Using wearable sensors to measure listening-related stress: a validation study

Kristie Roberts, Jani Johnson, Lipika Sarangi, & Miriam van Mersbergen
School of Communication Sciences & Disorders, University of Memphis, Memphis, TN

Presented at the Annual Meeting of the American Auditory Society, Scottsdale, AZ March 2020

INTRODUCTION

In addition to increased stress due to effortful listening[1], people with hearing loss deal with psychosocial stressors such as embarrassment, stigma, and other social consequences[2] related to hearing difficulties in daily life. Laboratory studies have used psychophysiological measures (e.g., changes in pupil dilation, heart rate, respiration, and skin sweat) to evaluate listening effort-related stress/arousal[3]; however, these assessments do not reflect effects of listening-related stress in daily life. Recent advances in wearable sensor technology might allow for similar assessments in daily listening. This research sought to evaluate whether commercially available, wearable sensors can provide valid information about psychophysiological effects associated with difficult listening in a controlled setting. Specifically, the following questions were asked:

1. Are wearable sensors sensitive to changes in listening difficulty?
2. Are wearable sensors sensitive to stressors related to negative consequences of incorrect understanding?
3. Do wearable sensors provide information that is consistent with laboratory-grade measures of psychophysiological states?

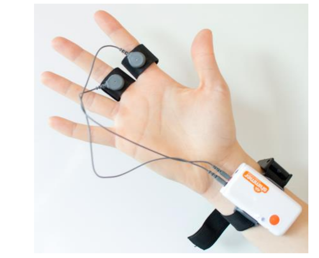
METHODS

Design: Within-subjects repeated measures
Participants: 16 young adults with normal hearing sensitivity
Stimuli: 16 Revised Speech Perception in Noise (RSPIN) lists of 25 sentences were presented at -4, 0, +4, and +8 dB SNR. Half of the lists' keywords were easily predictable from sentence context (HP) and half were not easy to predict (LP). Counterbalancing was used to determine list, predictability, and SNR orders.
Negative Feedback: To simulate negative consequences of misunderstandings in daily listening, participants were told that incorrect responses would result in a flash of light and a reduction in monetary compensation for half of the lists. Counterbalancing determined which lists resulted in negative feedback.
Procedure: Self-report measures were used to validate whether the testing conditions resulted in perceived differences in listening difficulty. Perceived task load, valence, and arousal were assessed. Participants repeated the last word of each RSPIN sentence. After each list, they reported their emotional state in terms of valence and arousal using the Self-assessment Manikin (SAM) and estimated their perceived task-load during the task using the National Aeronautics and Space Administration Task Load Index (NASA-TLX). Throughout the task participants wore 3 commercially available wearable sensors and were connected to lab-grade equipment to assess their physiological states.

BEHAVIORAL & SUBJECTIVE ASSESSMENT

R-SPIN: Behavioral measure of speech understanding
NASA-TLX: 6 questions that assess workload with a 21-gradient scale
SAM: Visual analog scales ranging from 1 to 9, one for Arousal and one for Valence

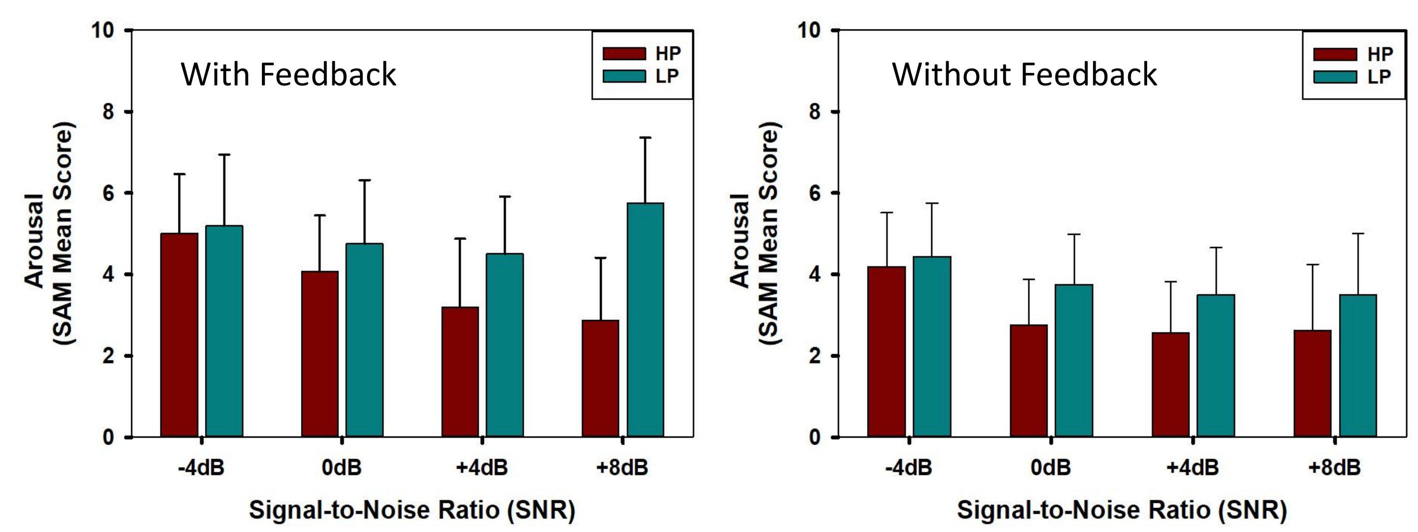
OBJECTIVE ASSESSMENT

Skin Conductance	Respiration Rate	Heart Rate
Shimmer 3+ GSR	Spire	Biostrap
		

Measurements were obtained with wearable sensors (shown below) and with laboratory-grade equipment.

RESULTS

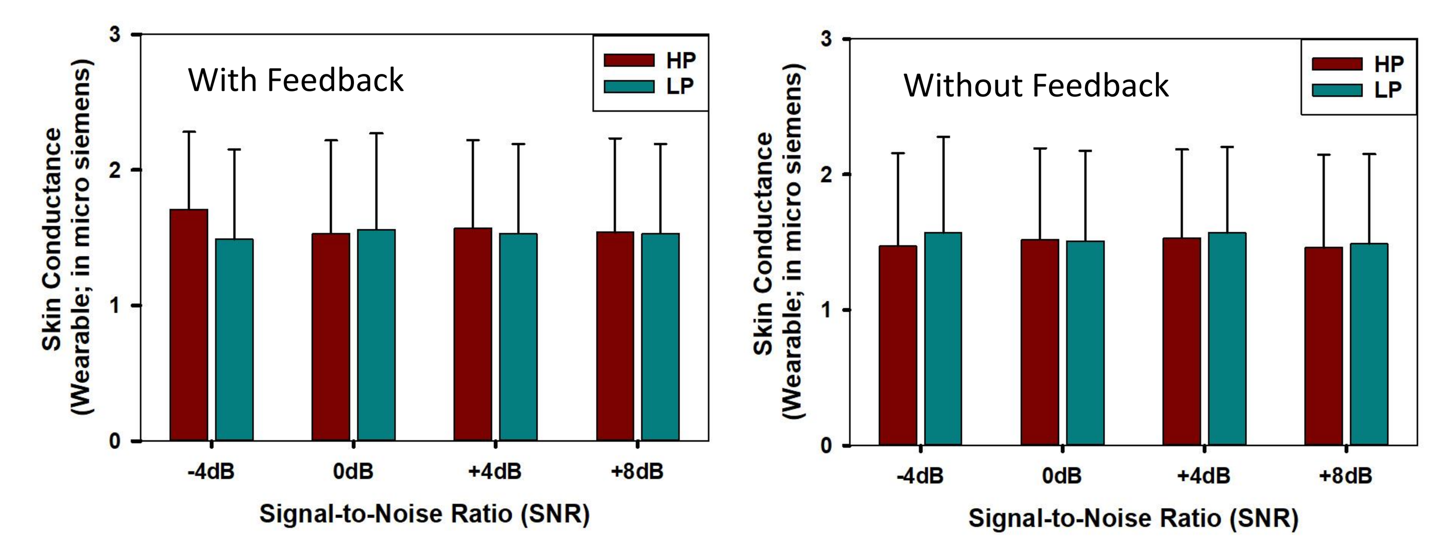
Behavioral and Self-report



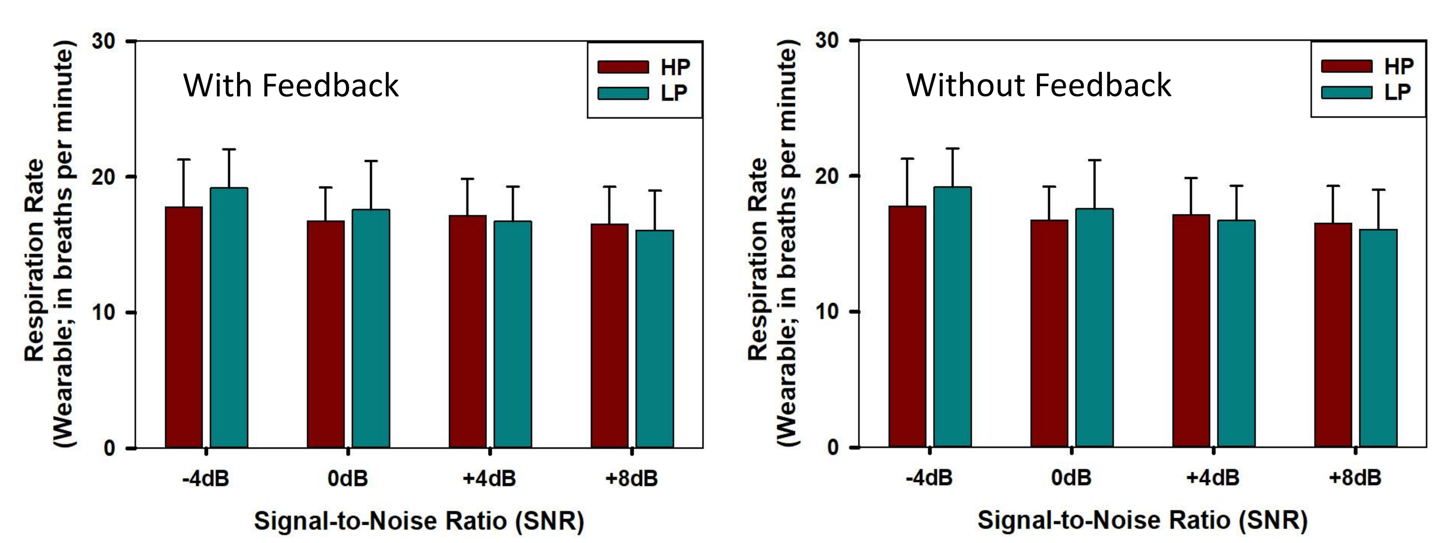
- The figure at left (self-reported arousal) demonstrates trends seen for the self-report and behavioral measures. Participants reported higher arousal with decreasing SNR, for low-predictability (LP) sentences, and when negative feedback was provided *($p \leq .001$)*.
- Poorer speech understanding performance was measured, and more negative valence (unpleasantness) and greater task load were reported with decreasing SNR and for LP sentences. ($p < .001$, Not shown.)

Skin Conductance

- No observable changes in galvanic skin response were noted across conditions for wearable or lab-grade equipment.



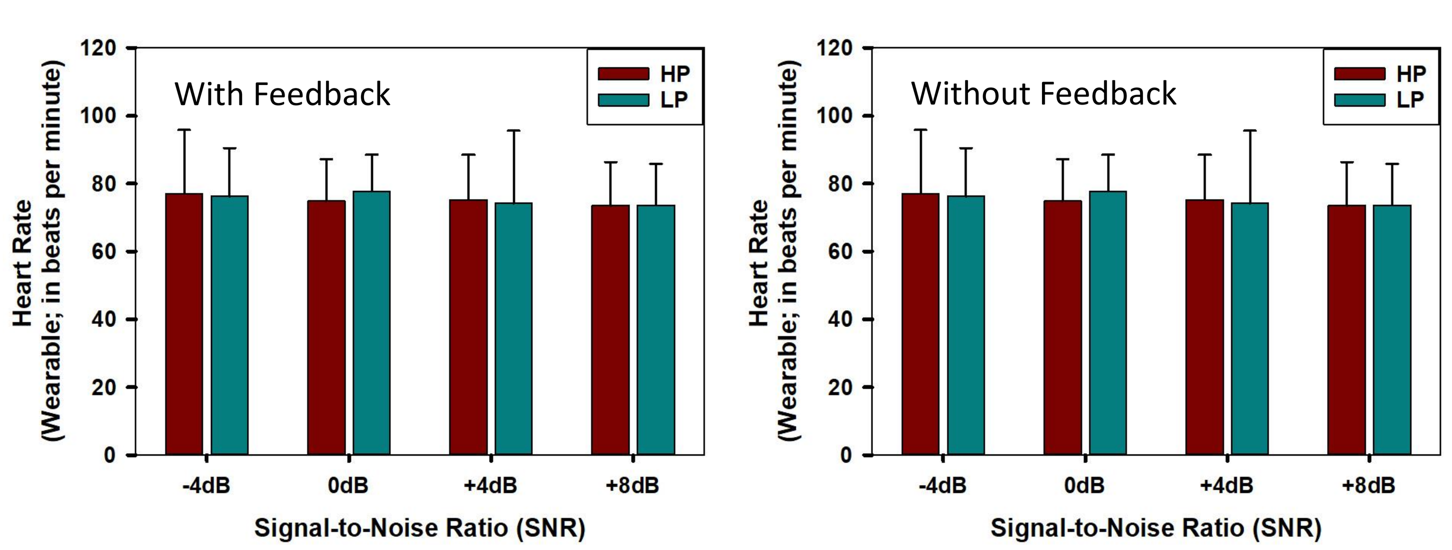
Respiration Rate



- There was a significant main effect of SNR when respiration was measured with a wearable sensor ($p < .001$; Higher BPM at -4 compared to all other SNRs). Moderate effect sizes were noted ($.4 \leq d \leq .6$).
- Laboratory-grade equipment reflected a difference across predictability conditions ($p = .01$, $d = .4$).

Heart Rate

- When measured with a wearable sensor, there was a statistically significant main effect of sentence predictability on heart rate ($p = .04$). However, this effect size was negligible ($d = .08$).
- Laboratory-grade equipment did not show any notable differences in heart rate across conditions.



Q&A

1. Were the wearable sensors sensitive to changes in listening difficulty?

Behavioral and self-report measures validated that the experimental conditions differed in difficulty. Participants reported expected changes in arousal, valence, and effort across conditions. However, the only psychophysiological measure that reflected expected changes in arousal was respiration. It is worth noting that participants were required to vocalize during this experiment. Factors related to speech breathing might have impacted these results.

2. Were the wearable sensors sensitive to stressors related to negative consequences of incorrect understanding?

Although participants reported increased arousal in the presence of negative feedback, this was not reflected by any of the psychophysiological measures.

3. Did the wearable sensors provide information that is consistent with laboratory-grade measures of psychophysiological states?

Not really. Data points for this experimental set-up were averaged across 25 items over about a 2-minute period. Although this might reflect episodic changes in real-world listening conditions, this type of gestalt measurement is not typical of laboratory psychophysiological experiments. Posthoc comparisons across run orders suggested that expected changes in laboratory heart rate measures might have been impacted by participants' habituation to each condition. Conversely, wearable sensors are intended to reflect physiological states across a wider time window in daily life. It is possible that a 2-minute average might not have allowed for sufficient data to reflect meaningful differences for most of these devices.

DISCUSSION

These data show that methodological choices in both measures and design appear to affect outcomes. Because respiration rate did vary with some conditions, this line of research does show promise. Future studies should focus on crafting designs appropriate to wearable devices.

REFERENCES

[1]Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K.J. (2016). Self-reported listening-related effort and fatigue in hearing impaired adults. Ear & Hearing, 38(1), e39-e48.
[2]Gunilla Wänström, Marie Öberg, Emelie Rydberg, Thomas Lunner, Ariane Laplante-Lévesque & Gerhard Andersson (2014) The psychological process from avoidance to acceptance in adults with acquired hearing impairment, Hearing, Balance and Communication, 12:1, 27-35.
[3]McGarrigle, R., Munro, K.J., Dawes, P., Stewart, A.J., Moore D.R., Barry, J.G., & Amitay, S. (2014). Listening effort and fatigue: what exactly are we measuring? A British society of audiology cognition in hearing special interest group 'white paper'. *International Journal of Audiology, 53*(7), 433-445.

Acknowledgements: Rachel Huber, AuD helped with data collection and data analysis. Emma Brothers, B.A. helped with data collection. The University of Memphis: Dean's Award Travel Award.

A pdf of this poster can be found at www.harlmemphis.org