Research Article

Psychological Science
2022, Vol. 33(7) 1027–1039

DOI: 10.1177/09567976211068110
www.psychologicalscience.org/PS
SAGE

Vocal-Stress Diary: A Longitudinal Investigation of the Association of Everyday Work Stressors and Human Voice Features

Markus Langer[1], Cornelius J. König[1], Rudolf Siegel[1], Therese Fredenhagen[1], Alexander G. Schunck[1], Viviane Hähne[1], and Tobias Baur[2]
[1]Industrial and Organizational Psychology, Saarland University, and [2]Human-Centered Artificial Intelligence, Augsburg University

Abstract
The human voice conveys plenty of information about the speaker. A prevalent assumption is that stress-related changes in the human body affect speech production, thus affecting voice features. This suggests that voice data may be an easy-to-capture measure of everyday stress levels and can thus serve as a warning signal of stress-related health consequences. However, previous research is limited (i.e., has induced stress only through artificial tasks or has investigated only short-term or extreme stressors), leaving it open whether everyday work stressors are associated with voice features. Thus, our participants (111 adult working individuals) took part in a 1-week diary study (Sunday until Sunday), in which they provided voice messages and self-report data on daily work stressors. Results showed that work stressors were associated with voice features such as increased speech rate and voice intensity. We discuss theoretical, practical, and ethical implications regarding the voice as an indicator of psychological states.

Keywords
stress, human voice, diary study, work stressors, well-being, multivariate Bayesian analyses, open data, open materials

Received 2/19/21; Revision accepted 11/19/21

The voice conveys rich information about the speaker. Voice features such as pitch and speech rate are associated with speakers' age (Ptacek & Sander, 1966), sex (Lass et al., 1980), emotional states (e.g., Scherer, 2003), and disorders (e.g., depression; Mundt et al., 2007). Particularly prevalent is the assumption that changes in voice features allow inferences about an individual's stress levels (Schuller, 2018; Slavich et al., 2019), thus providing a potentially objective measure of stress, stress being broadly defined as emotional, mental, and physiological strain in response to a real or perceived demand or threat (Goodday & Friend, 2019). Given the ubiquity of wearable technologies and microphone sensors in smartphones and smart speakers (e.g., Amazon's Echo), it could be possible to gather longitudinal voice data that provide insight on human stress levels (Goodday & Friend, 2019; Slavich et al., 2019). Considering that stress is a universal cause of health problems (Hassard et al., 2018), this could help to monitor the everyday impact of stressors and facilitate early stress detection, potentially contributing to better well-being (Slavich et al., 2019).

The predominant theoretical rationale behind the assumption that stress influences voice features is that stress, resulting from stressors, impacts the sympathetic and parasympathetic systems and the hypothalamic-pituitary-adrenal (HPA) axis (Giddens et al., 2013).

Corresponding Author:
Markus Langer, Saarland University, Industrial and Organizational Psychology
Email: markus.langer@uni-saarland.de

Stress is characterized by high physiological arousal associated with activation of the sympathetic system and the HPA axis stimulating production of adrenalin and cortisol (Peifer et al., 2014). In response, this should increase muscle tension (e.g., in the neck and throat), which affects the tension of the vocal cords, and can lead to a higher respiratory rate and bronchodilation, which increases the pressure with which air flows through the glottis. This should influence speech production and thus voice features (Giddens et al., 2013; Scherer, 1986). For instance, a higher tension in the vocal cords increases the voice's fundamental frequency (or pitch); bronchodilation and a higher respiratory rate can increase voice intensity and speech rate (for details on the mechanics of speech production, see Zhang, 2016).

Several studies support associations between stressors, stress, and voice features (for reviews, see Giannakakis et al., 2022; Giddens et al., 2013). For instance, Giddens et al. (2010) induced stress in 12 participants through pain (putting participants' hands in ice water) and found a higher speech rate in a speech-production task (i.e., repeating syllables). Most previous work induced stress through artificial tasks—for example, Mendoza and Carballo (1998) had 82 students perform different cognitive tasks (e.g., spelling backward) and told the students that performance in these tasks would affect their course grades. The authors found increased fundamental frequency and lower jitter and shimmer (measures of voice quality) compared with baseline tasks. Scherer et al. (2002) captured 100 male speakers' voices while they spoke standardized sentences during a multitasking experiment (conducting a deduction task while having to react to auditory signals) and found increases in fundamental frequency and speech rate. Beyond artificial tasks, there have been few attempts to capture stress in more realistic settings. Pisanski et al. (2018) induced stress by having 80 participants give a speech and interrupting them to count backward. They found a higher fundamental frequency during the task compared with baseline voice measures. Lu et al. (2012) induced stress in 14 participants by having them first conduct a job interview and then recruit strangers for studies. The authors trained machine-learning classifiers to separate baseline voice from the voice in the stressful tasks and reported a classification accuracy between 66% and 83% (depending on the feature set), with fundamental frequency and speech rate being two of the most important features for classification. Finally, previous research examined extreme stress situations—for example, Ruiz et al. (1990) investigated conversations between pilots before a plane crash and found increases in fundamental frequency.

Statement of Relevance

With the ubiquity of microphones in our environment (e.g., in smartphones and smart speakers), the human voice becomes a rich source of information. In fact, it can convey information about people ranging from personality to emotional states. There is also reason to believe that it is possible to detect stress—a universal cause for health issues—in the human voice, which could then help to prevent problematic stress levels. We show that everyday stressors such as time pressure or conflicts with colleagues are indeed related to voice features. Thus, capturing the way we speak may, in the future, help stress detection and provide an early signal of high stress levels. We emphasize that it is necessary to consider the ethical and legal implications of our findings because voice data are easily accessible and can provide insights about individuals—insights that should be drawn with consent and caution.

Whereas such findings have been used to draw the optimistic conclusion that the voice is a valid indicator of everyday stress levels (Schuller, 2018; Slavich et al., 2019), some researchers (e.g., Goodday & Friend, 2019; Mohr et al., 2017) have cautioned that there is crucial work to be done before generalizing to everyday stress experiences. We agree that it is a leap to conclude that, on the basis of research so far, the voice can be used for everyday stress assessment. Although Giddens et al.'s (2013) and Giannakakis et al.'s (2022) reviews report that research consistently finds increases in fundamental frequency, their results are less conclusive regarding speech rate and voice intensity. Furthermore, the relation between stress and measures of voice quality (e.g., jitter) is even less clear. Moreover, these reviews reveal research gaps. First, previous studies predominantly evoked stress through stressors that might not be comparable with everyday stress experiences, by inducing pain (Giddens et al., 2010) or having participants complete artificial tasks (Mendoza & Carballo, 1998). Second, most previous studies evoked short-term stress (e.g., Lu et al., 2012; Tolkmitt & Scherer, 1986; for a similar criticism, see Goodday & Friend, 2019), whereas everyday stressors affect people over days and weeks (Bakker & Demerouti, 2017). Third, previous studies investigated effects of extreme stressors (Ruiz et al., 1990)—stressors that only seldom occur in everyday life (DeLongis et al., 1982). Finally, research has investigated isolated contexts and experimentally manipulated specific stressors (e.g., Pisanski et al., 2018), whereas everyday stress is caused by various stressors (e.g., time related, interpersonal). Thus, an analysis of the association of everyday stressors and voice features is needed.

The next crucial step for research on the association between stress and the voice is to show whether everyday stressors captured longitudinally and in naturalistic settings are associated with voice features (Giddens et al., 2013; Goodday & Friend, 2019). Only if we can show this association outside the laboratory, in low-risk situations and with everyday stressors, can we conclude that voice can indeed become a useful indicator of everyday stress experiences. Thus, we conducted a 1-week diary study with working individuals in which we posited that work-related stressors, beyond other intra-individual variations related to changes in voice features (e.g., mood-induced effects on physiological arousal; Scherer, 2003), would be associated with changes in voice features. Specifically, in line with possible stress-related influences on the body (e.g., higher respiratory rate, bronchodilation, higher muscle tension), work stressors would be associated with higher fundamental frequency (similar to what has been previously found; Giannakakis et al., 2022; Giddens et al., 2013), voice intensity, and speech rate. Given that the research to date on the relationship between voice quality (e.g., jitter, shimmer) and stress is inconclusive (Giannakakis et al., 2022; Giddens et al., 2013), we captured voice-quality measures without clear hypotheses regarding their relation to work stressors.

Method

The procedure, measures, and analysis plans of the data-collection waves for this study were preregistered at AsPredicted (https://aspredicted.org/we52g.pdf and https://aspredicted.org/jp8sa.pdf). We set up two different preregistrations because each wave of data collection differed slightly. Specifically, the study procedure was parallel for each wave, except that participants responded to additional exploratory questions in the second wave (i.e., daily hours of sports, sports activity, and participants' coping style). Furthermore, the second wave of data collection was extended for an additional week. In this additional week, participants were randomly assigned to two experimental conditions. In one condition, participants were provided with information on stress, and in the other condition, they were instructed to progressively relax their muscles before recording the voice messages every day. However, the second week of the second wave of data collection resulted in a low number of participants completing the study; thus, we decided to use data from only the first week of data collection.

Sample

Over both waves of data collection, we intended to collect data from as many participants as possible, with a minimum of 90 participants. This study was advertised as a diary study on the topic of well-being at work and explicitly targeted working individuals who work at least part-time. We advertised this study on different social media platforms, on the campus of a medium-size German university, and in the downtown area of a medium-size German city, and we contacted participant pools of working individuals via email.

Overall, 166 people agreed to participate. For the final data set, we included only participants who indicated that they usually work more than 18 hr per week (indicating that they work at least part-time; we excluded 14 participants because of this criterion) and participants who provided data on 4 or more days (we excluded 37 participants because of this criterion). Additionally, we excluded four participants because they reported that they had filled out the study on separate days when they had actually responded to the study several times on a single day. Furthermore, we checked whether participants followed the instructions to respond after work (or whether they responded, for instance, in the morning before work) and excluded 56 days from the final data set where participants did not follow the instructions. The final sample consisted of 111 German participants (65.77% female). Within this sample, responses were provided by 10 participants on 4 days, 13 participants on 5 days, 28 participants on 6 days, and 60 participants on all 7 days.

Participants' mean age was 32.96 years (SD = 10.90, range = 19–59), and they worked on average 37.30 hr per week (SD = 8.89). Participants were from a variety of professional backgrounds: 24% worked in health care and medical professions, 11% in administration, 11% in consulting, 10% in research, 6% in the service sector, and 5% in engineering. For the rest of the participants, less than 5% worked in areas such as human resources, information technology, automotive, banking, or insurance, respectively. Participants were offered €25 for participation.

Procedure

This study was approved by the first author's research institution's ethics committee. The study followed a longitudinal design over 1 week (Sunday until Sunday) with seven measurement occasions for each participant, excluding Saturdays, because those are more likely to be working days compared with Sundays.[1]

The entire study was conducted online, and participants used their personal smartphone devices. Each day of the study consisted of two parts: Participants sent voice messages in which they responded to a set of four questions and completed a set of daily self-report measures. After participants stated their interest in participating via email, we provided them with an overview of the study procedure and the technical details to be aware of within this study. Specifically, we provided participants with a set of instructions to follow each day. First, participants were informed that within this study, they would be asked to respond to a series of questions after work on every day of the week, optimally soon after finishing work. Second, they were instructed to start this study on a Sunday, during a week when they did not have vacation, and optimally during a week when there was no official holiday. Third, they were instructed to set a daily reminder in their calendar to notify them of the study after they finished work. Fourth, participants received instructions on how to prepare themselves every day prior to study participation. Participants were informed that before they start the study each day, they should find a quiet place at home where they can take part in the study. They were instructed to use the same room each time and to use a room in which there is no obvious echo. They were also instructed on how to generate the voice messages (i.e., put their mobile phone about 10 cm from their mouth and speak directly into the microphone).

Fifth, participants received a mobile phone number. With this number, they reached a chatbot through a messenger app. We used the app AutoResponder (Version 0.9.7; Kosmala, 2018) to program this chatbot. Sixth, participants received information on how to interact with the chatbot (together with a picture showing a sample conversation with the chatbot). The chatbot guided participants through a set of four questions, which participants were asked to respond to via voice messages. Those questions were designed and pretested so that responses to all four questions would result in about 5 min of voice recordings per day. With this, we followed the methodology of Lu et al. (2012), who analyzed about 4 min of voice recordings in their study to automatically derive voice features from those recordings. Accordingly, we decided that aiming for 5 min of voice recordings would balance the trade-off between receiving enough voice data to reliably analyze daily voice features and not overtaxing participants. At the start of participants' conversations with the chatbot, they were instructed to generate an individual key phrase and send this phrase to the chatbot via a text message in the messenger app. The chatbot then automatically responded to participants with further instructions. The chatbot instructed participants to send a keyword that triggered the first question of the day for participants.

The first question each day was, "How was your day? Please respond to this question with as much detail as possible. Report about your day from the time you woke up until the time you started the study." (Note that all questions and instructions are translated from German.) Participants were instructed to respond to this question via a voice message that was at least 2 min long. Additionally, participants were asked, after sending the voice message, to send another keyword to the chatbot. This keyword triggered the second question: "What was the highlight of your day?" The same procedure was repeated for the third question, "What didn't you enjoy about your day? If you cannot think about a negative experience, please describe today's most demanding event," and for the fourth question, "What are your plans for tomorrow? It can also be simple plans (e.g., go to bed early). Additionally, please report on whether you were able to follow through on your plans for today." For Questions 2 through 4, participants were instructed to record a 1-min voice message. After responding to the last question, participants were instructed to send another keyword to the chatbot to which the chatbot responded with a link to an online survey platform where participants responded to the self-report measures. This procedure was repeated each day. Within the online survey tool, participants first reported the key phrase they generated in the beginning of the study, which enabled us to match voice recordings and survey information. Only on the first Sunday, participants reported their age, sex, usual working hours per week, profession, and long-term stressors. On each day, participants reported their positive and negative mood, everyday work stressors, hours worked that day, and perceived stress.

Measures

Items used in this study were administered in German. The English version of the items can be found in Table S1 in the Supplemental Material available online. We chose these measures because the respective constructs have been shown to be associated with human voice features in previous research.

Measure on only the first Sunday. Because long-term stressors (like everyday stressors) might contribute to changes in voice features, we measured participants' long-term stressors once on the first Sunday. Specifically, participants, on a scale from 1 (*not at all*) to 5 (*extremely*), responded to six items taken from the German Stress Coping Inventory (Satow, 2012) with the instruction, "How much strain did you experience because of the following issues over the last months or weeks?" The issues covered debt and financial problems; health issues; housing and moving; performance pressure or problems at work; demands of friends, family, or their partner; and their own expectations and demands.

Daily measures.

Positive and negative mood. Mood can affect voice features (Scherer, 2003). We thus measured participants' mood using 10 items from the German short form of the International Positive and Negative Affect Schedule Short Form (Krohne et al., 1996; Thompson, 2007; Watson, 1988). For both positive and negative mood, the scale includes five items. Participants responded on a scale from 1 (*not at all*) to 5 (*extremely*), and we used the following instruction: "This scale consists of a number of words that describe different feelings and emotions. Read each item and indicate to what extent you feel this way at the present moment." A sample item for positive mood was "active"; a sample item for negative mood was "upset."

Perceived stress. We measured perceived stress[2] with the item, "At this moment, I feel stressed" (see Metzenthin et al., 2009, or Watson, 1988, who used similar one-item measures for perceived stress). Participants responded to this item on a scale from 1 (*not at all*) to 7 (*extraordinarily*).

Everyday work stressors. To measure everyday work stressors, we used the German version of the Daily Stress Inventory (Brantley et al., 1987; Traue et al., 2000). The original version of this inventory consists of 58 statements. To not overtax participants, we used only 16 items, which we adapted to increase their relevance to everyday work experiences. However, after the first wave of data collection, we realized that for the items "I was at work although I was sick," "I was late for work," "Someone did not fulfill an agreement," "I was competing with someone else," and "I received bad news," there was nearly no variance in responses to reasonably use them within the scale for everyday work stressors. To reduce everyday study duration, we decided to remove those items for the second wave of data collection. Therefore, our final measure of everyday work stressors used in our analyses included 11 items. Sample items were "I had to work longer than intended," "I did not understand a work task," and "I was criticized or insulted." Participants responded to these items on the German version of this scale from 1 (*did not occur*) to 8 (*it was unbearable*).

Daily working hours. Participants were instructed to report how many hours they worked on the respective day.

Voice features. To extract participants' voice features, we used the software *PRAAT* (Version 6.1.38; Boersma & Van Heuven, 2001; Boersma & Weenink, 2019), which is commonly used for voice analysis throughout disciplines, including the context of the association of stress and voice (e.g., Sondhi et al., 2015). First, we extracted participants' voice recordings from the mobile phone on which the chatbot was installed. We added the key phrase that participants generated in the beginning of their individual study procedure to folders where we saved the voice recordings to match the recordings to online self-report data. Using *PRAAT*, we then automatically extracted voice features from the recordings. We chose to extract the following voice features that have commonly been measured in previous studies on voice, stressors, and stress (Giannakakis et al., 2022; Giddens et al., 2013): fundamental frequency, voice intensity, speech rate, jitter, and shimmer. Fundamental frequency was measured in hertz. Voice intensity was measured in decibels. Speech rate was measured in syllables per second. Jitter measures short-term changes and irregularities in fundamental frequency. Shimmer measures short-term changes and irregularities in voice intensity. For each of those features, we calculated the weighted mean (weighted for the duration of the respective voice recording) over the four daily voice recordings.

Data analyses

All measures were included in the main analyses reported in the Results section. Of focal interest for our study was the relation between everyday work stressors and voice features as well as perceived stress and voice features. All other variables were included because they are biological determinants of voice features (sex, age; Lass et al., 1980; Ptacek & Sander, 1966), because they have been shown to affect voice features (positive and negative mood; Scherer, 2003), or because we considered them to be control variables (long-term stressors, working hours, Sunday vs. workday).

For all measures consisting of more than one item, we calculated the mean of those items and used the resulting mean in the analyses. Our data were longitudinal and multilevel, and we had multiple dependent variables (i.e., voice features). For this longitudinal, multivariate, multilevel structure of our data, we determined that the most appropriate option to conduct our analyses was a Bayesian mixed model. Analyses were conducted in the R programming environment (Version 3.6.3; R Core Team, 2020) using the *brms* package (Version 2.14.4; Bürkner, 2017) to fit the Bayesian mixed model. In this Bayesian mixed model, we included fundamental frequency, voice intensity, speech rate, jitter, and shimmer as dependent variables. We included participants' sex, age, daily working hours, whether it was a workday or a Sunday, long-term stressors, positive and negative mood, perceived stress, and everyday work stressors as predictors on the population level (or in frequentist terms, as fixed effects). We also included the interaction between everyday work stressors and perceived stress as a predictor in our analyses because this additional subjective appraisal of stressors could fundamentally change the association between stressors and voice features (Scherer, 1986). Thus, including this interaction as a predictor should account for the fact that the effects of everyday stressors on voice features might be exacerbated among participants with high perceived stress. We estimated varying intercepts for the voice features for each participant (random intercepts).

For predictors measured on a continuous scale that were gathered only once (i.e., age, long-term stressors), we used the grand-mean-centered version of those predictors. For predictors measured on a continuous scale that were measured every day, we included the person-mean-centered version of this measure. In this way, results can be interpreted as deviations from the person mean for a respective predictor and criterion. We additionally included the grand-mean-centered person mean of the predictors measured on a continuous scale that were measured every day. This was done to also account for between-person deviations in those predictors over the course of the week. All of these variables were z-standardized before we estimated the model.

Results

Table 1 presents descriptive statistics and correlations for the measured variables. As a measure of reliability, we report McDonald's ω.

The following reporting of results is based on the analyses shown in Table 2 and Figure 1. We report on findings that relate to our main research question (i.e., relating to the association between stressors, stress, and voice features). Specifically, we report on findings where the 95% credibility interval (CrI) does not include zero and on findings where zero is not included in the 66% CrI. In the latter case, we emphasize that those results show weaker evidence for an association by reporting that there was only a "tendency." Note that unlike frequentist confidence intervals, CrIs indicate the range in which the respective estimates lie with a 95% probability based on the current data and prior distributions (for information on the prior distributions, see the Supplemental Material).

Our main goal in this study was to investigate the relation between everyday work stressors, stress, and voice features. With regard to everyday work stressors, the clearest associations were found for voice intensity (b = 0.38, 95% CrI = [0.18, 0.58]) and speech rate (b = 0.03, 95% CrI = [0.01, 0.04]); both were higher with more intense work stressors. There was also a tendency toward a higher fundamental frequency (b = 0.69, 95% CrI = –0.23, 1.61) and a tendency toward a negative association with shimmer (b = –0.08, 95% CrI = [–0.16, 0.01]). Furthermore, perceived stress showed tendencies toward being associated with lower voice intensity (b = –0.19, 95% CrI = [–0.39, 0.01]) and higher shimmer (b = 0.07, 95% CrI = [–0.02, 0.15]). Regarding the interaction between everyday stressors and perceived stress, results indicated that there was a tendency toward a higher speech rate (b = 0.01, 95% CrI = [0.00,0.02]). This means that on days on which participants experienced more intense work stressors and at the same time perceived more stress, speech rate was even higher. In summary, these results support our proposition that everyday work stressors are associated with voice features. Moreover, the results are mostly in line with what can be expected with respect to stress-related consequences on speech production, especially with regard to a higher voice intensity, higher speech rate, and, less clearly, higher fundamental frequency. Additionally, these results indicate that there is a clearer association between everyday work stressors and voice features compared with perceived stress and voice features.

As an additionally noteworthy result, the clearest association with voice features was found for positive mood. Specifically, positive mood was associated with higher fundamental frequency (b = 1.39, 95% CrI = [0.66, 2.15]), higher voice intensity (b = 0.40, 95% CrI = [0.24, 0.57]), and higher speech rate (b = 0.03, 95% CrI = [0.01, 0.04]), as well as with lower jitter (b = –0.05, 95% CrI = [–0.09, –0.02]) and lower shimmer (b = –0.09, 95% CrI = [–0.16, –0.02]). Initially, the findings for positive mood might seem contradictory when one considers that everyday work stressors were also associated with higher speech rate, voice intensity, and fundamental frequency. However, these voice features are associated not only with possible stress-related consequences on physiological activity but also with elevated physiological arousal in positive mood (Scherer, 2003) that is also reflected in the positive mood items (e.g., items such as "active").

Discussion

Our diary study shows that everyday work stressors (e.g., time related, interpersonal) are indeed associated with voice features. Thus, our study goes beyond studies manipulating single (artificial) stressors, studies investigating extreme stressors, or studies investigating short-term stress induction and lends support to the

Table 1. Means, Standard Deviations, Reliabilities, and Correlations for Key Study Variables Using a Frequentist Approach

				Correlations											
Variable	*M*	*SD*	ω	2	3	4	5	6	7	8	9	10	11	12	13
1. Sex	1.66	0.48		.01 [–18, .20]	.35 [.18, .51]	–.09 [–.27, .10]	.26 [.08, .42]	–.04 [–.22, .15]	.26 [.08, .43]	.26 [.08, .42]	.90 [.86, .93]	.15 [–.04, .33]	.09 [–.10, .27]	–.67 [–.76, –.55]	–.75 [–.82, –.65]
2. Age	32.96	10.90		—	–.08 [–.26, .11]	.02 [–.16, .21]	–.07 [–.26, .12]	.08 [–.11, .26]	–.05 [–.23, .14]	–.11 [–.29, .08]	–.14 [–.31, .05]	–.11 [–.29, .08]	.10 [–.09, .28]	.14 [–.05, .32]	.27 [.09, .44]
3. Long-term stressors	2.33	0.62	.69		—	–.18 [–.36, .00]	.50 [.35, .63]	–.04 [–.23, .14]	.53 [.38, .65]	.54 [.40, .66]	.40 [.23, .55]	.04 [–.14, .23]	.15 [–.04, .32]	–.24 [–.41, –.06]	–.17 [–.34, .02]
4. Positive mood	2.58	0.56	.89 (.81–.92)			—	–.04 [–.22, .15]	–.08 [–.15, –.01]	–.11 [–.18, –.03]	–.07 [–.26, .11]	.15 [.07, .22]	.19 [.12, .26]	.17 [.09, .24]	–.10 [–.18, –.03]	–.09 [–.17, –.02]
5. Negative mood	1.38	0.35	.82 (.71–.86)			–.06 [–.13, .02]	—	.10 [.03, .18]	.56 [.51, .61]	.64 [.52, .74]	–.06 [–.14, .01]	.00 [–.08, .07]	.09 [.02, .16]	–.04 [–.12, .03]	.03 [–.05, .10]
6. Working hours	5.29	1.80						—	.08 [–.11, .26]	.15 [–.03, .33]	.10 [.03, .17]	.04 [–.04, .11]	.06 [–.02, .13]	.04 [–.04, .11]	.05 [–.02, .12]
7. Perceived stress	2.23	0.98						.21 [.14, .28]	—	.62 [.48, .72]	.00 [–.07, .08]	–.06 [–.13, .01]	.05 [–.03, .12]	.01 [–.07, .08]	.08 [.00, .15]
8. Everyday work stressors	1.80	0.69	.89 (.84–.91)			–.09 [–.16, –.02]	.40 [.34, .46]	.42 [.36, .48]	.34 [.27, .41]	—	.06 [–.02, .13]	.13 [.06, .20]	.18 [.11, .26]	–.03 [–.11, .04]	–.03 [–.10, .05]
9. Fundamental frequency	163.55	43.53									—	.17 [–.02, .34]	.03 [–.04, .11]	–.70 [–.78, –.59]	–.06 [–.14, .01]
10. Voice intensity	61.15	4.49									.15 [.07, .22]	—	.37 [.31, .43]	–.44 [–.50, –.38]	–.48 [–.54, –.42]
11. Speech rate	3.05	0.28											—	.05 [–.14, .24]	.18 [–.01, .35]
12. Jitter	2.58	0.60									.20 [.12, .27]		–.22 [–.29, –.15]	—	.68 [.64, .72]
13. Shimmer	9.93	1.93											.05 [–.02, .13]		—

Note: Correlations in the upper triangle are between participants (multiple measurements were averaged). Correlations within the lower triangle are within participants. For each correlation, a 95% confidence interval is provided in brackets. For reliability (given as McDonald's ω), values in parentheses show the range of reliability values between multiple measurements.

Table 2. Results of the Bayesian Mixed Model Predicting Speech Characteristics From Measures of Work-Related Stress

	Dependent variable									
	Fundamental frequency		Voice intensity		Speech rate		Jitter		Shimmer	
Predictor	*b*	95% CrI	*b*	95% CrI	*b*	95% CrI	*b*	95% CrI	*b*	95% CrI
Between-participant parameters										
Female	190.57	[185.49, 195.42]	61.39	[60.11, 62.68]	3.09	[3.01, 3.17]	2.31	[2.17, 2.45]	8.89	[8.54, 9.25]
Male	111.42	[105.17, 117.36]	59.96	[58.35, 61.56]	3.04	[2.94, 3.14]	3.15	[2.98, 3.32]	12.04	[11.61, 12.47]
Age	–4.59	[–7.47, –1.69]	–0.50	[–1.35, 0.32]	0.03	[–0.02, 0.09]	0.08	[-0.00, 0.17]	0.54	[0.32, 0.75]
Working hours	0.97	[–1.83, 3.74]	–0.27	[–1.12, 0.58]	0.01	[–0.05, 0.06]	–0.07	[-0.16, 0.01]	0.06	[–0.16, 0.28]
Long-term stressors	1.34	[–2.03, 4.65]	–0.01	[–1.11, 1.06]	0.09	[0.03, 0.16]	0.04	[-0.07, 0.15]	0.39	[0.11, 0.67]
Positive mood	0.46	[–2.58, 3.35]	0.72	[–0.17, 1.58]	0.08	[0.02, 0.13]	0.02	[-0.07, 0.10]	0.03	[–0.19, 0.26]
Negative mood	1.40	[–2.16, 4.95]	–1.26	[–2.52, –0.02]	–0.07	[–0.15, 0.00]	–0.02	[-0.15, 0.10]	0.11	[–0.21, 0.43]
Perceived stress	–0.22	[–3.90, 3.51]	0.60	[–0.58, 1.86]	–0.02	[–0.09, 0.06]	0.02	[-0.10, 0.14]	–0.09	[–0.41, 0.22]
EWS	3.09	[–0.70, 6.85]	0.59	[–0.73, 1.94]	0.02	[–0.07, 0.10]	–0.07	[-0.21, 0.07]	–0.21	[–0.57, 0.13]
EWS × Perceived Stress	0.57	[–2.12, 3.25]	0.27	[–0.54, 1.10]	–0.05	[–0.10, 0.00]	0.01	[-0.08, 0.09]	–0.12	[–0.33, 0.09]
Within-participant parameters										
Working hours	0.83	[–0.36, 2.00]	–0.07	[–0.34, 0.20]	–0.00	[–0.02, 0.01]	0.02	[–0.03, 0.07]	0.02	[–0.09, 0.14]
Sunday vs. workday	–0.16	[–2.55, 2.21]	0.17	[–0.39, 0.74]	0.01	[–0.03, 0.05]	–0.03	[–0.14, 0.08]	0.04	[–0.21, 0.29]
Positive mood	1.39	[0.66, 2.15]	0.40	[0.24, 0.57]	0.03	[0.01, 0.04]	–0.05	[–0.09, –0.02]	–0.09	[–0.16, –0.02]
Negative mood	–0.97	[–1.89, –0.03]	–0.03	[–0.23, 0.17]	0.00	[–0.01, 0.02]	–0.01	[–0.04, 0.03]	0.03	[–0.06, 0.12]
Perceived stress	0.33	[–0.56, 1.24]	–0.18	[–0.39, 0.01]	–0.00	[–0.01, 0.01]	0.01	[–0.03, 0.05]	0.07	[–0.02, 0.15]
EWS	0.69	[–0.23, 1.61]	0.38	[0.18, 0.58]	0.03	[0.01, 0.04]	–0.02	[–0.06, 0.02]	–0.08	[–0.16, 0.01]
EWS × Perceived Stress	–0.20	[–0.85, 0.44]	–0.01	[–0.15, 0.13]	0.01	[–0.00, 0.02]	0.01	[–0.02, 0.04]	–0.00	[–0.06, 0.06]

Note: Random effects: $\sigma^2 = 0.07$, $\tau_{00} = 0.03$, intraclass correlation coefficient = .32, $N = 111$, observations = 691. Including a categorical variable such as participants' sex in a Bayesian analysis can lead to a higher uncertainty in estimation for one sex category over the other (for an in-depth discussion on this issue, see McElreath, 2020). To prevent this, we included each sex category as a baseline, which resulted in one estimate for male participants that acted as an intercept for male participants and another estimate for female participants that acted as an intercept for female participants. EWS = everyday work stressors; CrI = credibility interval.

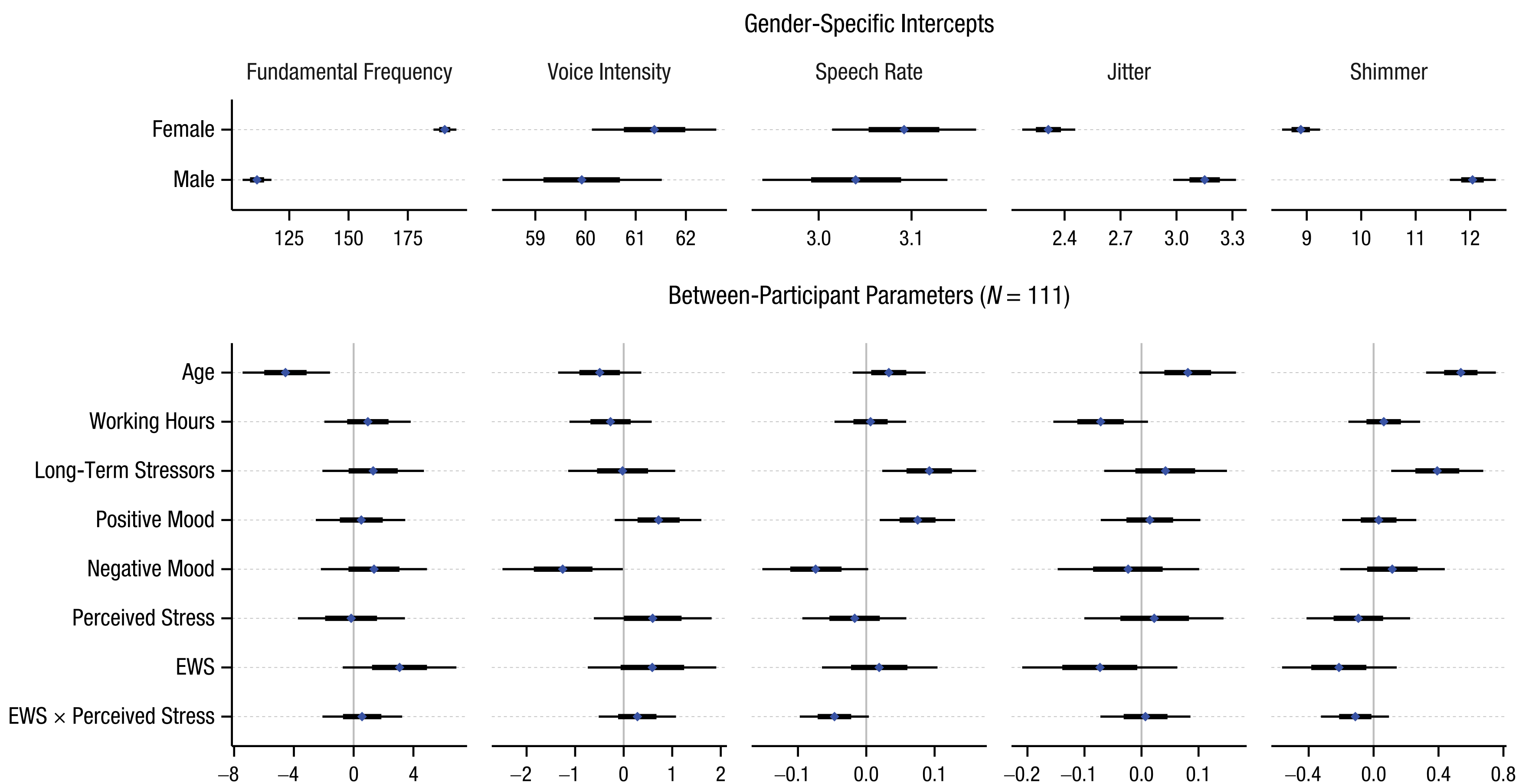

Fig. 1. *(continued on next page)*

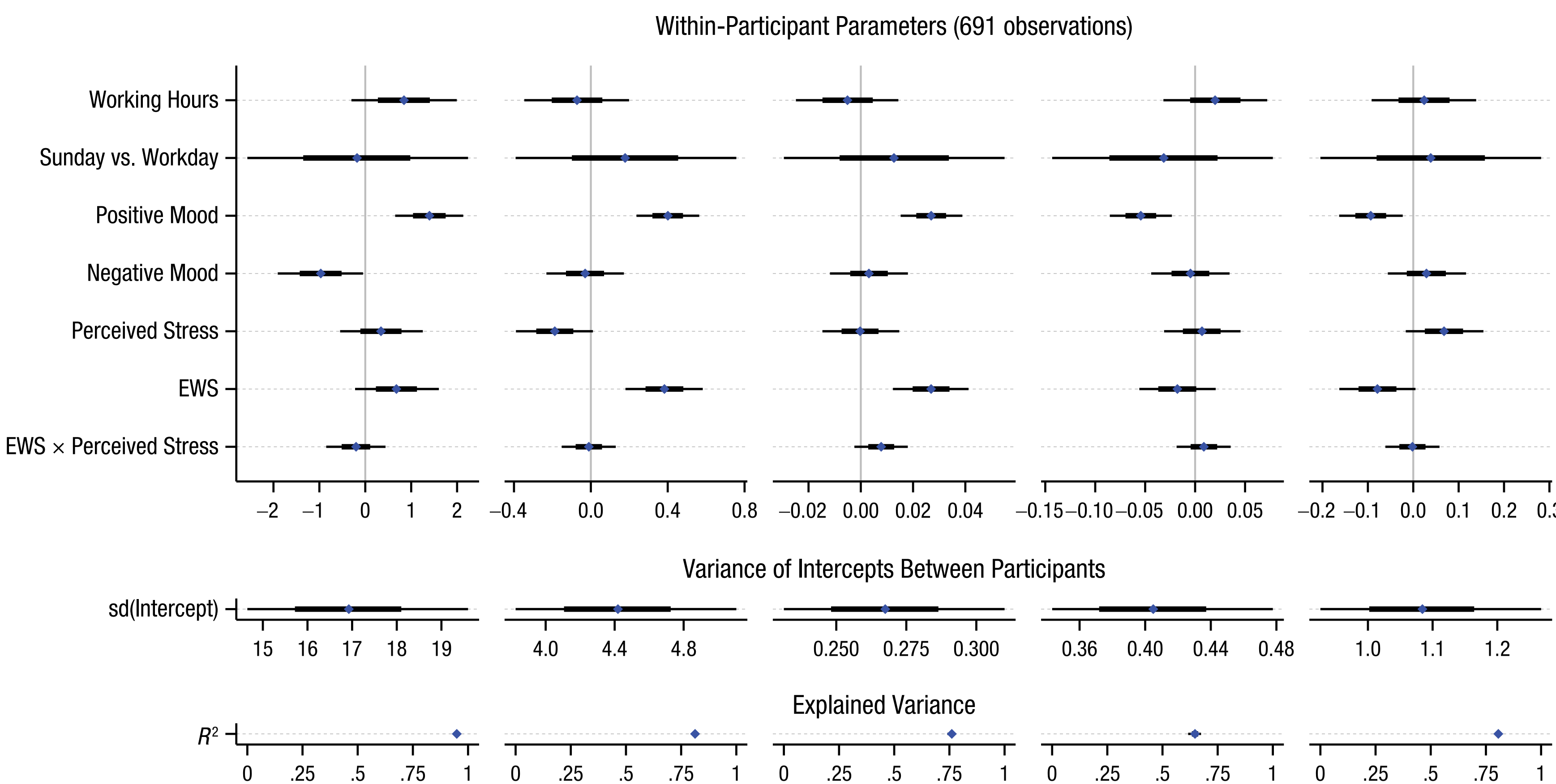

Fig. 1. Results and parameter distributions of the Bayesian mixed model predicting speech characteristics from measures of work-related stress. Each column shows results for one of the five dependent variables. The top graph shows intercept values for women and men in the study. The next two graphs show parameter estimates for each between-participant and within-participant parameter, respectively. The fourth graph shows the variation of intercepts between participants, and the fifth graph shows the variance explained by the model. In the top four graphs, blue triangles show the mean of the posterior distribution, thick lines indicate the 66% quantiles, and thin lines indicate the 95% quantiles. In the fifth graph, blue triangles indicate the proportion of variance. The parameter Sunday vs. workday shows the effect of a workday against a Sunday. EWS = everyday work stressors.

notion that everyday work stressors can also lead to changes in the human body that affect speech production (Zhang, 2016). This is the crucial, and previously missing, next step for research on the potential use of the voice as an indicator of everyday stress (Goodday & Friend, 2019; Slavich et al., 2019).

Our findings imply that everyday work stressors were associated with higher voice intensity, higher speech rate, and a tendency toward a higher fundamental frequency and lower shimmer. First, this supports the hypothesis that everyday stress-related changes in the human body (i.e., high physiological arousal associated with activation of the sympathetic system and the HPA axis) influence the mechanics of speech production (e.g., muscle tension, bronchodilation, higher respiratory rate) in a way that may be measurable in voice features. Second, this supports previously found associations of various stressors and higher fundamental frequency (Giannakakis et al., 2022; Giddens et al., 2013). However, we went beyond previous work by showing this association for everyday stressors in a naturalistic setting and by showing that even under the consideration of other variables influencing the voice (e.g., mood), associations between voice features and everyday stressors remain stable. Third, our results support the hypothesis that voice intensity and speech rate may increase through stressors, which is in line with assumed stress-related consequences on the body (Zhang, 2016). However, given ambiguous previous findings (e.g., regarding the associations between stressors and voice features other than fundamental frequency; Giannakakis et al., 2022), research needs to further uncover relations between stressors and voice features as well as possible situational (e.g., display rules) and individual moderators (e.g., coping style; Scherer, 2003). Fourth, our findings imply that everyday stressors were more strongly associated with voice features than perceived stress. A tentative interpretation of this finding is that the voice may reflect bodily changes due to everyday stressors and may be less related to subjective stress perceptions. If this interpretation proves to be valid, this would support the potential of the voice as a possibly objective indicator of stress.

Before discussing further implications, we highlight one central limitation: We assumed that stress leads to bodily changes that affect speech production but did not measure these changes (e.g., effect of stressors on cortisol). We draw the conclusion that everyday stressors affect bodily changes in ways that align with our theoretical rationale because we found changes in voice features that speak in favor of the proposed changes. Because of the already demanding study design, we decided not to measure cortisol, although this could have increased the clarity of our results (see also Pisanski et al., 2016).

Nevertheless, our study demonstrates that everyday stress is associated with the voice. Specifically, we emphasize that there might be potential to use the voice as an indicator of people's everyday stress levels and as a possible easy-to-capture source of data to detect and prevent stress-related consequences on well-being (Slavich et al., 2019). Clearly, voice data will not be the only or the best indicator for stress—it may rather be used in combination with other stress-related measures (e.g., electrodermal activity, cortisol; Giannakakis et al., 2022) as an inobtrusive and easy-to-capture additional measure.

Several future steps need to be taken to make voice analysis more viable. For instance, our study investigated only a small range of possible voice features. We see potential in analyzing longitudinal data sets with machine-learning approaches, which might detect more complex patterns than the ones we analyzed (Slavich et al., 2019). This would support Lu et al.'s (2012) preliminary results that the accuracy of machine-learning-based stress classification using voice features may be above 80% (although based on only 14 participants). Moreover, we restricted our analyses to only voice features captured in a sample of adult speakers of one language (German). This may limit the generalizability of our findings because research has shown that relations between voice features and human states (e.g., emotions) can also depend on people's language and/or cultural background (Kamiloğlu et al., 2020). In addition, research that additionally decodes what people say will yield further insights into stress experiences (Schuller, 2018). This could provide information on coping mechanisms or interpretations of daily stressors (Scherer, 1986).

Finally, if we want to advance the assessment of psychological states via voice data, it is crucial to consider ethical and legal implications (Mohr et al., 2017; Slavich et al., 2019). For instance, regarding data privacy, we need to determine the circumstances under which smart speakers are allowed to analyze the voice for inferences about psychological states (Slavich et al., 2019). Furthermore, we need to discuss what inferences can or should be drawn from voice features. Even application scenarios that are intended to benefit well-being can be perceived as undermining human autonomy. For example, smart speakers could constantly measure stress without people being in control of what information they want to provide to the party that analyzes their data.

Successfully taking the next steps toward using speech and voice information as stress indicators could help prevent stress-related negative consequences for well-being (Hassard et al., 2018). We imagine sensor-based stress monitoring that serves to capture individuals' everyday stress levels. These data could be shared with therapists to enable them to ask about specific days or events when patients experienced stress, as reflected in their voice. Several companies (e.g., www.audeering

.com; Engel, 2017) are working on applications using the voice as a window to psychological states. Our results support the potential of such applications for stress measurement and will hopefully serve to start discussions about the beneficial use of such applications.

Transparency

Action Editor: Daniela Schiller
Editor: Patricia J. Bauer
Author Contributions
M. Langer developed the study concept. All the authors contributed to the study design. M. Langer, T. Fredenhagen, A. G. Schunck, and V. Hähne collected the data. M. Langer, R. Siegel, and T. Baur analyzed the data. M. Langer, R. Siegel, and C. J. König interpreted the data. M. Langer and R. Siegel drafted the manuscript, and C. J. König provided critical revisions. All the authors approved the final manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This article presents research partly funded by the project ForDigitHealth. The project is part of the Bavarian Research Association on Healthy Use of Digital Technologies and Media (ForDigitHealth) funded by the Bavarian Ministry of Science and Arts.
Open Practices
All data, analysis scripts, and materials have been made publicly available via OSF and can be accessed at https://osf.io/byjef/. The design and analysis plans for each wave of data collection were preregistered on AsPredicted at https://aspredicted.org/we52g.pdf and https://aspredicted.org/jp8sa.pdf. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

Markus Langer https://orcid.org/0000-0002-8165-1803
Cornelius J. König https://orcid.org/0000-0003-0477-8293

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211068110

Notes

1. For one wave of data collection, we gathered data on Saturdays but decided to not use Saturday measurements in the analyses to keep the waves of data collection comparable.
2. We also measured perceived relaxedness but used only perceived stress in our analyses because this more directly reflects perceived stress.

References

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22*(3), 273–285. https://doi.org/10.1037/ocp0000056

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9–10), 341–347.

Boersma, P., & Weenink, D. (2019). *PRAAT: Doing phonetics by computer* (Version 6.1.38) [Computer software]. http://www.praat.org/

Brantley, P. J., Waggoner, C. D., Jones, G. N., & Rappaport, N. B. (1987). A daily stress inventory: Development, reliability, and validity. *Journal of Behavioral Medicine*, *10*(1), 61–73. https://doi.org/10.1007/BF00845128

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

DeLongis, A., Coyne, J. C., Dakof, G., Folkman, S., & Lazarus, R. S. (1982). Relationship of daily hassles, uplifts, and major life events to health status. *Health Psychology*, *1*(2), 119–136. https://doi.org/10.1037/0278-6133.1.2.119

Engel, J. B. (2017, September 13). MIT spinout Affectiva adds voice analysis to its emotion-sensing tech. *Xconomy*. https://xconomy.com/boston/2017/09/13/mit-spinout-affectiva-adds-voice-analysis-to-its-emotion-sensing-tech/

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2022). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440–460. https://doi.org/10.1109/TAFFC.2019.2927337

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

Goodday, S. M., & Friend, S. (2019). Unlocking stress and forecasting its consequences with digital technology. *npj Digital Medicine*, *2*(1), Article 75. https://doi.org/10.1038/s41746-019-0151-8

Hassard, J., Teoh, K. R. H., Visockaite, G., Dewe, P., & Cox, T. (2018). The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology*, *23*(1), 1–17. https://doi.org/10.1037/ocp0000069

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. A. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic Bulletin & Review*, *27*(2), 237–265. https://doi.org/10.3758/s13423-019-01701-x

Kosmala, T. (2018). *AutoResponder* (Version 0.9.7) [Mobile app]. https://www.autoresponder.ai/

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer Deutschen Version der "Positive and Negative Affect Schedule" (PANAS) [Investigations with a German version of the Positive and Negative Affect Schedule (PANAS)]. *Diagnostica*, *42*(2), 139–156.

Lass, N. J., Almerino, C. A., Jordan, L. F., & Walsh, J. M. (1980). The effect of filtered speech on speaker race and sex identifications. *Journal of Phonetics*, *8*(1), 101–112. https://doi.org/10.1016/S0095-4470(19)31445-7

Lu, H., Frauendorfer, D., Rabbi, M., Schmid Mast, M., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., & Choudhury, T. (2012). StressSense: Detecting stress in unconstrained acoustic environments using smartphones. In H.-H. Chu & G. Hayes (Program Chairs), *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 351–360). Association for Computing Machinery. https://doi.org/10.1145/2370216.2370270

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Chapman and Hall. https://doi.org/10.1201/9780429029608

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

Metzenthin, P., Helfricht, S., Loerbroks, A., Terris, D. D., Haug, H. J., Subramanian, S. V., & Fischer, J. E. (2009). A one-item subjective work stress assessment tool is associated with cortisol secretion levels in critical care nurses. *Preventive Medicine*, *48*(5), 462–466. https://doi.org/10.1016/j.ypmed.2009.02.001

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, *13*, 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics*, *20*(1), 50–64. https://doi.org/10.1016/j.jneuroling.2006.04.001

Peifer, C., Schulz, A., Schächinger, H., Baumann, N., & Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress—Can u shape it? *Journal of Experimental Social Psychology*, *53*, 62–69. https://doi.org/10.1016/j.jesp.2014.01.009

Pisanski, K., Kobylarek, A., Jakubowska, L., Nowak, J., Walter, A., Błaszczyński, K., Kasprzyka, M., Łysenkoa, K., Sukiennike, I., Piąteka, K., Frackowiak, T., & Sorokowski, P. (2018). Multimodal stress detection: Testing for covariation in vocal, hormonal and physiological responses to Trier Social Stress Test. *Hormones and Behavior*, *106*, 52–61. https://doi.org/10.1016/j.yhbeh.2018.08.014

Pisanski, K., Nowak, J., & Sorokowski, P. (2016). Individual differences in cortisol stress response predict increases in voice pitch during exam stress. *Physiology & Behavior*, *163*, 234–238. https://doi.org/10.1016/j.physbeh.2016.05.018

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech and Hearing Research*, *9*(2), 273–277. https://doi.org/10.1044/jshr.0902.273

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 3.6.3) [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org

Ruiz, R., Legros, C., & Guell, A. (1990). Voice analysis to predict the psychological or physical state of a speaker. *Aviation, Space, and Environmental Medicine*, *61*(3), 266–271.

Satow, L. (2012). SCI - Stress- und Coping-Inventar [SCI – Stress and coping inventory]. *Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Elektronisches Testarchiv (Psyndex Tests-Nr. 9006508)*. https://doi.org/10.23668/psycharchives.424

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion. *Speech Communication*, *40*(1–2), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., & Bänziger, T. (2002). Acoustic correlates of task load and stress. In J. H. L. Hansen, & B. Pellom (Eds.), *Proceedings of the International Conference on Spoken Language Processing* (pp. 2017–2020). International Speech Communication Association.

Schuller, B. W. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Communications of the ACM*, *61*(5), 90–99. https://doi.org/10.1145/3129340

Slavich, G. M., Taylor, S., & Picard, R. W. (2019). Stress measurement using speech: Recent advancements, validation issues, and ethical and privacy considerations. *Stress*, *22*(4), 408–413. https://doi.org/10.1080/10253890.2019.1584180

Sondhi, S., Khan, M., Vijay, R., & Salhan, A. K. (2015). Vocal indicators of emotional stress. *International Journal of Computer Applications*, *122*(15), 38–43. https://doi.org/10.5120/21780-5056

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242. https://doi.org/10.1177/0022022106297301

Tolkmitt, F. J., & Scherer, K. R. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 302–313. https://doi.org/10.1037/0096-1523.12.3.302

Traue, H. C., Hrabal, V., & Kosarz, P. (2000). AlltagsBelastungs-Fragebogen (ABF): Zur inneren Konsistenz, Validierung und Stressdiagnostik mit dem deutschsprachigen Daily Stress Inventory [Daily Stress Inventory: Reliability, validity and stress diagnostic with a German version of the Daily Stress Inventory]. *Verhaltenstherapie und Verhaltensmedizin*, *21*(1), 15–38.

Watson, D. (1988). Intraindividual and interindividual analyses of positive and negative affect: Their relation to health complaints, perceived stress, and daily activities. *Journal of Personality and Social Psychology*, *54*(6), 1020–1030. https://doi.org/10.1037/0022-3514.54.6.1020

Zhang, Z. (2016). Mechanics of human voice production and control. *Journal of the Acoustical Society of America*, *140*(4), 2614–2635. https://doi.org/10.1121/1.4964509