# Title

Acoustic and Prosodic Speech Features Reflect Physiological Stress but not Isolated Negative Affect: A Multi-Paradigm Study on Psychosocial Stressors

# Running/Short title

Speech Features in a Multi-Paradigm Study on Psychosocial Stressors

# Authors

Mitchel Kappen*[123], Gert Vanhollebeke[125], Jonas Van Der Donckt[45], Sofie Van Hoecke[45], Marie-Anne Vanderhasselt[12]

# Affiliations

[1]Department of Head and Skin, Ghent University, University Hospital Ghent (UZ Ghent), Department of Psychiatry and Medical Psychology, Ghent, Belgium

[2]Ghent Experimental Psychiatry (GHEP) Lab, Ghent University, Ghent, Belgium

[3]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

[4]IDLab, Ghent University - imec, Ghent, Belgium

[5]Department of Electronics and Information Systems, Ghent University, Ghent, Belgium

# Contact

Correspondence concerning this article should be addressed to Mitchel Kappen

* Mitchel Kappen: E-mail: Mitchel.Kappen@UGent.be

**Classification:** Social sciences, Psychological and Cognitive Sciences

# Abstract

Heterogeneity in speech under stress has been a recurring issue in stress research, potentially due to varied stress induction paradigms. This study investigated speech features in semi-guided speech following two distinct psychosocial stress paradigms (Cyberball and MIST) and their respective control conditions. Only negative affect increased during Cyberball, while self-reported stress, skin conductance response rate, and negative affect increased during MIST. Fundamental frequency (F0), speech rate, and jitter significantly changed during MIST, but not Cyberball; HNR and shimmer showed no expected changes. The results indicate that observed speech features are robust in semi-guided speech and sensitive to stressors eliciting additional

physiological stress responses, not solely decreases in negative affect. These differences between stressors may explain literature heterogeneity. Our findings support the potential of speech as a stress level biomarker, especially when stress elicits physiological reactions, similar to other biomarkers. This highlights its promise as a tool for measuring stress in everyday settings, considering its affordability, non-intrusiveness, and ease of collection. Future research should test these results' robustness and specificity in naturalistic settings, such as freely spoken speech and noisy environments while exploring and validating a broader range of informative speech features in the context of stress.

# Competing Interests

The authors declare no conflicts of interest.

# Data Availability Statement

All data and code are publicly available at https://osf.io/qf6ck/ and https://github.com/mitchelkappen/stress_cyberball-mist. Supplemental materials, analyses, and figures can be found in the same directories.

# Funding

*Preprint: Submitted for Appearance in Scientific Reports - 2023/03/29 - v1*

# Introduction

Stress is a physiological and psychological response to internal or external stimuli that are perceived as threatening to an individual's well-being (Lazarus & Folkman, 1984; McEwen, 2007). Whether it be personal or professional, acute or chronic, stress is a common aspect of modern life that impacts people of all ages and backgrounds. Acute stress is a normal part of the human experience and can be adaptive in the short term by enabling individuals to respond to challenges and adapt to their environment (McEwen, 2007). However, when stress becomes chronic, it can have serious and long-lasting effects on an individual's physical and mental health, such as cardiovascular disease, cognitive impairment, depression, anxiety, and other (mental) health disorders (Slavich, 2016; Yaribeygi et al., 2017). As a result, accurately measuring and regularly monitoring stress levels is crucial for maintaining optimal health and well-being (Crosswell & Lockwood, 2020; Epel et al., 2018).

Given chronic stress's effects on mental and physical health, various methods have been developed for assessing peoples' stress levels including physiological, self-report, and behavioral methods (Allen et al., 2014). While each method has advantages, they also have unresolved limitations, such as cost, validity, intrusiveness, or lack of accuracy in natural settings (Slavich et al., 2019). Consequently, speech has been suggested as a non-intrusive and cost-effective method capable of measuring stress over extended periods. Speech recordings can be obtained from various sources, such as phone calls or meetings, without the need for specialized equipment, making it cost-effective and allowing for data collection in naturalistic settings,

which reduces intrusiveness (Giddens et al., 2013; Kappen, Hoorelbeke, et al., 2022; Slavich et al., 2019).

Acoustic changes in speech have been observed in response to (acute) stress (Giddens et al., 2013; Kappen, Van Der Donckt, et al., 2022; Van Puyvelde et al., 2018). However, whereas most former analyses were done in studies that used either voice actors or read-out-loud speech paradigms, it is important to shift towards evaluating the potential of freely spoken speech as it would occur in daily life (Van Der Donckt et al., 2023) to not limit the ecological validity of the results. In addition, whereas former studies rarely used validated stress paradigms or failed to validate the stress experience of participants, our recent studies addressed these limitations by utilizing validated stress induction techniques and gathering self-reports. However, they are still limited by 1) the use of one single stress induction paradigm and 2) read-out-loud speech (Kappen, Hoorelbeke, et al., 2022; Kappen, Van Der Donckt, et al., 2022).

In the current study, we focus on a key set of features that, although varying in the frequency of appearance in the literature, have consistently demonstrated their relevance across studies. We refer to these features as acoustic (physical properties of speech) and prosodic (suprasegmental aspects of speech contributing to the overall rhythm, intonation, and stress patterns), with all chosen features belonging to either one or both categories. These include the Fundamental Frequency (F0), a measure of the vocal cord's vibration frequency, that generally increases with stress (Giddens et al., 2013; Kappen, Van Der Donckt, et al., 2022; Van Puyvelde et al., 2018), jitter (vocal frequency variation) and shimmer (vocal intensity variation) which have been observed to decrease due to stress (Giddens et al., 2013; Kappen, Van Der

Donckt, et al., 2022; Van Puyvelde et al., 2018), and the Harmonics-to-Noise Ratio (HNR; relative amount of noise in comparison to harmonics in the voice) which has been shown to decrease in the context of a physical stressor (any physical event or stimulus that elicits stress) and has mixed results in the context of psychological stress (Giddens et al., 2013; Godin et al., 2012; Godin & Hansen, 2015; Kappen, Van Der Donckt, et al., 2022; Mendoza & Carballo, 1998). Additionally, we will investigate the effect of stress on changes in speech rate (talking speed) which has been shown to increase during stress in free speech samples (Giddens et al., 2010, 2013; Rothkrantz et al., 2004).

While previous studies have identified links between specific (acoustic) features of speech and stress, more research is needed to fully understand the current heterogeneity, robustness, and sensitivity of these relationships (Giddens et al., 2013; Kappen, Hoorelbeke, et al., 2022; Van Puyvelde et al., 2018). This can only be done if we do not limit our studies to single stress paradigms, especially considering that different stressors used in these paradigms elicit different stress responses. Therefore, exposing participants to different stress paradigms, but with similar experimental setups (i.e., active control task vs stress task), will allow us to better understand the basis of the observed effects on speech under stress. We aim to understand whether the observed changes in speech features that occur are related to one's changes in mood (e.g., increased negative affect) or to physiological reactions (by activation of the hypothalamic-pituitary-adrenocortical (HPA) axis), by using two well-established stress induction paradigms that specifically elicit these changes. We employed the Cyberball (Williams et al., 2000) and the Montreal Imaging Stress Task (MIST; Dedovic et al., 2005) paradigms to address these limitations and further uncover the sensitivity and robustness of speech features

under stress. Both paradigms use a psychosocial stressor, include an active control condition, and unscripted speech will be collected by having participants describe screenshots from the paradigm (see Van Der Donckt et al., 2023 for a thorough comparison of speech styles and considerations in speech collection paradigms).

The main difference between these two paradigms is that the Cyberball induces stress in the form of feelings of negative mood by means of *ostracism* due to excluding the participant from the task (Williams, 2007; Williams et al., 2000), whereas the MIST induces stress by adding components of *social evaluative threat* (SET) to a cognitively challenging task (Dedovic et al., 2005). Ostracism has been shown to worsen one's mood but is mostly limited to psychological responses and does not show a neuroendocrine (cortisol) response (Helpman et al., 2017; Williams, 2007; Zwolinski, 2012), whereas SET elicits a strong physiological, neuroendocrine (cortisol) response in addition to a decreased mood (Allen et al., 2014; Bosch et al., 2009; Dickerson, 2008; Dickerson & Kemeny, 2004).

**Research Objectives & Hypotheses**: We will gauge the stress response based on increased skin conductance response rate (SCRR), as well as self-reports on increased experienced stress and negative affect during the stress block as compared to the control block. Moreover, this is the first study to use a picture-describe paradigm to capture semi-guided speech that closely resembles natural speech in order to yield ecologically valid results (for more details, see: Van Der Donckt et al., 2023). *Negative Affect*. We expect increases in Negative Affect after the stress blocks compared to the control blocks for both paradigms. *Self-reported Stress*. We expect increased self-reported stress during the stress block for the MIST. However, we do not expect increases in self-reported stress for the Cyberball, as its effect is inconsistent and strongly

mediated by traits such as the need to belong, limiting the observance of this effect in a general population (Beekman et al., 2016). *Skin Conductance Response Rate (SCRR).* We expect an increase in SCRR during the stress block for the MIST, but not for the Cyberball. *Speech features.* We expect similar results for speech features as observed in earlier studies that used read-out-loud protocols (Giddens et al., 2013; Kappen, Van Der Donckt, et al., 2022; Van Puyvelde et al., 2018) for the MIST. That is, increases in Fundamental Frequency (F0; Giddens et al., 2013; Kappen, Van Der Donckt, et al., 2022; Van Puyvelde et al., 2018) and measures of changes in speech rate (Giddens et al., 2013; Rothkrantz et al., 2004), decreases in Jitter (Giddens et al., 2013; Van Puyvelde et al., 2018) and Shimmer (Giddens et al., 2013; Kappen, Van Der Donckt, et al., 2022; Van Puyvelde et al., 2018), and changes in Harmonics to Noise (HNR; added noise in the voice), but the direction is unclear due to mixed results in the context of psychological stressors (Giddens et al., 2013; Godin et al., 2012; Godin & Hansen, 2015; Kappen, Van Der Donckt, et al., 2022; Mendoza & Carballo, 1998). In the Cyberball paradigm, the occurrence and direction of significant speech feature changes will reveal the sensitivity and heterogeneity of speech as a biomarker for (psychological) stress. Considering the expected difference in the stress reaction in the Cyberball (negative mood) compared to the MIST (negative mood + physiological reaction), the occurrence of significant changes in speech features would show that speech is responsive to mere changes in mood due to stress (therefore occurring in both paradigms). A lack of changes in speech features would, however, illustrate that speech (features) are merely related to physiological stress responses and therefore follow the patterns (i.e., effects in MIST, but not in Cyberball) observed in other biomarkers such as cortisol (Allen et al., 2014; Dickerson, 2008; Dickerson & Kemeny, 2004).

Lastly, it is possible that different speech features have varying sensitivity, where some might just be responsive to combined mood and physiological stress responses, and others might be responsive to mere changes in mood. For example, one of the most homogeneously reported speech features to change under stress is F0, which could indicate that this is a sensitive feature to any change in experienced stress (i.e., mood or physiological) since it occurs in many different studies and stress paradigms. Other features have shown to be more heterogeneous (e.g., HNR, Jitter), which could be explained by the use of different paradigms and stressors. The combination of two different stress induction paradigms, which elicit different stress responses by calling on different psychological constructs (Cyberball; ostracism, MIST; social evaluative threat), tested in the same group of participants, will give unique insights into the robustness (by using semi-guided speech), sensitivity (by comparing a mood only to a mood plus physiological stress reaction), and up to this point heterogeneity (by comparing two commonly used stress paradigms) in a variety of speech features under stress.

# Methods

## Participants

A convenience sample of 66 healthy subjects (13 women, 53 men, age M = 21.29, SD = 2.82) was recruited through social media. Upon registration, participants were checked for exclusion criteria (see supplemental material). The study was conducted in accordance with the declaration of Helsinki and received ethical approval from the Ghent University hospital ethical committee (registration number: B6702020000676). Another part of the study investigates the effects of (psychosocial) stressors on neural correlates. Results of electrophysiological correlates will be published elsewhere. Other collected data that were not part of the current paper's research objectives will only be described in the supplemental materials.

All participants gave written informed consent before participating and were debriefed afterward on the true purpose of the study. A 40 Euro compensation fee was awarded upon completion of both testing days through bank transfer.

## Procedure

### *On-site experimental session*

Participants completed online informed consent and trait questionnaires (beyond this paper's scope) prior to two in-person experimental sessions, which were conducted in a dedicated room in the Department of Adult Psychiatry at Ghent University Hospital. At the start of the first

on-site session, participants signed a paper consent form, and experimenters reviewed the cover story (see Cover Story).

The experiment was designed in OpenSesame version 3.2.8 and was carried out on a dedicated computer (Dell, Windows 10). Participants came in on two different days, at least 7 days apart. The experimental sessions were identically structured, but only the task contents differed (Day1; see Cyberball, Day2; see MIST). Prior to the task, electrodes (ECG, EDA; see Physiological Data) and an EEG cap were placed (duration 10-30 minutes; beyond this paper's scope). The experiment started with a 10-minute (5 minutes eyes closed, 5 minutes eyes open) resting block to achieve habituation. After this, the Control task started. After this condition, there was another 10-minute resting block, followed by a Stress task and another 10-minute resting block. Subsequent to each task block (i.e., Control task and Stress task), participants were prompted to do a speech trial (see Speech Data) and respond to self-report psychological state questionnaires (see Self-Report Data). See figure 1 for a flowchart figure.

**Figure 1.**

Flowchart of experimental design
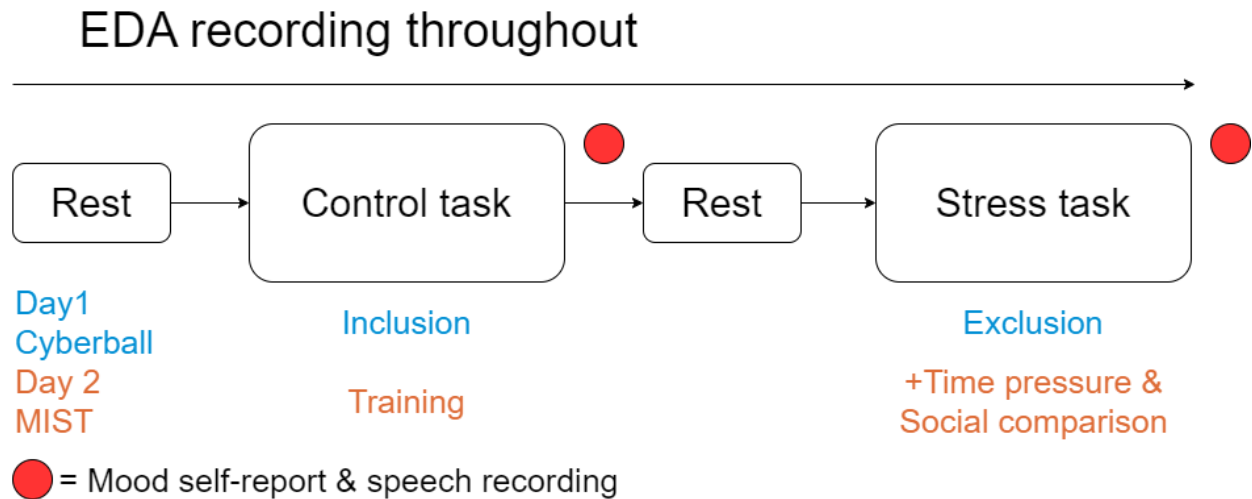


*Note.* The two days follow a similar structure except for the presented paradigm and respective control and stress tasks. EDA (electrodermal activity) is collected throughout the paradigm. Speech recordings (picture description) and self-report questionnaires are collected directly after task execution. The colors used in the figure are congruent with the colors presented in the results section.

## Cyberball - Day 1

The Cyberball paradigm involved a ball-tossing game in which participants played with two computer-generated confederates (one man, one woman, placement counterbalanced across participants), represented by pictures (from Allaert et al., 2022). However, participants were told that the other players were humans participating at other universities. The confederates' behaviors were predefined and the game was visualized with a picture of the participant at the bottom center, while the confederate pictures were placed at the top left and right (Williams et al., 2000). Participants could throw the ball to either confederate by pressing an arrow key (right

hand), and the ball's movement took 1500 ms. The confederates held possession between two and three seconds (randomly generated) to increase the credibility that they were human. During the **control task** (inclusion phase), participants received the ball 33% of the time over 150 throws, while in the **stress task** (exclusion phase), participants were excluded in a probabilistic manner from receiving the ball after an initial normal phase of 30 throws. The chance of retrieving the ball increased with each subsequent throw not directed at the participant, with chances ranging from 0 to 100% (adapted from Williams et al., 2000).

## *Montreal Imaging Stress Task (MIST) - Day 2*

During the Montreal Imaging Stress Task (MIST), participants solved mathematical equations of increasing difficulty (Dedovic et al., 2005). Equations were displayed in black on a white background, and the correct answer was always a number between zero and nine, with participants answering each question using the corresponding number on the keyboard's numpad (right hand). The difficulty scales and equation-generating code was identical to the original study, as supplied by prof. Pruessner (Dedovic et al., 2005).

The **control task** included seven difficulty scales, with participants solving up to ten equations per scale. After each equation, feedback was given in the form of *"Correct!"*, *"Incorrect!"*, or *"Timeout!",* shown in black. The **stress task** employed the same difficulty scales but introduced changes to the task and feedback. Participants were informed that their performance would be compared to that of a group and that they should perform at least on par with the average. Equations were presented with a shrinking bar indicating the remaining time to solve the equation, and the allowed time was set to be 90% of their average response

time during the control task. After every three successive correct or incorrect answers, the allowed time was adjusted by 10% to increase/decrease difficulty. Participants also saw a performance bar with two arrows indicating their personal and group average scores. Their personal arrow moved in steps of 5% of the bar's length after each equation (incorrect/timeout; left, correct; right), whereas the group average arrow was stationary at 83%. If a participant's performance fell below the average group performance after completing five difficulty scales, the experimenter would inform them that their data might not be usable and urge them to improve.

## Data Collection

In this study, several types of data were collected for analysis. While our main hypotheses focused on specific data modalities, we also collected additional self-report and cardiac data. To ensure transparency, we have provided an overview of these data modalities and a complete study flowchart in the supplemental materials.

### *Speech Data*

On both days, after completion of either task (i.e., control/stress task), participants were prompted to describe a picture out loud, see Figure 1. The image was a screenshot of the task they had just completed to avoid introducing noise to our self-report measures by having their minds wander (for a similar approach and considerations, see: Van Der Donckt et al., 2023). The participants were instructed to describe the images based on what they saw, as well as how it made them feel. See supplemental materials for screenshots.

*Preprint: Submitted for Appearance in Scientific Reports - 2023/03/29 - v1*

## *Self-report Data*

On both days, after completion of either task (i.e., control/stress task), participants were asked to rate their current levels of stress and negative affect ("Right now, how much do you feel…") using 6 negative affect (NA) prompts and 1 stress prompt question, each with a 0-100 sliding scale (0 = Not at all - 100 = Very much), see Figure 1. The six negative affect (NA) prompts are: upset, distressed, scared, angry, anxious, and sad, whereas stress was a single item asking "Right now, how much do you feel stressed?". Positive activating and soothing affect were also collected but not part of the primary hypotheses, thus only described in the supplemental materials. These scales were adopted from Petrocchi and colleagues (2017). Given the high internal consistency between the prompts in the NA category, these responses were aggregated to compute a single mean score to be used in the analysis.

## *Physiological Data*

We collected electrocardiography (ECG; see supplemental material for more info) and electrodermal activity (EDA) data throughout the paradigms using the VU-AMS ambulatory monitor[1], which was specifically designed for this purpose. To collect EDA data, we placed two velcro electrodes with isotonic electrode gel (Biopac) on the middle phalanges of the left index and middle finger.

---

[1] VU University Amsterdam, www.vu-ams.nl, Amsterdam, the Netherlands

## Data Analysis

### *Physiological Data*

EDA data was preprocessed using specific Python code, which can be found in the supplemental material ([0.2_EDA.ipynb](#) & [scl_processing.py](#)). To prepare the data for analysis, a 2Hz low-pass filter was applied to the raw signal, which was then decomposed into a tonic and phasic component. From the phasic component, the Skin Conductance Response Rate (SCRR) was extracted by identifying peaks with the SciPy toolkit. The thresholds for rise and fall time, as well as peak parameters, were determined based on established guidelines from the literature (e.g., see: Posada-Quintero & Chon, 2020).

### *Extraction of Speech Features*

To ensure data quality, we manually checked all recordings whether they were complete, clear, with limited background noise, and no excessive clipping. In addition, recordings were dropped if there was no complementary self-report scales (due to technical issues). Eight recordings were removed, resulting in 120 control recordings (64 Cyberball, 56 MIST) of 66 out of 66 participants, and 119 stress recordings (64 Cyberball, 55 MIST) of 66 out of 66 participants. Prior to feature extraction, we downsampled the speech samples to 16Khz and applied dithering. These steps were performed in order to make the extracted OpenSMILE metrics less sensitive to environmental harmonics at the voiced boundaries (Van Der Donckt et al., 2023). To extract features from the recordings, we used OpenSmile 2.3.0 (Eyben et al., 2010) with the

GeMAPSv01b configuration (Eyben et al., 2015), a widely-used acoustic feature set in voice research and affective computing. From this feature set (feature names as described in GeMAPS added between brackets), we selected Fundamental Frequency (F0; F0semitoneFrom27.5Hz_sma3nz_amean), Jitter (jitterLocal_sma3nz_amean), Shimmer (shimmerLocaldB_sma3nz_amean), Harmonics-to-Noise Ratio (HNR; HNRdBACF_sma3nz_amean), and Voiced Segment Length (MeanVoicedSegmentLengthSec) and Mean Voiced Segments per Second (a proxy for speech speed; VoicedSegmentsPerSec) to capture changes in speech rate. All features were computed using Python 3.9.6 for a sliding window and then mean-aggregated over the whole recording, thus not displaying high temporal changes. For detailed information regarding feature calculation and extraction procedure, we refer the reader to Eyben et al. (2010) and Section 6.1 of Eyben et al. (2015).

## Statistical Analysis

Statistical analyses were performed using R4.1.1 (for detailed version information of the software and packages used, please refer to the supplemental materials).

We used the 'lme4' (Bates et al., 2014) package to fit linear mixed models (LMMs) to each of the dependent variables. The sum of squares for each model was estimated using a partial sum of squares (Anova type III approach), and the statistical significance level was set to $p < .05$ (these results are only reported in the supplemental materials). Tests for pairwise comparisons of the EMMs (estimated marginal means) were performed with the 'emmeans' package (Lenth, 2018). A false discovery rate (FDR) was used to correct for multiple comparisons correction for each data modality (e.g., all speech comparisons pooled together and penalized

accordingly) to minimize the risk of Type 1 errors (Benjamini & Hochberg, 1995) using the 'p.adjust()' function from the 'stats' package. In the results section, only corrected p-values will be reported. Moreover, effect sizes (Cohen's D) and their 95% confidence intervals (CI) are estimated with the 'eff_size()' function from the 'emmeans' package (Lenth, 2018). Results are only reported using the effect sizes of within-paradigm comparisons to make a comparison between different dependent variables and data modalities. Note that between-paradigm comparisons (i.e., Cyberbal vs MIST) should be avoided for each respective task (i.e., control and stress task) as the performed tasks and the described pictures are inherently different.

To control for the potential effect of gender on the different dependent variables, gender was considered as a fixed effect for each individual model prior to statistical inference. However, to make sure our models were parsimonious, we bottom-up tested whether adding *gender* as an independent variable to the model improved each model's fit (Bates et al., 2018). For each dependent variable, we compared models that included and excluded *gender*, and it was only included in the model if it showed to be a significant contributor after comparing models with reducing complexity using $\chi^2$ goodness-of-fit tests within the 'anova()' function (Fox et al., 2012). The statistical significance level was set to $p < .05$ and based on this, gender was included in the models for F0, Shimmer, HNR, Voiced segment length, and self-reported stress. As such, each model followed the following structure; *DependentVariable ~ Phase * Task + Gender + (1|ID)* or *DependentVariable ~ Phase * Task + (1|ID)*. With *Phase* having 2 levels (control vs stress), *Task* having 2 levels (Cyberball vs MIST), and participant as a random intercept.

# Results

Throughout the different paradigms, we focus in our analyses on three main modalities of which two are already more validated in literature (i.e., self-reports and physiological measures) and one is our novel addition to the state-of-the-art (i.e., speech). The models reported will contain the dependent variable, taskPhase (control vs stress task), taskType (MIST vs Cyberball), Gender (man vs woman) if showing to be a significant contributor, and '(1|ID)'; a random intercept for each participant. Per category (i.e., self-reports, physiological, and speech), for each feature, we will only describe the pairwise comparisons of stress vs control block for each individual paradigm as these are directly related to our research questions. All effect sizes and corresponding 95% confidence intervals for the control-stress comparisons per feature, per paradigm, are also displayed in Figure 2. Full model information and corresponding statistics are described in the analyses section of the supplemental materials.

## Physiological

*Skin Conductance Response Rate (SCRR).*

A significant increase in *SCRR* was observed during the stress task in the MIST, $b = 1.30$, $SE = .34$, $t = 3.77$, $p < .001$, $d = .72.$, 95% CI [.33, 1.10], but not in the Cyberball, $b = .53$, $SE = .32$, $t = 1.66$, $p = .098$, $d = .29$, 95% CI [-.06, .64].

## Self-reports

*Negative Affect.*

A significant increase in *Negative Affect* was observed during the stress task in the MIST, $b = 4.29$, $SE = 1.69$, $t = 2.54$, $p = .016$, $d = .48.$, 95% CI [.11, .86], as well as in the Cyberball, $b = 5.37$, $SE = 1.57$, $t = 3.42$, $p = .002$, $d = .60$, 95% CI [.25, .96].

*Stress.*

A significant increase in self-reported *Stress* was observed during the stress task in the MIST, $b = 17.26$, $SE = 3.84$, $t = 4.49$, $p < .001$, $d = .85.$, 95% CI [.47, 1.24], but not in the Cyberball, $b = 2.60$, $SE = 3.58$, $t = .73$, $p = .469$, $d = .13$, 95% CI [-.22, .48].

## Speech/Acoustic

*Fundamental Frequency (F0).*

A significant increase in *F0* was observed during the stress task in the MIST, $b = .42$, $SE = .15$, $t = 2.76$, $p = .026$, $d = .52.$, 95% CI [.15, .90], but not in the Cyberball, $b = .12$, $SE = .14$, $t = .85$, $p = .531$, $d = .15$, 95% CI [-.20, .50].

*Voiced segments per second.*

A significant increase in *voiced segments per second* was observed during the stress task in the MIST, $b$ = .23, $SE$ = .04, $t$ = 5.44, $p$ < .001, $d$ = 1.03., 95% CI [.65, 1.42], but not in the Cyberball, $b$ = .02, $SE$ = .04, $t$ = .52,  $p$ = .729, $d$ = .09, 95% CI [-.26, .44].

*Voiced segment length.*

A significant increase in *voiced segment length* was observed during the stress task in the MIST, $b$ = .01, $SE$ = .005, $t$ = 2.62, $p$ = .029, $d$ = .50., 95% CI [.12, .88], but not in the Cyberball, $b$ = -.001, $SE$ = .005, $t$ = -.27,  $p$ = .858, $d$ = -.05, 95% CI [-.40, .30].

*Harmonics-to-noise ratio (HNR).*

No significant change in *HNR* was observed during the stress task in the MIST, $b$ = .21, $SE$ = .09, $t$ = 2.31, $p$ = .053, $d$ = .44., 95% CI [.06, .82], nor in the Cyberball, $b$ = .10, $SE$ = .08, $t$ = 1.19,  $p$ = .354, $d$ = .21, 95% CI [-.14, .56].

*Shimmer.*

No change in *Shimmer* was observed during the stress task in the MIST, $b$ = -.02, $SE$ = .01, $t$ = -1.65, $p$ = .201, $d$ = -.31., 95% CI [-.69, .06], nor in the Cyberball, $b$ = 0, $SE$ = .01, $t$ = -.07,  $p$ = .942, $d$ = -.01, 95% CI [-.36, .34].

*Jitter.*

A significant decrease in *Jitter* was observed during the stress task in the MIST, $b$ = -.003, $SE$ = .0009, $t$ = -3.28, $p$ = .007, $d$ = -.62., 95% CI [-1.00, -.24], but not in the Cyberball, $b$ = -.001, $SE$ = .0009, $t$ = -1.38,  $p$ = .292, $d$ = -.24, 95% CI [-.59, .11].

**Figure 2.**

Forest plot of effect sizes and confidence intervals for all control-stress comparisons per paradigm
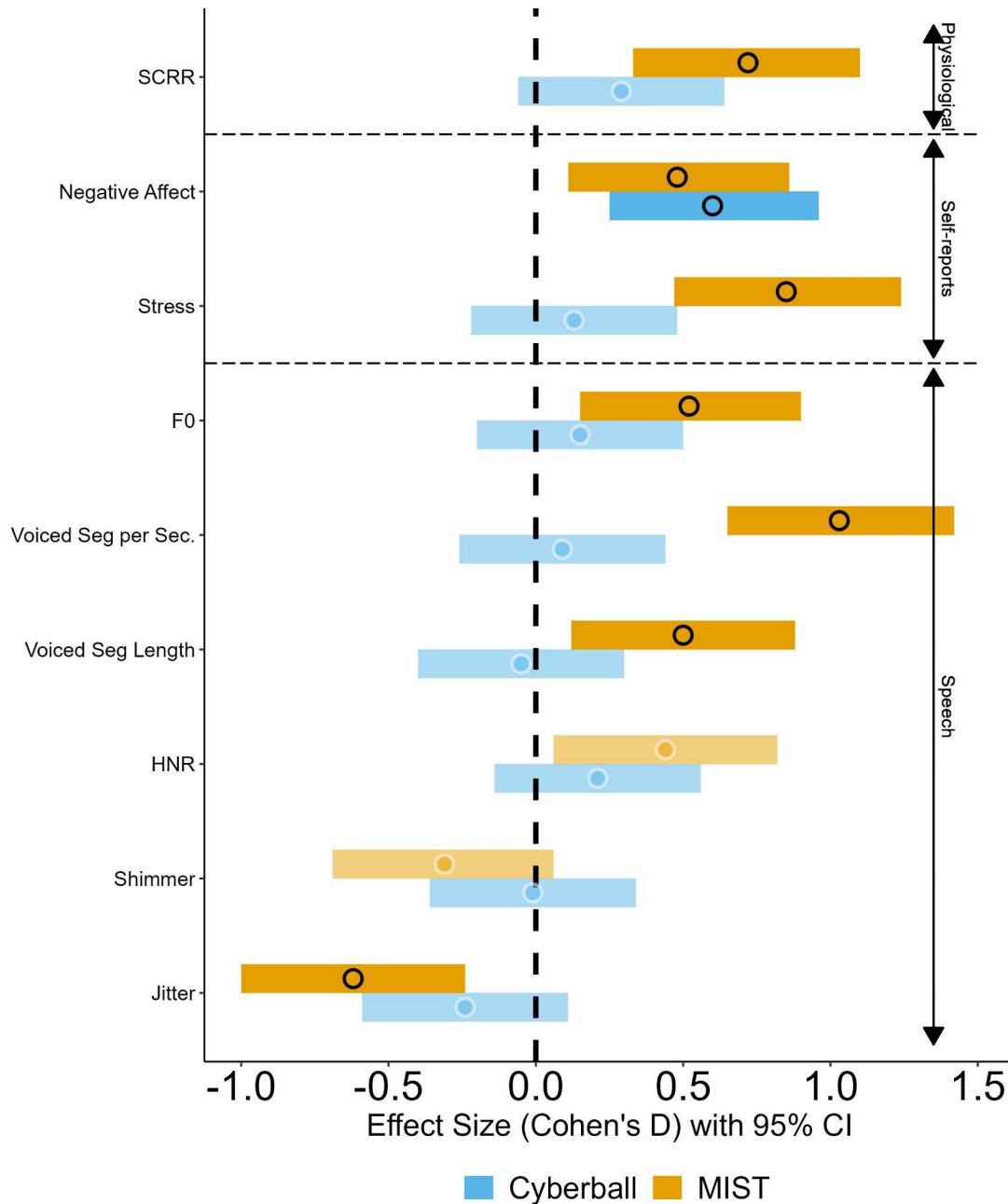


*Note.* Effect sizes (dots; Cohen's D) and 95% confidence intervals (bars) for each control vs stress task comparison per stress induction paradigm (Cyberball - MIST). Dependent variables are grouped on their categories (i.e., physiological, self-report, speech) and are FDR corrected within their respective categories. Dots are circled black and ranges are saturated (i.e. non-transparent) if a significant effect is observed after correction.

**Preprint:** *Submitted for Appearance in Scientific Reports - 2023/03/29 - v1*

# Discussion

In this study, we aimed to gain insights into the robustness (by using semi-guided instead of read-out-loud speech), sensitivity (by comparing a mood only to a mood plus physiological stress reaction), and up to this point observed heterogeneity (by comparing two different commonly used stress paradigms) of the effects of stress on acoustic and prosodic speech features. On two different days, participants were exposed to two different stress induction paradigms (i.e., Cyberball and Montreal Imaging Stress Task; MIST) with an expected different stress reaction (i.e., Cyberball; changes in mood, MIST; changes in mood and physiological response). Both paradigms included an active control condition in order to isolate the effects of added stress on their speech. Speech samples were collected directly after each paradigm phase (i.e., control or stress phase) using a picture-describe paradigm (prompting participants to describe a screenshot from the paradigm) to capture semi-guided speech that closely resembles natural speech in order to yield ecologically valid results (for more details, see: Van Der Donckt et al., 2023).

First, we used validated measures to gauge how the stress responses elicited by the different stress paradigms (Cyberball and MIST), differed. As such, we observed that when considering physiological responses, there was only an increase in skin conductance response rate (SCRR) during the stress phase of the MIST, but not for the Cyberball, which corresponds to our prior hypotheses. In addition to the physiological responses, we also assessed participants' moods using self-reported measures. In line with our expectations, we observed an increase in self-reported negative affect during the stress task of both paradigms. Moreover, participants only reported increased self-reported stress during the MIST. These results are in line with the

literature, as previously mentioned. The Cyberball task affects one's mood due to feelings of ostracization, but it only elicits psychological responses and does not elicit a physiological, neuroendocrine stress response (Helpman et al., 2017; Williams, 2007; Zwolinski, 2012). However, the MIST, using social evaluative threat by means of (negative) social comparison, elicits both physiological and neuroendocrine stress responses, alongside a decrease in mood (Allen et al., 2014; Bosch et al., 2009; Dedovic et al., 2005; Dickerson, 2008; Dickerson & Kemeny, 2004).

Several key acoustic features, described in the literature to be responsive to stress, were extracted from the speech samples (F0; fundamental frequency, HNR; harmonics-to-noise ratio, jitter, shimmer, speech rate, and voiced segment length). Prior results have been heterogeneous, which is possibly due to the use of many different paradigms and stressors which introduce noise rather than robustness in this new modality's early, exploratory stages. We tackle this by doing explicit, side by side analysis of two often used stress paradigms in the same sample. In the current study, during the Cyberball task, none of the tested acoustic speech features changed significantly during the stress, compared to the control phase. On the other hand, however, all features except HNR and shimmer changed in the expected direction during the stress phase of the MIST. We observe increases in F0 (consistent with Giddens et al., 2013; Kappen, Van Der Donckt, et al., 2022; Van Puyvelde et al., 2018), speech rate and voiced segment length (consistent with Giddens et al., 2010, 2013; Rothkrantz et al., 2004), and a decrease in Jitter (consistent with Giddens et al., 2013; Van Puyvelde et al., 2018). The observed increase in HNR, related to stress in the MIST, did not survive multiple comparison corrections, which indicates that the observed effect was rather small. This is consistent with the literature,

as previous studies have reported mixed results or conflicting findings in HNR changes (Giddens et al., 2013; Godin et al., 2012; Godin & Hansen, 2015; Mendoza & Carballo, 1998). However, the observed effect in the current study follows the same direction as our former study, which did show a significant increase (Kappen, Van Der Donckt, et al., 2022). This consistency might be indicative of the true direction of the effect, despite the small effect size in the present study. Additionally, no decrease was observed for shimmer during the stress task, whereas mixed results have been observed in the literature (Giddens et al., 2013; Kappen, Van Der Donckt, et al., 2022; Van Puyvelde et al., 2018). The absence of a significant change in shimmer during the stress phase of the MIST can be related to two things. First, it could be related to the speech collection paradigm used in the current study. We used a semi-guided speech paradigm in which participants were shown a screenshot from the task they just completed and were prompted to describe it. For the MIST, that means participants were shown a mathematical puzzle, similar to the earlier task, which many participants would try and solve out loud. This speech follows a less natural flow than naturally spoken speech, and as such could affect the amount of changes in shimmer. Second, which is arguably related to the first, the absolute observed values for shimmer were rather low as compared to former studies (Kappen, Van Der Donckt, et al., 2022) indicating potential floor effects. However, it should be noted that the observed absolute shimmer values are consistently lower for the recordings in the Cyberball paradigm (see supplemental materials). Nonetheless, due to the presented images being different between the two speech collection paradigms, no formal comparisons between the two should be done, and our interpretations are limited to within-paradigm changes. The methodological choice, to have participants describe screenshots from the task rather than

other, off-topic images, should also be noted as the study's biggest limitation. The study's objective was, first and foremost, to elicit stress using two different psychosocial stress induction paradigms. By having participants describe unrelated images directly after completing the task, it can be argued that they could be distracted from the stressor and thus decrease its potency, confounding the final results. As such, follow-up studies should collect speech samples using a standardized semi-guided speech paradigm consisting of validated images that are congruent to psychosocial paradigms, to describe and keep consistent throughout longitudinal designs as described in Van der Donckt and colleagues (2023). It should also be noted that the sample consisted of predominantly young adults, therefore possibly limiting the generalization of our results to the general population.

The current study used two different psychosocial stress paradigms that are different in their stress responses (i.e., Cyberball; negative mood, MIST; negative mood + physiological reaction). As such, we were able to relate promising acoustic and prosodic speech features to these distinct stress responses. We demonstrate that most features that are described in the literature in relation to stress only changed in the MIST (social evaluative threat paradigm), and not in the Cyberball (ostracism paradigm). These results follow our observed changes in self-report and physiological measures and as such, we conclude that speech as a biomarker is indeed a promising method for detecting changes in stress levels. Speech is comparable to other validated methods (i.e., skin conductance response rate & self-reported stress) as illustrated by the observed effect sizes, in that it does not respond to mere changes in negative affect, but only when physiological changes occur.

This study is the first to demonstrate how speech features change due to different stress paradigms and corresponding stressors, using a within-participant design. These results might explain the current heterogeneity in the literature with regard to these speech features. We conclude that semi-freely spoken speech (features) are promising for stress detection, and are not affected by stressors that only evoke changes in negative mood. This further outlines its potential in real-world applications, where it appears increasingly promising in passive, remote, non-intrusive tracking of major stressors in daily life that can have severe health implications (Kappen et al., 2023).

## Conclusions

To conclude, we collected repeated semi-guided speech fragments from participants in two different psychosocial stress paradigms, both including active control conditions. We observed distinct stress reactions in the two paradigms through self-reports and psychophysiological responses. A change in self-reported negative affect during the Cyberball, and an additional physiological and self-reported stress reaction during the MIST were found. Similar effects (i.e., effect during MIST, but not during Cyberball) were found for most speech features of interest; F0, voiced segments per second, mean voiced segment length, and jitter, but not for HNR and shimmer. Therefore, we conclude that these effects are robust in (semi-)freely spoken speech (as compared to earlier studies using read-out-loud speech), and are sensitive to stressors that activate the HPA axis, but not to changes in negative affect alone. The difference in observed effects between the two stressors possibly explains the current heterogeneity in the literature. These results further solidify the potential use of speech as a

biomarker for stress level assessment in everyday settings, given its affordability, non-intrusiveness, and ease of collection. Future studies should focus on further testing the robustness of these results in increasingly naturalistic settings, such as completely freely spoken speech and noisy environments while exploring a broader range of speech features that can be informative in the context of stress.

# References

Allaert, J., De Raedt, R., Sanchez-Lopez, A., Baeken, C., & Vanderhasselt, M.-A. (2022). Mind the social feedback: Effects of tDCS applied to the left DLPFC on psychophysiological responses during the anticipation and reception of social evaluations. *Social Cognitive and Affective Neuroscience*, *17*(1), 131–141. https://doi.org/10.1093/scan/nsaa066

Allen, A. P., Kennedy, P. J., Cryan, J. F., Dinan, T. G., & Clarke, G. (2014). Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. *Neuroscience & Biobehavioral Reviews*, *38*, 94–124. https://doi.org/10.1016/j.neubiorev.2013.11.005

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). *Parsimonious Mixed Models* (arXiv:1506.04967). arXiv. http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beekman, J. B., Stock, M. L., & Marcus, T. (2016). Need to Belong, Not Rejection Sensitivity, Moderates Cortisol Response, Self-Reported Stress, and Negative Affect Following Social Exclusion. *The Journal of Social Psychology*, *156*(2), 131–138. https://doi.org/10.1080/00224545.2015.1071767

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bosch, J. A., de Geus, E. J. C., Carroll, D., Goedhart, A. D., Anane, L. A., van Zanten, J. J. V.,

Helmerhorst, E. J., & Edwards, K. M. (2009). A general enhancement of autonomic and cortisol responses during social evaluative threat. *Psychosomatic Medicine*, *71*(8), 877–885. https://doi.org/10.1097/PSY.0b013e3181baef05

Crosswell, A. D., & Lockwood, K. G. (2020). Best practices for stress measurement: How to measure psychological stress in health research. *Health Psychology Open*, *7*(2), 2055102920933072. https://doi.org/10.1177/2055102920933072

Dedovic, K., Renwick, R., Mahani, N. K., Engert, V., Lupien, S. J., & Pruessner, J. C. (2005). The Montreal Imaging Stress Task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience*, *30*(5), 319–325.

Dickerson, S. S. (2008). Emotional and Physiological Responses to Social-Evaluative Threat. *Social and Personality Psychology Compass*, *2*(3), 1362–1378. https://doi.org/10.1111/j.1751-9004.2008.00095.x

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355

Epel, E. S., Crosswell, A. D., Mayer, S. E., Prather, A. A., Slavich, G. M., Puterman, E., & Mendes, W. B. (2018). More than a feeling: A unified view of stress measurement for population science. *Frontiers in Neuroendocrinology*, *49*(December 2017), 146–169. https://doi.org/10.1016/j.yfrne.2018.03.001

Eyben, F., Scherer, K., Schuller, B., Sundberg, J., André, E., Busso, C., Devillers, L., Epps, J., Laukka, P., Narayanan, S., & Truong, K. (2015). The Geneva Minimalistic Acoustic Parameter Set (

GeMAPS ) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing*, *7*(2), 190–202. https://doi.org/10.1109/TAFFC.2015.2457417

Eyben, F., Wöllmer, M., & Schuller, B. (2010). OpenSMILE - The Munich versatile and fast open-source audio feature extractor. *MM'10 - Proceedings of the ACM Multimedia 2010 International Conference*, 1459–1462. https://doi.org/10.1145/1873951.1874246

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., ..., & Heiberger, R. (2012). Package 'car.' *Vienna: R Foundation for Statistical Computing.*

Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: A review. *Journal of Voice*, *27*(3), 390.e21-390.e29. https://doi.org/10.1016/j.jvoice.2012.12.010

Giddens, C. L., Barron, K. W., Clark, K. F., & Warde, W. D. (2010). Beta-Adrenergic Blockade and Voice: A Double-Blind, Placebo-Controlled Trial. *Journal of Voice*, *24*(4), 477–489. https://doi.org/10.1016/j.jvoice.2008.12.002

Godin, K. W., & Hansen, J. H. L. (2015). Physical task stress and speaker variability in voice quality. *Eurasip Journal on Audio, Speech, and Music Processing*, *2015*(1). https://doi.org/10.1186/s13636-015-0072-7

Godin, K. W., Hasan, T., & Hansen, J. H. L. (2012). Glottal waveform analysis of physical task stress speech. *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012*, *2*(January 2012), 1646–1649.

Helpman, L., Penso, J., Zagoory-Sharon, O., Feldman, R., & Gilboa-Schechtman, E. (2017). Endocrine and emotional response to exclusion among women and men; cortisol, salivary alpha amylase, and mood. *Anxiety, Stress, & Coping*, *30*(3), 253–263.

https://doi.org/10.1080/10615806.2016.1269323

Kappen, M., Hoorelbeke, K., Madhu, N., Demuynck, K., & Vanderhasselt, M.-A. (2022). Speech as an indicator for psychosocial stress: A network analytic approach. *Behavior Research Methods*, *54*(2), 910–921. https://doi.org/10.3758/s13428-021-01670-x

Kappen, M., Van Der Donckt, J., Vanhollebeke, G., Allaert, J., Degraeve, V., Madhu, N., Van Hoecke, S., & Vanderhasselt, M.-A. (2022). Acoustic speech features in social comparison: How stress impacts the way you sound. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-26375-9

Kappen, M., Vanderhasselt, M.-A., & Slavich, G. M. (2023). Speech as a promising biosignal in precision psychiatry. *Neuroscience & Biobehavioral Reviews*, *148*, 105121. https://doi.org/10.1016/j.neubiorev.2023.105121

Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. In *Spirit and Capital in an Age of Inequality*. Springer publishing company. https://doi.org/10.4324/9781315413532

Lenth, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means*.

McEwen, B. S. (2007). Physiology and Neurobiology of Stress and Adaptation: Central Role of the Brain. *Physiological Reviews*, *87*(3), 873–904. https://doi.org/10.1152/physrev.00041.2006

Mendoza, E., & Carballo, G. (1998). Acoustic analysis of induced vocal stress by means of cognitive workload tasks. *Journal of Voice*, *12*(3), 263–273. https://doi.org/10.1016/S0892-1997(98)80017-9

Petrocchi, N., Piccirillo, G., Fiorucci, C., Moscucci, F., Di Iorio, C., Mastropietri, F., Parrotta, I., Pascucci, M., Magrì, D., & Ottaviani, C. (2017). Transcranial direct current stimulation

enhances soothing positive affect and vagal tone. *Neuropsychologia*, *96*, 256–261. https://doi.org/10.1016/j.neuropsychologia.2017.01.028

Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors*, *20*(2), Article 2. https://doi.org/10.3390/s20020479

Rothkrantz, L. J. M., Wiggers, P., Van Wees, J. W. A., & Van Vark, R. J. (2004). Voice stress analysis. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, *3206*, 449–456. https://doi.org/10.4135/9781452229300.n1969

Slavich, G. M. (2016). Life Stress and Health: A Review of Conceptual Issues and Recent Findings. *Teaching of Psychology*, *43*(4), 346–355. https://doi.org/10.1177/0098628316662768

Slavich, G. M., Taylor, S., Picard, R. W., Slavich, G. M., Taylor, S., & Stress, R. W. P. (2019). *Stress measurement using speech: Recent advancements , validation issues , and ethical and privacy considerations*. *3890*. https://doi.org/10.1080/10253890.2019.1584180

Van Der Donckt, J., Kappen, M., Degraeve, V., Demuynck, K., Vanderhasselt, M.-A., & Hoecke, S. V. (2023). *Ecologically Valid Speech Collection in Behavioral Research: The Ghent Semi-spontaneous Speech Paradigm (GSSP)*. PsyArXiv. https://doi.org/10.31234/osf.io/e2qxw

Van Puyvelde, M., Neyt, X., McGlone, F., & Pattyn, N. (2018). Voice stress analysis: A new framework for voice and effort in human performance. *Frontiers in Psychology*, *9*, 1994. https://doi.org/10.3389/fpsyg.2018.01994

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, *58*(1), 425–452. https://doi.org/10.1146/annurev.psych.58.110405.085641

Williams, K. D., Cheung, C. K. T., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. *Journal of Personality and Social Psychology*, *79*(5), 748–762. https://doi.org/10.1037/0022-3514.79.5.748

Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI Journal*, *16*, 1057–1072. https://doi.org/10.17179/excli2017-480

Zwolinski, J. (2012). Psychological and Neuroendocrine Reactivity to Ostracism. *Aggressive Behavior*, *38*(2), 108–125. https://doi.org/10.1002/ab.21411