ASSESSING VIBROACOUSTIC SOUND MASSAGE THROUGH THE BIOSIGNAL OF HUMAN SPEECH: EVIDENCE OF IMPROVED WELLBEING

Charlotte Fooks, Oliver Niebuhr

Centre for Industrial Electronics, University of Southern Denmark, Sonderborg, Denmark

ABSTRACT

Stress has notorious and debilitating effects on individuals and entire industries alike, with instances of stress continuing to rise post-pandemic. We investigate here (1) if the new technology of Vibroacoustic Sound Massage (VSM) has beneficial effects on user wellbeing and (2) if we can measure these effects based on the biosignal of speech prosody. Forty participants read a text before and after VSM treatment (45 min). The 80 readings were subjected to a multi-parameteric acoustic-prosodic analysis. Results provided positive answers to (1) and (2), showing that timbre and pitch features were most sensitive to VSM treatment, followed by loudness and pausing. Overall, participants spoke more deeply, softly, and slowly after VSM, suggesting that they were in a more relaxed state. Practical implications and future research perspectives are discussed.

Index Terms— prosody, VSM, biosignals, wellbeing.

1. INTRODUCTION

The destigmatisation of mental health has birthed a global stress-management movement and subsequent boom in the health and wellness industries. The corporate wellbeing market grows by about 10% every year, and the COVID-19 pandemic "has radically shifted the focus in terms of benefits offerings", according to Bersin [1]. Today, "not only do we need to keep employees healthy, safe, and protected but also, we now need to help them adapt to constant disruption. All over the world employees are under stress, and the uncertainty and stress of change has caused burnout. So, a new paradigm has emerged: a corporate focus on resilience" [1].

Yoga is a popular physical workplace intervention to improve this resilience, well-acclaimed for its relaxation and stress-alleviating properties [2]. Over and above such traditional approaches, "a growing market tends to attract new ideas" [1]. Consistent with this axiomatic statement, technological advances including wearables, mobile and desktop applications, AI and new hardware constantly infiltrate the corporate wellbeing market, and are endorsed by organisations and workers as effective ways to monitor and mediate stress levels [3].

The prospects that contemporary technology poses to the workplace wellness industry are central to our research, which assesses the effects of a vibroacoustic intervention on stress.

1.1. Vibroacoustic Sound Massage (VSM)

Vibroacoustic Sound Massage (VSM) is a multi-modal approach combing purposefully designed audio soundscapes (wind, waterfalls, temple sounds etc.) with low-frequency tactile vibration stimulation synchronised to these soundscapes. Research posits that vibroacoustic stimulation may be pertinent to stress management [4]. Bracketed as an alternative treatment, its uses range from inducing a relaxation response [5], to pain relief [6], mood management [7], and symptom reduction for dementia [8] and depression [9]. It has a multitude of alias's, such as physioacoustics or physioacoustic therapy or treatment, and rhythmic sensory stimulation, in addition to a myriad of physical vibration-transducing devices including chairs, beds, boards, mattresses, tables and other equipment. All hardware has built-in speakers that emit kinetic low-frequency sonic vibrations.

The present study exploits one of the most sensitive human biosignals, i.e. speech, to analyse and assess the efficacy of VSM and draw conclusions on how it affects user wellbeing. Unlike questionnaire data or rating tasks, biosignals can be collected without having to instruct participants or requiring certain actions. This makes studies on perception easier, faster, and less prone to interpretation biases [10]. Analysing and interpreting biosignals also avoids conscious participant reflection as to correct, appropriate response behavior and/or the supposed goals of the study [11]. Within the domain of speech, we focus on a specific set of prosodic parameters.

1.2. Speech Prosody: A useful biosignal

The melody and rhythm of speech, known as speech prosody, is crucial in communication and comprehension. Prosodic parameters are systematised in terms of four dimensions: pitch, loudness, voice quality or timbre, and duration or tempo, e.g., of pauses, syllables, and speech breathing [12]. The timbre dimension has long been neglected but recently gained increasing attention with the research and modeling of speech moving beyond syntactic units such as words and phrases into the realm of emotions, attitudes and speaker traits [13, 14]. Gobl & Chasaide's study [15] showed "that voice quality changes alone can evoke differences in [perceived] speaker affect" (p.204); while Patel et al. [13] found timbre parameters to be as powerful as the more traditionally popular pitch parameters to distinguish emotional speaker states. Consistent with this, more than half of the parameters included in our speech analysis are also timbre parameters.

In general, prosodic parameters are subtle, powerful markers that modulate speech both consciously and unconsciously. They can be used intentionally as in therapeutic contexts, strategically in public speaking, and are oftentimes non-negotiable when expressing emotional states [16]. For example, emotion identification solely through prosodic parameters often results in accuracy levels of 80-90% [17]. With similar precision, depression and mental stress can be detected through prosody [14] – while prosody is also a major predictive factor in success indication, as in oral-exam grades or investor-pitch performances [18, 19]. Thus, we assumed that prosody would prove to be a useful biosignal in our study as well.

1.3. Sound and vibration

The therapeutic efficacy of music is recognised and well established. As a multifaceted resource, it can be utilised for warcrime victim reintegration [20], paediatric care [21], substance abuse rehabilitation [22], psychiatric treatment [23], and workplace stress management [24] (this list is non-exhaustive). Vibration is the atomic core of all existence. On a human level, vibration regulates bodily functioning from the cellular, all the way up to sensory perception. Indeed, without vibration in the vocal tract we'd be unable to speak, and without vibrating cochlear hair cells, unable to hear. Yet despite the pillars of sound and vibration as salient to our wellbeing, welfare and inner-workings, the vibroacoustic field which pairs them, remains remarkably juvenile.

1.4. Goals and hypotheses

The main goal of our study is to test the effectiveness of VSM technology through further analysis. Particularly original is that we use biosignals to that end or, more precisely, the acoustic speech signal. Doing so, we avoid a possible social-desirability response bias that can occur with explicit rating or questionnaire/interview tasks as they were conducted in previous VSM studies. As a secondary goal, we assess whether speech prosody parameters are suitable biosignals for testing the effectiveness of VSM and similar wellbeing-related therapies and treatments. Speech signals are relatively easy and robust to collect and can be elicited intuitively – and would therefore have many practical advantages over other biosignals, such as EEG, skin-conductance response, heart rate, etc.

Acknowledging the calming effect of the VSM treatment physiologically and psychologically [25], our study's main hypothesis (H1) is that acoustic-prosodic parameters will be affected by VSM in a comparative before-after-treatment experimental design.

As to the nature of these effects, it is reasonable to assume a reduction in arousal and vocal effort from before to after VSM treatment, and perhaps also a change from a more other-directed to a more self-directed way of speaking.

Altogether, with reference to the literature in 1.2, such effects would have the following hypothesized manifestations: lower and less variable levels of (H2) pitch and (H3) loudness, (H4) a slower speaking rate, (H5) a less tense and strained voice timbre, and less structured speech in terms of fewer and/or shorter silent pauses (H6).

In the following, we explain how these measurements were carried out, how and from whom speech material was elicited, and how the elicitation procedure was embedded in the study design.

2. METHOD

2.1. Experimental procedure

Forty voluntary participants between 20-35 years old (13 males/27 females) were recruited to partake in the study.

The study itself consisted of three stages: The before-treatment reading of a short text (BEF), the VSM treatment (VSM), and the after-treatment reading of the same short text (AFT). BEF served as the prosodic baseline condition and AFT as the test condition against which the baseline condition was compared. We used a within-subjects experimental design in which all 40 participants went through all three stages. BEF and AFT were both allocated 5 minutes in the study design, and took approx. 2 minutes to complete. The VSM treatment lasted for 45 minutes. A total session took about 90 minutes as it additionally included extensive briefing and debriefing (incl. rating and interview) stages that are not reported here.

Fig. 1: *Example of the VSM experimental setting.*

The experiment was carried out in individual sessions at the National Institute of Public Health in Copenhagen, Denmark. The study environment was designed as a calming setting with soft light and furnishings. There were two sound absorbing and sight-blocking mobile wall panels between the researcher and the participant during each stage to reduce priming. Participants wore a Muse S (Gen 2) headset at all times throughout the experiment (the data of which is presented in a different paper), see Figure 1.

2.2. Elicited text

Participants read aloud "The Rainbow Passage", which was developed to be both phonetically balanced and comprehensive with respect to the phonotactic structures of English. It consists of a total of 1,096 phonemes (108 content words) and is widely used in the speech sciences. Besides being a standard text facilitating comparisons of our speech data with that of other studies, we chose this due to its potential for eliciting fluent, emotionally colored speech. That is, compared to other texts, it includes a greater amount of high-arousal words (17.5%) and a lower amount of less familiar words (5.5%) [26].

To avoid unintended contextualisations and biases, participants received no instruction other than to read the text aloud. For both BEF and AFT, they sat behind a free-standing sound-attenuated wall when reading the passage. Audio-only was recorded with a ZOOM H6 hi-fi audio recorder which was placed out of participant view and at a distance to ensure VSM-related intensity effects would stand out from smaller intensity fluctuations, e.g., caused by head movements. Recordings were made at 48 kHz, 24-bit.

2.3. VSM Equipment

We used the latest Bass Module of the Danish startup vibroacoustics.dk [27]. It is a wooden box the size of a one-person mattress housing three transducer speakers. Participants lay in the supine position on the module and listened with Bose QuietComfort 35 II noise cancelling headphones to a soundscape designed specifically for the study. Low-frequency bass vibrations in the music are tactile when in contact with the module. Participants were required to keep their eyes closed throughout the VSM, with some opting to wear an eye mask, see Figure 1.

The soundscape designed for the study featured cognitively-harmonious sonic elements known to induce relaxation and enhance the vibroacoustic experience. Immersive auditory attributes included bird song, water and wind [28], isochronic tones (single repetitive tones) that naturally occur in nature, and frequencies between 40-80Hz [29, 30].

2.4. Acoustic-prosodic analyses

All audio recordings were normalized to 99% of the maximum amplitude. Then, the acoustic-prosodic analysis was conducted automatically in PRAAT [31] using ProsodyPro [32] and a script of the second author. To prepare the automatic analysis, another script [33] was run to mark and separate speech sections from intermittent pauses. For the latter, we determined per participant the number of silent pauses and their average duration.

Prosodic features measured in the speech were: the mean levels of pitch (i.e. fundamental frequency, f0, Hz) and loudness (i.e. RMS, dB), the pitch minimum and maximum (Hz, 10th and 90th values to exclude measurement/octave errors), the loudness variability (std.dev. RMS, dB), and mean tempo (syll/s). Voice timbre representatives analysed were: the two spectral-tilt estimates mean H1-H2 (the acoustic open quotient [34], dB) and mean H1-A3 (a vocal effort correlate [35], dB) and the mean levels of the first, second and third formant frequencies (F1, F2, F3, Hz). Additionally, to estimate the variability in vocal effort, we determined the standard deviation of H1-A3 (dB) per person and reading. Taking sex of speaker into account, all Hz measurements were normalised to semitones prior to statistical analysis.

3. RESULTS

As multiple prosodic parameters were assessed in a within-subjects design, we analyzed the measurements statistically in a one-way repeated-measures MANOVA with VSM (BEF vs. AFT) as fixed factors. In addition, we conducted a linear discriminant analysis (LDA) to estimate and quantify the degree to which prosodic measurements alone distinguish before-VSM from after-VSM readings of "The Rainbow Passage". The RM-MANOVA showed a highly significant main effect of VSM on reader prosody before and after treatment ($F[19,19] = 8.263$, $p<0.001$, $\eta_p^2 = 0.892$). Based on that, we conducted a series of univariate analyses to determine how much each measurement contributed to this main effect of VSM.

Summarised in Table 1, results show the strongest effects of VSM (in terms of ηp^2) outside timbre, concerned decreases in pitch, loudness and tempo from before to after treatment. That is, participants performed their text passages after VSM with a deeper voice, at lower and less variable loudness levels, and at a slower pace.

Within the domain of timbre, we found strong increases of F2, F3 and the acoustic open quotient H1-H2. Less strongly but in line with decreasing loudness and variability, we found decreases in the acoustic vocal effort measures of H1-A3. Overall, results are indicative of a softer (perhaps breathier), more relaxed voice timbre. Increases in F2 and F3 levels do not contradict this characterization, while F1 levels did not increase. Increases in F2 and F3 without conformity in F1, indicates that they cannot be an acoustic manifestation of higher vocal-intensity or sentence-stress levels [36]. Thus, they indicate stronger participant smiling during the after-VSM performance [37]. With regards to pausing behavior while reading, results show VSM effects in the direction of fewer and shorter pauses that approach significance.

Prosodic parameter	*F*	*p*	η_p^2	*change BEF to AFT*
pitch level	25.996	<0.001	0.413	↓
pitch range	2.634	0.113	0.066	→
min pitch	19.108	<0.001	0.341	↓
max pitch	16.333	<0.001	0.306	↓
loudness level	19.412	<0.001	0.344	↓
loudness varia.	13.285	0.001	0.264	↓
mean tempo	25.485	<0.001	0.408	↓
H1-H2	8.629	0.006	0.189	↑
H1-A3	7.840	0.008	0.175	↓
F1 level	0.792	0.379	0.021	→
F2 level	45.273	<0.001	0.550	↑
F3 level	62.403	<0.001	0.628	↑
H1-A3 varia.	7.840	0.008	0.175	↓
pause count	3.359	0.075	0.083	(↓)
pause duration	3.086	0.087	0.077	(↓)

Table 1: Univariate test results based on the RM-MANOVA and their underlying prosodic changes (df_1=1, df_2=37).

LDA's stand. canon. discr. coeff.	
H1-A3	1.373
pitch level	1.222
H1-H2	-0.846
min pitch	-0.510
F3 level	-0.434
loudness level	0.230
loudness varia.	0.133
F2 level	0.099
H1-A3 varia.	-0.085
pause count	-0.081

Table 2: The top-10 most important standardised canonical discriminant coefficients of the LDA (in descending order).

With respect to how well the BEF and AFT text readings can be discriminated based on the measured acoustic-prosodic parameters alone, the LDA showed a significant prediction performance with an overall success rate of 85.9% (Wilks-λ = 0.565, χ^2[13] = 39.618, p<0.001). Only 15% (6/40) of the BEF readings were misclassified by the LDA as AFT readings. In the opposite direction, the confusion rate was 13.2% (5/40). Thus, it is possible to infer significantly above chance level from participants' prosody alone whether a text was read before or after VSM.

Table 2 presents the top-10 standardized canonical discriminant coefficients in descending order of importance (absolute values). We see that the discrimination performance mainly relied on timbre and pitch, followed by loudness. Pausing behavior played just a minor role, as was expected from the RM-MANOVA results.

4. DISCUSSION AND CONCLUSION

Speech is a highly complex signal encoding information not only at a word level, but also at numerous other levels, from emotions and attitudes to social hierarchies and turn-taking. These additional levels are to a large extent, encoded in prosodic patterns – patterns that are neither represented in writing nor explicitly learned, e.g., at school, but used unconsciously. All this endows speech (specifically speech prosody) as a sensitive biosignal, e.g., for diagnosis, deception detection and success, see 1.2.

On this basis, we tested as our main hypothesis H1 whether effects of VSM treatment also manifest themselves in speech prosody. Our results provide positive evidence for H1. Furthermore, in hypotheses H2-H6, we assumed that VSM effects would manifest themselves in the form of lower and less-variable pitch and loudness (H2-H3), a slower tempo (H4), a less tense and effortful voice timbre (H5), and a less other-directed way of speaking with fewer and/or shorter pauses (H6). Our results support hypotheses H2-H6, albeit for H6 only as significant trend (p<0.1).

From the observed prosodic changes, we can draw two conclusions. First, VSM treatment does lead to increased wellbeing of its users. Secondly, speech prosody is a suitable biosignal to measure this increased wellbeing. Pitch and timbre emerged as the two key signal components in this measurement, which is consistent with their predictive weighting in other studies [38].

These conclusions are of high practical value and offer promising perspectives for future research. Unlike EEG, SCR and other biosignals, speech-prosody data can be collected without major technical effort or physical contact. Wellbeing effects can therefore be determined relatively easily. Moreover, research on acoustic charisma shows, among other things, that speech prosody signals are suitable to quantify and scale stimulation effects. Transferring that to wellbeing would pave the way for numerous therapy and test measures based on speech prosody, beyond VSM.

Developing such quantifying and evaluating measures is the long-term goal of our line of research. The next short-term goal is to examine inter-individual VSM effects, and analyse efficiency parameters e.g., the minimum stimulation duration and the bandwidth of effective soundscapes.

5. ACKNOWLEDGEMENTS

This study was kindly supported by a research grant from the Danish Sound Cluster (1022-00015B). Gratitude is shared with the University of Southern Denmark Copenhagen for the study location and Lily Baastrup for the data collection assistance. Extensive thanks to all participants.

REFERENCES

[1] J. Bersin. Business Resilience Arrives: The Wellbeing Market Matures. http://hdl.handle.net/10713/18708 (accessed Sept 2023).

[2] T. Shinba *et al.*, "Changes in Heart Rate Variability after Yoga Are Dependent on Heart Rate Variability at Baseline and During Yoga," Int J Yoga, vol. 13, no. 2, pp. 160-167, 2020.

[3] G. Paganin *et al.*, "New Technologies in the Workplace: Can Personal and Organizational Variables Affect the Employees' Intention to Use a Work-Stress Management App?," Int J Environ Res Public Health, vol. 18, no. 17, 2021.

[4] J. Kantor *et al.*, "Effect of Low Frequency Sound Vibration on Acute Stress Response in University Students-Pilot Randomized Controlled Trial," Front Psychol, vol. 13, p. 980756, 2022.

[5] G. Patrick, "The Effects of Vibroacoustic Music on Symptom Reduction," IEEE engineering in medicine and biology magazine, vol. 18, no. 2, pp. 97-100, 1999.

[6] E. A. Campbell *et al.*, "Exploring the Use of Vibroacoustic Treatment for Managing Chronic Pain and Comorbid Mood Disorders: A Mixed Methods Study," Nordic Journal of Music Therapy, vol. 28, no. 4, pp. 291-314, 2019.

[7] E. Rüütel *et al.*, "The Experience of Vibroacoustic Therapy in the Therapeutic Intervention of Adolescent Girls," Nordic Journal of Music Therapy, vol. 13, no. 1, pp. 33-46, 2004.

[8] E. A. Campbell *et al.*, "Tactile Low Frequency Vibration in Dementia Management: A Scoping Review," Front Psychol, vol. 13, p. 854794, 2022.

[9] T. Braun Janzen *et al.*, "A Pilot Study Investigating the Effect of Music-Based Intervention on Depression and Anhedonia," Frontiers in Psychology, Original Research vol. 10, 2019.

[10] O. Niebuhr *et al.*, "Assessing Hate-Speech Perception through Bio-Signal Measurements: A Pilot Study," in *Proc. Biosignale Workshop*, Kiel, pp. 1-2, 2020.

[11] I. Krumpal, "Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review," Quality & Quantity, vol. 47, no. 4, pp. 2025-2047, 2013.

[12] N. Campbell *et al.*, "Voice Quality: The 4th Prosodic Dimension," in *15th International Congress of Phonetic Sciences*, Barcelona, Spain, vol. 2417–2420, 2003.

[13] S. Patel *et al.*, "Mapping Emotions into Acoustic Space: The Role of Voice Production," Biological Psychology, vol. 87, no. 1, pp. 93-98, 2011.

[14] S. A. Almaghrabi *et al.*, "Bio-Acoustic Features of Depression: A Review," Biomedical Signal Processing and Control, vol. 85, p. 105020, 2023.

[15] C. Gobl *et al.*, "The Role of Voice Quality in Communicating Emotion, Mood and Attitude," Speech Communication, vol. 40, no. 1-2, pp. 189-212, 2003.

[16] S. Paulmann *et al.*, "How Psychological Stress Affects Emotional Prosody," PLoS One, vol. 11, no. 11, p. e0165022, 2016.

[17] P. van Rijn *et al.*, "Contribution of Pitch Measures over Time to Emotion Classification Accuracy," ed: PsyArXiv, 2023.

[18] O. Niebuhr, "Advancing Higher-Education Practice by Analyzing and Training Students' Vocal Charisma: Evidence from a Danish Field Study," in *Proc. 7th International Conference on Higher Education Advances*, Valencia, Spain, pp. 743-751, 2021.

[19] J. Niehof *et al.*, "What's Pitch Got to Do with It? – the Connection between Speech Prosody and Investor-Pitch Success," in *Proc. 1st International Conference on Tone and Intonation*, Sonderborg, Denmark, pp. 288-292 2021.

[20] N. Osborne, "Neuroscience and "Real World" Practice: Music as a Therapeutic Resource for Children in Zones of Conflict," Annals of the New York Academy of Sciences, vol. 1252, no. 1, pp. 69-76, 2012.

[21] C. Hendon *et al.*, "Hospitalized Children's Mood Differences During Play and Music Therapy," Child: Care, Health and Development, vol. 34, no. 2, pp. 141-144, 2008.

[22] C. Ghetti *et al.*, "Music Therapy for People with Substance Use Disorders," Cochrane Database Syst Rev, vol. 5, no. 5, p. Cd012576, 2022.

[23] C. Freitas *et al.*, "Music Therapy for Adolescents with Psychiatric Disorders: An Overview," Clinical Child Psychology and Psychiatry, vol. 27, no. 3, pp. 895-910, 2022.

[24] N. Mao, "The Role of Music Therapy in the Emotional Regulation and Psychological Stress Relief of Employees in the Workplace," J Healthc Eng, vol. 2022, p. 4260904, 2022.

[25] C. Boyd-Brewer *et al.*, "Vibroacoustic Sound Therapy Improves Pain Management and More," Holistic Nursing Practice, vol. 18, no. 3, pp. 111-118, 2004.

[26] B. M. Ben-David *et al.*, "Linguistic and Emotional-Valence Characteristics of Reading Passages for Clinical Use and Research," J Fluency Disord, vol. 49, pp. 1-12, Sep 2016.

[27] VibroAcoustics.dk. "The Vibroacoustics Bass Module." URL (accessed 08/05/23).

[28] R. T. Buxton *et al.*, "A Synthesis of Health Benefits of Natural Sounds and Their Distribution in National Parks," Proceedings of the National Academy of Sciences, vol. 118, no. 14, p. e2013097118, 2021.

[29] A. Mosabbir *et al.*, "The Effects of Long-Term 40-Hz Physioacoustic Vibrations on Motor Impairments in Parkinson's Disease: A Double-Blinded Randomized Control Trial," Healthcare, vol. 8, no. 2, 2020.

[30] L. R. Bartel *et al.*, "Vibroacoustic Stimulation and Brain Oscillation: From Basic Research to Clinical Application," Music and Medicine, vol. 9, no. 3, p. 153, 2017.

[31] C. J. Seong, "Guidance to the Praat, a Software for Speech and Acoustic Analysis," J Korean Soc Laryngol Phoniatr Logop, vol. 33, no. 2, pp. 64-76, 2022.

[32] Y. Xu, "ProsodyPro - a Tool for Large-Scale Systematic Prosody Analysis," Proceedings of the TRASP Conference, pp. 7-10, 2013.

[33] N. H. de Jong *et al.*, "Praat Script to Detect Syllable Nuclei and Measure Speech Rate Automatically," Behavior Research Methods, vol. 41, no. 2, pp. 385-390, 2009.

[34] C. Mooshammer, "Acoustic and Laryngographic Measures of the Laryngeal Reflexes of Linguistic Prominence and Vocal Effort in German," J Ac Soc Am, vol. 127, no. 2, pp. 1047-1058, 2010.

[35] A. M. C. Sluijter *et al.*, "Spectral Balance as an Acoustic Correlate of Linguistic Stress," J Ac Soc Am, vol. 100, no. 4, pp. 2471-2485, 1996.

[36] L. L. Koenig *et al.*, "Vowel Formants in Normal and Loud Speech," J Speech Lang Hear Res, vol. 62, no. 5, pp. 1278-1295, 2019.

[37] D. Erickson *et al.*, "Are You Laughing, Smiling or Crying?," APSIPA ASC 2009 - Asia-Pacific Signal and Information Processing Association 2009 Annual Summit and Conference, 2009.

[38] P. A. Barbosa, "Pleasantness and Wellbeing in Poem Declamation in European and Brazilian Portuguese Depends Mostly on Pausing and Voice Quality," Frontiers in Communication, Original Research vol. 7, 2022.