DESIGNING THE DATABASE OF SPEECH UNDER STRESS

RÓBERT SABO[1] – JAKUB RAJČÁNI[2]
[1]Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
[2]Faculty of Arts, Comenius University, Bratislava, Slovakia

SABO, Róbert – RAJČÁNI, Jakub: Designing the Database of Speech Under Stress. Journal of Linguistics, 2017, Vol. 68, No 2, pp. 326 – 335.

Abstract: This study describes the methodology used for designing a database of speech under real stress. Based on limits of existing stress databases, we used a communication task via a computer game to collect speech data. To validate the presence of stress, known psychophysiological indicators such as heart rate and electrodermal activity, as well as subjective self-assessment were used. This paper presents the data from first 5 speakers (3 men, 2 women) who participated in initial tests of the proposed design. In 4 out of 5 speakers increases in fundamental frequency and intensity of speech were registered. Similarly, in 4 out of 5 speakers heart rate was significantly increased during the task, when compared with reference measurement from before the task. These first results show that proposed design might be appropriate for building a speech under stress database. However, there are still considerations that need to be addressed.

Keywords: stress, arousal, stress detection, heart rate, speech under stress, speech database

1 INTRODUCTION

Research in the field of speech processing is increasingly drawn to specific manifestations of speech such as speech under stress. This area of speech research is closely linked with psychology and physiology, which should answer the question, what is stress and how to identify it in speech. Our goal in this study is to establish methodology for creating a database of speech under real stress, which may be used in other experiments investigating speech under stress in future.

2 RESEARCH OF STRESS IN SPEECH, EXISTING SPEECH DATABASES

One of the most widely used speech databases in speech under stress is the SUSAS database – Speech Under Simulated and Actual Stress [1], [2]. The database consists of four domains, encompassing a wide variety of stresses and emotions. It contains 32 speakers (13 female, 19 male), with ages ranging from 22 to 76 years who have made more than 16 000 utterances. SUSAS also contains several longer speech files from four Apache helicopter pilots and a common highly confusable vocabulary set of 35 aircraft communication words. Unfortunately, in carrying out acoustic analyses, researchers are limited by noisy channel and the 8 kHz sampling frequency.

Speech database containing speech under stress with high quality recordings is the CRISIS database [3]. This database contains acted expressive speech from 15

DOI 10.1515/jazcas-2017-0042

speakers. Each speaker records a set of 150 sentences, each in different arousal level. Once in a neutral manner (referred to as level 1 of tense arousal), then with higher imperativeness, like a serious command or directive (level 2), and finally like an extremely urgent command or statement being declared in a situation when human lives are directly in danger (level 3). Even though high-quality recordings allow to perform a number of acoustic analyses [4], [5], database is a missing part with speech under realistic stress.

In our approach, we propose a method to obtain high-quality recordings of speech under real stress. One of the important questions, that needs to be answered first, is: What is stress and how it can be measured?

2.1 Definition of Stress

Proposing a scientific definition of stress is a difficult problem, in a large part, due to the term being too general and hardly usable in different contexts [6]. In the general sense, stress is a state in which internal integrity (or homeostasis) of an individual is challenged via external or internal means – called stressors [7], [8].

Stress results in a complex physiological reaction, which can be marked by changes in bodily systems, such as autonomic nervous system (ANS), endocrine and immune system [8]. Sympathetic branch of ANS becomes predominant during stress reaction, which leads to acceleration of heart rate (HR), secretion of noradrenaline and adrenaline, as well as inhibition of gastrointestinal function, changes in electrodermal activity (EDA) and many other physiological changes. All these bodily reactions serve as preparatory measures for behavioral reaction to stress and successful adaptation. Increased preparatory physiological activation, may be labelled by term "arousal" which is also used in context of emotions as a level of overall physiological activation.

Investigating stress biomarkers, such as heart rate, electrodermal changes, stress hormones, etc., are a large part of current stress research. On the other hand, speech changes in stress are not so well examined. Though, there are studies investigating speech changes, they differ in proposed understanding of stress and used methods and therefore yield different results.

Current research shows that speech changes that are a result of both involuntary bodily changes and voluntary effort, are also dependent on a particular type of stressor. Hansen [9] proposed a taxonomy of stressors and their impact on speech, based on the mechanism in which they perturbate speech process. Stressors were sorted to several categories such as: "zero order" – stressors with direct physical impact on speech (e.g. acceleration), "first order" – biological or chemical stressors (e.g. dehydration), "second order" which involves perception (e.g. Lombard effect) and "third order" – psychological, emotional and social stressors.

Besides a lot of research findings on stress detection from studies using acted stress, studies of real-life stress also show a detectable difference in speech. Lu et al. [10] obtained stress identification accuracy of 71.3% when comparing job interview with indoor neutral speech, and accuracy 82.9% when personalized model was used. Increased skin conductance level as an electrodermal stress related phenomenon was used to validate stress during job interview. Similarly, Luig, et al. [11] proposed heart rate and heart rate variability as relevant physiological correlates to speech analysis.

Presence of stress may be detected via physiological, but also from speech parameters. This study follows findings of prior research on analysis of speech under stress. Our aim is to develop a database of speech under realistic stress and to further validate it by using physiological indicators, such as heart rate and electrodermal activity.

For obtaining relevant data, we have chosen laboratory setting aiming to implement these issues: 1. It must induce a strong enough stress reaction, 2. Person under stress needs to speak as much as possible, 3. The setting must be relevant to real-life applications.

One of the problems with speech databases of real stress is that speakers may speak very little, or that utterances included in the database are too short to yield any notable results. Our previous work with acted stress enabled us to detect stress with high accuracy, however, conducting a study of realistic stress is necessary [3]. For these reasons the following design was proposed.

3 METHOD

3.1 Research Setting

Based on the mentioned aims, we decided to use a communication-based task, in which the research subject must give instructions on solving the task to their partner via microphone. In a setup like this, subject is forced to speak as much as possible, however, inducing a strong enough stress reaction is also essential. This was realized considering following stress factors:

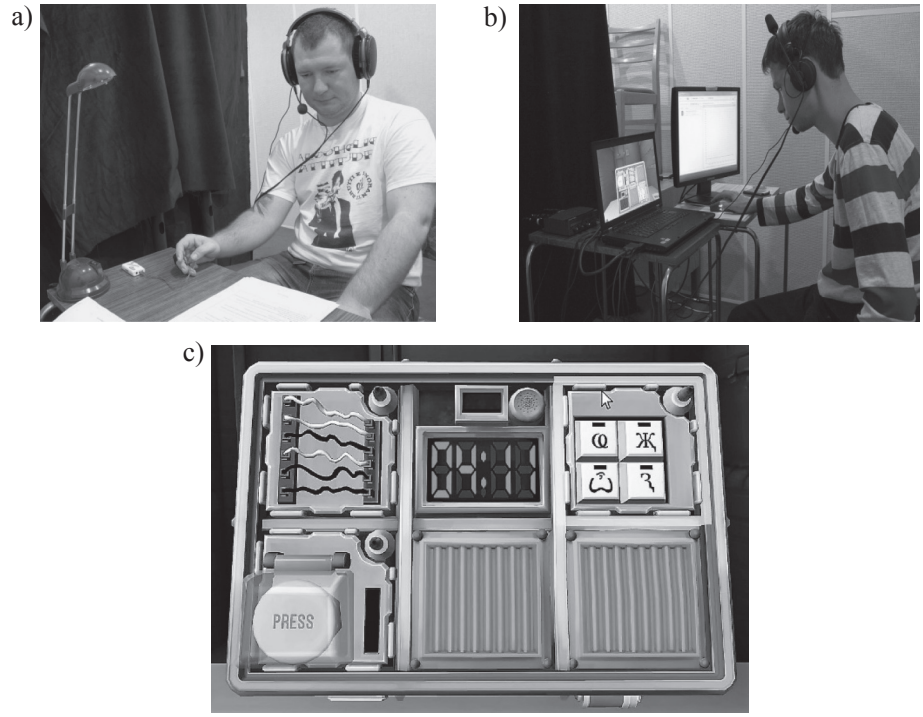

Fig. 1. Photo of a) research subject, b) researcher controlling the bomb on screen, c) bomb interface.

A) Task itself. For the purpose of data acquisition, we adapted a commercially available computer game "Keep talking and nobody explodes" [12]. The game itself is a moderate stress inducing task, in which two players dismantle a bomb composed of several modules, each representing a logical puzzle (Fig. 1c). While one player sees the bomb on screen (in our case a member of research team, to provide standard conditions for all subjects) (Fig. 1b), the other (a research subject) sees a manual with detailed instructions on solving individual puzzles (Fig. 1a). Two players do not see each other and they communicate only via microphone.

B) Time pressure. The game itself has a countdown timer, which can be adjusted to the task. Subject in our task sees the timer and hears beeping sounds in the headphones. Time pressure of 10 min for solving entire bomb composed of 6 modules provides a very hard, yet solvable task.

C) Environmental factors. During the task, subject is sitting in recording studio with lights off, only using a table lamp. At random times during dismantling the bomb, subject is disturbed by a siren in the headphones.

D) A reward and a set "best score." It is expected, that a research subject who solves the task has at least some degree of motivation to achieve a good result. For a task like this to become a stressor, it needs to be important and consequential to the subject. Therefore, we added incentives to enhance subjects' motivation. One incentive is financial reward. Subjects are instructed, that both they and their co-player (to increase their feeling of responsibility) will receive reward depending on their performance. They are told, they both receive 10€ for dismantling a bomb successfully, if they fail, they receive 1€ for each successfully solved module (Entire testing consists of three consecutive bombs so the reward can go up to 30€). Second incentive is information, that if players break the record, which is set to 8 min for solving the bomb, their reward doubles (20€ for solving a bomb).

To meet the conditions for quality of the recording testing was realized in an acoustically treated recording studio. Recordings were obtained via head-mounted close-talk microphone Sennheiser ME3 and Emu Tracker Pre USB audio interface with 48 KHz sampling frequency and 16 bit resolution. Participants used high-quality closed headphones Sennheiser HD 650. Each speaker was recorded in separate channel.

3.2 Participants and Procedure

The first recorded sample of the speech database contains speech from five speakers in the Slovak language. Subject A: female, 47 years; subject B: male, 28 years; subject C: male, 29 years; subject D: male, 45 years; subject E: female, 47 years.

Participants were contacted with a request to participate in a communication experiment, in which their voice and psychophysiology (heart rate – HR and electrodermal activity – EDA) will be recorded. All subjects were informed of the research procedure and signed informed consent. First, subjects were given bomb manual to study for 20 minutes to become acquainted with the game mechanics. Before studying the manual subjects were told they will communicate with a co-player, who is also a subject playing for the same reward. During debriefing after the test, subjects were explained, that the co-player was a member of research team, they could talk together and all subjects' questions about the research were answered.

The recording consisted of 10-minutes training game, which was realized using the same task with the experimenter in an easygoing manner. Training was used to collect reference values; stress factors were not present during training. Subsequently, three trials using the described procedure with the co-player were realized.

In this first test, the selection of speakers was not strictly limited of age and sex. Number of subjects in the research sample is only preliminary for initial tests of the research setting.

3.3 Analyzed Speech Features

The fundamental frequency (F0) and intensity values are specific for neutral speech of each speaker. Changes in frequency and intensity of speech can point to changes in speaker's emotional state. In the first data analysis, we evaluated mean fundamental frequency and mean intensity of speech for each task (training, trial 1, trial 2, trial 3), which represent approximately 8 minutes of speech for each task. When analyzing such long period of time, impact of various phonetic content and non-speech events such as hesitations should not be significant.

3.4 Physiological Measures

Beat to beat heart rate signal was obtained from all test subjects using FAROS 90° ECG device (Fig. 2a). Measurement of ECG was carried out using two electrodes, one positioned under right clavicle, the other on the left under ribs. Sampling rate for ECG was 250Hz, which is appropriate for high precision ECG and HR analysis.

a) b)

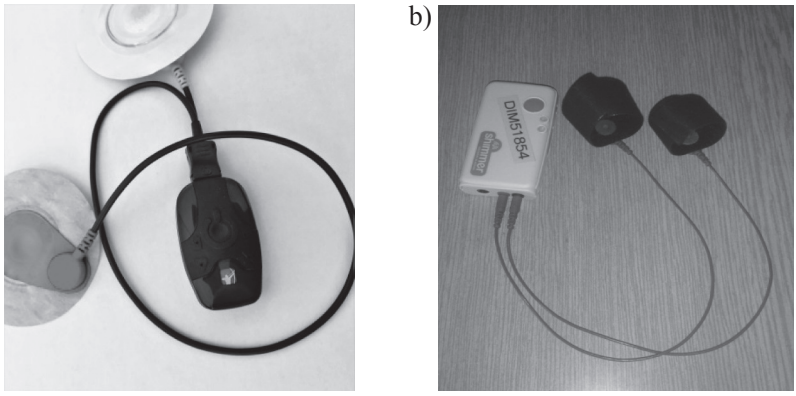

Fig. 2. a) FAROS 90° ECG device b) Consensys Shimmer GSR device

Though heart rate can be used as a reliable index of overall arousal and sympathetic activity, we may also calculate heart rate variability (HRV), which offers more information on autonomic nervous system influences of heart [11], [13].

Although Heart rate can be calculated from duration of single beat to beat interval, it is necessary to take HR changes during breathing cycle to consideration. Therefore, we analyzed 10s HR intervals, corresponding to events which occurred during data acquisition. On the other hand, heart rate variability measures can be reliably calculated only from longer segments of HR (at least 2–4 min) [13].

Electrodermal activity (EDA) is another useful indicator in stress research, which was previously used as a reference measure in study of speech [10]. From possible electrodermal phenomena, we measured skin resistance (in kΩ) via Consensys, Shimmer device (Fig. 2b). Both tonic, relatively stable skin resistance level and phasic, skin resistance responses can be further analyzed as stress indicators. This study will not include results from EDA analysis.

3.5 Subjective Stress Assessment

For assessment of subjective experience of stress and anxiety, we administered Slovak version of state anxiety inventory (STAI-X) [14]. STAI-X inventory is composed of 20 statements to which subjects answer on a 4-point scale. Test is used to describe an extent, to which a person feels anxiety at the given time. This inventory may be used for repeated measurements; we administered it before and after recording.

Moreover, after the recording, subjects also answered several standard questions regarding their motivation, feelings of stress and satisfaction with the achieved result.

4 RESULTS

Of all the subjects in the research sample, none could dismantle any of the given bombs in time, however two subjects were able to solve 5 of 6 modules before the bomb exploded.

Data analysis showed differences in both speech features and heart rate. Table 1 summarized increases in F0 and intensity of speech.

Speaker ID	Task ID	F0 [Hz]	Intensity[dB]
A	Training	173	60.9
	Trial 1	195	67.6
	Trial 2	198	68.3
	Trial 3	198	68.3
B	Training	144	64.8
	Trial 1	153	68.4
	Trial 2	153	69
	Trial 3	153	68.5
C	Training	126	52
	Trial 1	128	57
	Trial 2	126	57.8
D	Training	129	66
	Trial 1	153	79.9
	Trial 2	173	83.6
	Trial 3	170	82.8
E	Training	240	62.5
	Trial 1	246	52.5
	Trial 2	247	52.6
	Trial 3	242	52.6

Tab. 1. Average values of F0 and Intensity for training and each trial

4 of 5 speakers (except C) proved a significant increase of speaker's fundamental frequency in average of 14% (Fig. 3).

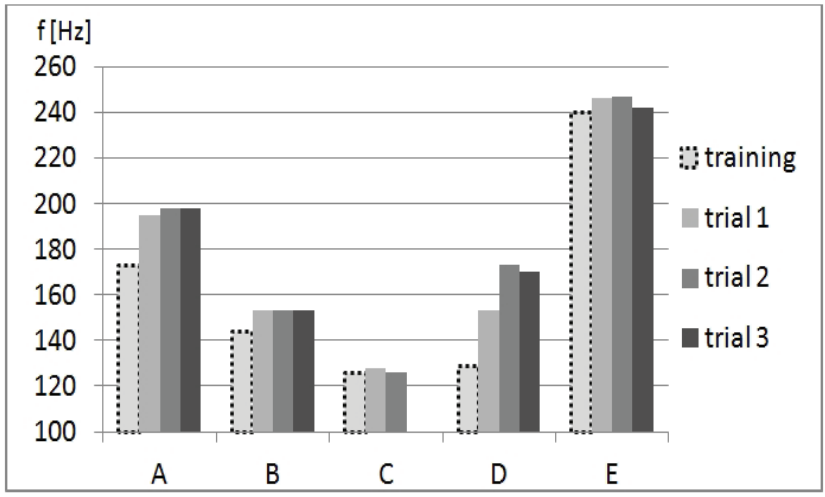

Fig. 3. Fundamental frequency for each speaker and each task

4 of 5 speakers (except E) proved an increase of the speech intensity in average of 16%.

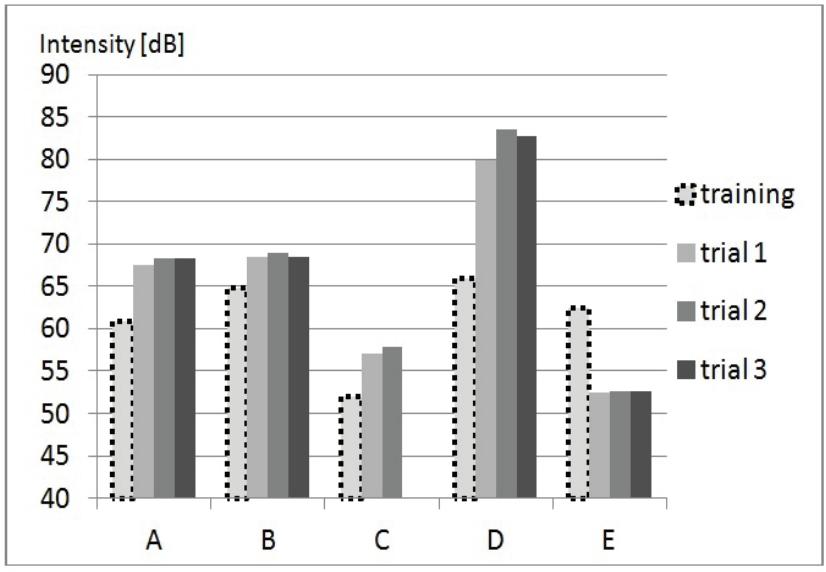

Fig. 4. Intensity of speech signal for each speaker and each task

Following Table 2 shows changes in heart rate (HR) expressed in beats per minute between training and three trials for each subject.

Speaker ID	Heart rate [bpm] Training	Trial
A	82.21	87.35 82.79 83.18
B	87.04	93.96 96.64 90.30
C	67.87	69.57 74.85
D	80.92	84.92 85.94 85.72
E	108.20	110.98 109.36 106.51

Tab. 2. Average values heart rate (in beats per minute) for each subject and trial

Figure 5 illustrates changes in HR during entire recording. It contains detailed analysis of 10s HR windows from the recording (data from subject "D" were chosen for illustration). Increases in HR during individual trials (dismantling of bomb 1, 2 & 3) may be observed in Figure 5.

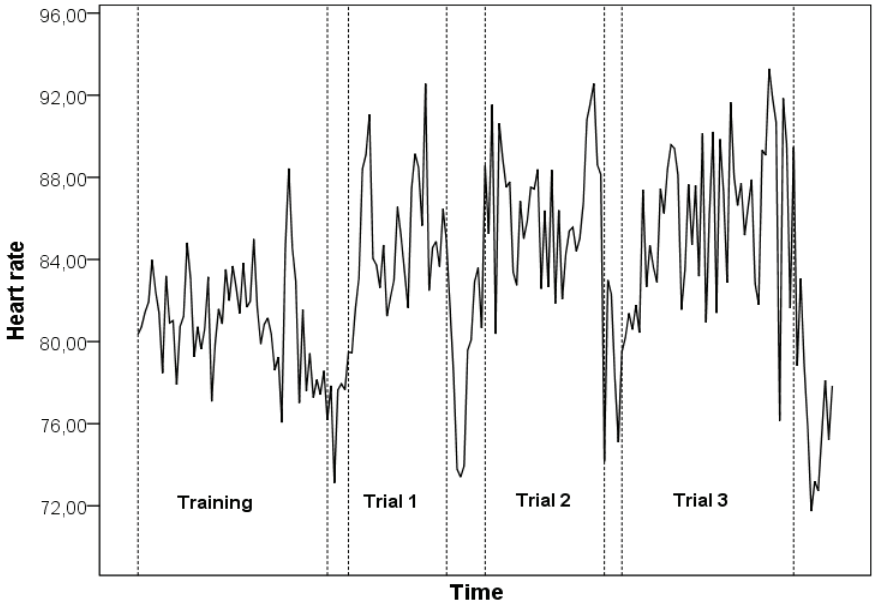

Fig. 5. Changes in HR during recording – differences between test and three trials (subject "D"). HR was sampled from 10s windows.

5 DISCUSSION

5.1 Speech Analysis and Physiological Findings

The initial analysis of five obtained recordings of speech under real stress show that proposed design should be appropriate for data acquisition. Even though a significant increases of F0 and intensity were observed only in 4 out of 5 speakers, balanced values between the trials point out, that speech obtained by the proposed method in trials is acoustically different from speech obtained in training. To identify whether this acoustic difference was induced by stress, analysis of psychophysiological correlates of stress was performed.

Findings from heart rate clearly indicate increase of physiological distress during trials. Moreover, as showed in Figure 5, increases of HR peaked in the last moments before the bomb exploded. However, due to low number of subjects so far, we cannot statistically evaluate these differences for the whole sample.

It is also important to note, that heart rate differs between individuals in a large extent. Factors such as age and sex must be taken into consideration when interpreting the HR data.

5.2 Further Methodological Issues and Considerations

Research Subjects. In the following data collection using this design, it is important to test at least 20 subjects, all of which fall into one age category. Because of using a computer game as a task interface, young subjects (age 18–30) would be optimal. Secondly, if we want to compare men and women, larger sample with balanced groups will be necessary. It is important that test subjects are naive to the task before participating in the study.

Design changes. In the first test, we needed to minimalize variables considered; therefore, we decided to use a member of the research team to stand in the role of co-player. The main advantage of this setup is that every subject had similar conditions during the game as their co-player responded in a standard manner (as somebody who sees this task for the first time). Moreover, if the subject spoke very little, or the utterances were very short, co-player encouraged them with asking more questions about the task at hand. However, there is a possibility of using two groups of subjects for both player positions in the game. This alternative may be useful to collect more speech data, however, subjects from different player positions will hardly be comparable.

Linguistic point of view. The speech in the database contained interesting linguistic, phonetic phenomena such as hesitations, repetitions, changes in speech rate, etc. High-quality stereo recordings allow us to perform precise analysis of overlapping speech patterns. If a design with real subjects on both player positions were used, the database might be a suitable for research on turn taking in speech.

In future, we plan to evaluate also other relevant acoustic features such as F0 maximum, intensity maximum, root mean square, spectral energy distribution etc., and also evaluate shorter speech segments, possibly related to annotated events during the game. A detailed phonetic annotation at the level of statements, words, phonemes will be carried out.

Other possible expansion of the database might be inclusion of another language, besides Slovak, if the prepared Slovak database yields good results in obtaining speech under stress.

ACKNOWLEDGEMENTS

The research leading to the results presented in this paper has received funding from the European Union FP7 under grant agreement n° 312382 (GAMMA – Global ATM Security Management project [15]. This study was also supported by grants APVV-0496-12, VEGA 1/0739/17.

References

[1] Hansen, J. H., Bou-Ghazale, S. E., Sarikaya, R., and Pellom, B. (1997). Getting started with SUSAS: a speech under simulated and actual stress database. *Eurospeech*, 97(4):1743–1746.

[2] Hansen, J. H. SUSAS LDC99S78. Web Download. Philadelphia: Linguistic Data Consortium, 1999. Accessible at: https://catalog.ldc.upenn.edu/LDC99S78.

[3] Sabo, R., Rusko, M., Ridzik, A., and Rajčáni, J. (2016). Stress, Arousal, and Stress Detector Trained on Acted Speech Database. In *International Conference on Speech and Computer*, pages 675–682.

[4] Rusko, M., Darjaa, S., Trnka, M., Sabo, R., and Ritomský, M. (2014). Expressive Speech Synthesis for Critical Situations. *Computing and Informatics*, 33(6):1312–1332.

[5] Rusko, M., Darjaa, S., Trnka, M., Ritomský, M., and Sabo, R. (2014). Alert!... Calm Down, There is Nothing to Worry About. Warning and Soothing Speech Synthesis. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 1182–1187, Reykjavik, Iceland.

[6] Newport, D. J., and Nemeroff, C. B. (2002). Stress. In Ramachandran, V. et al., editors, *Encyclopedia of Human Brain*. vol. 4, pages 129–139, Academic Press.

[7] Mc Ewen, B., and Lupien, S. (2002). Stress: Hormonal and Neural Aspects. In Ramachandran, V. et al., editors, *Encyclopedia of Human Brain*. vol. 4, pages 129–139, Academic Press.

[8] Chrousos, G. P. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology*, 5(7):374–381. Accessible at: http://doi.org/10.1038/nrendo.2009.106.

[9] Hansen, J. H. L. et al. (2000). The Impact of Speech Under 'Stress' on Military Speech Technology. NATO PROJECT 4 REPORT.

[10] Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., and Choudhury, T. (2012). Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 351–360, ACM New York, NY, USA.

[11] Luig, J., Sontacchi, A., Goswami, N., Moser, M., and Shaw, C. (2010). Conception and Realization of Speech Recordings for Instantaneous Stress Level Assessment. In *9th EUROCONTROL Innovative Research Workshop and Exhibition*, Nice, France.

[12] Computer game *Keep talking and nobody explodes*. Accessible at: http://www.keeptalking-game.com.

[13] Berntson, G. G., Thomas Bigger Jr. J., Eckberg, D. L., Grossman, P., Kaufmann, P. et al. (1997). Heart rate variability: Origins methods, and interpretive caveats. *Psychophysiology*, 34(6):623–648.

[14] Müllner, J., Ruisel, I., and Farkaš, G. (1980). Príručka pre administráciu, interpretáciu a vyhodnocovanie dotazníka na meranie úzkosti a úzkostlivosti. Psychodiagnostické a didaktické testy. 93, Bratislava.

[15] GAMMA – Global ATM Security Management project. Accessible at: http://www.gamma-project.eu.