Therapy Progress Indicator (TPI):
Combining speech parameters and the subjective unit of distress

Egon L. van den Broek

Center for Telematics and Information Technology (CTIT), University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands

`vandenbroek@acm.org`

Frans van der Sluis

Human-Media Interaction (HMI), University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands

`f.vandersluis@utwente.nl`

Ton Dijkstra

Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen
P.O. Box 9104, 6500 HE Nijmegen , The Netherlands

`t.dijkstra@donders.ru.nl`

Abstract

A posttraumatic stress disorder (PTSD) is a severe handicap in daily life and its treatment is complex. To evaluate the success of treatments, an objective and unobtrusive expert system was envisioned: an therapy progress indicator (TPI). Speech was considered as an excellent candidate for providing an objective, unobtrusive emotion measure. Speech of 26 PTSD patients was recorded while they participated in two reliving sessions: re-experiencing their last panic attack and their last joyful occasion. As a subjective measure, the subjective unit of distress was determined, which enabled the validation of derived speech features. A set of parameters of the speech features: signal, power, zero crossing ratio, and pitch, was found to discriminate between the two sessions. A regression model involving these parameters was able to distinguish between positive and negative distress. This model lays the foundation for an TPI for patients with PTSD, which enables objective and unobtrusive evaluations of therapies.

No laga duele bieu:
  Skavisábo di nobo.

Let not woes of old
  enslave you anew.

– Nydia Ecury –

1. Introduction

The interest in emotion originates from 'strange' or 'abnormal' behavior of people that were previously simply denoted as 'crazy'. While various forms of 'craziness' were identified, diagnostic criteria were determined. This enabled the classification of both behavior and personality. With the further exploration of these criteria, research on emotions slowly matured [21].

In the last decades, however, another branch of emotion research emerged, investigating the emotions of 'normal' people, who could not be classified using diagnostic criteria. Moreover, the interaction of people in their natural habitat with other people, objects, machines, computers, or ambient intelligence was studied in advance of envisioned futures [5, 12, 21–23]. Although studies with clear clinical relevance nowadays are even outnumbered by studies of 'normal' people's emotions, some forces drive emotion research back to its origin.

Rosalind W. Picard currently focuses on the analysis and diagnosis of autistic behavior [17]. On the one hand, a narrow target group limits the general use of the research results. On the other hand, conducting research on such specific groups has (at least) three advantages:

1. real-world, naturally occurring emotions,

2. clinical relevance, and

3. more controlled research, which facilitates results that can be used in practice

This paper will illustrate each of these advantages.

A major problem from which psychiatry and clinical psychology suffers is the struggle to objectively diagnose progress of patients under treatment. Therapists have their own approaches to treatment. Although suppressed by their professional attitude, there is often a bias for applying certain methods. This troubles their judgment of the (lack of) progress achieved.

To tackle this problem, a range of questionnaires and diagnostic criteria have been developed; e.g., [14]. Regrettably, the former are often a burden for the patients, because it takes some time to complete them. Moreover, they rely on the willingness of the patient to communicate [19]. The latter is especially important for diagnostic purposes but not sensitive for detecting changes in a patient's condition. Thus, in spite of efforts to standardize their professional approaches, therapists are still seeking for true objectivity in diagnosis and, possibly even more, in a therapy progress indicator (TPI).

The aim of the present research was to develop such a TPI; i.e., an expert system. To do so, we must take into account a broad range of requirements specified by the therapists. For purposes of brevity, we will not provide an exhaustive list of these requirements. In good negotiation with the therapists, all requirements were met. However, one requirement is in particular noteworthy: ultimate unobtrusiveness. It was beyond therapists' acceptance that the indicator would interfere with the therapies themselves.

In this study, patients with a posttraumatic stress disorder (PTSD) participated. Table 1 provides a concise description of this disorder. This table is included to help the reader in understanding the clinical setting in which this research was applied. Moreover, it helps to understand some of the methodological choices made.

The second diagnostic criterion for PTSD (see Table 1) is that the traumatic event is persistently reexperienced. Such reexperiences can happen in five ways. One of these is the feeling that the traumatic events were recurring, which included a sense of reliving the experience, illusions, flashbacks, etc. In this research, we will exploit this characteristic, as will be discussed later.

The following sections will further guide the reader through the quest of our research towards an unobtrusive TPI for patients with a PTSD. As core of the eventual expert system, techniques were developed to discern different states of emotion, in particular their intensity; i.e., a reduced emotional intensity of a re-living experience is a sign of successful treatment. This expert system should be able to serve as a standard second opinion concerning patient's (therapy) progress.

In the next section (Section 2), we outline the various methods that are frequently used for emotion detection. In addition, we explain the approach chosen for this research. Subsequently, the subjective measure for patients experience of distress is introduced in Section 3.

Section 4 explains the clinical study that was conducted. Section 5 presents the feature extraction from the speech signal, the ground truth, the results on the clinical sessions, and an overview of the features. In Section 6, we end with a discussion of the research and its implications.

2. Emotion detection

A broad plethora of methods exist that enable the detection of signals reflecting emotions. A concise overview of these methods is provided. In addition, the subjective measure used in this research is presented.

The emotional state that people are in can be detected by processing various signals. A literature review reveals that these signals can be assigned to two groups:

1. A broad range of physiological measures / biosignals. For recent overviews, we refer to [7, 15, 21].

2. Specialized areas of signal processing: (a) movement analysis [2, 10], (b) computer vision techniques [6, 10], and (c) speech processing [6, 20, 24].

With the most important requirement for this research in mind (i.e., absolute unobtrusiveness), most signal processing schemes are judged as unsuitable.

Rapid progress in biomedical engineering now enables (nearly) unobtrusive measurement of biosignals; e.g., [8]. Biosignals can be monitored through ring-like and earring-like devices. However, also these devices have to be attached to a patient's body. All methods to record biosignals need sensors attached to a person. Hence, biosignals are not acceptable in our situation.

Both movement analysis and computer vision techniques need markers to secure sufficient reliability. Hence, they are either obtrusive or are not reliable due to various reasons (e.g., light, colors, and occlusion) [25]. Considering the problems with the aforementioned and the requirement of complete unobtrusiveness, speech processing is ideal for clinical applications, such as ours.

A vast amount of work has been done on speech processing. Various speech features are shown to be sensitive to experienced emotions; e.g., see [6, 20, 24]. However, note that in emotion research emotions are often mimicked or faked; see [24] for an overview. Even when this is done by professional actors, such emotions are distinct from true emotions. This is reflected in most speech parameters.

A big advantage of speech as a signal reflecting emotions is that the communication in therapy sessions is often recorded anyway. Hence, obtrusiveness plays no role with speech processing and no additional technological effort has

Table 1. Introduction on (the DSM-IV TR [1] criteria for) posttraumatic stress disorder (PTSD).

Trauma can cause long-term physiological and psychological problems. This has been recognized for centuries. Such suffering accompanying, e.g., a posttraumatic stress disorder (PTSD), can be characterized in terms of series of symptoms and causes. Trauma's can originate from a range of situations, either short or long lasting; e.g., warfare, natural disasters such as earthquakes, interpersonal violence such as sexual, physical, and emotional abuse, intimate partner violence, and collective violence.

Diagnostic criteria as defined by the DSM-IV TR [1] comprise six categories, each denoting their various indicators: 1) Exposure of the person to a traumatic event; 2) Persistent reexperience of the traumatic event; 3) Persistent avoidance of stimuli, associated with the trauma, and numbing of general responsiveness (not present before the trauma); 4) Persistent symptoms of increased arousal, not present before the trauma; 5) Duration of the disturbance (symptoms in criteria 2, 3, and 4) is more than one month; and 6) The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning. Many other symptoms have also been mentioned; e.g., weakness, fatigue, loss of will power, and psychophysiological reactions such as gastrointestinal disturbances. However, these are not included in the DSM-IV TR diagnostic criteria.

Taken together, PTSD includes a broad variety of symptoms and diagnostic criteria. Consequently, the diagnosis is hard to make, as is also the case for various other mental disorders.

---

to be made on the side of the therapists. Moreover, because therapy sessions are generally held under controlled conditions in rooms shielded from noise, the degree of noise that distorts the speech signal is limited. Therefore, speech was chosen to assess a patient's emotional state.

3. Subject unit of distress (SUD)

To validate parameters that are derived from speech, a subjective measurement is needed. The subject unit of distress (SUD) is optimally suited. It is a Likert scale that simply registers how much stress or tension a person says to experience at a certain moment (in our case, on a scale with a range between 0-10).

The SUD has been introduced more than half a century ago [26]. Since then, the SUD has shown itself as a reliable measure for a person's emotional state. Like all Likert scales, the SUD simply requires to put a dot or a cross on a linear scale. This can be asked for every minute; hence, throughout a session it becomes a routine. In this research, the SUD will serve as the ground truth for further analysis.

The data was specifically gathered with the aims of the current research in mind. Hence, the database obtained through this research consists of a unique set of validated speech samples; cf. the overview provided in [24]. It can be used to test many more speech processing paradigms for emotion detection.

4. Research setup

In this study, 26 female Dutch PTSD patients voluntarily participated. All patients signed an informed consent and all were aware of the tasks included.

The research consisted of four phases. The first and last phase involved the recording of a baseline for both speech and the SUD. The second and third phase were reliving sessions. In one of these sessions, the patients were asked to reexperience their last panic attack. In the other session, the patients were asked to tell about the last happy event they could recall. The order of sessions was counterbalanced over the participants.

The therapists assured us that real emotions would be triggered in the reliving sessions with PTSD patients (see also Table 1), in particular in reliving the last panic attack. Because the reliving sessions were expected to have a high impact on the patient's emotional state, with each patient and during all four sessions, a therapist was present.

The two reliving sessions were chosen to resemble two phases in therapy: its start and end. Reliving a panic attack resembles the trauma in its full strength, as with the intake of a patient. Telling about the last happy event a patient experienced, resembles a patient who is relaxed or (at least) in a 'normal' emotional condition. This should resemble the end of the therapy sessions, when the PTSD has disappeared or is diminished.

Recording speech was done using a personal computer, a microphone preamplifier, and a microphone. The recording's sample rate was 44.1 kHz, mono channel, with a resolution of 16 bits. The recordings of the sessions were divided in samples of approximately one minute of speech. This enabled a one-on-one mapping of speech features in the SUD scores.

Table 2. Pearson's correlations between speech parameters and the subjective unit of distress (SUD), including the level of significance (p).

feature	parameter												
	iqr10	iqr25	max	mean	median	min	q10	q25	q75	q90	range	std	var
pitch				$-0.34^{\dagger}$	$-0.31^{\dagger}$	$-0.19^{\S}$	$-0.24^{\dagger}$	$-0.27^{\dagger}$	$-0.32^{\dagger}$	$-0.33^{\dagger}$			
power	$-0.41^{\dagger}$	$-0.33^{\dagger}$				$0.21^{\S}$	$0.18^{\S}$		$-0.17^{\S}$	$-0.18^{\S}$	$-0.34^{\dagger}$	$-0.40^{\dagger}$	$-0.37^{\dagger}$
cross	$-0.17^{\S}$	$-0.17^{\S}$		$-0.24^{\dagger}$		$-0.22^{\dagger}$			$-0.19^{\S}$	$-0.20^{\S}$		$-0.19^{\S}$	$-0.19^{\S}$
ampl.	$-0.20^{\S}$		$-0.18^{\S}$	$0.19^{\S}$	$0.19^{\S}$		$0.21^{\dagger}$	0.14^{*}		$-0.19^{\S}$	-0.15^{*}	$-0.22^{\dagger}$	$-0.21^{\dagger}$

Note. $^{*}p < .05$, $^{\S}p < .01$, $^{\ddagger}p < .001$. *Abbreviations:* power: power (dB); cross: signal crossings; and ampl.: amplitude.

5. Analysis

5.1. Feature extraction

Although there is no general consensus concerning the best speech parameters for emotion detection, much evidence is present in favor of the power or intensity of the speech signal and for its pitch [6, 20, 24]. In addition, we measured two less often used parameter: the zero crossings rate [13] and the raw amplitude [20]. They are considered to be useful for measuring experienced emotions.

The amplitude or sound pressure of the signal is denoted in Pa (Pascal) as $x(t)$. For a domain $[0, T]$, the power of the speech signal is defined as:

$$10 \ \log_{10} \frac{1}{T \, P_0^2} \int_0^T x^2(t) \, \mathrm{d}t, \qquad (1)$$

where $P_0 = 2 \cdot 10^{-5}$ Pa is the auditory threshold [4]. The power is computed over the discrete signal in following manner:

$$10 \ \log_{10} \frac{1}{N \, P_0^2} \sum_{i=0}^{N-1} x^2(t_i). \qquad (2)$$

where N is the number of samples. It is expressed in dB (decibels) relative to the auditory threshold P_0.

The F0 of pitch was extracted from the corrected speech signal through a fast Fourier transform over the signal. We refer to [3], for a more detailed description of the extraction of F0 of pitch from the speech signal.

The zero crossings rate of the corrected speech signal was computed. This is defined as:

$$\frac{1}{N-1} \sum_{i=1}^{N-1} \mathbb{I} \{x(t_i)x(t_{i-1}) < 0\}, \qquad (3)$$

where N is the number of samples of the signal amplitude x. The $\mathbb{I} \{\alpha\}$ serves as a logical function [13]. This feature will be called signal crossings.

Of each of these features, a number of statistical parameters were derived: *mean*, *median*, standard deviation (*std*), variance (*var*), minimum value (*min*), maximum value (*max*), range (*max* $-$ *min*), the quantiles at $10\%(q10)$, $90\%(q90)$, $25\%(q25)$, and $75\%(q75)$, the inter-quantile-range $10\% - 90\%$ ($iqr10$, $q90 - q10$), and the inter-quantile-range $25\% - 75\%$ ($iqr25$, $q75 - q25$).

Except for the feature amplitude, the features were computed over a time window of 40 msec., using a step length of 10 msec.; i.e., computing each feature every 10 msec. over the next 40 msec. of the signal. Subsequently, the statistical parameters were derived over time chunks of 60 sec., allowing a one-on-one comparison with the SUD data.

5.2. Subjective Unit of Distress (SUD)

In order to establish a ground truth, several audio features were compared to the SUD scores. All available SUD scores were used, consisting of the reexperience conditions as well as the baseline conditions. The comparison was made on audio chunks of one minute, using a Pearson's correlation to test a linear relationship. The resulting significant correlations are given in Table 2.

A model containing all significant features was tested with linear regression analysis. This resulted in an explained variance $R^2 = .52, F(32, 207) = 6.91, p < .001$.

5.3. Clinical sessions

The TPI should be able to differentiate between the clinical sessions involving reexperience of respectively the trauma and a happy event. Both SUD and the speech signal were analyzed and compared with each other. A series of multiple analyses of variance (MANOVAs) were conducted in which SUD and clinical session were dependent variables and the four speech features (pitch, power, signal crossings, and amplitude) were independent variables; see also Section 5.1.

The SUD was unable to distinguish directly between the two session types. However, when taking into account session order, a significant interaction effect was found ($F(3, 136) = 5.78, p < .001$). Next, for speech, all four features will be discussed. Since most effects did dependent on the order of the sessions, similar to the SUD results, only MANOVAs including the order factor will be reported.

Concerning pitch, multiple interaction effects with the factor order were found: mean ($F(3, 158) = 4.44, p < .01$), median ($F(3, 158) = 3.03, p < .05$), q10 ($F(3, 158) = 5.51, p < .001$), q75 ($F(3, 158) = 3.01, p < .05$), and q90 ($F(3, 158) = 2.86, p < .05$). Please notice the similar pattern of significant parameters with the SUD.

The power (dB) also showed significant results, when the

Table 3. Top-10 features and their parameters that map on the SUD and the differentiation between both clinical sessions.

SUD		clinical sessions	
feature	parameter	feature	parameter
power (dB)	iqr10	power (dB)	min
power (dB)	std	signal crossings	q10
power (dB)	var	power (dB)	q10
pitch	mean	pitch	q10
power (dB)	range	amplitude	median
power (dB)	iqr25	amplitude	iqr25
pitch	q90	amplitude	q75
pitch	q75	pitch	mean
pitch	median	power (dB)	mean
pitch	q25	amplitude	q10

factor order was taken into account: mean ($F(3, 159) = 4.19, p < .01$), median ($F(3, 159) = 3.82, p < .05$), min ($F(3, 159) = 6.98, p < .001$), q10 ($F(3, 159) = 5.65, p < .001$), and q75 ($F(3, 159) = 3.10, p < .05$).

Signal crossings, again, showed two significant effects when order was taken as an interacting factor: iqr25 ($F(3, 159) = 3.00, p < .05$) and q10 ($F(3, 159) = 6.14, p < .001$).

Finally, amplitude, showed first and second order effects. The median was a significant determinant for the clinical session, ($F(3, 159) = 5.51, p < .05$). Similar, the iqr25 ($F(1, 159) = 4.42, p < .05$) and q75 ($F(1, 159) = 4.40, p < .05$) were found to be significant. Moreover, several interaction effects were found, namely: iqr10 ($F(3, 159) = 3.98, p < .01$), iqr25 ($F(3, 159) = 4.71, p < .01$), median ($F(3, 159) = 4.87, p < .01$), q10 ($F(3, 159) = 4.10, p < .01$), q75 ($F(3, 159) = 4.51, p < .01$), and q90 ($F(3, 159) = 3.81, p < .05$).

5.4. Overview of the features

Table 3 gives an overview of the 10 most significant features and their parameters, for SUD as well as for the two clinical sessions. However, note that for the clinical sessions all shown features had Order as interacting factor. For SUD, the data are already time-specific.

Table 3 clearly shows a difference between the two dependent variables. SUD is differentiated by power (dB) and pitch, whereas the clinical sessions rely more on amplitude and other parameters of the power (dB) feature. Thus, there seems to be a difference between subjectively reported stress and clinically induced stress.

Clinical stress detection is possible with features that are fairly easy to obtain, as is shown in both Table 2 and Table 3. Note that order and, thereby, time, needs to be taken into account. The values of most features, as well as the subjectively reported stress ($F(3, 136) = 2.99, p < .033$), change with time.

6. Discussion

This research presents unique data, containing very rare displays of intense, real, emotions; hence, a data set with a high ecological validity. As such, the data is relevant for research into speech features underlying stress, as well as for research towards a TPI in clinical psychology.

The primary speech features that indicated subjective stress were pitch and power (dB). The features corresponding to the clinical sessions, which simulated the beginning and end of a therapy of PTSD, were primarily power (dB) and amplitude. These results show the ease with which a reasonable estimate of stress can be achieved, i.e., an estimate explaining almost half of the variance in subjectively reported stress. However, a complication arose from the data. In order to successfully differentiate between clinical sessions, the timing of the data needs to be taken into account. The stress displayed seemed to increase over time. This is in line with what is known on emotions and their accompanying physiological reactions that (indeed) can accumulate over time [9, 12, 23].

The difference between features, on the one hand, representing clinical sessions and, on the other hand, relating to subjective stress hint at another insight into the experience of stress. Viewed from a valence and arousal model of emotion [18], the clinical condition is mainly concerned with valence, because it queries happiness and fear. Instead, the SUD is concerned with arousal, because it measures the intensity of experienced stress. As such, the true emotions recorded in this experiment can be used to verify aspects of a more general model of emotion.

The current approach could also be used to determine both the coping style of clients and the severity of clinical disorders. Note that the main strength of this research lies not in making the distinction between emotional states, but in measuring the intensity of emotions. In combination with measurements indicating the type of experienced emotion, this approach could provide an important step forward towards a more complete and reliable diagnosis of both the coping style and the emotional well being of people.

The speech analysis approach could be linked to approaches that measure physiological responsiveness of PTSD in more obtrusive ways; e.g., [8, 17]. See [7, 15, 21–23] for overviews of possible physiological measurement methods. This would facilitate a triangulation of the construct under investigation, providing more reliability on the results obtained [21]. Moreover, more specific analyses can be conducted; for example, in terms of either the valence and arousal model or discrete emotion categories [7, 15].

Apart from being unobtrusive, the speech-driven approach as applied in the current research has another major advantage. It enables remote determination of people's emotional state. This can be used, for example, in telepsychiatry [11] and call-centers [16] that frequently have to

cope with highly agitated customers.

Taken together, our unique research setup resulted in a database of both experimentally controlled and ecological valid speech samples. Using this database, a model was constructed that uses speech parameters for emotion measurement. This model showed the feasibility of developing an objective, easily usable, unobtrusive, and powerful therapy progress indicator (TPI) in the near future. Although we still have a long way to go in assessing long-term progress in therapy, the assessment of affect within individual sessions and the dedicated tracking of progress for specific disorders, may be coming within reach.

Acknowledgments

References

[1] American Psychiatric Association. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, USA: American Psychiatric Publishing, Inc., 4 (Text Revision) edition, 2000.

[2] L. S. S. Bialoskorski, J. H. D. M. Westerink, and E. L. van den Broek. Mood Swings: An affective interactive art system. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (Intelligent Technologies for Interactive Entertainment)*, 9:181–186, 2009.

[3] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to- noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, volume 17, pages 97–110. University of Amsterdam, 1993.

[4] P. P. G. Boersma and D. J. M. Weenink. Praat 4.0.4, 2006. URL: http://www.praat.org [Last accessed on June 30, 2009].

[5] J. F. Cohn. Foundations of human computing: Facial expression and emotion. *Lecture notes in Artificial Intelligence (Artifical Intelligence for Human Computing)*, 4451:1–16, 2007.

[6] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. Emotion recognition in human–computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.

[7] S. H. Fairclough. Fundamentals of physiological computing. *Interacting with Computers*, 21(1–2):133–145, 2009.

[8] H. Gamboa, F. Silva, H. Silva, and R. Falcão. PLUX – Biosignals Acquisition and Processing, 2009. URL: http://www.plux.info [Last accessed on June 30, 2009].

[9] R. Geenen and F. J. R. van de Vijver. A simple test of the law of initial values. *Psychophysiology*, 30(5):525–530, 1993.

[10] H. Gunes and M. Piccardi. Automatic temporal segment detection and affect recognition from face and body display. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 39(1):64–84, 2009.

[11] D. M. Hilty, S. L. Marks, D. Urness, P. M. Yellowlees, and T. S. Nesbitt. Clinical and educational telepsychiatry applications: A review. *The Canadian Journal of Psychiatry*, 49(1):12–23, 2004.

[12] J. H. Janssen, E. L. van den Broek, and J. H. D. M. Westerink. Personalized affective music player. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume [this volume], 2009.

[13] B. Kedem. Spectral analysis and discrimination by zero-crossings. *Proceedings of the IEEE*, 74(11):1477–1493, 1986.

[14] S. Knapp and L. VandeCreek. *Anxiety disorders: A scientific approach for selecting the most effective treatment*. Sarasota, FL, US: Professional Resource Press/Professional Resource Exchange, Inc, 1994.

[15] I. B. Mauss and M. D. Robinson. Measures of emotion: A review. *Cognition and Emotion*, 23(2):209–237, 2009.

[16] D. Morrison, R. Wang, and L. C. De Silva. Ensemble methods for spoken emotion recognition in call-centres. *Speech Communication*, 49(2):98–112, 2007.

[17] R. W. Picard and M. Goodwin. Developing innovative technology for future personalized autism research and treatment. *Autism Advocate*, 50(1):32–39, 2008.

[18] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[19] H. A. Sackheim and R. C. Gur. *Self deception, self confrontation, and consciousness*, volume 2, pages 117–129. New York, NY, USA: Plenum Press, 1978.

[20] K. R. Scherer. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1–2):227–256, 2003.

[21] E. L. van den Broek, J. H. Janssen, and J. H. D. M. Westerink. Guidelines for Affective Signal Processing (ASP): From lab to life. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume [this volume], 2009.

[22] E. L. van den Broek, M. H. Schut, J. H. D. M. Westerink, and K. Tuinenbreijer. Unobtrusive Sensing of Emotions (USE). *Journal of Ambient Intelligence and Smart Environments*, 1(3):287–299, 2009.

[23] E. L. van den Broek and J. H. D. M. Westerink. Considerations for emotion-aware consumer products. *Applied Ergonomics*, 40:[in press; online available], 2009.

[24] D. Ververidis and C. Kotropoulos. Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9):1162–1181, 2006.

[25] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Towards practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:[in press], 2009.

[26] J. Wolpe. *Psychotherapy by reciprocal inhibition*. Stanford, California: Stanford University Press, 1958.