Biol Res 41: 389-395, 2008

Measuring emotion in the voice during psychotherapy interventions: A pilot study

MARÍA E MONETA[1, 3, *], MARIO PENNA[1], HUGO LOYOLA[1], ANNA BUCHHEIM, and HORST KÄCHELE[2]

[1] Faculty of Medicine, University of Chile
[2] Department of Psychosomatic Medicine and Psychotherapy, University of Ulm, Germany
[3] Universidad Diego Portales
* To whom all correspondence should be addressed.

ABSTRACT

The voice as a representation of the psychic world of patients in psychotherapeutic interventions has not been studied thoroughly. To explore speech prosody in relation to the emotional content of words, voices recorded during a semi-structured interview were analyzed. The subjects had been classified according to their childhood emotional experiences with caregivers and their different attachment representations. In this pilot study, voice quality as spectral parameters extracted from vowels of the key word "mother" (German: "Mutter") were analyzed. The amplitude of the second harmonic was large relative to the amplitude of the third harmonic for the vowel "u" in the secure group as compared to the preoccupied group. Such differences might be related to the subjects' emotional involvement during an interview eliciting reconstructed childhood memories.

Key terms: mother, attachment, vocal cues, emotion.

INTRODUCTION

Spoken language is considered to be the main psychotherapeutic tool (Russel, 1993); however, the relevance of voice quality for the communicative process, i.e. the effects of emotion on speech intonation and tempo, has received little attention. Dittman and Wynne (1961) proposed that para-linguistic variables like pitch and speech rate may capture emotional signals from discourse, but they could be independent of the subject's emotional expression in an interview. Studies on speech structure and emotion imply various components, spanning from the emotions underlying cognitive activity to the accompanying physiological responses during different emotions (Scherer, 1986; Gobl and Chasiade, 2003; Campbell, 2004).

Modern techniques for voice analysis are accurate enough to unveil shifts in the emotional involvement of speakers. Vocal patterns denote involuntary physiological changes in the speaker's speech production system (Scherer, 1986), as well as culturally accepted speaking styles (Scherer et al., 2001; Campbell and Erickson, 2004; Erickson 2005). Physiological arousal and the appraisal of the emotion experienced exert a particularly strong influence on the configuration of vocal cues during discourse (Siegmann and Boyle, 1993; Pell, 2001).

A number of studies have shown that different emotions are cued by various combinations of acoustic parameters; speech rate and the fundamental frequency presumably exerting the strongest effect (Banse and Scherer, 1996, Murray and Arnolt, 1993; Pell, 2001).The mean fundamental frequency and speech rate are generally higher for emotions associated with high sympathetic arousal, like anger

Correspondence concerning this article should be addressed to Dr. Maria Eugenia Moneta, Faculty of Medicine, University of Chile, Santiago de Chile. E-mail adress: mmoneta@med.uchile.cl

Received: August 4, 2008. In Revised form: January 13, 2009. Accepted: January 19, 2009

and fear, as well as happiness and anxiousness (Banse and Scherer,1996; Ellgring and Scherer, 1996; Johnston and Scherer, 1999; Scherer 2003; Hachizaga et al., 2004; Erickson 2005). Fundamental frequency and speech rate are the parameters that give a better assessment of the emotional state of speakers as evaluated by listeners (Banse & Scherer, 1996; Breitenstein et al., 2001).

An important, but difficult-to-analyze aspect of expressive speech, is voice quality, defined as the quality of "a sound by which a listener can tell that two sounds of the same loudness and pitch are dissimilar" (ANSI, Psycho-acoustical terminology, 1973; Campbell and Mokhtari, 2003). Changes in voice quality can signal both paralinguistic information in terms of changes in the speaker's emotional state, mood or attitude to the message and listener, and non-linguistic information in terms of the speaker's social or geographical background, as well as personal characteristics related to the speaker's physical constitution or health (Mokthari, 2003; Erickson, 2005). Changes in voice quality are the result of changes in the configuration of the vocal tract, laryngeal and glottal source (Mokthari, 2003; Skakibara and Imagawa, 2004). These changes can be quantified by comparing the amplitude levels of different spectral components, among which the first and second harmonic and the first formant have been used (Gordon and Ladfoged, 2001). In addition, spectral slope has been measured by splitting the spectrum in third-octave bands and fitting a line to their respective energies as an indication of harshness or softness (Schroeder, 2004).

Rice & Kerr (1986) developed a qualitative approach for studying vocal expression during client therapist interaction and outcome in psychotherapy. However, no quantitative analysis of the structure of vocal patterns of either patient or therapist has yet been attempted. In contrast to the lack of studies on vocal correlates of emotion, facial expression during psychotherapy has received considerable attention in recent years (Bänninger-Huber, 1992; Krause et al. 1992; Krause, 1998; Dreher et al., 2001).

In this study, presented here as a preliminary report, we explored voice quality through vocal correlates of emotional styles evoked by the word

"mother" (German: Mutter) in subjects selected for psychotherapy. A standardized interview, the Adult Attachment Interview (AAI; George et al., 1994), was used to classify the subjects in three categories according to three attachment representations (Secure, Dismissing and Preoccupied). Secure individuals give an open, coherent and consistent account of their childhood memories, regardless if they were positive or negative. These persons can easily address the topics asked about and convey an emotional balance about the relationships with their parents. Adults with the dismissing classification give incoherent, incomplete accounts of the experience with their parents and often show gaps in memory. As a defense against painful memories, they minimize the importance of attachment experiences. These people insist on the normality of their affections and on their inner independence from their parents.

Preoccupied adults recall childhood experiences in an angry, excessive and non-objective way. A characteristic of this group is the oscillation between positive and negative evaluations without being conscious of this contradiction. The language employed seems confused, unclear and vague. We predicted that the spectral structure of the word "Mutter" during recall of childhood memories differs among subjects of these three categories, affecting their voice quality.

METHODS

Subjects

The study is based on the Adult Attachment Interview (AAI) conducted during the first therapeutic session with ten female subjects at the Psychosomatic Clinic of the University of Ulm. The subjects had all had their first baby in the last two months and were between the ages of 27 and 32 years old. They were instructed about the AAI, in

which they are asked to remember personal emotional issues in a psychotherapeutic setting. Patients were asked about their willingness to collaborate in this study. All subjects were interviewed by the same therapist (A.B.) and were native-German speakers. The AAI measures the current representations of past and present attachment experiences based on narrative accounts. The inter-individual differences in the assessed attachment representations form three main categories: "secure", "dismissing", "preoccupied" (Main & Goldwyn, 1994). Three secure, four dismissing and four preoccupied subjects were analyzed according to their vocal spectrum.

Procedure

For the present study, the word "mother" (German: Mutter) was selected from the audiotapes of AAIs of subjects of the three attachment groups to compare the affective quality of speech. Recordings of the interviews containing the word "Mutter" in response to specific questions were acquired at a sampling rate of 22050 Hz with a Macintosh computer (Power PC 7100), using the Sound Edit 16 software. The acquired sounds were further analyzed with the Signalyze 3.12 software. Twelve repetitions of the word "mother" in response to the first half of the interview, containing emotionally loaded questions from the AAI in relation to the mothers were analyzed for each subject. Power spectra (0-5500 Hz, 20 Hz resolution) of the two vowels of the word "mother" were obtained at the midpoint of the vowels u and e ("Mutter"). The amplitudes and frequencies of the fundamental frequency of the following spectral peaks were measured, fundamental frequency (Fo) and the second and third (H2 and H3, respectively). The differences in amplitudes and frequencies of these three spectral peaks were computed, and averages of these measurements calculated for each individual and compared among individuals of the three groups with a One-way ANOVA and the Duncan's statistical test ($P < 0.05$). In addition to power spectra, oscillograms and sonograms of the complete word were obtained for graphic representation of this sound.

RESULTS

Fig. 1 shows the oscillogram, sonogram and power spectra of the word "Mutter" from one individual of the secure group. Peaks corresponding to Fo, H2 and H3 from the power spectra were taken from the middle part of vowels u and e.

Power spectra showed that Fo was about 206.20 ±12.27 Hz for vowel u and 205.92 ±18.15 Hz for vowel e when all the subjects were pooled for analysis, and no significant differences occurred between groups (ANOVA: $F_2 = 0.659$, $P = 0.719$). However, the amplitude differences between Fo, H2 and H3 varied considerably among individuals and groups. Amplitude differences in dB between Fo and H3 (Fo-H3) and between H2 and H3 (H2-H3) were always positive for vowel u, i.e.: Fo and H2 always had larger amplitudes relative to H3. A tendency to higher values occurred for some of these amplitude differences in the Secure group relative to the Preoccupied group, yielding significant differences in the amplitude of H2 and H3 for vowel u between both groups (H2-H3: ANOVA: $F_2 = 6.727$, $P = 0.0346$; Duncan's test, $P = 0.045$, Fig. 2). The Dismissing group showed intermediate values for these measurements and did not differ significantly from the values of the Preoccupied and Secure groups (Duncan's test, $P = 0.640$).

DISCUSSION

In this preliminary study on voice quality measurements related to a psychotherapeutic interview, we have found differences in the relative amplitudes of harmonics H2 and H3 between subjects of the Preoccupied and Secure groups for the vowel "u", but not for the vowel "e". The spectral variation in vowel "u" is probably related to its emphasis and longer duration in the phonetics of the German word

"mutter" (mother). Our data suggest subtle variations in the voice for the Preoccupied group relative to the Secure and Dismissing groups, while pronouncing the word "mutter". As mentioned in the Introduction, this could imply a difference in voice quality due to emotional arousal. Changes in voice quality are related to modifications of the vocal tract; involuntary changes in tonicity and thus, we believe, speech production is a vehicle for emotion and mood.

Although preliminary, our data point to the usefulness of measuring the amplitude of harmonics in detecting affective attributes of speech. Measurements of the first harmonics could indicate more subtle changes in the voice than those detected by variations of the fundamental frequency and formants.

The lack of differences in Fo among subjects in our study could be related to a similar level of arousal for all individuals interviewed by the same therapist. Such invariance could account for a convergence in fundamental frequency between speakers during conversation, as reported by Gregory et al. (2000).

Although a number of studies have focused on pitch variables, in particular fundamental frequency (Scherer, 1986; Tolkmitt and Scherer, 1986; Murray and Arnolt 1993), little is known about the role of voice quality in communicating affection. As pointed out by Scherer (1986) and later on by Gobl and Chasaide (2003), the tendency has been to concentrate on those parameters that are relatively easy to measure, such as fundamental frequency and timing variables. Johnston and Scherer

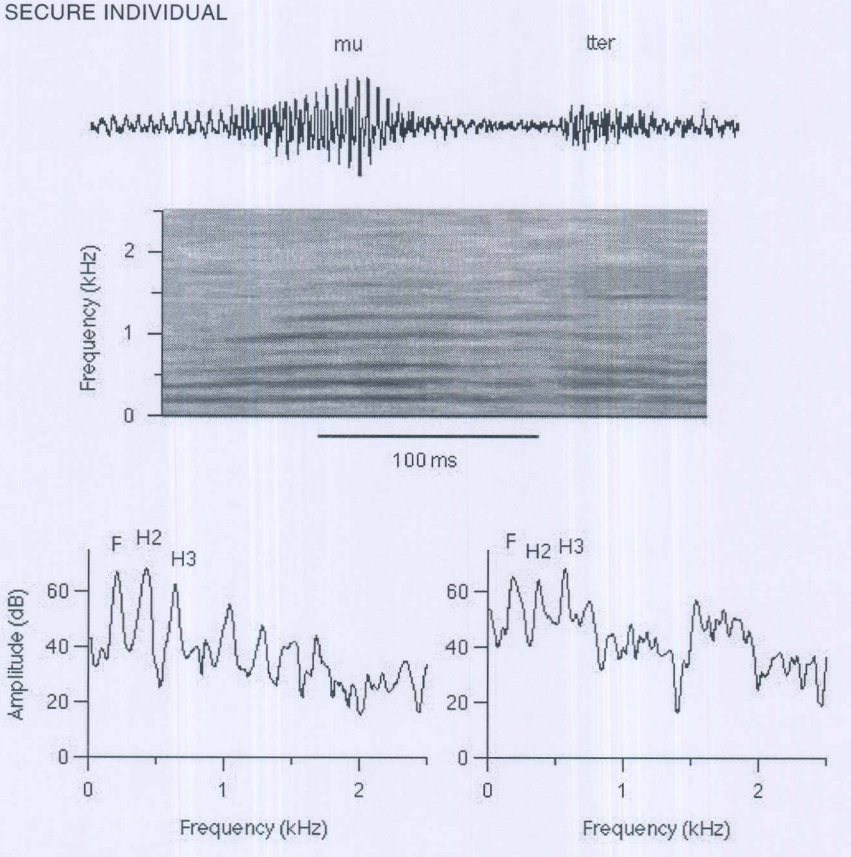

Figure 1: Oscillogram, sonagram and power spectra of the word "Mother" (German: Mutter) from one individual of the secure (S) group. Peaks corresponding to the fundamental frequency (Fo), second (H2) and third (H3) harmonics are shown in the power spectra of the vowels u and e.

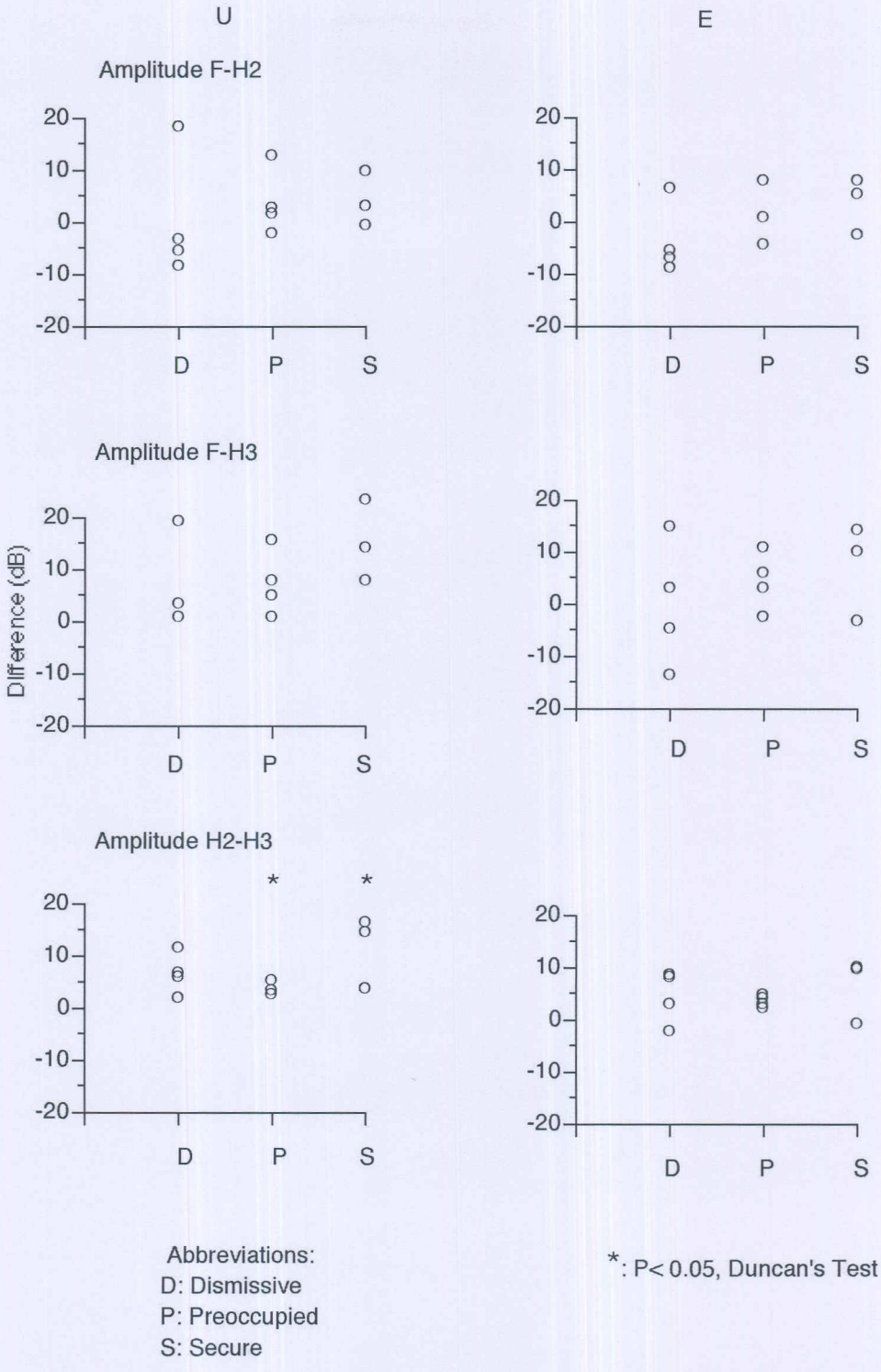

Figure 2: Differences between the amplitudes of the spectral peaks of the fundamental frequency (Fo) and second (H2) and third (H3) harmonics for the vowels u and e for the three attachment groups. Each circle indicates the average value for an individual. Significant differences (Duncan test, $P < 0.05$) occurred only for vowel u (amplitude H2-amplitude H3) between the preoccupied (P) and secure (S) groups.

(2001) have found that fundamental frequency varies with emotional state, being lower for states related to boredom and depression and higher for happy and anxious states. However, the study of voice quality implying more detailed analysis of spectral contents has been restricted because of methodological difficulties involved.

Voice quality signals information concerning the speaker's attitude towards the interlocutor, the subject matter and the situation. Furthermore, the listener's evaluations of emotions appeared to be primarily determined by voice quality cues other than fundamental frequency (Gobl, 1989; Lee and Childres, 1991; Gobl and Chasaide, 2003). Gobl and Chasaide (2003) in particular, have emphasized that voice modifications are more related to attitudes, states and mood rather than to specific emotions. A human's ability to "listen between-the-lines" is heavily dependent on voice quality (Campbell, 2004). In terms of subjective perception of the word "mother" (mutter) by non-instructed listeners, they could not appreciate any audible differences between the subjects in our study.

Changes in the spectral structure of speech that may denote emotional differences have been reported by measuring formants at different stages of therapeutic interventions (Tolkmit & Scherer, 1986). Fujimura and Erickson (2004) have proposed a model that incorporates speech rhythm aspects, as well as the linguistic and sociolinguistics aspects of expressive speech during conversation. Nevertheless, a comprehensive model of expressive speech is still lacking.

It has been established that the infant brain, as early as seven months after birth, detects emotionally loaded words and shows differential attentional responses, depending on their emotional valence (Grossmann et al., 2005). Infants are able to interpret prosody and recognize and imitate vocal patterns and rhythms, discriminating intonational aspects of human voice (Fernald, 1993; Spence and Freeman, 1996; Floccia et al., 2000). They can also respond to variations in frequency, intensity and temporal patterning of sounds (implicit knowledge) signaling affective states (Bebbe et al., 1997; Kuhl et al., 1997; Papousek and Papousek 1981). Such an early disposition to react to the emotional contents of speech in an implicit way suggests the need for further research on prosody influences on interpersonal encounters in the therapeutic context. As Beebe points out, words are not enough; music is needed (Bebbe, 2007).

REFERENCES

ANSI (American National Standard report, 1973) Psychoacoustical Technical Report, S. 3: 30

BÄNNINGER-HUBER E (1992) Prototypical affective microsequences in psychotherapeutic interaction. Psychotherapy Research, 4: 291-306

BANSE R, SCHERER K (1996) Acoustic profiles in vocal emotion expressions. Journal of Personality and Social Psychology, 70: 614-636

BEEBE B, JAFFE J, LACHMANN F (1997) Mother-Infant interaction structures and pre-symbolic self and object representations. Journal of Relational Perspectives, 7: 133-182

BEBBE B, (2007) On Intersubjective Theories: The implicit and the explicit in interpersonal relationships. Carli L. Rodini C (Eds)

BREITENSTEIN C, VAN LACKER D, DAUM I (2001) The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. Cognition and Emotion, 15, 1: 57-79

DITTMANN A, WYNNE L (1961). Linguistic techniques and the analysis of emotionality in interviews. Journal of Abnormal Psychology, 63: 201-204

DREHER M, MENGELE U, KRAUSE R, KAEMMERER A (2001) Affective Indicators of the Psychotherapeutic Process. An empirical case study. Psychotherapy Research, 11: 99-117

CAMPBELL N (2004) Listening between the lines: A study of paralinguistic information carried by tone of voice. Proc. Int. Symp.Tonal Aspects of Language with Emphasis on Tone Languages, Beijing (TAL) 13-16

CAMPBELL N, ERICKSON D (2004) What do people hear? A study on the perception of nonverbal affective information in conversational speech. J. Phonet. Soc. Jpn. 8: 9-28

CAMPBELL N, MOKHTARI P (2003) Voice quality: The 4th prosodic parameter. Proc. 15th Int. Congr. Phonetic Sciences: 2417-2420

CHILDERS D G, LEE K (1991) Vocal quality factors: Analysis, Synthesis and perception. J. Acoust. Soc. Am. 90: 2394-2410

ELLRING H, SCHERER K (1996) Vocal indicators of mood change in depression. Journal of Non-verbal Behavior, 20, 2: 83-110

ERICKSON D (2005) Expressive speech: production, perception and application to speech synthesis. Acoust. Sci. and Tech. 26, 4: 317-325

ERICKSON D, YOSHIDA K, MOCHIDA T, SHIBUYA Y (2004) Acoustic and articulatory analysis of sad Japanese speech. Phonet. Soc. Jpn. Fall Meet.: 113-118

ERICKSON D, MOKHTARI P, MENEZES C, FUJINO A (2003) Voice quality and other acoustic changes in sad Japanese speech. Tech. Rep: Inst. Electron. Inf.Commun. Eng.: 43-48

FERNALD A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. Child development, 64: 657-674

FLOCCIA C, NAZZI T, BERTONCINI J (2000). Unfamiliar voice discrimination for short stimuli in newborns. Developmental Science, 3, 3: 333-343

FUJIMURA H, ERICKSON D (2004) The C/D Model for prosodic representation of expressive speech in English. Proc. Autumn Meet. Acoust. Jpn. 271-272

GEORGE C, KAPLAN N, MAIN M (1984) The Adult Attachment Interview.
Unpublished manuscript, University of California, Berkeley.

GOBL C (1989) A preliminary study on acoustic voice quality correlates. STL-QPSR1, Speech, Music and Hearing, Royal Institute of technology, Stockholm 9-21

GOBL C, CHASIADE NI (2003) The role of voice quality in communicating emotion, mood and attitude. Speech communication 40: 189-212

GREGORY S W, GREEN B E, CARROTHERS R, DAGAN K, WEBSTER S (2000) Verifying the primacy of voice fundamental frequency in social status accommodation. Language and Communication, 21: 37-60

GROSSMANN T, STRIANO T, FEDERICCI A (2005) Infant's electric brain activity responses to emotional prosody. NeuroReport 16: 1825-1828

GORDON M, LADFOGED P (2001) Phonation types: A cross-linguistic overview. J. Phonet., 29: 383-406

HASHIZAWA Y, TAKEDA M, HAMZAH M, OHBYAMA G (2004) On the difference of prosodic features of emotional expressions in Japanese speech according to the degree of emotion. Proc. Speech prosody Nara, 655-658

JOHNSTONE T, SCHERER K (2001) Vocal Communication of emotion. In M.Lewis and J. Haviland (Eds.) The handbook of emotions (2nd ed.) (p. 226-235) New York: Guilford Press

KRAUSE R, STEIMER-KRAUSE, E ULRICH B (1992) Use of affect research in dynamic psychotherapy. In M. Leuzinger-Bohleber, H. Schneider, & R. Pfeifer (Eds.), Two_Butterflies on my Head. Psychoanalysis in the Interdisciplinary Dialogue (p. 277-291) Berlin, Springer

KUHL P, ANDRUSKI J E, CHISTOVICH I A, KOZHEVNIKOVA EV, RYSKINA V L, STOLYAROVA E I, SUNDBERG U, LACERDA F (1997) Cross-language analysis of phonetics units in language addressed to infants. *Sciences 227:* 684-686

MAIN M, GOLDWYN R (1994) Adult Attachment Scoring and Classification Systems. Unpublished manuscript, University of California, Berkeley (1st edition 1985)

MOKTHARI P (2003) Parameterization and control of laryngeal voice quality by principal components of glottal waveforms (2003) J. Phonet. Soc.Jpn.7: 40-54

MURRAY I, ARNOLT J I (1993) Towards a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. Journal of the Acoustic Society of America, 93: 1097-1108

PAPOUSEK M, PAPOUSEK H (1981). Die Bedeutung musikalischer Elemente in der frühen Kommunikation zwischen Eltern und Kind. Sozialpädiatrie in Praxis und Klinik, 3: 412-415

PELL M (2001). Influence of emotion and focus location on prosody in matched statements and questions. Journal of the Acoustic Society of America, 109: 1668-1680

RICE L, KERR G (1986) Measures of client and therapist vocal quality In L.Greenberg & W. Pinsof (Eds.) The psychotherapeutic process. A research Handbook (p.73-105). New York: Guilford Press.

RUSSELL R L (Ed.) (1993) Language in Psychotherapy: Strategies of Discovery. New York: Plenum Press.

SAKAKIBARA K, IMAGAWA H (2004) Acoustical interpretation of certain laryngeal settings using a physical model. Proc. Speech Prosody, Nara, 637-640

SCHERER K, JOHNSTON T, BANZINGER T (1998) Automatic_verification of emotionally_stressed speakers: The problem of individual differences. Proceedings of the International Workshop on Speech and computer, St. Petersburg.

SCHERER K (1986) Voice, stress and emotion. In M. H. Appley & Trumbull (Eds.) Dynamic of stress (p159-181). New York: Plenum Press

SCHERER K, BANSE R, WALLBOTT HG (2001). Emotion inferences from vocal expression correlate across languages and cultures. Journal of Cross-cultural Psychol.32, 1: 76-92

SCHERER K (2003) Vocal communication of emotion: "A review of research Paradigms." Speech Commun., 40: 227-256

SCHROEDER M (2004) Speech and Emotion research: an overview of research framework and a dimensional approach to emotional speech synthesis. Doctoral Thesis, Phonus 7, Res. Inst. Phonetics Saarland University

SIEGMAN A W, BOYLE S (1993) Voices of fear and anxiety and sadness and depression: The effects of speech rate and loudness on fear and anxiety and sadness and depression. Journal of Abnormal Psychology, 10: 430-437

SPENCE M J, FREEMAN MS (1996) Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. Infant Behavior and Development, 18, 15: 727-735

TOLKMITT P, SCHERER K (1986) Effects of experimentally induced stress on vocal patterns. Journal of experimental Psychology: Human Perception and Performance, 12: 302-313