



OPEN

DATA DESCRIPTOR

Speech production under stress for machine learning: multimodal dataset of 79 cases and 8 signals

Jan Pešán^{1,6}, Vojtěch Juřík^{2,3,6}✉, Alexandra Ružičková^{2,6}, Vojtěch Svoboda^{2,6}, Oto Janoušek^{4,6}, Andrea Němcová⁴, Hana Bojanovská², Jasmína Aldabaghová², Filip Kyslík², Kateřina Vodičková², Adéla Sodomová², Patrik Bartys², Peter Chudý^{1,5} & Jan Černocký¹

Early identification of cognitive or physical overload is critical in fields where human decision making matters when preventing threats to safety and property. Pilots, drivers, surgeons, and operators of nuclear plants are among those affected by this challenge, as acute stress can impair their cognition. In this context, the significance of paralinguistic automatic speech processing increases for early stress detection. The intensity, intonation, and cadence of an utterance are examples of paralinguistic traits that determine the meaning of a sentence and are often lost in the verbatim transcript. To address this issue, tools are being developed to recognize paralinguistic traits effectively. However, a data bottleneck still exists in the training of paralinguistic speech traits, and the lack of high-quality reference data for the training of artificial systems persists. Regarding this, we present an original empirical dataset collected using the BESST experimental protocol for capturing speech signals under induced stress. With this data, our aim is to promote the development of pre-emptive intervention systems based on stress estimation from speech.

Background & Summary

In recent decades, Machine Learning (ML) has rapidly grown as an industry sector, promoting significant advancements in the recognition and classification of human speech. Despite early research on speech processing dating back to the 1960s, technological limitations hindered widespread adoption until cheap and accessible Graphics Processing Units (GPUs) became available. The wide availability of GPUs opened the research avenues in Neural Networks (NNs). This progress led to a general improvement in ML performance, namely Natural Language Processing (NLP), image processing and also speech processing. The most prominent field in speech processing is Automatic Speech Recognition (ASR), which transcribes speech from audio recordings. In addition to ASR, supplemental tasks such as Gender Identification (GID), Language Identification (LID), and Speaker Identification (SID) can be provided. The growing importance of metadata related to speech transcriptions has created a market to recognize emotions, health, age, and other information from speech. Stress detection, however, is rapidly developing due to its relevance in key areas of human activity. The concept of stress has been known since ancient Rome¹, but its systematic study in a physiological sense did not begin until the 19th century with Claude Bernard's theory of "milieu intérieur"² and Walter Cannon's extension of this concept to a theory of homeostasis³. Cannon also linked psychological and psychosomatic symptoms and proposed that prolonged exposure to fear could result in death⁴. The Fight-or-Flight response, which he developed with Philip Bard, is a widely accepted theory that a mix of different physiological processes prepares the body for fighting or fleeing in response to an acute stressor. Based on this work, more research has been conducted on the application of speech-based features to stress estimation and the development of multimodal datasets. John Hansen's⁵ early work explored the features of stressed speech, but limited data hindered progress. Hansen later collected data

¹Speech@FIT, Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic. ²Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czech Republic. ³Institute of Computer Aided Engineering and Computer Science, Faculty of Civil Engineering, Brno University of Technology, Brno, Czech Republic. ⁴Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic. ⁵Institute of Computer Science, University of Würzburg, Würzburg, Germany. ⁶These authors contributed equally: Jan Pešán, Vojtěch Juřík, Alexandra Ružičková, Vojtěch Svoboda, Oto Janoušek. ✉e-mail: jurik.vojtech@mail.muni.cz

in cooperation with North Atlantic Treaty Organization (NATO) to establish initial stress-related databases. He identified four main features for stress estimation: intensity, pitch, duration of words, and vocal tract spectrum. Lombard effect needs to be taken into account w.r.t. these features, as it may negatively impact its descriptive power. Tet Fei Yap's doctoral thesis⁶ explored the effects of cognitive load on speech and found that formant frequencies, although lower-dimensional than Mel-Frequency Cepstral Coefficients (MFCC), were comparable in performance for cognitive load classification systems. These advances in research help to study the very nature of stress demonstration in speech, but there is still a considerable lack of datasets to support the research efforts. Currently, the available data sources for stress analysis from speech face two main problems: data availability and data accuracy. Only the Speech Under Simulated & Actual Stress Database (SUSAS)^{7,8} database from the Linguistic Data Consortium (LDC) is easily accessible for academic purposes, but its nature (particularly its small number of samples and high variance of the stress scenarios) makes it unsuitable for NN training. The other databases (Cognitive Load Corpus with Speech and Performance Data from a Symbol-Digit Dual-Task (CoLoSS), Speech under Stress Conditions-0 (SUSC-0), Speech under Stress Conditions-1 (SUSC-1) and partly Cognitive Load with Speech and EGG (CLSE)⁶) are difficult to access and also do not meet the requirements for conversational and spontaneous sentences with appropriate stress load labels. To address this need, we created a new database with subjective self-determined and objective biological based ground truth labels. In this paper, we present a dataset collected using a specific methodology and experimental protocol called Brno Extended Stress and Speech Test (BESST)⁹ and which aims to provide empirical data for the development of pre-emptive intervention systems based on stress estimation from speech.

Methods

For the collection of necessary data, the BESST, which is an adaptation of the original Maastricht Acute Stress Test (MAST)¹⁰, was developed for the Czech context. It involves participants performing stress-inducing tasks while being recorded by multiple cameras and microphones. The first task involves immersion of a non-dominant hand in ice water, while the second is a dual task based on Reading Span task (RSPAN). The data presented in this study was collected in April 2022 by the Faculty of Information Technology - Brno University of Technology (FIT-BUT) and the Department of Psychology at the Faculty of Arts at Masaryk University (PSYCH-PHIL-MUNI).

Participants. The research sample was represented by 90 people (21F, 69M; Caucasian young adults). Due to technical issues and data loss, 79 cases (19F, 60M) were eventually included in the final dataset. All participants taking part in the study had to be between the ages of 19 and 26 years old. Participants were recruited through an ad on social networks, university websites, and posters placed around the FIT-BUT. Some were also recruited through the snowball method or by an email invitation from Masaryk University (MUNI) to students from the Faculty of Arts. A reward was promised for participation (irrespective of whether they finished the whole session), in the form of a food voucher worth about 100 CZK (around 4 EUR). Participants were informed of exclusion criteria (epilepsy, current medications, acute psychological or physiological issues, treatments or illnesses, heart conditions), and checked if Czech was their native language. Any heart conditions could have been dangerous for participants exposed to icy water, and the Czech language was imposed to increase the homogeneity of the speech data for analysis. Participants were given their unique ID to allow pseudonymization and later complete anonymization of the data.

Ethics statement. This study was assessed and approved by the Ethical Committee Faculty of Electrical Engineering and Communication For Biomedical Research, Brno University of Technology, Brno, Czech Republic under number EK:02b/2022 ("Stress and Cognitive Loads Effects on Speech"). The study was carried out according to relevant guidelines and regulations following the principles of the Declaration of Helsinki.

Participation was voluntary, participants were explicitly informed that they could leave from the experiment at any time without any negative consequences or loss of promised remuneration. The details about the experiment, recorded modalities (e.g. face, bio, and voice recordings) and overall conditions of the experiment were specified in the informed consent. Informed consent was read and signed by the participant with three different signatures: 1) Experiment participation consent 2) GDPR clause 3) Consent with the data distribution.

Procedure. The goal of the research was to explore speech production under stress and to use the findings of this work to create an automatic speech classifier. Each subject recorded a 45-60 minute session at FIT-BUT. During the session, noninvasive measurements of participants' physical state were performed. Specifically, 1) heart electrical activity (Electrocardiogram (ECG)) using the FAROS sensor and 2) electrodermal activity (Electro-Dermal Activity (EDA)) using the EMPATICA bracelet were measured. The session was 3) recorded with multiple cameras, and the voice signal was 4) recorded with several microphones. All introductory information about the experiment was provided to participants upon arrival in the experimental room, and then informed consent was signed. Prior to the experimental run, ECG electrodes and a bracelet to measure EDA, skin temperature and heart rate were attached to the participants. Electrodes were placed on the anterior chest, one under the right clavicle, and the other on the left side of the rib cage, at the farthest point from the heart. The participants then moved to the testing area, where instructions were presented on a computer screen and they were asked to complete the long-term anxiety level Perceived Stress Scale 14 (PSS14)¹¹ questionnaire on a tablet. Cameras were turned on, recording the participant from different angles (face-frontal, bucket, posture-left, posture-right), and a headset was turned on to record the voice. Another microphone recorded the participant's voice on the table in front of him/her in a distance of approximately 50cm. After filling in the questionnaire, the session started with a 3-minute relaxation segment, during which the participants were instructed to stay calm. All lights in the room were turned off for the duration of relaxation. Baseline measurements of participants' resting state (baseline:

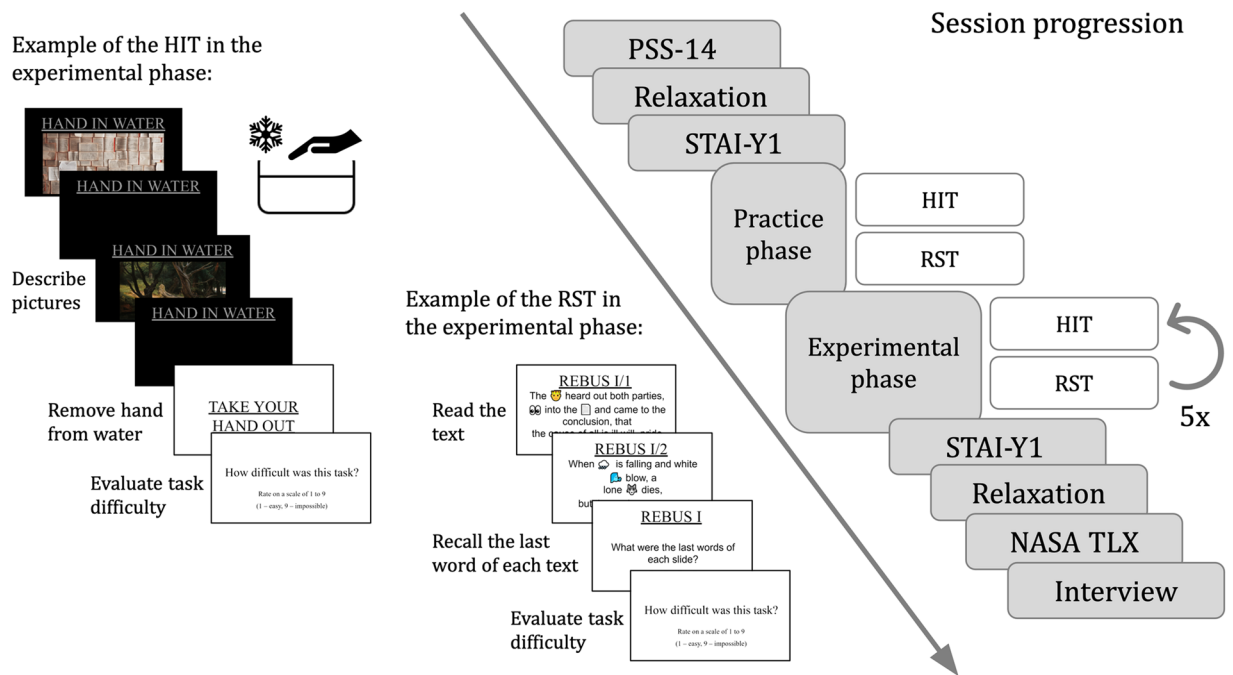


Fig. 1 Scheme of the BESST experimental session; figured materials have been translated for the purpose of this paper, during the data collection, they were presented in Czech language.

resting ECG, EDA) were taken. After the resting segment, participants were asked to complete a questionnaire about their current anxiety level State-Trait Anxiety Inventory-Y1 (STAI-Y1). To better manipulate the stress level of the participants, one of the researchers responsible for preparing the participant for the experiment and conducting the final interview behaved in a friendly and warm manner, while the other researcher leading the experimental part behaved coldly and directly toward the participants. Another factor intended to raise the level of psychosocial stress of the participants was the output of the cameras visible on computer screens in front of the participant throughout the experimental phase.

The experimental phase consisted of two alternating tasks as can be seen in Fig. 1:

- Hand Immersion Task (HIT) - Participant's hand was immersed in cold water while they were asked to describe the pictures they saw on the screen.
- Reading Span task (RSPAN) - It included sentences in the form of rebuses presented to participants with the assignment to read them out loud and remember the pre-defined words.

In HIT, participants had to immerse their hand up to their elbow in a bucket of ice cold water (approximately 5–10 °C), while describing what they saw in the pictures presented to them on the computer screen. At the end of each segment, participants rated how subjectively difficult the task was on a scale from 1 (“Easy”) to 9 (“Impossible to complete”). The RSPAN task required participants to read short texts in which some words were replaced with an image of the same meaning. Participants were asked to verbally express these words and read the full paragraphs. To increase cognitive load, participants had to memorize the last word of each text and at the end of each segment, the words had to be reproduced in the correct order. The number and difficulty of the texts increased with each block. Again, after each block, participants were asked to rate the difficulty of this task on a scale from 1 to 9. Both tasks alternated, HIT five times, RSPAN four times, totaling 9 segments. The experimental phase was preceded by a practice presentation of both tasks, during which participants did not immerse their hand in cold water. At the end of the experimental procedure, participants completed the second STAI-Y1 questionnaire, relaxed again for 3 minutes, and answered the NASA Task Load Experience (NASA-TLX) questionnaire. A short interview was conducted on their experience of the experiment. Each participant subsequently received his/her reward and was de-instrumented.

For more details on the acquisition process, see¹².

Sensors & Instruments

Self-reported measures.

1. Perceived Stress Scale 14 (PSS14)¹¹ - Participants were asked to report on their trait stress level before the experimental session started. The PSS-14 questionnaire contains 14 items and it was designed to measure the degree to which situations in one's life are appraised as stressful. The questions focus on assessing subjective feelings of uncontrollability, unpredictability, and overload of the participants' lives. Participants are asked to respond using a five-point Likert scale from “never” (0) to “very often” (4) to items such as “In

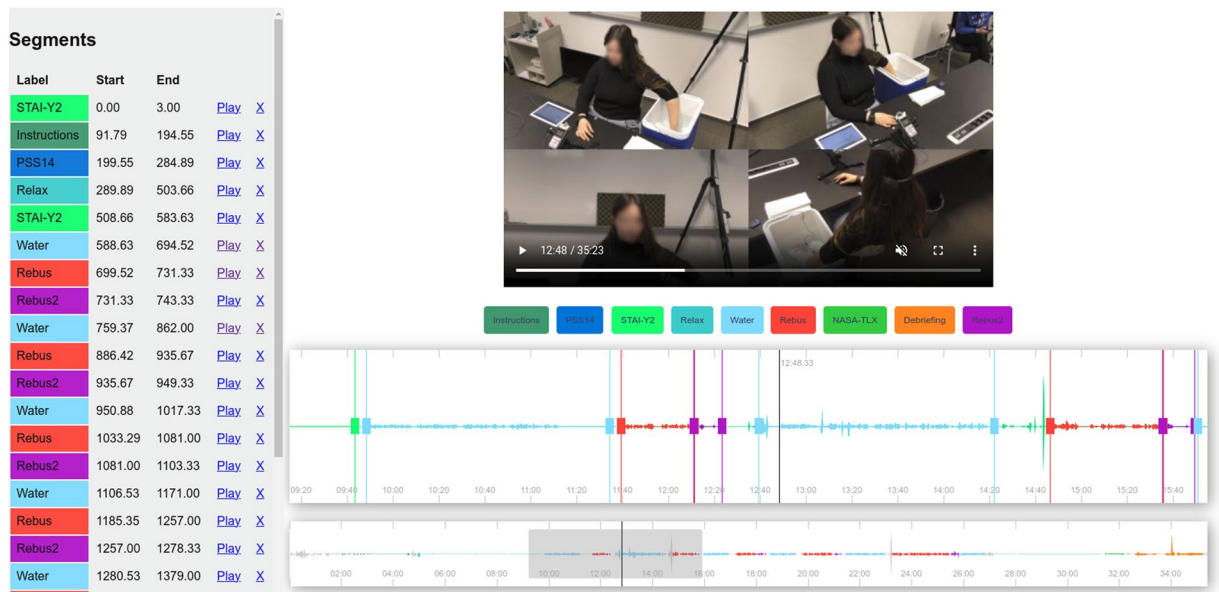


Fig. 2 Screenshot of one annotated session within the BESSTiANNO tool.

the last month, how often have you felt anxious for something that happened unexpectedly?” or “In the last month, how often have you felt that you were on top of things?”.

2. State-Trait Anxiety Inventory (STAI)¹³ - Participants were asked to report their current stress level before the experimental session started. The State-Trait Anxiety Inventory Questionnaire is a widely used tool for measuring anxiety. It consists of two parts, each with 20 items. In this research, the section that measures state anxiety (STAI-Y1) was used to assess participants’ current/state feelings. The participants rated their agreement with statements such as “I feel tense” or “I feel excited” on a scale of “not at all” (1), “a little” (2), “rather yes” (3), or “very” (4).
3. Difficulty rating (Likert scale¹⁴) - After each block (HIT/RSPAN), participants were asked to rate the subjective load of the block they just completed on a scale from 1 (“Easy”) to 9 (“Impossible to complete”).
4. NASA Task Load Experience (NASA-TLX)¹⁵ - The questionnaire explored the cognitive load experienced during the entire experimental session, as reported by the participants. It consists of six 20-point Likert rating scales, with their edge values marked as “low” or “high”, and the rest of the scale without labels. Participants are to rank the completed task on these six scales, respectively, concerning mental demand, physical demand, temporal demand (time required to perform the task), performance, effort, and frustration level.

Objective measures.

1. Heart electrical activity captured by the Faros 180 (<https://www.bittium.com/medical/bittium-faros>) sensor
2. Electro-Dermal Activity (EDA) captured by the Empatica E4 (<https://www.empatica.com/research/e4/>) bracelet
3. Audio recordings captured by multiple microphones (1x Shure SM-35XLR7 headset; 1x XY microphone through the Zoom H4n recorder)
4. Video recordings using 4x Panasonic HC-VX9805 camera.

Data Acquisition and Pre-processing

Data segmentation tool. The BESSTiANNO tool was designed to provide precise segmentation and manual verification of audio and video streams. It was inspired by the VEST: Video Event Segmentation Tool (<https://github.com/jcheong0428/vest>), an open-source software used for annotating events in video clips. However, the original VEST tool was limited as a browser-based JavaScript application, which was not suitable for larger and distributed annotation tasks. To address this issue, the BESSTiANNO implementation was improved by incorporating a Python Flask backend and an SQLite database, making it a more reliable solution for larger scale annotation operations. An example of the BESSTiANNO interface is shown in Fig. 2.

Data Records

High-level data summary is in Table 1. The dataset is available on⁹.

Data streams. The captured data was represented by the streams specified in the Table 2.

Measurement(s)	Heart electrical activity (electrocardiogram, ECG); skin conductivity (electro dermal activity, EDA); speech; video; Stress level (initial current/acute);
Technology Types(s)	ECG - FAROS sensor; EDA - EMPATICA E4 bracelet; speech - microphones SHURE SM-35XLR7 headset/XY microphone connected to Zoom H4n recorder; Video sequences - Panasonic HC-VX9805 cameras; Stress - PSS14/STAI Y2; Cognitive load - NASA TLX
Factor Type(s)	Physiological stress; Psychological (Cognitive load) stress
Sample Characteristic - Organism	Homo sapiens
Sample Characteristic - Environment	Laboratory environment - controlled experimental context
Sample Characteristic - Location	Czech Republic

Table 1. Dataset summary.

Modality	Details	Count	Device	File format
Audio	Main stream, 48000 Hz, 2 channels, 24 bit, Pulse Code Modulation (PCM)	2	Zoom H4n	WAV
Audio	Auxiliary stream 48000 Hz, 2 channels, 16 bit, AAC	4	Panasonic HC-VX9805	WAV
Video	(face, left posture, right posture, back), 1080p, 50 fps, h264	4	Panasonic HC-VX9805	MP4
EDA	Auxiliary skin conductance, 4 Hz	1	Empatica E4	CSV
Temperature	Auxiliary skin temperature, 4 Hz	1	Empatica E4	CSV
Acceleration	Acceleration (x,y,z), 32 Hz	3	Empatica E4	CSV
Heart activity	RR (Interbeat) intervals	1	Empatica E4	CSV
Acceleration	Acceleration (x,y,z), 250 Hz	3	Faros 180	CSV
Heart activity	ECG, 1000 Hz, 24 bit	1	Faros 180	CSV

Table 2. Available data streams.

Per participant state attributes.

- Self reported difficulty perception (NASA-TLX)
- Current level of stress (STAI-Y1)
- General level of stress (PSS14)
- Gender

Dataset structure. The dataset is partitioned into subfolders with different modalities in each folder. The directory structure of the dataset is described in detail on Fig. 3.

Dataset descriptive statistics. The duration statistics of the speech segments can be seen in Table 3. Histograms for these segments are in Fig. 4.

Participants self-reported scores statistics are in Table 4 with histograms in Fig. 5. The score histogram for the rebus part is shown in Fig. 6.

Technical Validation

Regarding the technical validation, the presented data were meticulously collected using established psychological experimental methodologies and state-of-the-art technologies encompassing the entire data lifecycle, which includes data collection, pre-processing, and post-processing stages.

All data streams are time synchronized. The time t_0 is set at the moment when the audio recording from the Zoom recorder started. The Zoom recorder and video cameras have no explicit way to programmatically set the correct time from an external source (i.e. Network Time Protocol (NTP)). Instead, we set each of the device's correct time manually with an acceptable margin of error of 1 second.

The Empatica E4 device has an implicit time synchronization of the internal Real-Time Clock (RTC) every time the data is downloaded from the bracelet. The Faros 180 has an explicit way to synchronize the internal RTC when the device is connected to the provided software.

In our pilot trial, the Empatica E4 bracelet was our single-source of the correct time. The bracelet has a way to mark an event in the data stream by pushing a button. This method was not used to mark the beginning of the experiment, but there were a few accidental event markers created that were visible on the video streams. A relative shift of the camera internal clocks was computed from these events, that allowed us to compute real-time alignment of video streams to the Empatica datastream.

Subsequently, the alignment of the audio to the video streams was computed using the audio track from the video as the ground truth time source. Using the Audalign (<https://github.com/benfmiller/audalign>) project, we obtained a relative lag for the audio streams. We used the cross-correlation and fingerprinting method from the Audalign package.

Finally, we relied on the internal RTC of the Faros 180 device to obtain the alignment with the rest of the data streams. Unfortunately, we found that the Faros 180 RTC has a non-trivial time drift issue. This led to incorrect timestamps in our Faros 180 streams. As we had already extracted RR-intervals from the ECG from the Faros

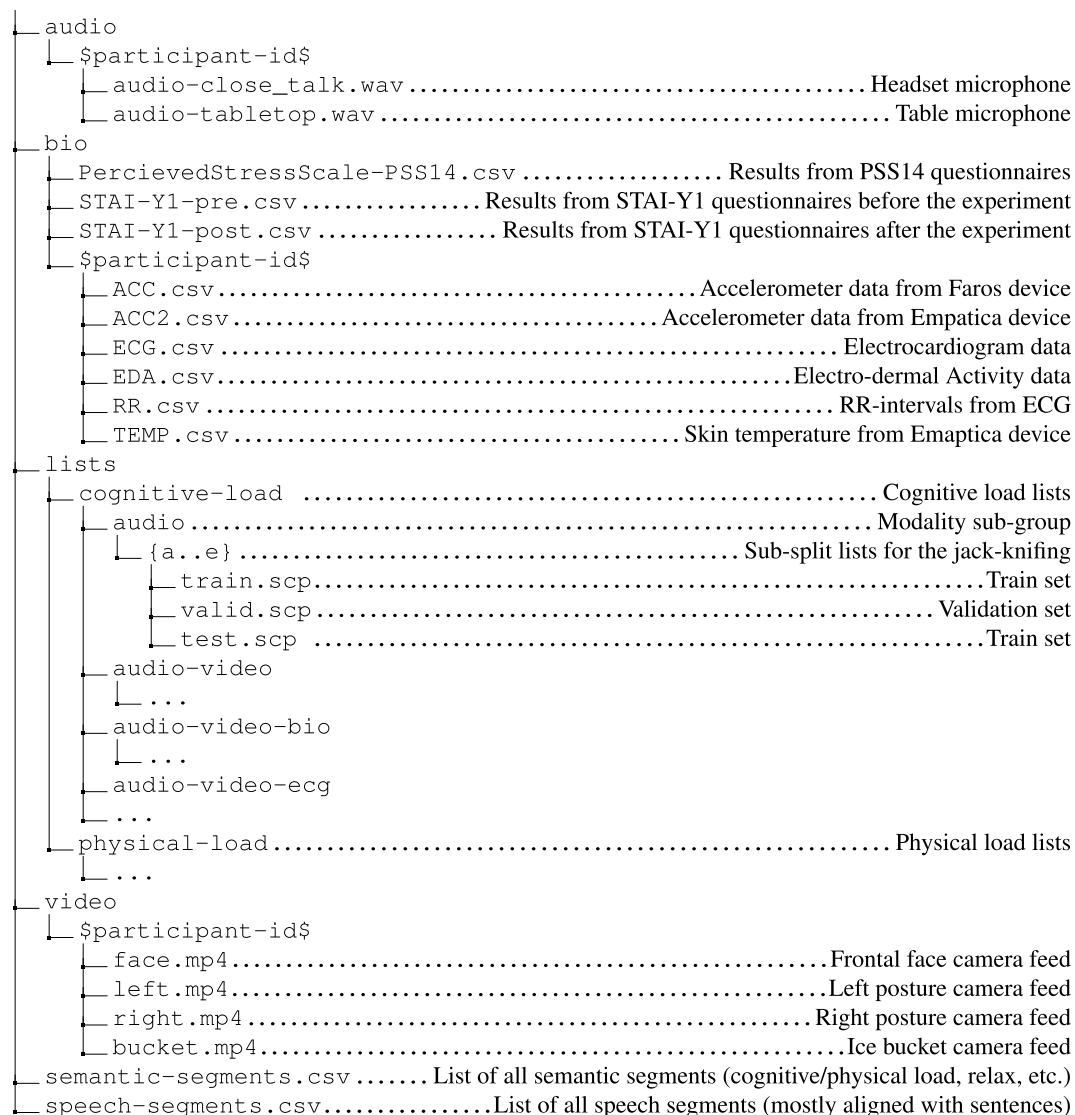


Fig. 3 Dataset directory structure.

	Mean	Std	Count
Whole session	12.51	7.93	5236
Physical Load	15.23	7.87	1837
Cognitive Load	10.93	3.44	1774

Table 3. Speech durations.

device and Empatica E4 had one auxiliary RR-interval stream, we computed cross-correlation between those two streams per participant and obtained corrected Faros ECG timestamps.

To help with manual annotations and corrections of the data, the BESTiANNO tool was developed and employed. This tool enables precise segmentation and manual verification of both audio and video streams, contributing to the reliability of the dataset. Researchers interested in utilizing or inspecting the BESTiANNO tool can access it online.

Further, the RR-interval extraction was applied as follows: In all Electrocardiogram (ECG) recordings, QRS complexes were detected and outliers were manually verified and corrected. At first, a robust QRS complex detection algorithm¹⁶ was used. This algorithm is based on an ensemble of three independent QRS detectors, each based on a different principle, the continuous wavelet transform, the Stockwell transform and the phasor transform, and individual adaptive thresholding¹⁶. The positions of QRS complexes were automatically refined to be on top of the R waves. The detections and positions of the QRS complexes were visually checked by an ECG expert beat-by-beat. Missing and additional detections and inaccurate positions were manually corrected by the expert using SignalPlant software¹⁷. The RR intervals were extracted from QRS position series and ectopic beats and erroneously long RR intervals were carefully visually inspected and removed from further analysis.

	Mean	Std	Count
Physical Load	3.81	1.84	474
Cognitive Load	6.45	1.86	1695

Table 4. Self-reported loads.

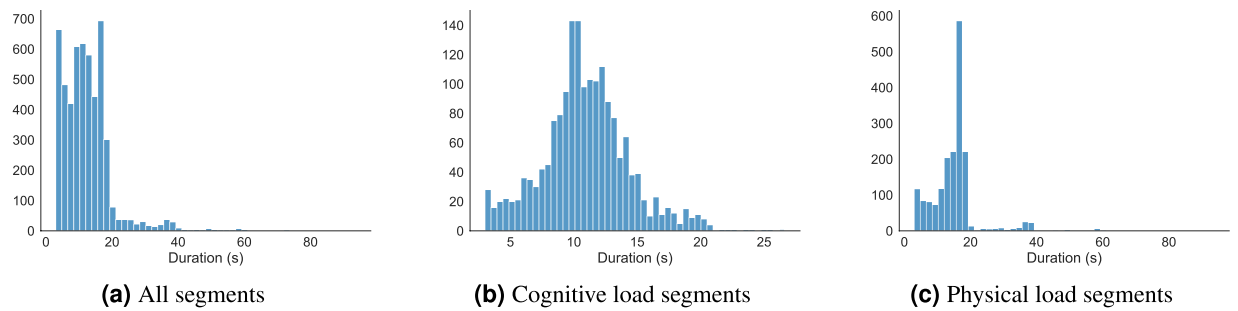


Fig. 4 Final check screen for all modalities.

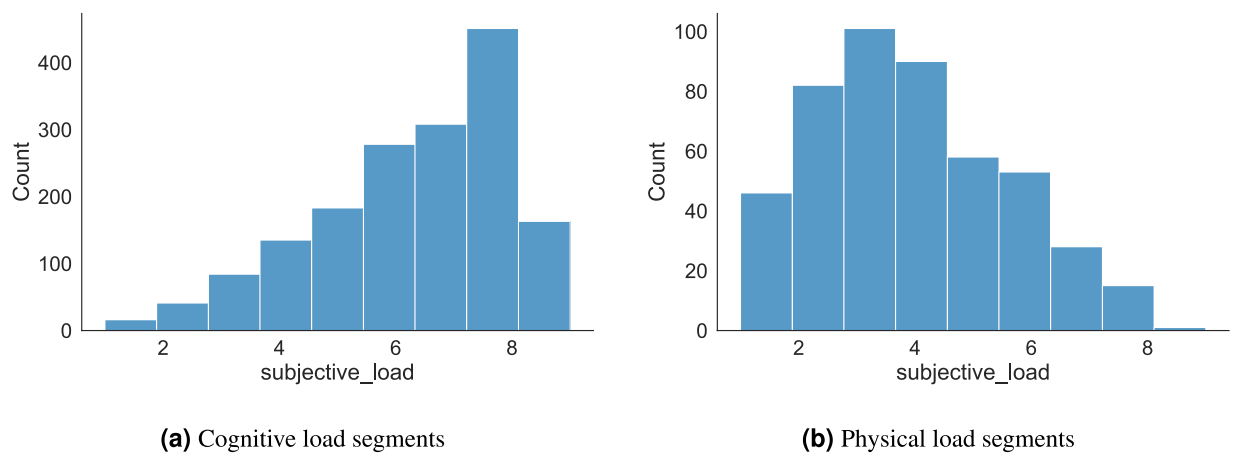


Fig. 5 Speech segment durations.

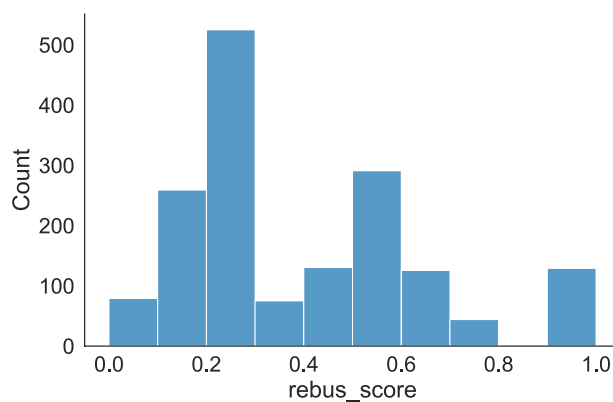


Fig. 6 Self-reported subjective loads.

After filtration and pre-processing, each participant’s data went through a manual check of alignment and dynamic range. Potential discrepancies in the data were subsequently fixed. In case, the data were corrupted beyond repair (e.g. some Empatica RR intervals), the modality was discarded from the participant’s dataset. An example of a final check screen is shown in Fig. 7.

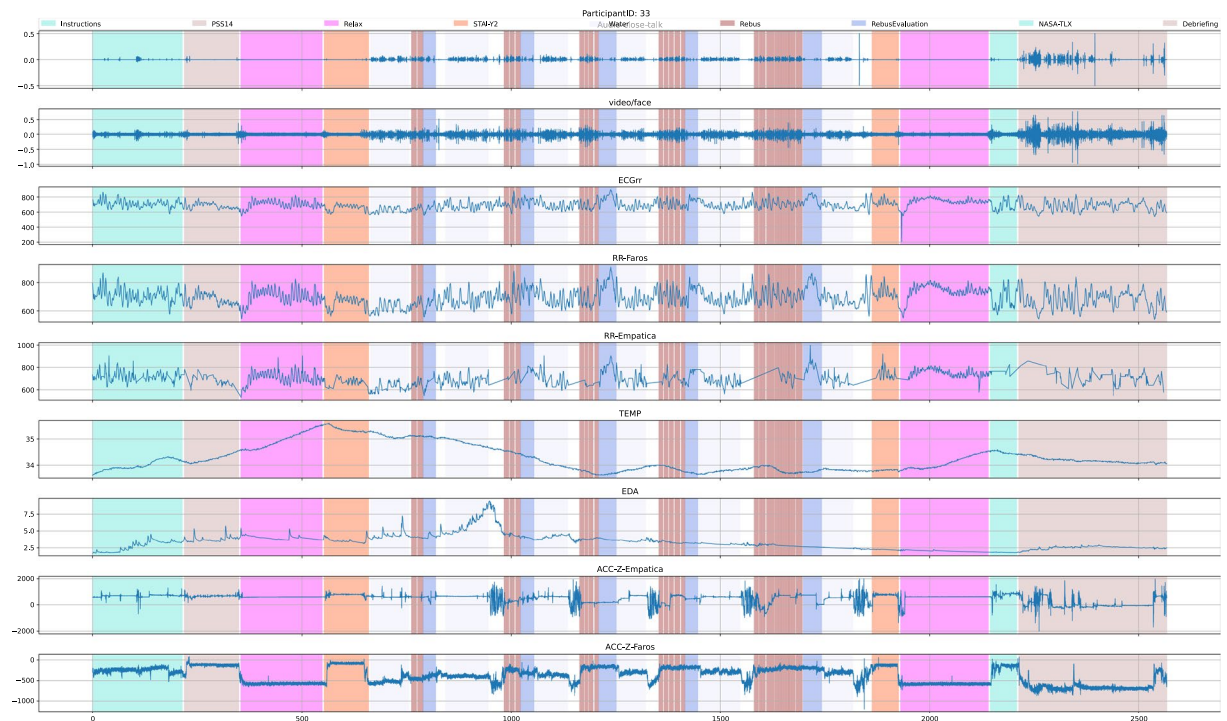


Fig. 7 Rebus score histogram.

Regarding the validation of stress induction, participants were asked to report their anxiety levels as a measure of self-experienced discomfort. These levels were measured before and after the stress induction phase. To assess the expected effect of stress induction tasks, the results of the STAI-Y1 questionnaire were compared. Given the non-normal distribution of the data, the Wilcoxon Signed Rank Test was used. There was a statistically significant difference ($V = 2436.5$, $p < 0.001$) between the levels of self-reported anxiety measured before (mean = 38.76; median = 37.00; SD = 8.95) and after (mean = 43.80; median = 42.00; SD = 11.24) the stress induction, indicating that the intervention significantly increased subjectively perceived anxiety.

Usage Notes and Limitations

The BESST dataset can be used for research and development by academic and commercial entities. The license agreement must be signed before the download link is provided. The results and demonstration of how the data corpus can be used are provided in the solitary article¹⁸. The final data may be affected by the overall stressful experimental situation - the experimental context may have increased the stress level in participants generally, and we consider this as a potential limitation, which should be further considered when the dataset is used. Losses in ecological validity of presented stress situations can be also seen as limitation of this experimental approach. BESST protocol can reliably provide speech signals affected by stress; however, the protocol cannot effectively simulate the whole variety of real-world situations and scenarios, so the generalizations made upon the provided data should reflect the experimental context of the measurements.

Future Work

Provided dataset includes experimentally controlled measurements of the human reaction to direct physiological stress and cognitive load. BESST protocol originally aimed to collect speech stress data to train the stress classifier from speech samples. However, the entire procedure and setting can be used to explore various research questions related to stress or stressful situations. The relationships between physiological and psychological responses in stressful contexts should be further explored. The entire dataset, which includes various measures, can also be investigated beyond the context of speech stress. The original dataset contained 90 participants, data from participants with missing signals or protocol errors were removed so the final provided dataset includes 79 cases with the full records (all participants were Caucasian young adults), the provided dataset includes only a Czech language sample. Regarding this, larger datasets should be collected, possibly using experimental protocol BESST.

Code availability

All code relevant to the BESST dataset collection and processing is available at <https://speech.fit.vutbr.cz/software/besst>.

Received: 27 February 2024; Accepted: 11 October 2024;

Published online: 12 November 2024

References

1. Baratta, A. & Colletta, T. *A Roman viaduct-bridge in Campania: History, structure and maintenance* (CRC Press, 2020).
2. Bernard, C. *Lectures on the phenomena of life common to animals and plants* (Thomas, Springfield, Ill, 1974).
3. Cannon, W. B. Organization for physiological homeostasis. *Physiological Reviews* **9**, 399–431, <https://doi.org/10.1152/physrev.1929.9.3.399> (1929).
4. CANNON, W. B. “voodoo” death. *American Anthropologist* **44**, 169–181, <https://doi.org/10.1525/aa.1942.44.2.02a00010> (1942).
5. Hansen, J. H. L. & Clements, M. A. Evaluation of speech under stress and emotional conditions. *The Journal of the Acoustical Society of America* **82**, S17–S18, <https://doi.org/10.1121/1.2024686> (1987).
6. Yap, T. F. *Speech production under cognitive load: Effects and classification*. Ph.D. thesis, The University of New South Wales (2012).
7. Hansen, J. H. L. Susas ldc99s78. <https://catalog.ldc.upenn.edu/LDC99S78> Last accessed on 2022-04-15 (1999).
8. Hansen, J. H. L. Susas transcripts ldc99t33. <https://catalog.ldc.upenn.edu/LDC99T33> Last accessed on 2022-04-15 (1999).
9. Pešán, J. *et al.* Besst: Brno extended stress and speech database, <https://doi.org/10.13164/data.fit.besst> (2023).
10. Smeets, T. *et al.* Introducing the maastricht acute stress test (mast): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology* **37**, 1998–2008, <https://doi.org/10.1016/j.psyneuen.2012.04.012> (2012).
11. Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress. *Journal of Health and Social Behavior* **24**, 385, <https://doi.org/10.2307/2136404> (1983).
12. Pešán, J., Juřík, V., Kolářová, J., Chudý, P. & Černocký, J. Besst: An experimental protocol for collecting speech stress datasets for machine learning (2023). Manuscript submitted for publication.
13. Hedberg, A. G. Review of state-trait anxiety inventory. *Professional Psychology* **3**, 389–390, <https://doi.org/10.1037/h0020743> (1972).
14. Likert, R. *A Technique for the Measurement of Attitudes*. No. nos. 136–165 in *A Technique for the Measurement of Attitudes* (Archives of Psychology, 1932).
15. Hart, S. G. & Staveland, L. E. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology*, 139–183, [https://doi.org/10.1016/s0166-4115\(08\)62386-9](https://doi.org/10.1016/s0166-4115(08)62386-9) (Elsevier, 1988).
16. Smital, L., Marsanova, L., Smisek, R., Nemcova, A. & Vitek, M. Robust qrs detection using combination of three independent methods. In *2020 Computing in Cardiology*, 1–4, <https://doi.org/10.22489/CinC.2020.100> (2020).
17. Plesinger, F., Jurco, J., Halamek, J. & Jurak, P. Signalplant: an open signal processing software platform. *Physiological Measurement* **37**, N38, <https://doi.org/10.1088/0967-3334/37/7/N38> (2016).
18. Pešán, J., Kesiraju, S., Burget, L. & Černocký, J. Beyond the labels: Unveiling text-dependency in paralinguistic speech recognition datasets (2024).

Acknowledgements

The work was supported by European Union’s Horizon Europe project No. SEP-210943216 “ELOQUENCE” and Czech Ministry of Interior project No. VJ01010108 “ROZKAZ”, and by the United States Office of Naval Research (ONR) Global, award number N62909-23-1-2050. Computing on IT4I supercomputer was supported by the Ministry of Education, Youth and Sports of the Czech Republic through e-INFRA CZ (ID:90254). Results presented in this paper were co-financed from the state budget by the Technology agency of the Czech Republic under the Programme of applied research and experimental development EPSILON, project TH04010325 TREVAL - Flight Training Evaluation Software. The study was also supported by the Gray Lab research infrastructure held at the Department of Psychology, Faculty of Arts, Masaryk University, Brno and by the project of Masaryk University under grant n. MUNI/A/1519/2023. We would like to express gratitude to Tom Smeets, the author of the MAST protocol, for generously providing his materials and guidelines for executing the protocol. Sincere gratitude is also given to the Aeroworks group at FIT BUT for their exploratory work on pilot state estimation from speech applicable to ground-based synthetic flight training platforms. The authors want to thank LCDR Geirid Morgan from ONR Code 342 and Dr. Martina Siwek from ONR Global Prague Office for their support.

Author contributions

J.P., P.CH. and J.Č. conceived the study; J.P., A.N., P.CH. and V.J. designed the methodology/experiment; J.P., V.J., A.R., V.S., H.B., J.A., F.K., K.V., A.S., P.B. conducted the experiment; J.P., O.J., A.R. and V.J. analyzed the data; J.C. and P.CH. supervised the study. J.P., V.J., O.J., A.R., V.S. drafted the original article, and all authors reviewed and edited the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to V.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.