COMPUTER AIDED DIAGNOSIS FOR MENTAL HEALTH CARE
On the Clinical Validation of Sensitive Machines

Frans van der Sluis[1,2], Ton Dijkstra[3] and Egon L. van den Broek[1,2,4]

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

[2] *Karakter University Center, Radboud University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands*

[3] *Donders Institute for Brain, Cognition, and Behaviour, Radboud University, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands*

[4] *TNO Technical Sciences, P.O. Box 5050, 2600 GB Delft, The Netherlands*

Keywords: Stress, Mental health care, Speech, Computer aided diagnostics (CAD), Artificial neural network, Validation.

Abstract: This study explores the feasibility of sensitive machines; that is, machines with empathic abilities, at least to some extent. A signal processing and machine learning pipeline is presented that is used to analyze data from two studies in which 25 Post-Traumatic Stress Disorder (PTSD) patients participated. The feasibility of speech as a stress detector was validated in a clinical setting, using the Subjective Unit of Distress (SUD). 13 statistical parameters were derived from five speech features, namely: amplitude, zero crossings, power, high-frequency power, and pitch. To achieve a low dimensional representation, a subset of 28 parameters was selected and, subsequently, compressed into 11 principal components (PC). Using a Multi-Layer Perceptron neural network (MLP), the set of 11 PC were mapped upon 9 distinct quantizations of the SUD. The MLP was able to discriminate between 2 stress levels with 82.4% accuracy and up to 10 stress levels with 36.3% accuracy. With stress baptized as being the black death of the 21st century, this work can be conceived as an important step towards computer aided mental health care.

1 INTRODUCTION

In contrast to animals, humans have the ability to make cognitive representations of events, both from the past and for the future. Although such representations aid our daily work and living, they have their down side. In the case of stressful life events, cognitive representations can catalyze worrying and, hence, facilitate chronic stress, unknown to animal species (Brosschot, 2010). This effect is strengthened by the influence of stress on our memory systems (Schwabe et al., 2010). Chronic stress often produces similar physiological responses to those occurring during the stressful events from which it originates. In turn, this can cause pervasive and structural chemical imbalances in our physiological systems, including the autonomic nervous, central nervous, neuroendocrine, immune system, and even in the brain (Brosschot, 2010). Due to the complexity of our physiological systems, their continuous interaction, and their inherent dynamic nature, a thorough understanding of 'chronic stress' is still missing.

Current day practice in treatment of chronic stress focusses on the treatment of either cognitive representations, our habit memory system, or both (Schwabe et al., 2010). In general, under stressful events, the habit memory system tends to dominate over the cognitive memory (or representations) system; however, their precise relation remains unknown (Schwabe et al., 2010). This lack of understanding makes treatment inherently complex and requires a very high level of expertise from the clinician. Moreover, most indicators of patients' progress rely on behavior measures and the clinician's expertise.

This article presents the development of the backend of a computer aided diagnosis (CAD) for mental health care, in particular for the treatment of chronic stress related disorders. This backend will be validated with two previously gathered sets of clinical data (van den Broek et al., 2009; van den Broek et al., 2011; van der Sluis et al., 2010; van der Sluis et al., 2011). Its foundation lay in speech signal analysis and the processing of the signal's features by a Multi-Layer Perceptron neural network (MLP). The envi-

van der Sluis F., Dijkstra T. and L. van den Broek E..
COMPUTER AIDED DIAGNOSIS FOR MENTAL HEALTH CARE - On the Clinical Validation of Sensitive Machines.
DOI: 10.5220/0003891404930498
In *Proceedings of the International Conference on Health Informatics* (BSSS-2012), pages 493-498
ISBN: 978-989-8425-88-1

sioned CAD system can be used as a decision support system in everyday life (e.g., at work) and in mental health care settings.

The remaining article is composed as follows. Next, we will present a brief section that will describe the data set with which the CAD framework will be clinically validated. Section 3 will describe relevant speech signal features and their parameters. Section 4 will present the results of the actual stress classification and validation employed in this study. Last, Section 5 will present a general discussion, with which this article end.

2 DATA SET

Various stimuli have been applied in the endeavor to trigger stress in a controlled manner, including, images, sounds, (fragments of) films (van den Broek and Westerink, 2009), and real-world experiences (Healey and Picard, 2005). However, how do we know which methods actually triggered participants' true stress? This is a typical concern of validity, which is a crucial issue for stress assessment.

Although understandable from a measurement-feasibility perspective, stress measurements are often done in controlled lab settings (cf. the Trier Social Stress Test (Kirschbaum et al., 1993)). This makes results poorly generalizable to real-world applications (van den Broek, 2010). Moreover, under normal circumstances, in our every day lives, bursts of significant stress are sparse, which makes it even more difficult to obtain such data (in a limited time frame) (Picard, 2010). However, luckily, previously we already obtained two data sets of clinically validated data that comprises bursts of authentic chronic stress (van den Broek et al., 2009; van den Broek et al., 2011; van der Sluis et al., 2010; van der Sluis et al., 2011), which we will briefly introduced here.

In total, 25 female Dutch Post-Traumatic Stress Disorder (PTSD) patients (mean age: 36; SD: 11.32) participated of their own free will. All patients suffered from panic attacks, agoraphobia, and panic disorder with agoraphobia (Sánchez-Meca et al., 2010). Before the start of the studies, all patients signed an informed consent form and all were informed of the tasks they could expect. The data from one patient with problems in both studies were omitted from further analysis. Hence, the data of 24 patients were used for further analysis.

All patients took part in two studies: a storytelling (ST) study and a reliving (RL) study. Possible factors of influence (e.g., location, apparatus, therapist, and experiment leader) were kept constant. In the ST study, the participants read aloud both a stress-provoking and a positive story; see also (van den Broek et al., 2011; van der Sluis et al., 2010; van der Sluis et al., 2011). This procedure allows considerable methodological control over the invoked stress, in the sense that every patient reads exactly the same stories. The fictive stories were constructed in such a way that they would induce certain relevant emotional associations. The complexity and syntactic structure of the two stories were controlled to exclude the effects of confounding factors. In the RL study, the participants re-experienced their last panic attack and their last joyful occasion; see also (van den Broek et al., 2009; van den Broek et al., 2011). For RL , a panic attack approximates the trauma in its full strength, as with the during admission of a patient. The condition of telling about the last experienced happy event resembles that of a patient who is relaxed or (at least) in a 'normal' stress condition.

To evaluate the quality of our speech analysis, we compared the results of our speech analysis to those obtained by means of a standard questionnaire: the Subjective Unit of Distress (SUD) (Wolpe, 1958). The SUD has repeatedly proven itself as a reliable measure of a person's experienced stress. The SUD will serve as the ground truth in our quest to develop a CAD system for mental health care. The CAD should be robust, enable to process a variety of data. For this purpose, we have chosen to treat both the speech and the SUD data of both studies as one set. Consequently, the assessment of their relation was put to the test by such a diverse data set and enabled the illustration of the robustness of this relation. For more detailed analyses, we refer to (van den Broek et al., 2012).

The SUD is measured by a Likert scale that registers the degree of distress a person experiences at a particular moment in time. In our case, we used a linear scale with a range between 0 and 10 on which the experienced degree of distress was indicated by a dot or cross. The participants in our study were asked to fill in the SUD test once every minute; consequently, it became routine during the experimental sessions.

3 SPEECH SIGNAL PROCESSING

Speech was recorded using a personal computer, an amplifier, and a microphone. The sample rate of the recordings was 44.1 kHz, mono channel, with a resolution of 16 bits. All recordings were divided into epochs of approximately one minute of speech. For each of the two conditions of both experiments, 3 epochs of one minute of speech were taken. Because the therapy sessions were held under controlled conditions in a room shielded from noise, high quality speech signals were collected.

This section will describe the features that were extracted from the speech signal, using the Praat software package [1]. Subsequently, the basic preprocessing conducted on the speech signal will be described, namely: outlier removal, data normalization, and parameter derivation from the complete set of features.

3.1 Speech Signal Features

From the speech signals, various features have been shown to be sensitive to experienced stress; for a recent survey, see (El Ayadi et al., 2011). In this research, we extracted five features from the speech signal, namely: *i*) its wave amplitude (Scherer, 2003); *ii*) power, often used interchangeably with energy and intensity, which is also described as the Sound Pressure Level (SPL) , relative to the auditory threshold P_0 (i.e., in decibel (dB) (SPL)) (Cowie et al., 2001); *iii*)the zero-crossings rate of the speech signal (Yang and Lugger, 2010); *iv*) the high-frequency power (Banse and Scherer, 1996; Cowie et al., 2001; Yang and Lugger, 2010): the power for the domain $[1000, 22000]$, denoted in Hz; and *v*) the fundamental frequency (F0 or perceived pitch) (Cowie et al., 2001; Scherer, 2003; Yang and Lugger, 2010), extracted using an autocorrelation function (i.e., the cross-correlation of the signal with itself).

13 statistical parameters were derived from the 5 speech signal features: mean, median, standard deviation, variance, minimum value, maximum value, range, the quantiles at 10%, 90%, 25%, and 75%, the inter-quantile-range $10\% - 90\%$, and the inter-quantile-range $25\% - 75\%$. The features and statistical parameters were computed for each minute of speech sample over a time window of 40 ms, using a step length of 10 ms (i.e., computing each feature every 10 ms over the next 40 ms of the signal). This short term processing approach takes care of time varying spectral information and is in line with the generally accepted standards (Rossing et al., 2007). Two variations of amplitude are reported, one in which the parameters are calculated from the mean amplitude per window of 40 ms, and one where the features are calculated over the full signal (reported as amplitude(full)). In total, 6×13) = 78 parameters were determined on the basis of the speech signal features.

The same procedure for outlier removal was executed for all speech features. It was founded on the inter-quartile range *iqR*, which we define as: $q_3 - q_1$, with q_1 being the 25th percentile and q_3 being the 75th percentile. Next, data points x were removed from the data set if $q_1 - 3iqR \geq x \geq q_3 + 3iqR$. All other data points (i.e., that satisfied this requirement) were kept in the data set.

To achieve good classification results with pattern recognition and machine learning methods, the set of selected input features is crucial. The same holds for classifying stress. To limit this enormous feature space, a Linear Regression Model (LRM) - based heuristic search was applied, using $\alpha \leq 0.1$, which can be considered as a soft threshold. An LRM was generated using all available data, starting with the full set of parameters, and then reducing it in 32 iterations by means of backward removal, to a set of 28 parameters. The LRM model explained 59% ($F(28, 351) = 18.22$, $p < .001$) of the variance.

4 CLASSIFICATION AND VALIDATION

A principal component analysis (PCA) was used to (further) reduce the dimensionality of the set of speech signal parameters, while preserving its variation as much as possible. 28 parameters were fed into the PCA transformation. Subsequently, the first 11 principal components from the PCA transformation were selected, covering 95% of the variance as was explained by the fill set of 28 parameters. These principal components provided a condensed representation of the LRM and, as such, served as input for the MLP that will be introduced next.

The MLP has been used as state-of-the-art machine learning technique. It are universal approximators, capable of modeling complex functions. Moreover, MLPs can adequately handle irrelevant inputs and noise and can adapt their weights and/or topology in response to environmental changes. They are used for classification, providing discrete outputs, but also for regression with numeric outputs and reinforcement learning when output is not perfectly known. For a proper introduction to this classifier, we refer to the many handbooks and survey articles that have been published. Here, we will only specify the MLP, for purpose of replication. We computed WEKA's (Hall et al., 2009) MLP trained by a backpropagation algorithm (in its binary mode). It used gradient descent with moment and adaptive training parameters. In our case, a MLP with 3 layers with 7 nodes in the hidden layer was shown to have optimal topology. This topology was trained with 500 cycles. The nodes in this network were all sigmoid. For all

[1] http://www.praat.org by P. P. G. Boersma and D. J. M. Weenink [Last accessed on December 09, 2011]

other settings, the defaults of WEKA were used (Hall et al., 2009).

We conducted a cross-validation of the (precision of the) SUD with the parameters of the speech signal features that are classified by the MLP. On the one hand, this verifies the validity of the SUD ; on the other hand, this determines the performance of the classifier in objective stress detection. The MLP was tested using 10-fold cross-validation. Its average performance is reported in Table 1.

4.1 Validation

The SUD ranged from 0 to 10, giving 11 classes to classify. However, none of the patients used SUD score 10; hence, this class was omitted from the classification process. We classified the SUD scores over both studies, including both conditions and their baselines; see also Section 2.

The SUD is an established instrument in psychology; nevertheless, to the authors' knowledge the instrument's precision has not been assessed. People's interoception is said to be unreliable (Craig, 2002), which calls for an assessment of the reliability of a SUD with a relatively high precision (i.e., range: $0-10$). However, although interoception is possibly indeed prone to errors, it has been reported that patients with anxiety disorders are (over)sensitive to interoception (Domschke et al., 2010). Please note that this can possibly be explained by the influence chronic stress has on human memory (Schwabe et al., 2010)

We used the SUD as a ground truth. To assess the precision of the SUD , the scale was quantized into all possible numbers of levels (i.e., from 2 to 10); see Table 1. This quantization is performed by reassigning the SUD responses into N steps, with a step size of r/N, where r is the range of the SUD values (i.e., $0-9$). This quantization allowed the assessment of the reliability of the SUD in relation to the LRM. Moreover, to provide a fair report on the MLP's classification, we provide both the correct classification rate (C_N), delta classification rate (dC_N), and the relative classification rate (rC_N) for each of the N bins. dC_N is a standard correction, also known as delta or reaction score/classification, which is defined as

$$C_N - \mu_N, \tag{1}$$

with μ_N being the chance level for N classes. rC_N expresses the improvement of the classification compared to chance level, which is defined as

$$\frac{C_N - \mu_N}{\mu_N} \times 100, \tag{2}$$

rC_N is also known as a range correction and used more often in health research (Fillingim et al., 1992).

The relative classification rate (see Eq. 2) enables the assessment of the true classification performance on each level of quantization of the SUD . Figure 1 shows that the MLP has an almost monotone linear increase in relative classification rate. Its linear fit follows the data in Table 1 nicely (explained variance: $R^2 = .95$). Moreover, Figure 1 shows that the classification rate of the MLP is almost constant, independent of the level of quantization of the SUD . These fits underline the validity of the SUD as an instrument to assess people's stress levels. The fits also illustrate the SUD 's high concurrent validity, with its ability to discriminate between up to 10 levels of stress. Moreover, the fits indicate that the use of the SUD as ground truth for stress assessment is adequate.

Table 1: The classification results (in %) of the Multi-Layer Perceptron neural network (MLP). Correct classification (C_N), chance level for classification (μ_N), delta classification rate (dC_N; see Eq. 1) and relative classification rate (rC_N; see Eq. 2) are reported. The Subjective Unit of Distress (SUD) was taken as ground truth. N indicates the number of SUD levels employed.

N	C_N	μ_N	dC_N	rC_N
2	82.4 %	50.0 %	32.4 %	64.7 %
3	72.4 %	33.3 %	39.0 %	117.1 %
4	57.4 %	25.0 %	32.4 %	129.5 %
5	49.0 %	20.0 %	29.0 %	144.7 %
6	47.6 %	16.7 %	31.0 %	185.8 %
7	42.4 %	14.3 %	28.1 %	196.6 %
8	41.6 %	12.5 %	29.1 %	232.6 %
9	34.7 %	11.1 %	23.6 %	212.6 %
10	36.3 %	10.0 %	26.3 %	263.2 %

5 DISCUSSION

We have explored the feasibility of CAD for mental health care, which can help both in daily life and in therapy. To assure a clinically valid assessment of stress, previously collected data of 25 PTSD patients was used, see also Section 2. The stress level of the patients was assessed by the SUD and their speech characteristics were mapped upon the SUD ; hence, a behavioral and an indirect physiological measure. The MLP neural network was used to classify the speech sample, with the SUD scores as ground truth. Correct classification rates of 82.4%, 72.4%, and 36.3% were achieved on, respectively, 2, 3, and 10 SUD levels. Given the fact that the complete research is conducted on patients in their regular clinical setting, this underlined the feasibility of a CAD for mental health care.

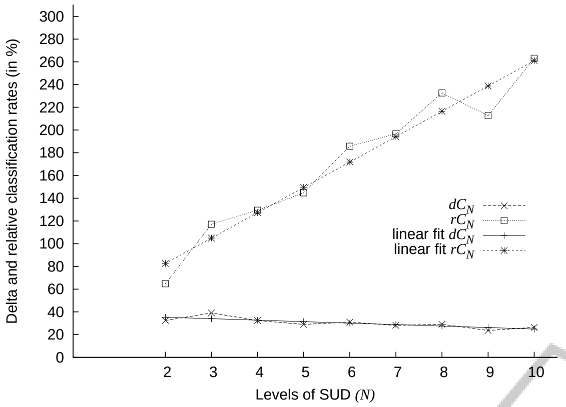

Figure 1: The relative classification results (in %) of 11 principal components based on 28 parameters of speech features using a Multi-Layer Perceptron neural network (MLP). The Subjective Unit of Distress (SUD) was taken as ground truth, using quantizations between 2 and 10. Both the delta classification rate (dC_N; see Eq. 1) and the relative classification rate (rC_N; see Eq. 2) are reported as well as the linear fit of their performance in relation to the level of quantization of the SUD.

A detailed report of the two studies conducted can be found in (van den Broek et al., 2011). Additional analyses in line with the analyses presented in this article are reported in (van den Broek et al., 2012). These analyses distinguish between the two studies conducted. Moreover, in addition to the MLP, *k*-nearest neighbors (*k*-NN) and a support vector machine were employed. The interested reader is kindly directed to (van den Broek et al., 2012). Moreover, it would be of interest to validate the current signal processing and pattern recognition pipeline on new (unseen) data sets. Such data sets could comprise, for example, different patient groups and/or different methods of emotion elicitation.

A limitation of the current research can be found in the unbalanced data set (He and Garcia, 2009). Not all SUD scores were chosen equally by the patients nor is their distribution Gaussian. Possibly, this issue is even more prominently present in the quantization of the SUD scores in a smaller number of bins. Hence, the chance level for classification (μ_N) as reported in Table 1 is not the actual chance level. The linear fits presented in Figure 1 are possibly not as strong as presented in this figure (He and Garcia, 2009). However, the unbalanced data set could also have declined the chance level and, hence, the to be expected classification level. Therefore, we feel it is justified to give this straightforward intuitive representation of the classification rates for the different quantizations.

The success of machines in sensing emotions by way of the speech signal ranges from 25% correct classification on 14 emotions (Banse and Scherer, 1996) to 73.5% correct classification on 6 emotions (Yang and Lugger, 2010). However, these results are not stable and have been shown to be hard to replicate. This is well illustrated by the structured benchmark conducted by (Schuller et al., 2011), who report up to 71% (2 classes) and 44% (5 classes) correct classification. Apart from the differences in classification rate and the number of classes to be distinguished, these studies can both be questioned with respect to their ecological validity of the experienced emotions. Often they employ methods to elicit stress that are not validated on their effectiveness. Exceptions to this, such as the Trier Social Stress Test (Kirschbaum et al., 1993), are seldom used in engineering and applied sciences. Therefore, we feel the need to stress that a careful interpretation of laboratory results is needed. A one-on-one mapping between lab and real-world results cannot be taken for granted (Picard, 2010; van den Broek, 2010). The current research deviates from common practice of speech-based stress recognition in its use of a clinically valid data set.

In sum, a leap was made towards modeling stress through an acoustic model, which can be applied in our daily lives and in mental health care settings. By the specific research design, it was ensured that bursts of authentic chronic stress were measured. The precision of the SUD as an instrument to assess experienced stress, as was claimed by clinical practitioners, was confirmed. Moreover, a set of features derived from the speech signal was shown to enable the detection of stress using an MLP neural network. This shows that unobtrusive and ubiquitous automatic assessment experienced stress is both possible and promising and can already be applied as a reliable instrument in clinical settings.

ACKNOWLEDGEMENTS

We thank all Post-Traumatic Stress Disorder (PTSD) patients who participated voluntarily in this research. We are grateful to four reviewers for their helpful suggestions on an earlier draft of this article. Moreover, we thank Lynn Packwood (HMI, University of Twente, NL) for proofreading this article.

REFERENCES

Banse, R. and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3):614–636.

Brosschot, J. F. (2010). Markers of chronic stress: Prolonged physiological activation and (un)conscious perseverative cognition. *Neuroscience & Biobehavioral Reviews*, 35(1):46–50.

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

Craig, D. A. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8):655–666.

Domschke, K., Stevens, S., Pfleiderer, B., and Gerlach, A. L. (2010). Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical Psychology Review*, 30(1):1–11.

El Ayadi, M., Kamel, M. S., and Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3):572–587.

Fillingim, R. B., Roth, D. L., and Cook III, E. W. (1992). The effects of aerobic exercise on cardiovascular, facial EMG, and self-report responses to emotional imagery. *Psychosomatic Medicine*, 54(1):109–120.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18.

He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.

Healey, J. A. and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166.

Kirschbaum, C., Pirke, K. M., and Hellhammer, D. H. (1993). The "Trier Social Stress Test"– A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1–2):76–81.

Picard, R. W. (2010). Emotion research by the people, for the people. *Emotion Review*, 2(3):250–254.

Rossing, T. D., Dunn, F., Hartmann, W. M., Campbell, D. M., and Fletcher, N. H. (2007). *Springer handbook of acoustics*. Berlin/Heidelberg, Germany: Springer-Verlag.

Sánchez-Meca, J., Rosa-Alcázar, A. I., Marín-Martínez, F., and Gómez-Conesa, A. (2010). Psychological treatment of panic disorder with or without agoraphobia: A meta-analysis. *Clinical Psychology Review*, 30(1):37–50.

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

Schuller, B., Batliner, A., Steidl, S., and Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, 53(9–10):1062–1087.

Schwabe, L., Wolf, O. T., and Oitzl, M. S. (2010). Memory formation under stress: Quantity and quality. *Neuroscience & Biobehavioral Reviews*, 34(4):584–591.

Van den Broek, E. L. (2010). Robot nannies: Future or fiction? *Interaction Studies*, 11(2):274–282.

Van den Broek, E. L., van der Sluis, F., and Dijkstra, T. (2009). Therapy Progress Indicator (TPI): Combining speech parameters and the subjective unit of distress. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume 1, pages 381–386, Amsterdam, The Netherlands. IEEE Press.

Van den Broek, E. L., van der Sluis, F., and Dijkstra, T. (2011). *Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients*, volume 12 of *Philips Research Book Series*, chapter 10, pages 153–180. Dordrecht, The Netherlands: Springer Science+Business Media B.V.

Van den Broek, E. L., van der Sluis, F., and Dijkstra, T. (2012). Cross-validation of bimodal health-related stress assessment. *Personal and Ubiquitous Computing*, 16:[in press].

Van den Broek, E. L. and Westerink, J. H. D. M. (2009). Considerations for emotion-aware consumer products. *Applied Ergonomics*, 40(6):1055–1064.

Van der Sluis, F., van den Broek, E. L., and Dijkstra, T. (2010). Towards semi-automated assistance for the treatment of stress disorders. In *HealthInf 2010: Proceedings of the Third International Conference on Health Informatics*, pages 446–449, Valencia, Spain. INSTICC – Institute for Systems and Technologies of Information, Control and Communication.

Van der Sluis, F., van den Broek, E. L., and Dijkstra, T. (2011). Towards an artificial therapy assistant: Measuring excessive stress from speech. In *Proceedings of the International Conference on Health Informatics*, pages 357–363, Rome, Italy. Portugal: SciTePress.

Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition*. Stanford, CA, USA: Stanford University Press.

Yang, B. and Lugger, M. (2010). Emotion recognition from speech signals using new harmony features. *Signal Processing*, 90(5):1415–1423.