding is that, though we tried to control for individual differences in pitch due to speaker gender in the analyses, we did not control for effects of possible individual differences in other variables, like STAIS-S scores or other voice cues. Therefore the results in Table 3 need to be complemented by analyses where individual differences with regard to the baseline values of the variables of interest are controlled for by, for example, conducting within-person comparisons where each speaker acts as his or her own control.

To that end we computed the correlations (Pearson *r*) between the changes (from pre- to post-treatment) in voice cues and the changes (from pre- to post-treatment) in both patients' STAI-S scores and listeners' mean ratings of nervousness (see Table 4). To calculate the change scores for the voice cues, STAI-S scores, and mean nervousness ratings, we subtracted the post-trial value for each variable from the baseline value for the same variable for each patient. A large positive change score thus indicated a large decrease from baseline to post-trial. In the following, change scores will be denoted by the Greek letter delta (Δ).

Table 4 Correlations (Pearson *r*) between changes (from baseline to post-trial) in voice cues and changes (from baseline to post-trial) in (a) speakers' self-reports of anxiety (STAI-S) and (b) listeners' ratings of nervousness

Voice cue	*N*	ΔSTAI-S	ΔNervousness
ΔF0 M	71	0.25*	0.16 ns
ΔF0 SD	71	−0.14 ns	−0.06 ns
ΔF0 max	71	0.26*	0.24*
ΔIntensity M	47	−0.21 ns	−0.35*
ΔHF 500	47	0.29*	0.24 ns
ΔSpeech rate	71	0.18 ns	0.02 ns
Δ% Silence	71	0.37***	0.29*

Note. * $p < 0.05$, *** $p < 0.001$

As can be seen from Table 4, changes in several voice cues were significantly correlated with changes in both experienced anxiety and perceived nervousness. A decrease in ΔF0 max and Δ% Silence was associated with decreases in both ΔSTAI-S and ΔNervousness. The positive correlation between ΔHF500 and ΔSTAI-S was also significant, whereas the correlation between ΔHF500 and ΔNervousness did not reach statistical significance ($p = 0.09$). Further, the positive correlations with ΔF0 M only reached significance for ΔSTAI-S but not for ΔNervousness. Also, the negative correlation between ΔIntensity M was only significant for ΔNervousness but not for ΔSTAI-S.

A comparison of Tables 3 and 4 reveals that stronger correlations between STAI-S scores and voice cues were obtained when individual differences were controlled for (except for F0 SD), suggesting that individual differences in baseline levels may have added variability to the analyses conducted on post-trial values only. However, the associations between nervousness ratings and voice cues were instead overall slightly lower for the change-score correlations than for the post-trial correlations (except for F0 max and HF500). Because the listeners rated the speech samples in random order and were unaware of which stimuli belonged to which speaker, they had to apply the same judgment strategy across all speech stimuli regardless of different speakers' individual baseline levels for the various voice cues. Therefore, the post-trial correlations may have better captured the way the listeners made their judgments.

Taken together, the above analyses suggest that experienced, expressed, and perceived affect were linked to a certain degree. The change-score correlations between STAI-S and voice cues (see Table 4) largely confirmed the results from the ANOVAs on voice cues in showing that a change in experienced anxiety was associated with corresponding changes in voice cues. Further, both the post-trial and the change-score correlations between nervousness ratings and the voice cues give clues about what cues listeners use when making inferences about speakers' nervousness. Generally, the correlations between voice cues and both STAI-S and perceived nervousness were in the same direction, which indicates that senders and perceivers utilized the same cues in a similar fashion.

Finally, we also calculated the correlation between ΔSTAI-S and ΔNervousness as a measure of strength of association between experienced and perceived anxiety. This correlation was significant and positive ($r = 0.24$, $p < 0.05$), which again suggests a link between experienced and perceived anxiety in our data.

Discussion

In summary, the results first showed that anxiety had an effect on several of the measured voice cues. Specifically, a decrease in experienced anxiety was accompanied by decreases in mean and maximum voice pitch, high-frequency components in the energy spectrum, and proportion of silent pauses. Thus, we could show that anxiety had an impact on the nonverbal aspects of speech, and these effects were in accordance with our hypotheses drawn from previous research. Second, the results from the listening test showed that listeners were able to perceive the speakers' level of anxiety from the nonverbal aspects of speech alone. Third, we found several statistically significant correlations between experienced, expressed, and perceived anxiety.

Effects of Anxiety on Nonverbal Aspects of Speech

Emotions are generally thought to cause changes in physiological responding, in order to increase the likelihood that the organism will successfully resolve adaptive challenges (e.g., Ekman *1992*; Levenson *1994*). Vocal emotion expressions are also generally thought to be partly the result of this physiological response. Based on his component process theory, Scherer (*1986*) made detailed predictions about the patterns of vocal cues associated with different emotions. The predictions were based on the idea that emotions involve sequential cognitive appraisals of stimulus features such as novelty, intrinsic pleasantness, goal significance, coping potential, and norm/self compatibility. The outcome of each specific cognitive appraisal is assumed to have an effect on the physiological responding in ways that influence voice production. For instance, physiological changes associated with anxiety may increase the tautness of laryngeal and vocal fold muscles which increases the pitch of the voice. Though the coupling between physiological emotional responding and resulting changes in the voice is, in theory, well established, there is yet relatively little direct empirical evidence from studies of authentic emotional speech to support it (e.g., Johnstone et al. *2007*).

The effects of anxiety on pitch cues (F0 M and F0 max) were in accord with our hypotheses based on previous research conducted on posed expressions (Juslin and Laukka *2003*), as well as with Scherer's (*1986*) theoretical predictions based on somatic alterations associated with emotional responses. This is consistent with the idea that posed expressions are at least partly based on spontaneous expressions, which in turn are at least partly based on physiological emotional responses. However, the largest effects of anxiety were found for proportion of pauses, with a decrease in anxiety being accompanied by a decrease in proportion of pauses. This voice cue has rarely been investigated in studies on posed expressions because these studies have mainly focused on short phrases (and pauses rarely occur in short utterances). Pausing, along with other measures of speech dysfluency, has more frequently been investigated in studies of spontaneous speech and the present results corroborate earlier findings of increased dysfluency in anxious speech (Siegman *1987*). Pausing in speech is often assumed to reflect cognitive activity (e.g., Zellner *1994*). Thus, the effect of anxiety on pausing indicates that anxiety may also have an impact on the nonverbal aspects of speech through an interference of cognitive activities such as word finding and discourse planning.

A longstanding question in research on vocal emotion has been whether the acoustic changes that result from emotional responses reflect specific emotions or a single dimension of arousal (e.g., Laukka *2005*). Empirical evidence from research on posed expressions strongly suggests that specific emotions are associated with characteristic patterns of voice cues (Juslin and Laukka *2003*). Also, Scherer's (*1986*) appraisal theory based predictions of the physiological changes associated with emotional responses posits similar differentiated vocal response patterns for specific emotions. However, evidence for the existence of specific acoustic characteristics for different emotions, coming from studies of authentic expressions, remains scarce. Unfortunately the present study is not ideally suited to address the question of whether vocal emotion response patterns are best explained by specific-emotions or dimensional accounts. Among the voice cues that were measured in the present study, the two different accounts make different predictions for only pitch variability (F0 SD). More specifically, an increase in anxiety/fear has repeatedly been associated with a decrease in pitch variability, whereas an increase in the activation or arousal level of emotions instead has been associated with increased pitch variability (e.g., Juslin and Laukka *2003*; Laukka *2008*). In our data, we found a significant negative correlation between speakers' self-reports of anxiety and F0 SD at post-trial, indicating that high anxiety was associated with low pitch variability. Thus, our results tentatively support a specific-emotions, rather than an arousal level, interpretation of the effects of anxiety on vocal expression, though obviously this finding needs to be replicated before any wider conclusions are drawn.

The Relationship Between Experienced, Expressed, and Perceived Anxiety

Previous studies on the linkage between experienced and expressed emotion conducted on facial expressions have concluded that subjective feelings and facial behavior often are correlated (see Adelman and Zajonc 1989, for a review), but have also shown that the links are often modest and sometimes inconsistent (e.g., Fernandez-Dols et al. 1997; Gross et al. 2000). The results from the current study paint a similar picture of vocal expressive behavior. This study is the first to show that a reduction in anxiety causes accompanying changes in nonverbal vocal behavior. Thus, the results lend support to the often hypothesized coupling between experienced emotion and expressive behavior. However, the anxiety manipulation had a larger effect on the patients' self-reports of anxiety than on the corresponding nonverbal behavior, as indicated by the sizes of the group and interaction effects from the ANOVAs on STAI-S scores and voice cues, respectively. This speaks against a one-to-one relationship between experienced and expressed emotion. There were also large individual differences among speakers for those voice cues for which significant effects of anxiety were found, and no effects of anxiety were observed for certain voice cues.

One possible explanation for the relatively weak evidence often found for the coupling between experienced emotion and expressive behavior is that possible 'push' effects of the emotional response may be obscured by 'pull' effects (Scherer 1989). In other words, expressors often willfully try to conceal, or mask, their expressive behavior. Suppression of emotion expressions is one of the most common emotion regulation strategies (e.g., Gross 2002), especially for people with excessive social anxiety (e.g., Kashdan and Steger 2006; Turk et al. 2005), and recent studies have shown that suppression can lead to less anxiety expression during evaluated speaking tasks (e.g., Egloff et al. 2006). In the present study, we did not instruct the patients to either show or hide their emotion expressions in the public speaking task, but it is possible that they utilized various emotion regulation strategies in order to keep up a non-nervous appearance in front of the audience. It could be speculated that because the most consistent and largest effects of anxiety were found for proportion of pauses, the indirect effects of emotion-related cognitive interference on vocal nonverbal behavior may be harder to mask than the effects of emotion-related physiological responses (see also Batliner et al. 2003). The possibility of expression suppression clearly presents a challenge for the study of authentic vocal expressions and should be directly addressed in future investigations.

This study is further the first to report correlations between voice cues and the level of nervousness perceived by listeners from the nonverbal aspects of authentic emotional speech. These correlations suggest that listeners utilized a wide variety of cues related to pitch, loudness, and temporal aspects for their nervousness judgments. The correlation between voice quality cues and nervousness did not reach significance, however, which probably reflects the fact that spectral information was not available to the listeners because of the low-pass filtering procedure. While the listeners were able to differentiate responders from non-responders from the nonverbal aspects of the speech, the effects of the anxiety manipulation were smaller for perceived nervousness than for self-reported anxiety. The low-pass filtering, together with the low sound quality of the speech samples, likely attenuated the effects of anxiety found in the listening test. However, it is also possible that the small effect of anxiety on perceived nervousness is the result of the listeners' attempts at masking their vocal anxiety expressions.

Limitations of the Current Study

The small to moderate magnitude of the effects of anxiety on voice cues may in part be due to methodological artifacts of the present study. For example, we did not control the linguistic content of the speech which probably introduced additional variability in the voice cue measurements. Also, the low sound quality of the recordings and the long time lag between the pre- and post-treatment recordings may have added extra variability in the data. Further, we did not obtain recordings of the patients in a non-anxious state. Including a baseline of emotionally neutral speech with which to compare the results from anxious speech would likely have lead to larger effects on the voice cues. As our study stands, the comparison was instead between speech in a state of relatively high anxiety and speech in a state of milder anxiety. Finally, our acoustic analyses only measured a small part of the acoustic variation in the voice signal that may be affected by emotional responding. Therefore it remains a possibility that some voice cues not measured in this study may co-vary with anxiety (e.g., jitter, see Smith *1977*). Nevertheless, in contrast to most previous studies, here we utilized a within-subjects design, where each speaker acted as his or her own control, and induced relatively strong affect in our speakers, which likely increased the power to detect effects of anxiety.

Another limitation of the present study is that all our speakers were social phobics. It can be argued that state anxiety experienced by social phobics is similar to state anxiety experienced by non-phobics, albeit phobics experience anxiety more often and more intensely. Nevertheless, further studies are required to confirm the generalizability of the findings to a non-clinical population. Finally, all speakers received pharmacological treatment (except for those who received placebo treatment) and there was no untreated control group. This raises the issue of possible effects of the pharmacological treatment on the voice. There is no evidence that the anxiolytics used in the present study have direct effects on laryngeal function, but they may exert an indirect influence on the vocal tract through their effect on the autonomic nervous system. For example, because SSRI drugs are weakly anticholinergic they may cause mild drying of the vocal tract mucosa (e.g., Thompson *1995*). However, it is unlikely that the differential voice effects for responders and non-responders were caused by the drugs because the different treatments were approximately equally distributed across responders and non-responders.

Designing an experimental study of authentic affect expressions is fraught with difficulties, especially concerning how to collect the expressions. However, we believe that such studies are necessary to yield a better understanding of emotional vocal production, and about how and when emotion experience and various aspects of vocal expression co-occur. Future studies should try to address some of the limitations of the present study and, for example, use more sensitive acoustic measures together with standardized speech material in order to minimize variability in voice cues caused by factors other than emotion. Ideally, studies should also collect authentic expressions of more than one affective state from each person and try to control the push and pull components of expression. Concerning future steps in the analysis of the present data, the data were collected in the framework of a large study of treatment of social phobia and the recordings were made while the speakers lay in a PET scanner. Thus an obvious next step will be to correlate measures of brain activity with aspects of vocal nonverbal behavior in order to study the neural correlates of emotional speech production.

Acknowledgments This research was supported by the Swedish Research Council and the Ryoichi Sasakawa Young Leaders Fellowship Fund (SYLFF) through grants to the first author.

References

Adelmann, P. K., & Zajonc, R. B. (1989). Facial efference and experience of emotion. *Annual Review of Psychology, 40*, 249–280.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274.

American Psychiatric Association. (1994). *Diagnostic and statistical manual for mental disorders* (4th ed.). Washington, DC: American Psychiatric Press.

Aubergé, V., Audibert, N., & Rilliard, A. (2006). De E-Wiz à C-Clone: Recueil, modélisation et synthèse d'expressions authentiques. *Revue d'Intelligence Artificielle, 20*, 499–527.

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustical properties of speech are associated with emotional intensity and context. *Psychological Science, 6*, 219–224.

Barrett, J., & Paus, T. (2002). Affect-induced changes in speech production. *Experimental Brain Research, 146*, 531–537.

Batliner, A., Fischer, K., Huber, R., Spilker, J., & Nöth, E. (2003). How to find trouble in communication. *Speech Communication, 40*, 117–143.

Biersack, S., & Kempe, V. (2005). Exploring the influence of vocal emotion expression on communicative effectiveness. *Phonetica, 62*, 106–119.

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.6.12) [Computer program]. http://www.praat.org/. Retrieved 27 July 27 2007.

Bonner, M. R. (1943). Changes in the speech pattern under emotional tension. *American Journal of Psychology, 56*, 262–273.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40*, 5–32.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., et al. (2001). Emotion recognition in human–computer interaction. *IEEE Signal Processing Magazine, 18*(1), 32–80.

Darwin, C. (1998). *The expression of the emotions in man and animals* (with introduction, afterword, and commentaries by P. Ekman). New York: Oxford University Press. (Original work published 1872).

Davitz, J. R. (Ed.). (1964). *The communication of emotional meaning*. New York: McGraw-Hill.

Devillers, L., Vidrascu, L., & Lamel, L. (2005). Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks, 18*, 407–422.

Egloff, B., Schmukle, S. C., Burns, L. R., & Schwerdtfeger, A. (2006). Spontaneous emotion regulation during evaluated speaking tasks: Associations with negative affect, anxiety expression, memory, and physiological responding. *Emotion, 6*, 356–366.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P. (2003). *Emotions revealed*. New York: Henry Holt.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49–98.

Eldred, S. H., & Price, D. B. (1958). A linguistic evaluation of feeling states in psychotherapy. *Psychiatry, 21*, 115–121.
PubMed

Fernandez-Dols, J., Sanchez, F., Carrera, P., & Ruiz-Belda, M. (1997). Are spontaneous expressions and emotions linked? An experimental test of coherence. *Journal of Nonverbal Behavior, 21*, 163–177.
[SpringerLink]

First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. W. (1998). *SCID-I: Interview protocol [Swedish]*. Stockholm: Pilgrim Press.

Forsell, M., Elenius, K., & Laukka, P. (2007). Acoustic correlates of frustration in spontaneous speech. *Speech, Music and Hearing: Quarterly Progress and Status Report, 50,* 37–40. Stockholm, Sweden: Department of Speech, Music and Hearing, KTH.

Fuller, B. F., Horii, Y., & Conner, D. A. (1992). Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. *Research in Nursing and Health, 15*, 379–389.
PubMed crossref

Furmark, T., Appel, L., Michelgård, Å., Wahlstedt, K., Åhs, F., Zancan, S., et al. (2005). Cerebral blood flow changes after treatment of social phobia with the neurokinin-1 antagonist GR205171, citalopram, or placebo. *Biological Psychiatry, 58*, 132–142.
PubMed crossref

Furmark, T., Tillfors, M., Everz, P.-O., Marteinsdottir, I., Gefvert, O., & Fredrikson, M. (1999). Social phobia in the general population: Prevalence and sociodemographic profile. *Social Psychiatry and Psychiatric Epidemiology, 34*, 416–424.

Furmark, T., Tillfors, M., Marteinsdottir, I., Fischer, H., Pissiota, A., Långström, B., et al. (2002). Common changes in cerebral blood flow in patients with social phobia treated with citalopram or cognitive-behavioral therapy. *Archives of General Psychiatry, 59*, 425–433.

Greasley, P., Sherrard, C., & Waterman, M. (2000). Emotion in language and speech: Methodological issues in naturalistic settings. *Language and Speech, 43*, 355–375.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology, 39*, 281–292.

Gross, J. J., John, O. P., & Richards, J. M. (2000). The dissociation of emotion expression from emotion experience: A personality perspective. *Personality and Social Psychology Bulletin, 26*, 712–726.

Hagenaars, M. A., & van Minnen, A. (2005). The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. *Journal of Anxiety Disorders, 19*, 521–537.

Harrigan, J. A., Wilson, K., & Rosenthal, R. (2004). Detecting state and trait anxiety from auditory and visual cues: A meta-analysis. *Personality and Social*

Psychology Bulletin, 30, 56–66.

Haskard, K. B., Williams, S. L., DiMatteo, M. R., Heritage, J., & Rosenthal, R. (2008). The provider's voice: Patient satisfaction and the content-filtered speech of nurses and physicians in primary medical care. *Journal of Nonverbal Behavior, 32*, 1–20.

Hofmann, S. G., Gerlach, A. L., Wender, A., & Roth, W. T. (1997). Speech disturbances and gaze behavior during public speaking in subtypes of social phobia. *Journal of Anxiety Disorders, 11*, 573–585.

Johnstone, T., van Reekum, C. M., Bänziger, T., Hird, K., Kirsner, K., & Scherer, K. R. (2007). The effects of difficulty and gain versus loss on vocal physiology and acoustics. *Psychophysiology, 44*, 827–837.

Johnstone, T., van Reekum, C. M., Hird, K., Kirsner, K., & Scherer, K. R. (2005). Affective speech elicited with a computer game. *Emotion, 5*, 513–518.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1*, 381–412.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770–814.

PubMed CrossRef

Kashdan, T. B., & Steger, M. F. (2006). Expanding the topography of social anxiety: An experience-sampling assessment of positive emotions, positive events, and emotion suppression. *Psychological Science, 17*, 120–128. PubMed CrossRef

Kasl, S. V., & Mahl, G. F. (1965). The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology, 1*, 425–433. CrossRef

Kessler, R. C., Stein, M. B., & Berglund, P. (1998). Social phobia subtypes in the National Comorbidity Survey. *American Journal of Psychiatry, 155*, 613–619. PubMed

Koerner, A. F., & Fitzpatrick, M. A. (2002). Nonverbal communication and marital adjustment and satisfaction: The role of decoding relationship relevant and relationship irrelevant affect. *Communication Monographs, 69*, 33–51. CrossRef

Kuroda, I., Fujiwara, O., Okamura, N., & Utsuki, N. (1976). Method for determining pilot stress through analysis of voice communication. *Aviation, Space, and Environmental Medicine, 47*, 528–533. PubMed

Lang, P. J. (1985). The cognitive psychophysiology of emotion: Fear and anxiety. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 131–170). Hillsdale, NJ: Lawrence Erlbaum.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion, 5*, 277–295.

Laukka, P. (2008). Research on vocal expression of emotion: State of the art and future directions. In K. Izdebski (Ed.), *Emotions in the human voice. Vol 1. Foundations* (pp. 153–169). San Diego, CA: Plural Publishing.

Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion, 19*, 633–653.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lee, C. M., & Narayanan, S. (2005). Towards detecting emotion in spoken dialogs. *IEEE Transactions on Speech and Audio Processing, 13*, 293–303.

Levenson, R. W. (1994). Human emotions: A functional view. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 123–126). New York: Oxford University Press.

Lewin, M. R., McNeil, D. W., & Lipson, J. M. (1996). Enduring without avoiding: Pauses and verbal dysfluencies in public speaking fear. *Journal of Psychopathology and Behavioral Assessment, 18*, 387–402.

Litman, D. J., & Forbes-Riley, K. (2006). Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors. *Speech Communication, 48*, 559–590.

Mahl, G. F. (1956). Disturbances and silences in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology, 53*, 1–15.

McNally, R. J., Otto, M. W., & Hornig, C. D. (2001). The voice of emotional memory: Content-filtered speech in panic disorder, social phobia, and major depressive disorder. *Behaviour Research and Therapy, 39*, 1329–1337.

Perry, C. K., Ingrisano, D. R., Palmer, M. A., & McDonald, E. J. (2000). Effects of environmental noise on computer-derived voice estimates from female speakers. *Journal of Voice, 14*, 146–153.

Planalp, S., DeFrancisco, V. L., & Rutherford, D. (1996). Varieties of cues to emotion in naturally occurring situations. *Cognition and Emotion, 10*, 137–153.

Pope, B., Blass, T., Siegman, A. W., & Raher, J. (1970). Anxiety and depression in speech. *Journal of Consulting and Clinical Psychology, 35*, 128–133.

Russell, J. A., Bachorowski, J.-A., & Fernandez-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*, 329–349.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143–165.

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 165–197). New York: Wiley.

Scherer, K. R., & Ceschi, G. (2000). Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and Social Psychology Bulletin, 26*, 327–339.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*, 123–148.

Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research, 1*, 269–285.

Siegman, A. W. (1987). The telltale voice: Nonverbal messages of verbal communication. In A. W. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication* (2nd ed., pp. 351–434). Hillsdale, NJ: Lawrence Erlbaum Associates.

Smith, G. A. (1977). Voice analysis for the measurement of anxiety. *British Journal of Medical Psychology, 50*, 367–373.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting

Psychologists Press.

Thompson, A. R. (1995). Pharmacological agents with effects on voice. *American Journal of Otolaryngology, 16*, 12–18.

Tillfors, M., Furmark, T., Marteinsdottir, I., Fischer, H., Pissiota, A., Långström, B., et al. (2001). Cerebral blood flow in subjects with social phobia during stressful speaking tasks: A PET study. *American Journal of Psychiatry, 158*, 1220–1226.

Turk, C. L., Heimberg, R. G., Luterek, J. A., Mennin, D. S., & Fresco, D. M. (2005). Emotion dysregulation in generalized anxiety disorder: A comparison with social anxiety disorder. *Cognitive Therapy and Research, 29*, 89–106.

van Bezooijen, R., & Boves, L. (1986). The effects of low-pass filtering and random splicing on the perception of speech. *Journal of Psycholinguistic Research, 15*, 403–417.

Williams, C. E., & Stevens, K. N. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine, 40*, 1369–1372.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 1238–1250.

Zaider, T. I., Heimberg, R. G., Fresco, D. M., Schneier, F. R., & Liebowitz, M. R.

(2003). Evaluation of the clinical global impression scale among individuals with social anxiety disorder. *Psychological Medicine, 33*, 611–622. PubMed crossref

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). New York: Wiley.