

TOWARDS SEMI-AUTOMATED ASSISTANCE FOR THE TREATMENT OF STRESS DISORDERS

Frans van der Sluis

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

Egon L. van den Broek

Human-Centered Computing Consultancy (H-CCC), The Netherlands

Ton Dijkstra

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

Keywords: Stress, Diagnosis, Indicator, Speech.

Abstract: People who suffer from a stress disorder have a severe handicap in daily life. In addition, stress disorders are complex and consequently, hard to define and hard to treat. Semi-automatic assistance was envisioned that helps in the treatment of a stress disorder. Speech was considered to provide an excellent tool for providing an objective, unobtrusive emotion measure. Speech from 25 patients suffering from a stress disorder was recorded while they participated in two storytelling sessions. The Subjective Unit of Distress (SUD) was determined as a subjective measure and enabled the validation of the derived speech features. A regression model with four speech parameters (i.e., signal, power, zero crossing ratio, and pitch), was able to explain 70% of the variance in the SUD measure. As such it lays the foundation for semi-automated assistance for the treatment of patients with stress disorders.

1 INTRODUCTION

Stress is indisputably a major factor in modern life. This is illustrated by the voluminous stress related literature that has appeared. In 1936, Hans Selye popularized the concept of stress by calling it the “*general adaptation syndrome*” (Selye, 1936); i.e., a problematic coping with noxious stimuli. Already more than half a century ago, stress was often mentioned together with life events and illness, where an inability to cope with the life events can lead to stress, and where stress can lead to illness. As such, stress has been recognized as one of the potential factors contributing to disease in general (Rabkin and Struening, 1976), making it a tremendously important construct from a health perspective.

A few prevalent stress-related psychiatric disorders are: Post-Traumatic Stress Disorder (PTSD), depression, and insomnia. The different disorders can be explained by different aspects of stress. However, the denominating factor seems to be a chronic stress

response; either in the onset of the illness (e.g., depression) or as a symptom of the illness (e.g., PTSD).

The diagnosis of stress-related psychiatric disorders is inherently difficult. Each disorder includes a broad variety of symptoms and diagnostic criteria. One of the key diagnostic criteria is the existence of excessive stress, whether or not in relation to a specific stressor. Moreover, for some disorders a repeated diagnosis of stress response can be used to indicate therapy progress (American Psychiatric Association, 2000).

However, the detection of excessive stress is complicated. A clinician has a range of questionnaires and diagnostic criteria available to support this aim. However, these methods rely on introspection and the expert opinion of the clinician. Inherently, subjective measures can be unreliable; e.g., when a patient is not straightforward in his answering or when a patient complies too much with other expectations. Moreover, standardized questionnaires are often a burden on the patient. An expert opinion is limited as well,

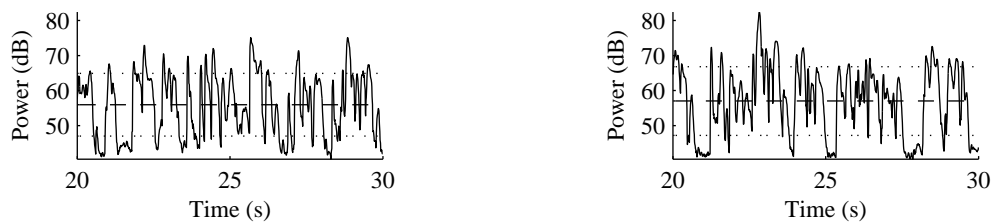


Figure 1: Energy of an illustrative part of the speech signal. Left: anxiety-inducing, right: happy-inducing condition. The dashed lines show the mean and standard deviation.

especially when the stress response is less profound or the stressor is less clear. This makes subtle differences, such as required for treatment progress, difficult to measure. Hence, (inter-expert) reliability can be an issue. In sum, in order to support measurement, assist in decision making, and help with tracking the treatment progress, therapists are still looking for an objective method not solely dependent upon introspection or expert opinion. The next section will discuss this challenge.

In the last few decades, emotion research received a lot of interest. In this period, the research areas of stress and emotion were to some extent found to be complementary and similar; e.g., as Lazarus (1993) stated it: *“Psychological stress should be considered part of a larger topic, the emotions”* (p. 10).

A broad range of methods exist for the automatic detection of emotions. A literature review reveals that these signals can be assigned to physiological measures, movement analysis, computer vision techniques, and speech processing (Cowie et al., 2001; Van den Broek et al., 2009).

This research focusses on speech since it has a number of advantages: 1) The communication in therapy sessions is often recorded anyway. Hence, no additional technological effort has to be made on the side of the therapists; 2) Obtrusiveness plays no role with speech processing; 3) The degree of noise that distorts the speech signal is limited, because therapy sessions are generally held under controlled conditions in rooms shielded from noise. Moreover, speech has been indicated to hold information about the psychophysiological state of speaker, with foreseen applications in other health-related as well as non-health related areas (Lutfi et al., 2009).

Regrettably, most research on stress detection through speech suffers from two problems, which makes it hard to compare previous studies and methods. First, many results are based upon mimicked emotions; i.e., acted vs. experienced emotions. Second, a ground truth is often lacking, making it unclear if the measured vocal cues actually represent an induced affective state. Please consult Scherer (2003) for a more elaborate view on the problems. Hence, we

present a feasibility study to indicate how well stress can be measured from speech in Section 2, followed by a discussion of the possibility of a diagnostic support system in Section 3.

2 FEASIBILITY STUDY

The goal of the feasibility study is to induce stress similar to how it is experienced in a therapy session, and using this to find speech features related to stress.

2.1 Method

In this study, 26 female PTSD patients (mean age: 38) voluntarily participated. All patients signed an informed consent. For several reasons, PTSD patients were used. Namely, this group of patients is relatively sensitive to stress and, thus, to stress inducing stimuli. They become earlier stressed and were expected to react better to emotion elicitation. Furthermore, considering the context of the study, using real patients increases its ecological validity.

The research consisted of four phases, each aimed at triggering an affective state at the patient. The first and last phase involved the recording of a neutral baseline for both speech and the ground truth. The second and third phase were aimed at triggering either a happy or an anxious state. Hence, anxiety was used to induce stress.

Story telling was used to elicit emotions. This method allows great methodological control over the invoked emotion; i.e., every patient reads exactly the same story. Moreover, contrary to many methods used in speech and emotion research (e.g., mimicking emotions), story telling is expected to yield true emotions. Furthermore, story telling automatically leads to speech.

The patients had to read aloud two stories, describing an anxious and a happy situation. The stories were controlled on their complexity and on their syntactic structure, as to prevent any interfering factors. The order of both stories was counterbalanced over

Table 1: Pearson's correlations between Subjective Unit of Distress (SUD) and derived speech features.

Feature	Parameters												
	IQR10	IQR25	Max	Mean	Median	Min	Q10	Q25	Q75	Q90	Range	Std	Var
F0				-0.276‡	-0.248‡		-0.173*	-0.283‡	-0.224†	-0.245‡			
ZC						-0.326‡	-0.228‡						
HFE	-0.440‡	-0.307‡	-0.209†	-0.147*	-0.221†	0.166*	0.142*		-0.239‡	-0.234‡	-0.347‡	-0.413‡	-0.387‡
E	-0.437‡	-0.374‡			-0.18†	0.168*	0.157*		-0.249‡	-0.223†	-0.306‡	-0.425‡	-0.402‡

Note. * $p < .05$. † $p < .01$. ‡ $p < .001$

the participants. Before the patients read the stories, they were asked to read a sample story to familiarize themselves with the task.

Two methods were used to measure stress: 1) speech processing and 2) a subjective measure, serving as a ground truth for 1).

In order to measure stress from speech, several steps were performed. First, the signal was recorded at a sample rate of 44.1 kHz, mono channel, and 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 on the ground truth, explained further on. Second, the recorded signal was 'cleaned': speckle noise and other voices were removed from the signal.

To enable the validation of the parameters derived from speech, a subjective measurement was needed. For this, the Subjective Unit of Distress (SUD) suited optimally. It is a Likert scale, which registers the amount of (dis)stress a person experiences at a certain moment. In our case, a linear scale with range 0-10 was used on which a dot or cross should be placed. In 1958, Wolpe introduced the SUD. Since then, the SUD has proved to be a reliable measure to determine a person's emotional state. The subjects were asked to use the SUD every minute; so, throughout the experiment it became a routine. The SUD served as the ground truth for further analysis; see also Section 2.2.

Using the clean signal, the following features were extracted and compared to the ground truth: pitch (F0), energy (E) (Cowie et al., 2001; Scherer, 2003; Ververidis and Kotropoulos, 2006), high-frequency energy (HFE) (Cowie et al., 2001; Rothkrantz et al., 2004), and zero-crossings rate (ZC) (Kedem, 1986; Rothkrantz et al., 2004). Although there is no general consensus regarding the best speech parameters for stress detection, there is a fair amount of evidence for the affective information in these features. Hence, these features were extracted from the audio signal; see Figure 1 for samples of the features.

All features were computed using a time window of 40msec. and a step length of 10msec. Several statistical parameters were calculated for each feature. The less common ones are the 10%, 25%,

75%, and 90% quartiles, further-on denoted by Q , and the inter-quartile ranges $Q90\% - Q10\%$ ($IQR10$) and $Q75\% - Q25\%$ ($IQR25$).

2.2 Results

Using a Multivariate Analysis of Variance (MANOVA), no direct effects of story telling condition (happy or anxiety) or time (first, second, or third minute of story telling) on SUD scores were found, nor did a significant interaction effect appear. Looking at only the anxiety condition, an Analysis of Variance (ANOVA) showed a trend for time on SUD scores ($F(2,56) = 2.726, p = .07$). This indicates that patients reported experienced stress later-on in the course of the story telling. Since there is a large amount of variance ($mean = 3.03, std = 2.56$), it is likely that inter-personal differences caused the non-significant result. Moreover, this variability is useful for the goal of this study; i.e., whether or not subjectively reported stress can be explained through speech features.

There was a strong relation between acoustic features and the SUD scores; Table 1 shows the significant Pearson's correlations. Furthermore, a linear regression model (\mathcal{M}) was created using only the emotion inducing conditions; i.e., the SUD scores of the anxiety and happy conditions. Here, a \mathcal{M} including all features and parameters (i.e., 40 predictors), explained 69.72% of the variance: $R^2 = .697, F(40,99) = 5.70, p < .001$.

3 DISCUSSION

Through a feasibility study, this research showed the possibility of assisting clinicians in the diagnosis of stress-related psychiatric disorders. Moreover, some generic speech features allowing the creation of an assistive system have been uncovered. Considering the various difficulties in the diagnosis and treatment of stress-related psychiatric disorders, such an assistive system can be expected to be an important step forward towards creating more objective clinical meth-

ods.

In the feasibility study, stress was successfully caused and reported by 26 subjects. By measuring speech and a subjective report of stress, acoustic features of stress in speech were determined. These features were able to explain 70% of variance of subjectively reported experienced stress. Hence, demonstrating the possible success of speech as an objective measure of experienced stress.

The reported stress, induced by story telling, was quite dispersed. Although this is partly due to inter-personal difference, this also indicates that overall the stories did have an influence. Moreover, a trend was found for the anxiety inducing story, corroborating this influence. These results not only suggest the value of story telling, but also its drawbacks. Two problems can be identified. First, stories are heavily dependent on their temporal course; i.e., a story needs a build-up before inducing an affective state. Second, there were substantial inter-personal differences in the experience of the stories. However, contrary to many other methods, this method is likely to create true emotions. The triangulation through various speech characteristics and the SUD did indicate that indeed true emotions were triggered through the story-telling.

Considering the number of patients used to create an acoustic profile of stress characteristics in speech, the achieved explained variance of 70% for the emotional conditions is high. In particular, being a non-personalized profile, some generic features of stressful speech seem to be uncovered. However, also some restrictions apply: a) only PTSD patients were used, other patient groups might show other stress responses; b) different kinds of stress may exist; and c) any restrictions applying to story telling as emotion elicitation method may have affected the results. This triplet can be considered as future research challenges. Namely, to use other patient groups, different stressful emotions, and different emotion elicitation techniques.

This study has demonstrated that giving a second opinion based on the speech signal is feasible. An assistive system can help the clinical setting through several ways: 1) to support the measurement of a stress response; 2) to assist in deciding whether or not the patient has excessive stress; and 3) to aid in the treatment of a stress disorder. Therefore, by making the diagnosis objective, the measurement is made more reliable; i.e., by no longer solely relying on introspection. Hence, objective measurement increases inter- and intra-expert reliability and helps diagnosis, decision-making, and treatment become more fine-grained.

ACKNOWLEDGEMENTS

The patients suffering from a post-traumatic stress disorder (PTSD), who voluntarily participated in this research, are gratefully acknowledged. Further, we thank the anonymous reviewers for their critical and constructive comments on the original manuscript. In addition, we would like to acknowledge Paul Boersma and David Weenink (Institute of Phonetic Sciences, University of Amsterdam, The Netherlands) for their work on Praat and the accompanying manual, tutorials, and articles.

REFERENCES

- American Psychiatric Association (2000). *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, USA: American Psychiatric Publishing, Inc., 4 (Text Revision) edition.
- 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.
- Kedem, B. (1986). Spectral analysis and discrimination by zero-crossings. *Proceedings of the IEEE*, 74(11):1477–1493.
- Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual Review of Psychology*, 44(1):1–22.
- Lutfi, S. L., Montero, J. M., Barra-Chicote, R., Lucas-Cuesta, J. M., and Gallardo-Antolín, A. (2009). Expressive speech identifications based on hidden markov model. In Azevedo, L. and Londral, A. R., editors, *HEALTHINF*, pages 488–494. INSTICC Press.
- Rabkin, J. G. and Struening, E. L. (1976). Life events, stress, and illness. *Science*, 194(4296):1013–1020.
- Rothkrantz, L. J. M., Wiggers, P., van Wees, J.-W. A., and van Vark, R. J. (2004). Voice stress analysis. *Lecture Notes in Computer Science (Text, Speech and Dialogue)*, 3206:449–456.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1–2):227–256.
- Selye, H. (1936). A syndrome produced by diverse noxious agents. *Nature*, 138(3479):32.
- Van den Broek, E. L., Janssen, J. H., Westerink, J. H. D. M., and Healey, J. A. (2009). Prerequisites for Affective Signal Processing (ASP). In Encarnação, P. and Veloso, A., editors, *Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 426–433, Porto – Portugal.
- Ververidis, D. and Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9):1162–1181.
- Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition*. Stanford, California: Stanford University Press.