**Plain, B, Pielage, H, Richter, M, Bhuiyan, TA, Lunner, T, Kramer, SE and Zekveld, AA**

# Social observation increases the cardiovascular response of hearing-impaired listeners during a speech reception task

**http://researchonline.ljmu.ac.uk/id/eprint/15403/**

**Article**

**Citation** (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Plain, B, Pielage, H, Richter, M, Bhuiyan, TA, Lunner, T, Kramer, SE and Zekveld, AA (2021) Social observation increases the cardiovascular response of hearing-impaired listeners during a speech reception task. Hearing Research. 410. ISSN 0378-5955**

Hearing Research 410 (2021) 108334

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

# Social observation increases the cardiovascular response of hearing-impaired listeners during a speech reception task

Bethany Plain[a,b,*], Hidde Pielage[a,b], Michael Richter[c], Tanveer A. Bhuiyan[d], Thomas Lunner[b], Sophia E. Kramer[a], Adriana A. Zekveld[a]

[a] Amsterdam UMC, Vrije Universiteit Amsterdam, Otolaryngology Head and Neck Surgery, Ear & Hearing, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands
[b] Eriksholm Research Centre, Snekkersten, Denmark
[c] School of Psychology, Liverpool John Moores University, Liverpool, United Kingdom
[d] Demant A/S, Kongebakken, Smørum, Denmark

ARTICLE INFO

*Article history:*
Received 18 February 2021
Revised 21 July 2021
Accepted 4 August 2021
Available online 12 August 2021

*Keywords:*
Cardiovascular
Listening effort
Autonomic nervous system
Speech reception
Social observation
Intelligibility

ABSTRACT

Certain cardiovascular measures allow for distinction between sympathetic and parasympathetic nervous system activity. Applied during listening, these measures may provide a novel and complementary insight into listening effort. To date, few studies have implemented cardiovascular measures of listening effort and seldom have these included hearing-impaired participants. These studies have generally measured changes in cardiovascular parameters while manipulating environmental factors, such as listening difficulty. Yet, listening effort is also known to be moderated by individual factors, including the importance of performing successfully. In this study, we aimed to manipulate success importance by adding observers to the traditional laboratory set-up. Twenty-nine hearing-impaired participants performed a speech reception task both alone and in the presence of two observers. Auditory stimuli consisted of Danish Hearing in Noise Test (HINT) sentences masked by four-talker babble. Sentences were delivered at two individually adapted signal-to-noise ratios, corresponding to 50 and 80% of sentences correct. We measured change scores, relative to baseline, of pre-ejection period, two indices of heart rate variability, heart rate and blood pressure (systolic, diastolic, and mean arterial pressure). After each condition, participants rated their effort investment, stress, tendency to give up and preference to change the situation to improve audibility. A multivariate analysis revealed that cardiovascular reactivity increased in the presence of the observers, compared to when the task was performed alone. More specifically, systolic, diastolic, and mean arterial blood pressure increased while observed. Interestingly, participants' subjective ratings were sensitive only to intelligibility level, not the observation state. This study was the first to report results from a range of different cardiovascular variables measured from hearing-impaired participants during a speech reception task. Due to the timing of the observers' presence, we were not able to conclusively attribute these physiological changes to being task related. Therefore, instead of representing listening effort, we suggest that the increased cardiovascular response detected during observation reveals increased physiological stress associated with potential evaluation.

## Introduction

Listening effort, or the "mental effort that occurs when a task involves listening" (Pichora-Fuller et al., 2016, p. 11S), has gained momentum in recent years as an important topic within hearing research. Driving this momentum is the knowledge that individuals with hearing impairment, even with adequately fitted amplification, expend more effort during day-to-day listening than their normal-hearing peers (Alhanbali et al., 2017; Ohlenforst, Zekveld, Jansma, et al., 2017). This disproportionate effort investment stems from the cognitive processing required to decipher a degraded auditory signal (McCoy et al., 2005; Rönnberg et al., 2008). In the long-term, the prolonged, effortful nature of listening with hearing loss is suspected to have negative health consequences, such as fatigue and stress (Bess & Hornsby, 2014; Hasson et al., 2011;

---

* Corresponding author at: Eriksholm Research Centre, Rørtangvej 20, 3070 Snekkersten, Denmark.
*E-mail address:* bepl@eriksholm.com (B. Plain).

https://doi.org/10.1016/j.heares.2021.108334

Nachtegaal et al., 2009). Excessive listening effort may also cause hearing-impaired individuals to withdraw from social situations (Holman et al., 2019) and thus contribute to their higher risk of social isolation and loneliness (Shukla et al., 2020). Despite its clear importance, listening effort is not currently accounted for during clinical audiology appointments. The topic warrants further investigation, to deepen our understanding of the cognitive and physiological processes underlying different aspects of listening effort and mediate the negative consequences for those with hearing impairment (McGarrigle et al., 2014).

To study listening effort in the laboratory setting, a body of work has focussed on manipulating the acoustic properties of listening tasks. For instance, researchers have varied the signal-to-noise ratio, the talker pace, level of reverberation and the type of masking noise presented during speech reception tasks (Mackersie & Calderon-Moultrie, 2016; Ohlenforst et al., 2018; Pichora-Fuller et al., 2016; Picou et al., 2019). Beyond the quality of the acoustic signal, the effort invested in a task is also moderated by how important it is for the listener to successfully complete the task (Brehm & Self, 1989; Pichora-Fuller et al., 2016; Richter, 2016). Attempts to manipulate this “success importance” experimentally have often involved performance-dependent monetary reward, as an incentive to be successful at completing the task. For instance, Richter (2016) demonstrated that during a demanding tone discrimination task, a high reward level was required to motivate participants to invest effort (shown by a shortened pre-ejection period (PEP), see details in section 1.2.1), whereas the reward had no significant effect at the easy condition. Another study presenting a speech-in-noise task showed that participants invested more effort, as shown by an increased peak pupil dilation, when a higher reward was offered (Koelewijn et al., 2018).

### *Social observation during listening*

In real life communication situations, rather than from monetary reward, the importance of successful listening arises more commonly from social relationships, expectations, and constructs. Depending on the content of the message being conveyed, or the relationship to the speaker, for example, one might be motivated to listen, even in very demanding auditory conditions. To this end, some studies have recently included manipulations of social factors during listening. For example, Zekveld’s participants received both verbal and visual feedback that suggested they were performing a speech perception task below the expected level (Zekveld et al., 2019). They were encouraged by the experimenter to “please try harder”. This explicit evaluation resulted in an increase in peak pupil dilation at two different intelligibility levels for both normal-hearing and hearing-impaired participants, compared to when no feedback was given. In another study, normal-hearing participants performed a speech-in-noise task concurrently with another participant. Simply the presence of another participant in the same room led to an increase in peak pupil dilation, compared to when the participant performed the task alone (Pielage et al., 2021).

Manipulations of social factors have also been applied during a listening effort study measuring cardiovascular responses, namely heart rate variability (see section 1.2.1). Mackersie and Kearney (2017) recruited hearing-impaired participants who listened to a narrative and either had to recall parts of the narrative or answer comprehension questions based upon it. This task was performed under two evaluative conditions: high or low. During the high evaluation condition, participants were recorded using a video camera, and were told that a panel of experts would review the footage and evaluate their performance. During the low evaluation condition, no video camera was present. Compared to baseline, all task conditions elicited a reduction in heart rate variability (HRV) relative to baseline, but surprisingly, no difference was found between the different task demands (recall or comprehension) or the evaluation conditions. The lack of an effect of evaluation is surprising, because the mere presence of observers during non-auditory tasks has previously been effectively reflected in a decrease in HRV, an increase in heart rate, a shortened PEP (Bosch et al., 2009), and an increase in blood pressure (Gendolla & Richter, 2006). Building upon the literature presented here, a primary aim of the present work was to manipulate success importance in a more ecologically valid way by adding two physically present observers to the traditional laboratory speech reception task.

### *Quantifying listening effort*

Measures of listening effort can be split into three general categories: behavioural, self-report and physiological measures (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). Behavioural measures include single and multi-tasking paradigms, where recall ability, reaction time or performance accuracy on a task are thought to reflect the amount of effort investment (Hällgren et al., 2001; McGarrigle et al., 2014; Strand et al., 2021; Wu et al., 2016). Such measures are beyond the scope of this paper and will not be discussed further here. Self-reported listening effort involves the participant reporting their perception of their listening effort investment. This typically is in the format of a closed set questionnaire or rating scale (Alhanbali et al., 2017). The clear benefit of the subjective rating approach is that it reveals the participant’s conscious awareness of their listening experience (Francis & Love, 2020), which is likely to closely relate to the difficulties reported to audiologists in clinic. The disadvantage of such scales is susceptibility to subjectivity, as people have different perceptions of effort and different internal ‘effort’ scales (Moore & Picou, 2018). Physiological measures will be discussed below.

Listening effort has been quantified using a range of physiological correlates of autonomic nervous system (ANS) activity, including pupil diameter, skin conductance and various heart-related parameters (McGarrigle et al., 2014; Pichora-Fuller et al., 2016; Zekveld & Kramer, 2014). The basis for measuring from such seemingly disparate organs is that mental effort—similar to other cognitive states and processes, including stress and emotional regulation (Levenson, 2014; Ziegler, 2012)—is accompanied by ANS activation, which can be detected by measuring changes in various organs or systems (Kahneman, 1973). By applying such measures during listening tasks, researchers have attempted to measure effort investment and thus better understand the cognitive processes occurring during listening.

It is of growing interest to disentangle the contributions arising from the two branches of the ANS, the sympathetic (SNS) and parasympathetic nervous systems (PNS), during effort investment. The SNS is responsible for “fight and flight”, preparing the body for action, whereas the PNS is responsible for “rest and digest”, allowing the body to restore and repair (Lovallo, 2005; McCorry, 2007). These two branches of the ANS interact in a complex, and not always inversely related (Berntson et al., 1991), balance to maintain homeostasis (McCorry, 2007). The response of the body to mental effort investment appears to be similar to that which occurs during physical effort (McArdle et al., 2010): Greater effort investment is reflected in an increase in SNS activity (Wright, 1996) and/or a withdrawal of PNS activity (Mackersie & Calderon-Moultrie, 2016).

By measuring changes in the two branches of the nervous system we can learn more about the underlying cognitive processes and emotional stress associated with hearing difficulties and hearing loss. Furthermore, a deeper knowledge of which ANS branch responds to effort-related manipulations will allow researchers to select and apply the most sensitive measures in future listening effort studies. Finally, in addition to providing information about transient processes, changes in SNS and PNS activity also provide an important association to longer term, chronic conditions such as chronic stress associated with noise (Francis et al., 2016).

One commonly applied measure of listening effort is the baseline-corrected peak pupil dilation, which has been shown to increase during greater effort investment (Zekveld et al., 2018). The pupil size is controlled by a complex interaction of SNS and PNS ANS activity (Kahneman, 1973; Loewenfeld & Lowenstein, 1999). The relative influences of the SNS and PNS on the pupil size depend on a range of factors, including environmental factors, such as illumination, and individual factors, such as arousal and fatigue (Hopstaken et al., 2015; McGarrigle et al., 2017; Steinhauer et al., 2004; Wang et al., 2018; Zekveld et al., 2018). For this reason, it can be difficult to elucidate whether changes to pupil diameter are a result of variations in SNS or PNS activity. Additional measures that provide insight into the individual branches of the ANS are needed.

### *Cardiovascular measures of listening effort*

Of the different physiological parameters that have been associated with effort investment, cardiovascular measurements remain relatively underexplored in the context of listening effort. This is surprising because certain cardiovascular measures allow for distinction between the SNS and PNS influences on the heart (Berntson et al., 1997; Giuliano et al., 2017; Sherwood et al., 1990). The few cardiovascular studies of listening effort to date have primarily implemented one of two measures: firstly, HRV, which, depending on the metric used, is an index of cardiac PNS activity (Cvijanović et al., 2017; Dorman et al., 2012; Mackersie et al., 2015; Mackersie & Calderon-Moultrie, 2016; Seeman & Sims, 2015), and secondly, PEP, which is an index of cardiac SNS activity (Plain et al., 2020; Richter, 2016). Below, we will introduce HRV and PEP and summarize the results of the listening effort studies that have applied these measures. Subsequently, other relevant cardiovascular parameters will be introduced.

### *Heart rate variability*

HRV refers to the natural fluctuation or changeability in the intervals between heart beats (Shaffer & Ginsberg, 2017). A key contributor to HRV is the respiratory sinus arrythmia. Respiratory sinus arrhythmia is the phenomenon whereby breathing affects the pace of the heart. During inspiration, the inter-beat interval becomes smaller (heart rate increases), whereas during expiration, the inter-beat interval becomes longer (heart rate slows). This fluctuation in heart beats is mediated by various mechanisms, including respiration-modulated cardiac vagal efferent activity. The vagus nerve provides parasympathetic (inhibitory) innervation to the sino-atrial node, which is the heart's internal pacemaker. Respiratory activity inhibits vagal nerve firing during inspiration, but not during expiration (Shaffer & Venner, 2013). HRV can be assessed using several different methods, including time domain or frequency domain analysis, amongst others. Two of these methods, the root mean square of successive differences (RMSSD; time domain) and high-frequency heart rate variability (HF-HRV; frequency domain), are popular measures of HRV as they are thought to reflect predominantly PNS, respiration-modulated activity, as detailed above (Malik et al., 1996; Shaffer & Ginsberg, 2017). A decrease in RMSSD and HF-HRV suggest a withdrawal of PNS activity, which is associated with effort investment (Byrd et al., 2015; Melis & van Boxtel, 2007).

HRV has been applied during several listening effort studies (Cvijanović et al., 2017; Mackersie et al., 2015; Mackersie & Calderon-Moultrie, 2016; Seeman & Sims, 2015). Three studies including only normal-hearing participants will be discussed first. For example, Seeman and Sims (2015) measured the standard deviation of normal-to-normal heart beat intervals (a time domain HRV method) during two different listening tasks: a speech-in-noise task and a diotic-dichotic task. The diotic-dichotic task involved recall of digits presented in three different configurations: (1) a single digit presented diotically; (2) a different, single digit presented to each ear; and (3) two different digits presented to each ear. Task demand was assumed to increase with increasing number of digits recalled. The authors demonstrated a reduction in HRV with increasing difficulty of both tasks. However, this index of HRV reflects contributions from both the PNS and SNS, not purely the PNS, hindering the interpretation about contributions from the two branches of the ANS. Similarly, the HRV indices applied in a study by Cvijanović et al., (2017), the low frequency power and the ratio of low/high frequency power, reflect contributions arising from both the PNS and SNS. Cvijanović's participants performed collaborative communication tasks with varying background noise levels. No significant differences in HRV were elicited across three demand levels (no background noise, at 6dB signal-to-noise ratio (SNR) and -6dB SNR). The final study involving only normal-hearing participants applied a speech-in-noise task at different talker rates (Mackersie & Calderon-Moultrie, 2016). Mackersie and Calderon-Moultrie (2016) demonstrated a decrease in HF-HRV at a faster talker rate, compared to a slower talker rate. These results revealed a reduction or withdrawal of PNS activity during the more challenging listening condition compared to at the easier condition.

One listening effort study measured HRV of both normal-hearing and hearing-impaired participants (Mackersie et al., 2015). Participants performed a speech-in-noise task, with adjusted SNRs for the two groups, to allow comparison between them. The hearing-impaired participants replicated the pattern reported above: the more challenging conditions (i.e. lower SNR) resulted in a significant decrease in HRV (HF-HRV), suggesting a withdrawal of PNS activity. Interestingly, the normal-hearing participants were not sensitive to this effect, showing no significant changes in HRV (HF-HRV), despite performance being adaptively matched in the normal-hearing and hearing-impaired groups. It is possible that even though performance level was matched, the different acoustical conditions presented to the two groups may have contributed to the differences demonstrated. This study suggests that the HRV response to reduced intelligibility level may differ between normal-hearing and hearing-impaired populations. Further work is needed, especially including hearing-impaired individuals, to clarify and replicate these findings.

### *Pre-ejection period*

The second cardiovascular measure that has been implemented to reveal listening effort is PEP. PEP consists of the time period between the onset of electrical depolarization of the heart's left ventricle and the opening of the aortic valve (Newlin & Levenson, 1979; Sherwood et al., 1986, 1990). This time period can be derived non-invasively, using the electrocardiogram (ECG) and impedance cardiogram (ICG). More specifically, PEP refers to the interval between Q-onset of the ECG and the B-point of the ICG. This time-interval is of interest as it provides an index of beta-adrenergic SNS activity on the heart. The effects of increased beta-adrenergic activity on the heart include increased heart rate, electrical conduction, and force of contraction. The force of contraction, in particular, is under mainly SNS control, as demonstrated by pharmacological studies (Ahmed et al., 1972; Harris et al., 1967). PEP is inversely related to cardiac contractility, i.e. increased beta-adrenergic activity causes the heart to beat with stronger force, which results in a shorter PEP (Newlin & Levenson, 1979). In this way, PEP duration is inversely related to effort mobilization: PEP becomes shorter as effort investment increases, indexing an increase in SNS activity on the heart (Richter, 2016; Richter et al., 2008; Wright, 1996).

PEP has been applied in two published listening effort studies, both including normal-hearing individuals. The first applied a tone discrimination task of two difficulty levels and with two reward levels (Richter, 2016). The main aim of the study was to demonstrate the moderating role of success importance (i.e. the reward manipulation) on listening effort investment. Participants were presented two pure tones that were either the same (difficult to discriminate), differed by 3Hz in frequency (difficult to discriminate) or 20Hz in frequency (easy to discriminate). The participants were required to specify whether the tones presented were identical in frequency or not. They were informed that by achieving 90% of trials correct in a block they could earn either a low reward (0.2 CHF) or a high reward (2.0 CHF). Effort was indexed by baseline corrected PEP scores, PEP reactivity. Richter (2016) revealed that at the difficult condition, participants required a high reward to motivate them to invest effort, whereas the low reward was not sufficiently motivating. The second study implemented PEP during a speech-in-noise task at six different SNRs and two reward levels (Plain et al., 2020). The authors found a linear, albeit weak, relationship between SNR and PEP reactivity, such that lower SNRs were associated with a more negative PEP reactivity, and no measurable impact of reward on PEP was demonstrated. To date, no studies have measured PEP in hearing-impaired participants, and no studies have reported the relationship between HRV and PEP during listening tasks.

### Heart rate and blood pressure

Although HRV and PEP provide “pure” metrics of cardiac PNS and SNS activity, respectively, other measures may also offer useful information about effort investment during listening. Heart rate and blood pressure rely on contributions from both branches of the ANS. Heart beats originate at the pacemaker cells in the sinoatrial node in the heart. The pace of firing of these cells is under both PNS and SNS influence and ultimately determines the heart rate. Increased SNS input to the heart increases heart rate and the contractile force of the heart, which in turn increases cardiac output, meaning that more blood is ejected from the heart (Gordan et al., 2015). Consequently, blood pressure increases (Richter et al., 2016). In contrast, PNS input to the heart’s pacemaker cells results in the opposite cascade: a decrease in heart rate and subsequently, a decrease in cardiac output and blood pressure.

Despite the more mixed and complex mechanisms underlying heart rate and blood pressure measures, there is evidence that they may still provide some useful information about listening. For example, a recent study of real-world data from hearing instrument data logging showed an association between heart rate and both the sound pressure level and signal to noise ratio (Christensen et al., 2020). In addition, a functional near-infrared spectroscopy study of normal-hearing individuals demonstrated that heart rate changes were mediated by different sound pressure levels, ranging from near-threshold to comfortably loud (Shoushtarian et al., 2019). At near-threshold sound pressure levels, a decrease in heart rate was demonstrated, whereas louder stimuli resulted in an increase in heart rate.

There is also some evidence from laboratory studies to suggest that heart rate might provide information about listening effort. For example, Richter (2016), also recorded heart rate and blood pressure in their study of PEP changes during listening (study design described in section 1.2.1 above). They found that heart rate followed the hypothesized pattern of results at a statistically significant level, confirming that heart rate was highest in the high demand, high reward condition. However, when applied to blood pressure, the planned contrast did not display this same relationship to a significant degree (Richter, 2016). Beyond the planned contrast, no additional statistical tests were conducted to determine if any alternative patterns were present. In another study, normal-hearing participants performed a dichotic digit test whereby they repeated digits presented to them at three separate demand levels: easy-demand, with single digits presented only to one ear, medium-demand, with single digits presented bilaterally and high-demand, with double digits presented bilaterally (Mackersie & Cones, 2011). This study found no statistically significant difference in heart rate across the three demand levels (Mackersie & Cones, 2011). Other studies reporting pulse rate (a measure of heart rate determined from photoplethysmography) and heart period (the inverse of heart rate) have also demonstrated no significant differences based upon various manipulations of listening demand (Francis et al., 2016, 2021).

Outside the listening effort literature, studies from the motivational psychophysiological domain have included both heart rate and blood pressure as dependent variables. Perhaps the most commonly presented has been systolic blood pressure (SBP), which refers to the maximal pressure in the vascular system during a heart cycle. SBP has been demonstrated to follow the hypothesized effort related predictions (Gendolla et al., 2019). Heart rate, diastolic blood pressure (DBP; the minimal pressure in the vascular system during a cycle) and mean arterial blood pressure (MAP; the average pressure in the vascular system during a cycle) have also been applied and reported as dependent variables corresponding to effort (Gendolla & Richter, 2006; Nolte et al., 2008).

### *Aims and hypotheses*

Relatively few papers have reported results of cardiovascular listening effort parameters in hearing-impaired groups and, of these, generally a single cardiovascular dependent variable is included (Mackersie et al., 2015; Mackersie & Kearney, 2017). Therefore, the picture of how the wider cardiovascular system reacts during listening in this population is currently incomplete. In particular, the impact of observation during listening has not been thoroughly investigated, despite being a situation that has obvious real-world applicability. To this end, this study investigated the effects of social observation on cardiovascular measures from hearing-impaired participants. We achieved this by adding two observers to the traditional laboratory set-up, while participants underwent a speech reception task at two intelligibility levels (corresponding to 50 and 80% correct). We measured several cardiovascular parameters, including two types of HRV (RMSSD and HF-HRV), PEP, heart rate, SBP, DBP and MAP (measures summarized in Table 1).

We based our hypotheses upon motivational intensity theory (Brehm & Self, 1989). It was expected that at the easier intelligibility level, effort investment would be relatively low and of a similar intensity between the alone and observed conditions. It was anticipated that, although not reaching ceiling performance levels, the task demand would be generally within the capabilities of the participant. At the harder intelligibility level, however, we expected that the observers’ presence would increase the participants’ desire to successfully complete the task, thereby increasing effort investment, compared to the alone condition (Brehm & Self, 1989; Richter et al., 2016). We therefore anticipated that the general cardiovascular response, as assessed by a multivariate analysis, would reveal an interaction between intelligibility and social state. We also expected to demonstrate this same pattern for each of the individual measures in subsequent univariate analyses. We deliberately selected a range of measures that would provide information about the contributions from the two branches of the ANS. Within our list of variables, we included PEP, to measure SNS activity (Newlin & Levenson, 1979), and HRV, to measure PNS activity (Malik et al., 1996; Shaffer & Ginsberg, 2017), as well as other variables that reflect contributions from both branches of the ANS. By

**Table 1**
Summary of cardiovascular measures included in study

| Measure | Abbreviation | Influenced by SNS, PNS or both | Expected effort-related change |
|---|---|---|---|
| Root mean square of successive differences | RMSSD | PNS | Decrease |
| High frequency heart rate variability | HF-HRV | PNS | Decrease |
| Pre-ejection period | PEP | SNS | Decrease |
| Heart rate | HR | Both | Increase |
| Systolic blood pressure | SBP | Both | Increase |
| Diastolic blood pressure | DBP | Both | Increase |
| Mean arterial pressure | MAP | Both | Increase |

applying these, we hoped to elucidate whether the cardiovascular response in this study was predominantly fuelled by the PNS or SNS, or a combination of both.

Finally, we anticipated that the subjective ratings would reflect a similar pattern to the cardiovascular measures. We expected participants to rate their effort investment, desire to improve the auditory situation and stress levels as highest in the observed, difficult intelligibility condition. In contrast, the tendency of participants to give up was anticipated to be highest in the alone, difficult intelligibility condition.

## Materials and methods

In this experiment, we aimed to determine the influence of social observation and SNR on hearing-impaired individuals during a speech-in-noise task. To investigate this, a two by two within-subject design was applied, including two social observation states (either alone or in the presence of observers) and two SNRs (corresponding to approximately 50 and 80% intelligibility levels). Both cardiovascular measures and pupil diameter were measured during the experiment. The pupillometry data are being analysed by the second author, who will report these data in a separate publication, in accordance with requirements for his doctoral thesis. These data will therefore not be described or presented here. Instead, this paper focuses exclusively on cardiovascular measures.

### *Participants*

The estimated sample size for the present study was calculated based upon reported univariate effect sizes for HRV and PEP, because multivariate effect sizes for the present experimental design and group of measures are unknown. A listening task of varying speaker pace (fast and slow) was demonstrated to induce changes in HRV of an effect size (partial $\eta^2$) of 0.29 (Mackersie & Calderon-Moultrie, 2016). A similar partial $\eta^2$ (0.28) was also deduced from a previous study showing that PEP was sensitive to changes in social evaluation from between one (partial $\eta^2 = 0.22$) and four individuals (partial $\eta^2 = 0.36$) (Bosch et al., 2009). Our power calculation, completed in G*Power 3.1.9.4 software, was based upon an estimated multivariate effect size of 0.26 (to provide a conservative estimate), an alpha error of 0.05, power of 0.8, and a correlation of 0.5 between repeated measures. The calculated recommended sample size was 26 participants.

To account for possible missing data, ultimately 29 hearing-impaired participants (17 males, 12 females) were tested at Eriksholm Research Centre. They were recruited from the Eriksholm Research Centre test person database. Inclusion criteria specified that participants were to be native Danish speakers, aged between 40 and 80 years of age. The mean age of participants was 64.55 years (SD = 9.10, range = 47 to 76 years). Participants were required to have a sensorineural hearing loss of at least 35dB HL four-frequency pure tone average (PTA) across 500, 1000, 2000 and 4000Hz in their poorer ear, as demonstrated by their most recent audiogram. Hearing thresholds were also required to be symmetrical, which was classified as the presence of less than 15dB difference between the left and right ears at 500, 1000 and 2000Hz, or less than 30dB at 3000, 4000 and 6000Hz. Mean audiometric threshold and standard deviations are plotted in Fig. 1. Group mean four-frequency PTAs were 50.17dB HL (SD = 8.87) for the right ear and 51.34 dB HL (SD = 8.68) for the left ear. All participants were bilateral hearing aid users with experience using Oticon hearing instruments.

In addition to hearing-related requirements, participants had to meet certain medical criteria in order to participate. These criteria were assessed by self-reported medical history. At the time of testing, participants reported no diagnosed psychiatric, neurological, ocular, or cardiovascular disease. A history of surgery on either the eyes or the cardiovascular system, in particular the insertion of a pacemaker, were strict exclusion criteria. As hypertension is relatively common in the older population, high blood pressure and the use of anti-hypertensive medication were not exclusion criteria. The medication lists of participants were not specifically requested or recorded by the research clinician or experimenters, but if voluntarily disclosed, this information was recorded in the study documentation. In total, five participants reported that they were taking antihypertensive medication and one participant disclosed that they had type II diabetes. Ultimately, one participant taking antihypertensive medication and the diabetic participant were removed from analysis due to data quality issues (see section 3). Ethical approval for the study was obtained from the Research Ethics Committees of the Capital Region of Denmark.

### *Hearing instruments*

For the duration of the experiment, each participant was issued with Oticon Opn hearing instruments programmed to their hearing thresholds using the manufacturer's first fit. Double-layered domes were applied for all participants, apart from one who reported uncomfortable occlusion so instead wore standard domes. One other participant with the same complaint retained the double domes but had the overall gain of the hearing instruments decreased slightly (2-3 dB maximum), to improve their comfort. To ensure uniformity across participants, the hearing instrument settings were as follows: experience level was set to “long-term”, noise reduction was deactivated, microphones were set to omnidirectional and all buttons (i.e. program and volume controls) were deactivated.

### *Procedure and apparatus*

Testing was conducted in a sound treated room. Upon arrival, participants were provided with an overview of the session structure, were checked for contraindications to testing and, if able to continue, gave their written informed consent. After this, otoscopy was performed, their height and weight were measured (to allow calculation of body mass index (BMI)), and their pre-programmed

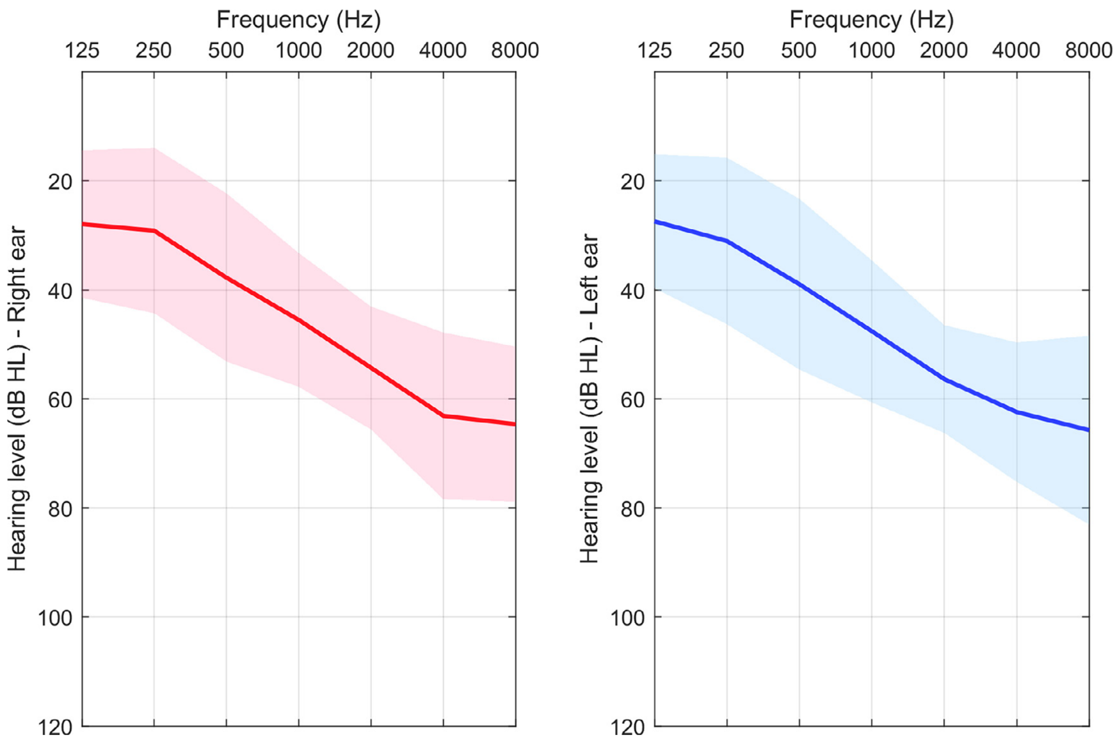

**Fig. 1.** Average audiogram of the right and left ears of participants. The standard deviation is displayed by the shaded region. dB HL = decibel hearing loss, Hz = hertz.

standard hearing aids (see section 2.2) were inserted. Next, participants were required to complete a self-efficacy questionnaire. The details and results of this questionnaire will not be presented here as they will be published in a separate paper.

Participants were then taken into the testing booth by the research clinician, where the electrodes and blood pressure cuff were applied, and the pupillometer calibrated. After instruction, a training list of sentences was conducted using an adaptive procedure aiming at 50% correct (see section 2.4.1 for details). Next, two adaptive procedures were completed, one adjusting towards 50% and the other towards 80% correct. The SNRs obtained by these adaptive procedures, which will be referred to here as the difficult and easy conditions respectively, were recorded, so as to be applied during the subsequent test blocks.

The test blocks consisted of speech-in-noise tasks under two social observation states (either alone or in the presence of observers) and at two SNRs (difficult and easy). Each of the four test blocks were preceded by a four-minute baseline video (details in section 2.4.4) and followed by the subjective ratings (details in section 2.5). During the observed half of the test blocks, the observers entered the test room prior to the video baseline, in order to introduce themselves to the participant. They were absent for the baseline itself, sitting out of the view of the participant in a different room, and returned to the test room at the beginning of the list of sentences. After they had observed the task blocks, the observers wrote some notes on paper about their observations, were thanked and dismissed. These notes were destroyed after the test session, due to ethical requirements.

After the participant had completed all four test blocks, the electrodes and blood pressure cuff were removed. The participant's own hearing aids were returned to them. They were then seated in the outer part of the laboratory in order to give the semi-structured interview. The list of interview questions and results of qualitative data analysis will be reported in a separate paper. At the end of the session, the research clinician discussed travel cost reimbursement with the participant and thanked them for their participation. The whole test session, including a five to ten-minute break after two HINT test blocks, lasted on average around 2.25 hours.

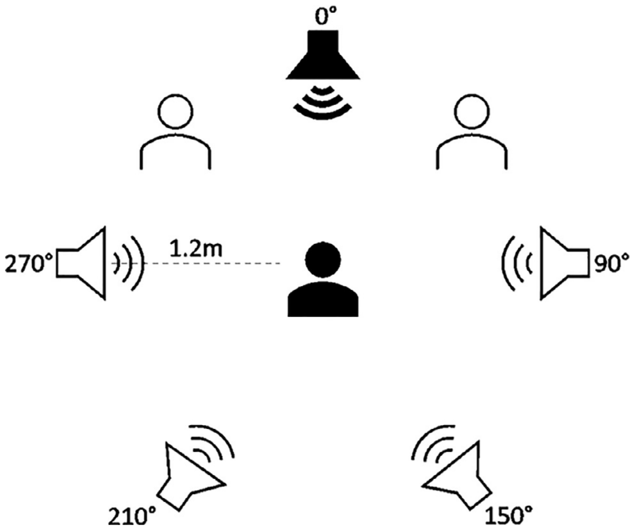

**Fig. 2.** Schematic of the experimental test set up, demonstrating the position of the participant (centre), 1.2 meters from the target loudspeaker (0°), the four masking loudspeakers (90°, 150°, 210°, 270°) and the observers.

*Speech-in-noise task*

Auditory stimuli consisted of the Danish version of the Hearing-in-Noise Test (HINT) (Nielsen and Dau, 2011). Target sentences were presented from a loudspeaker placed at 0° azimuth, and four-talker babble masking noise was provided by four additional loudspeakers, each positioned 1.2 meters from the participant at 90, 150, 210 and 270° azimuth. A schematic of the experimental set up is shown in Fig. 2. The masking noise consisted of audio files of individuals reading the newspaper. The clips were manipulated to have the same long-term average frequency spectrum as the target speaker. Each of the surrounding loudspeakers played a single masker voice, of which there were two females and two males. The positions of the masker voices were randomized across conditions.

Twenty sentences were presented in each condition of the training list, adaptive procedures, and fixed test blocks. The four-talker babble masker preceded and followed each target sentence by three seconds. This time duration was selected as it is thought to be the most optimal for finding the peak pupil dilation (Winn et al., 2018). After the masker offset, the participant was encouraged to repeat back what they heard, during an unfixed-duration response window. Scoring was completed live during the test session. Certain errors during sentence repetition were permitted. These included errors in verb tenses, singular vs. plural nouns, definite vs. indefinite articles (the / a), omission of single phonemes and the addition of words or phonemes. If the correct words were recalled in the wrong order, this was also tolerated. In addition to live scoring, responses were also audio recorded as a precaution.

### *Adaptive procedures*

The SNRs presented at the difficult and easy task blocks were determined using adaptive procedures for each participant. The order of the adaptive procedures was balanced across participants. During the adaptive procedures, the target level was adjusted, while the masking noise remained fixed at 70dB SPL. Scoring was sentence based; a correct response required all words of the sentence to be correctly repeated. To obtain the difficult SNR, the following procedure was used: for the first four sentences, the SNR was increased by 4dB for every incorrect answer and decreased by 4dB for every correct answer. The SNR of the fifth sentence was determined based upon whether the fourth sentence was correctly or incorrectly repeated and took into account the SNRs of the preceding sentences. If the fourth sentence was correctly recalled, the fifth sentence was presented at the average of the first four sentences' SNRs and the SNR of the fourth sentence minus 4dB. Whereas, if the fourth sentence was incorrectly recalled, the fifth sentence was presented at the average of the first four sentences' SNRs and the SNR of the fourth sentence plus 4dB. After this, the step size for the remaining 15 sentences was 2dB, i.e. the SNR was increased or decreased by 2dB if a sentence was incorrectly or correctly repeated, respectively. The average SNR of sentences five to twenty was recorded (the difficult SNR). The easy SNR was estimated using the same procedure as above, however for the first four sentences, the SNR was increased by 6.4dB step for incorrect sentences and decreased by 1.6dB for correct answers. For subsequent sentences, the SNR was increased by 3.2dB for every correct answer and decreased by 0.8dB for incorrect answers. The average SNR of sentences five to twenty was recorded as the easy SNR. Cardiovascular data were recorded during the adaptive procedures, but these data were not analysed.

### *Task blocks*

The order of presentation of the difficult and easy task blocks was balanced across participants. All 20 sentences presented during a task block were at the same SNR, and the masker signal was always presented at 70dB SPL, regardless of the condition. Scoring was completed based upon the number of words correctly repeated, but for the purposes of our analysis we considered sentence-based scores. Cardiovascular data from during the task blocks were recorded and analysed.

### *Social observation*

The participant performed the task both alone and in the presence of two other individuals. The observers were also hearing-impaired and were recruited through the Eriksholm test person database, although the participant was not explicitly told this. There were no specific requirements regarding the observers' configuration of hearing loss, hearing aid status, nor the gender or age of the observers. The observer pairs could therefore be two males, two females or one male and one female. We acknowledge that not controlling for the specific pairings may have introduced variability. We deliberately selected individuals who could feasibly be "social peers" of the participant, rather than being watched by young students, for example. Observers and participants did not know each other before the test session. Observers were scheduled to take part in the test sessions of two different participants. Their second session was never with the same "observation partner" as the first session. In total, around 30 observers took part in the experiment.

The observers were seated on chairs at 45° and 315° from the participant, at a distance of around 1.2 meters. They were facing inwards towards the participant and were within the peripheral vision of the participant. The participant and the two observers were instructed to imagine that they were socializing in a restaurant situation. The participant was told to envisage that the target sentence was being spoken by one of the observers. The observers were instructed to act in a non-threatening and pleasant demeanour when interacting with the test participant. They were told that their task was to judge how competent a communication partner the participant would be in real life. The order of the alone and observed conditions were randomized, although due to practical reasons the order was not fully equally balanced: ultimately 17 participants were tested while observed first and 12 were tested alone first.

### *Baseline videos*

In order to allow cardiovascular change scores to be calculated, a four-minute baseline was implemented before each task block. During this baseline, participants watched a video on a computer screen in front of them. They had no task during this time; they were instructed to rest and quietly watch the screen. The videos depicted footage shot from a drone slowly passing over parts of Edinburgh. The clips consisted of some countryside and some city-based footage and were carefully selected to be non-emotive and neutral in content. There were four videos, one to precede each of the four task blocks. The order of video presentation was randomized for each participant.

### *Subjective ratings*

After each task block, the participant was left alone in the booth to reflect upon the preceding task period and complete subjective ratings. They were asked to consider the following: 1) How much effort did it take you to understand the preceding sentences? 2) Imagine this was a real-life situation. How likely would you be to try to do something else to improve the situation (e.g. move to a quiet room, ask the speaker to speak louder)? 3) How likely were you to give up and stop trying? and 4) How stressful did you perceive the task to be? The answer model for these four questions consisted of a printed horizontal line, labelled from 0 to 10, with 1 decimal point precision. The participant was required to simply select a point on the line that corresponded to their answer. The extremes were labelled from "None/not at all" to "A lot" (e.g. of effort) or "Very" (e.g. stressful). The final subjective rating consisted of a seven-point Likert scale in answer to the question "Did you perceive the task to be challenging or threatening?". Participants were required to select one of seven answer options: extremely threatening, very threatening, slightly threatening, neutral, slightly challenging, very challenging and extremely challenging. Results of this scale are excluded from this paper, as participants did not fully understand the answer model (some selected two points on the same scale, for example).

The rationale behind including these five subjective rating scales will be discussed here. The first rating scale (effort) originates from the NASA task load index (TLX) (Hart & Staveland, 1988). This item was included to be consistent with a number of studies reporting subjective ratings of listening effort (Mackersie & Cones, 2011; Pielage et al., 2021; Plain et al., 2020; Seeman & Sims, 2015; Zekveld et al., 2010). However, ambiguity exists regarding whether people truly rate their effort, or instead substitute other aspects of the listening task, such as their performance (Moore & Picou, 2018; Picou et al., 2017). For this reason, and based upon the recommendation of Picou and Ricketts (2018), the second rating scale was included, asking people to consider their desire to change the situation. The third rating (giving up) has also been implemented in a number of studies, with a view to learning about disengagement from the task (Picou & Ricketts, 2018; Pielage et al., 2021; Plain et al., 2020). The stress rating was included on the premise that the social observation may elicit stress from the participant (Hellhammer & Schubert, 2012). The final rating scale (challenge / threat) was inspired by Blascovich and Tomaka's Biopsychosocial model, which uses cardiovascular markers to differentiate between challenge and threat (Blascovich and Tomaka, 1996).

### *Cardiovascular measures and data processing*

Cardiovascular measurements were collected during the speech-in-noise task using the Cardioscreen 2000 system (Medis, Ilmenau, Germany). Electrocardiography (ECG) and impedance cardiography (ICG) were measured at a sampling frequency of 1000Hz. This was achieved by placement of three disposable solid gel electrodes: one applied to the left side of the neck (dual sensor), one in the left mid-axillary line at the level of the xiphoid process level (beneath the left armpit, around half-way down the chest) and finally, one 10cm below this. The ECG and ICG were measured throughout the speech-in-noise task. A blood pressure cuff was applied to the participants' right arm, over the brachial artery and above the elbow. One blood pressure measurement was taken during each baseline video and one during each task block. The blood pressure was taken once in the middle of the baseline and once in the middle of the task block, with a gap between these measurements of approximately five minutes. Each blood pressure cuff inflation lasted around 30 seconds in duration and provided three measures of blood pressure: SBP, DBP and MAP (units of all: millimeters of mercury; mmHg).

### *Heart rate variability: RMSSD and HF-HRV*

In preparation for heart rate variability analysis, R peaks of the ECG signal were detected using a peak detection function in MATLAB (version R2018b). The data were visually inspected to confirm that peaks had been detected correctly. Any sections of data containing excessive artifacts were excluded. The inter-beat intervals were then loaded into Kubios software (Kubios HRV Standard 3.3.1) to be processed (Tarvainen et al., 2014). Kubios' artifact correction was set to low (threshold: 0.3). RMSSD and normalized HF-HRV were extracted from Kubios for each participant. Normalised HF-HRV consists of a ratio of the power in the high frequency band (0.15 – 0.4Hz) in relation to that in the low frequency band (0.05 – 0.15Hz). No manual averaging was required because Kubios produced single values representing the HRV indices across the whole task or baseline period (minus periods of artefacts that were removed). Delta change scores were calculated for both HRV measures by subtracting baseline values from task values. Subsequently, RMSSD change scores were log-transformed to correct for positive skew.

### *Pre-ejection period*

PEP was calculated using the method described by Richter (2016). R peaks were detected automatically in the ECG signal, and successful detection was confirmed visually. Then, the ICG signal was differentiated and filtered (low pass Butterworth filter, order 4, with a cut-off of 50Hz). Any cycles with artefacts were excluded. Ensemble averages consisted of 60 seconds of data. PEP, the period between the R-onset (ECG) and the B- point of the ICG, was scored from the ensemble averages using the method outlined by Sherwood et al. (1990). All data was scored by two separate scorers. Any PEP values with an inter-scorer difference of greater than 10 ms were reviewed and scoring errors were corrected accordingly. Agreement between scorers was high (intraclass correlation coefficient, two -way mixed, absolute agreement: 0.99). The final PEP values consisted of the average of scorer 1 and 2's PEP values.

In order to calculate a PEP change score for each condition, PEP values were averaged over time for each condition. The first ensemble average of the baseline was excluded as it was assumed that during the first minute, the participant may not yet have reached a resting state. Therefore, PEP values from minutes two to four of the baseline were averaged, to provide a single baseline PEP value for each condition. To obtain the single task PEP values for each condition, the first five minutes of data were averaged. On average, the task lasted around six minutes in total, with the shortest duration being just over five minutes. The first five minutes of the data were selected to ensure that all ensemble averages contained a full minute of ECG and ICG data. This was to avoid a situation as described here: for example, if a participant completed the task a few seconds into the sixth minute, the sixth PEP ensemble average would be calculated based upon just one or two cycles of data, which may reflect artifact and need excluding. By considering only the first five full minutes of data we could be more convinced of the reliability of the PEP values generated. A delta change score was calculated for each condition by subtracting the average baseline PEP from the average task PEP.

### *Heart rate and blood pressure*

Heart rate was determined from the inter-beat intervals obtained while processing the heart rate variability data (section 2.6.1). Average heart rate values were calculated for each baseline and task block. For heart rate and blood pressure, baseline values were subtracted from task values, to produce a delta change score. A positive score is associated with a higher heart rate and blood pressure during the task than the baseline.

### *Statistical analysis*

A two-way repeated measure analysis of variance (ANOVA) was used to determine the presence of main effects of social state or intelligibility, or interaction effects on performance. A multivariate analysis of variance (MANOVA) was conducted including all cardiovascular variables (RMSSD, HF-HRV, PEP, heart rate, SBP, DBP and MAP) to demonstrate whether these variables were together sensitive to the experimental manipulations. An additional multivariate analysis was conducted for the subjective rating data. Follow up univariate analyses were undertaken for both the cardiovascular and subjective rating data. Finally, a supplementary correlational analysis was undertaken to determine the relationships between the individual cardiovascular variables.

## Results

Twenty-nine participants took part in the experiment; however, some exclusions were necessary after data collection. Full datasets from three participants were excluded due to consistent

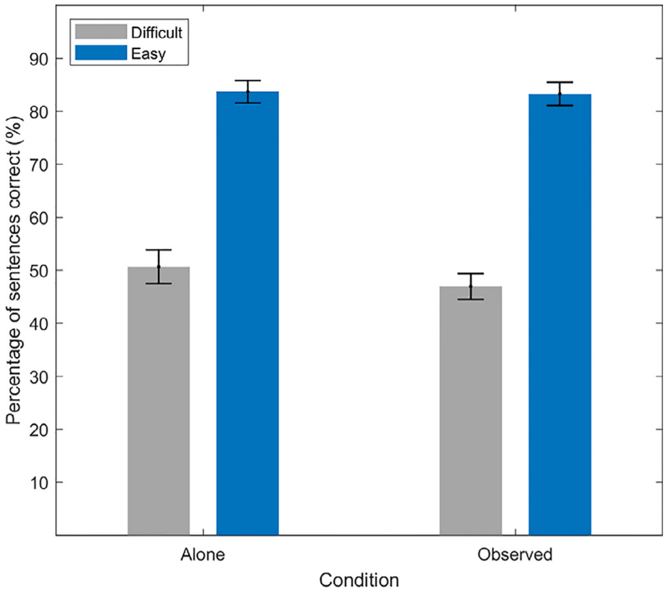

**Fig. 3.** Average percentage of sentences correctly repeated at the difficult and easy conditions while alone and observed. Error bars represent the standard error of the mean.

issues with data quality, rendering scoring of the ICG's B-point for PEP analysis inaccurate. Two participants had missing triggers in one test condition each, making identification of the baseline and task onset impossible; these conditions were therefore excluded. Finally, due to a testing error, one participant heard the same HINT list twice, so the repeated condition was excluded. As such, 23 full datasets were included in the following analyses.

*Performance*

To determine which fixed SNRs to apply during the test conditions, participants individually underwent adaptive procedures targeting 50 and 80% correct. The average SNR at the difficult, 50% condition was 5.18dB SNR (SD = 2.09) and the average SNR at the easy, 80% condition was 9.87dB SNR (SD = 3.00). The sentence-based performance data can be seen in Fig. 3. A two-way repeated measures ANOVA revealed a significant effect of intelligibility on performance (percentage of sentences correctly repeated) ($F_{[1,22]} = 185.54$, $p < 0.01$, $\eta p^2 = 0.89$), demonstrating that participants performed significantly better in the easy condition compared to the difficult condition, as expected. No effect of social state nor interaction between intelligibility and social state were demonstrated on performance (social state: $F_{[1,22]} = 1.00$, $p = 0.33$, $\eta p^2 = 0.04$; interaction: $F_{[1,22]} = 0.41$, $p = 0.53$, $\eta p^2 = 0.02$).

*Cardiovascular data*

*Baseline measures: multivariate analysis*

Group averaged baseline values and standard deviations for each of the cardiovascular variables were as follows: RMSSD, 3.77 (0.70) log milliseconds; HF-HRV, 50.22 (20.37) normalised units; PEP, 102.77 (15.94) milliseconds; heart rate, 63.26 (9.75) beats per minute; SBP, 139.05 (15.78) mmHg; DBP, 82.13 (10.44) mmHg; and MAP, 95.28 (10.98) mmHg. Analysis of baseline data is commonly reported in the psychophysiology literature (Richter et al., 2008; Richter & Gendolla, 2009). For the purposes of this study design, analysis of the baseline data was conducted to determine the presence of any effect of the experimental manipulations on the baseline periods. This was warranted because the upcoming social condition was known to the participant before the baseline period started (the intelligibility condition was not). Therefore, the baseline data for each of the cardiovascular measures was assessed using a multivariate repeated measures ANOVA. No significant overall main effects were demonstrated for social state (Wilks' lambda = 0.81, $F_{[7,16]} = 0.54$, $p = 0.80$, $\eta p^2 = 0.19$) or intelligibility (Wilks' lambda = 0.63, $F_{[7,16]} = 1.33$, $p = 0.30$, $\eta p^2 = 0.37$), and no interaction (Wilks' lambda = 0.73, $F_{[7,16]} = 0.87$, $p = 0.55$, $\eta p^2 = 0.28$) was demonstrated.

*Reactivity scores: analysis*

In some studies, cardiovascular measures are corrected based upon their relationship with BMI. For example, Richter and Gendolla (2009) and (Richter et al., 2008) applied BMI-correction to blood pressure measures. To determine if this was necessary in the present analysis, the relationship between the BMI and cardiovascular reactivity averages (i.e. the average across conditions) for each cardiovascular variable (RMSSD, HF-HRV, PEP, heart rate, DBP, SBP and MAP) was investigated using Pearson's correlations. A similar analysis was conducted using age and the reactivity averages. No significant correlations were found and therefore these two features (BMI and age) were not included or corrected for in the following analysis. Correlations between the cardiovascular measures themselves are reported in section 3.2.3 below.

Reactivity data (change scores from baseline) and standard errors for each of the cardiovascular variables are presented in Table 2 and Fig. 4. A repeated-measures multivariate analysis of variance (MANOVA) was undertaken to analyse the effects of social observation state and intelligibility level on all cardiovascular reactivity data, including the RMSSD, HF-HRV, PEP, heart rate, MAP, DBP and SBP. The multivariate analysis revealed a significant effect of social observation state on the cardiovascular variables (Wilks' lambda = 0.39, $F_{[7,16]} = 3.61$, $p = 0.02$, $\eta p^2 = 0.61$). No significant effect of intelligibility was found (Wilks' lambda = 0.65, $F_{[7,16]} = 1.25$, $p = 0.34$, $\eta p^2 = 0.35$) and no interaction was demonstrated (Wilks' lambda = 0.71, $F_{[7,16]} = 0.96$, $p = 0.49$, $\eta p^2 = 0.30$).

Based upon the significant effect of social observation state, follow-up univariate analyses were conducted to determine which of the cardiovascular measures demonstrated the effect of social observation. To account for multiple comparisons (Bird & Hadzi-Pavlovic, 2014), p values were corrected in MATLAB using the false discovery rate (FDR) Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; Martínez-Cagigal, 2021). The results of these analyses are presented in Table 3. As demonstrated in Fig. 4, SBP, DBP and MAP increased in the presence of the observers. There was no significant effect of the social manipulation on PEP, HF-HRV, RMSSD or heart rate.

*Supplementary correlational analysis*

As part of a supplementary analysis, the relationships between the cardiovascular change scores of the different measures were investigated. The rationale behind this exploratory analysis was that it may provide additional information to aid the interpretation of the ANS origins of the cardiovascular response. Pearson Correlation coefficients are presented in Tables 4. At the bottom of Table 4, average correlation coefficients are presented. These were calculated by first performing Fisher's r-to-z transforms on the individual correlation coefficients, then averaging across condition, before finally reverse transforming. Average correlation coefficients between variables ranged from very weak (for example, between PEP and RMSSD, r = -0.04) to strong relationships (for example, between DBP and MAP, r = 0.82) (Evans, 1996). Variables demonstrating average correlation coefficients greater than 0.5 are presented in scatter plots. Fig. 5 and 6 reveal the strong relationships between the blood pressure measures and PEP.

Three interesting observations from these correlation coefficients will be highlighted here. Firstly, the correlations between

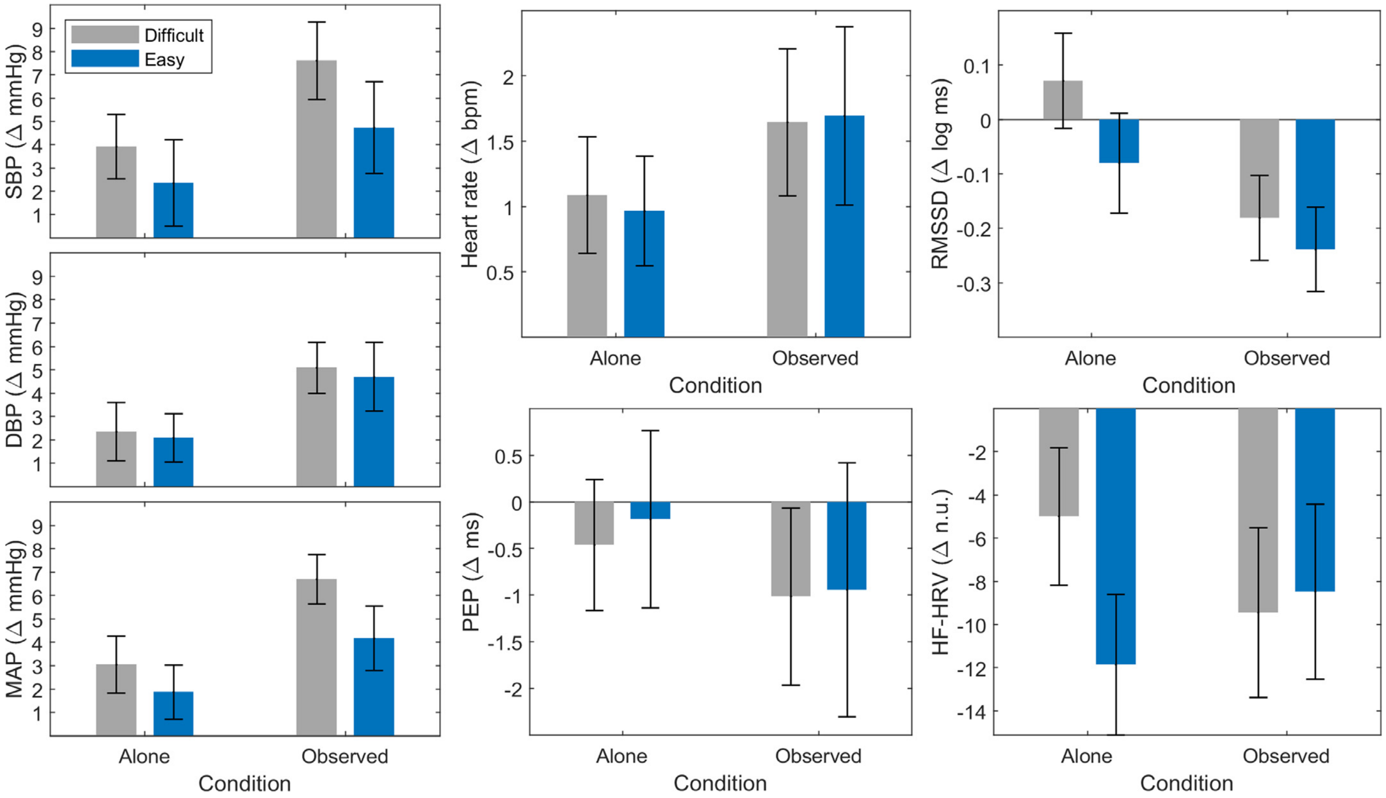

**Fig. 4.** Average change scores of the cardiovascular measures. The leftward panel demonstrates changes in blood pressure (upper, SBP, middle, DBP and lower, MAP). The middle panel demonstrates changes in heart rate (upper) and PEP (lower). The rightward panel demonstrates changes in heart rate variability (upper, RMSSD and lower, HF-HRV). Error bars represent the standard error of the mean. SBP = systolic blood pressure, DBP = diastolic blood pressure, MAP = mean arterial pressure, mmHg = millimeters of mercury, bpm = beats per minute, PEP = pre-ejection period, ms = milliseconds, RMSSD = root mean square of successive differences, HF-HRV = high-frequency heart rate variability, n.u. = normalised units.

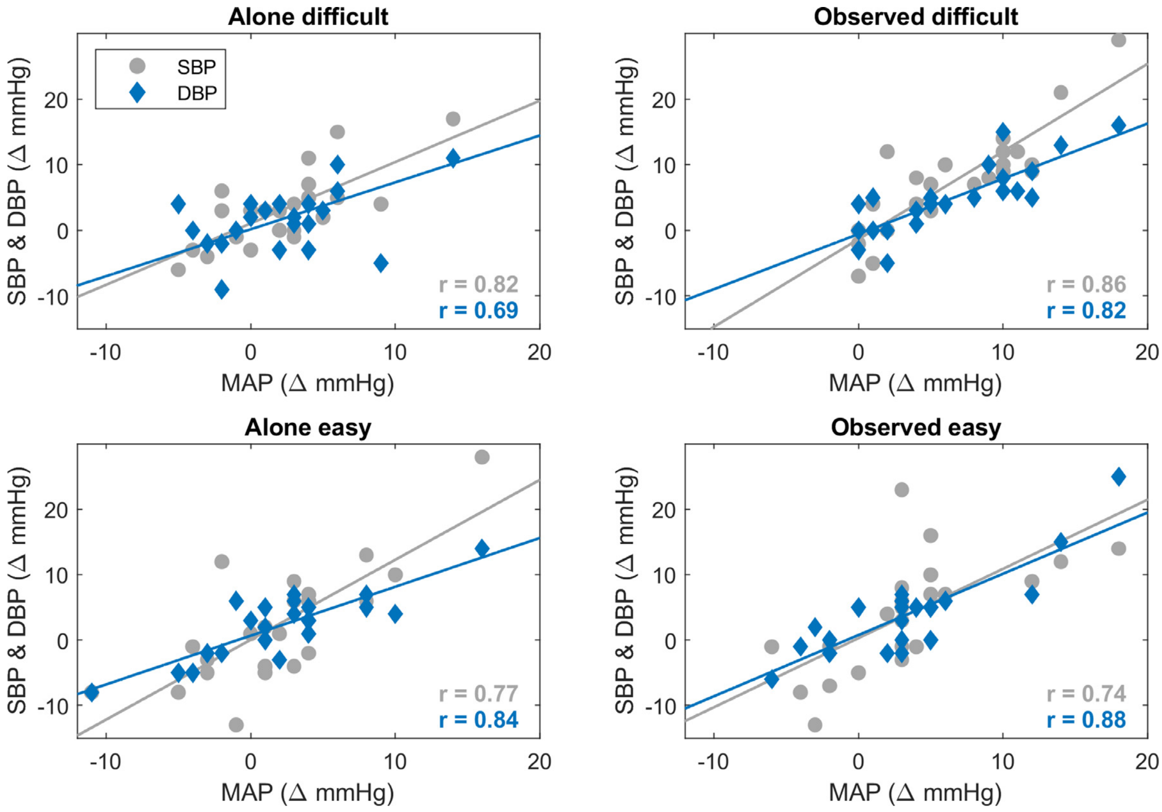

**Fig. 5.** Scatter plot demonstrating the relationship between MAP, SBP and DBP change scores. Lines represent lines of best fit. Pearson correlation coefficients are presented in the bottom righthand corner of each panel (grey text – SBP and blue text – DBP). MAP = mean aterial pressure, SBP = systolic blood pressure, DBP = diastolic blood pressure, mmHg = millimeters of mercury.

**Table 2**
Means (and SEMs) of cardiovascular change scores

| | Social observation condition | | | |
|---|---|---|---|---|
| | Alone | | Observed | |
| **Intelligibility condition** | **Difficult** | **Easy** | **Difficult** | **Easy** |
| RMSSD (log ms) | 0.07 (0.09) | -0.08 (0.09) | -0.18 (0.07) | -0.24 (0.08) |
| HF-HRV (n.u.) | -4.99 (3.18) | -11.86 (3.26) | -9.45 (3.92) | -8.48 (4.06) |
| PEP (ms) | -0.46 (0.70) | -0.19 (0.95) | -1.01 (0.95) | -0.94 (1.36) |
| HR (bpm) | 1.09 (0.45) | 0.97 (0.42) | 1.64 (0.56) | 1.69 (0.68) |
| SBP (mmHg) | 3.91 (1.38) | 2.35 (1.86) | 7.61 (1.66) | 4.74 (1.98) |
| DBP (mmHg) | 2.35 (1.26) | 2.09 (1.03) | 5.09 (1.09) | 4.70 (1.47) |
| MAP (mmHg) | 3.04 (1.21) | 1.87 (1.17) | 6.70 (1.06) | 4.17 (1.38) |

n = 23; change scores calculated by subtracting baseline value from task value. ms = milliseconds, n.u. = normalized units, bpm = beats per minute, mmHg = millimeters of mercury.

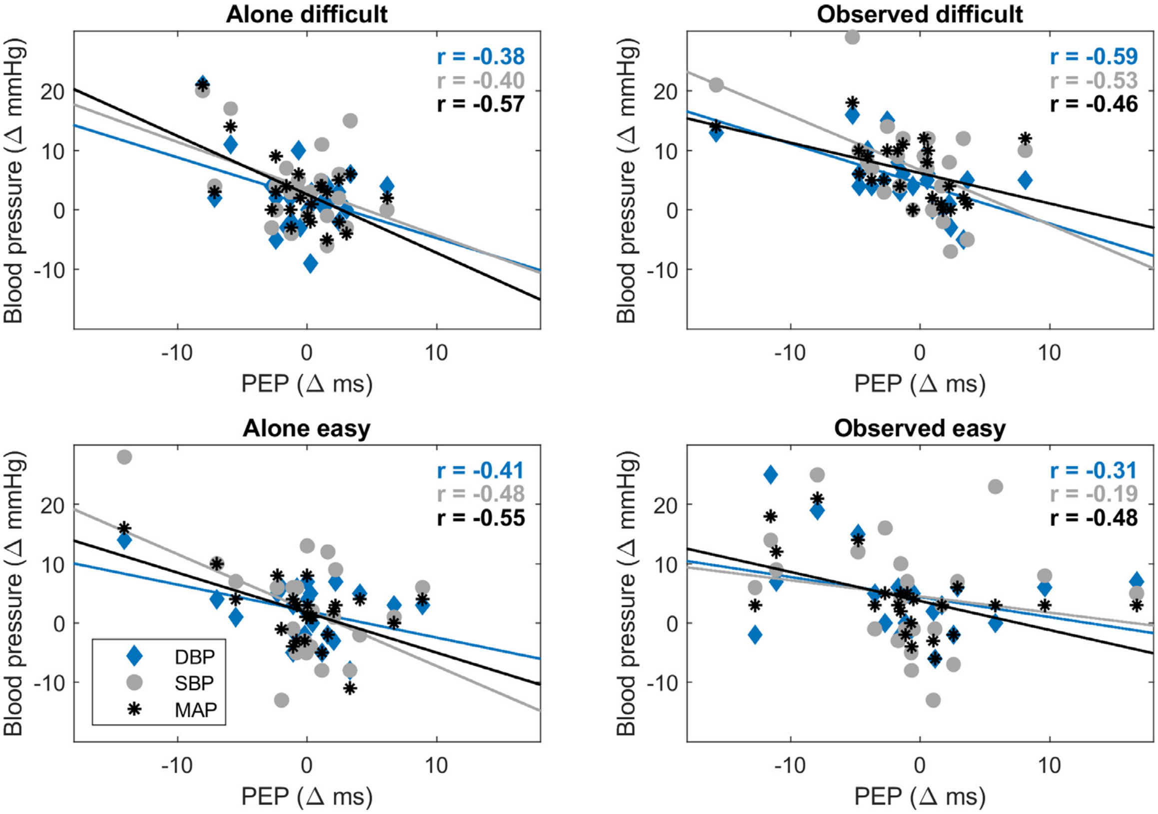

**Fig. 6.** Scatter plot demonstrating the relationship between PEP and the blood pressure measures' change scores. Lines represent lines of best fit. Pearson correlation coefficients are presented in the top righthand corner of each panel (black text – MAP, grey text – SBP and blue text – DBP). PEP = pre-ejection period, MAP = mean aterial pressure, SBP = systolic blood pressure, DBP = diastolic blood pressure, ms = milliseconds, mmHg = millimeters of mercury.

**Table 3**
Results of univariate tests on cardiovascular variables

| Effect tested | Measure | $F_{[1, 22]}$ | p value | Partial $\eta^2$ |
|---|---|---|---|---|
| **Social** | RMSSD | 5.34 | .054* | .20 |
| | HF HRV | 0.03 | .88 | <.01 |
| | PEP | 0.88 | .42 | .04 |
| | HR | 1.53 | .32 | .07 |
| | SBP | 6.70 | **.04** | .23 |
| | DBP | 10.00 | **.02** | .31 |
| | MAP | 14.08 | **<.01** | .39 |

Corrected p values presented. Those demonstrating significance (<0.05) are presented in bold. *p value presented to three decimal places due to the proximity to 0.05.

the two HRV measures themselves were weak (average correlation coefficient = 0.10). This result was unexpected because RMSSD and HF-HRV are generally highly correlated and are both thought to represent PNS activity (Kleiger et al., 1991). As a further exploration, additional correlations were run between RMSSD and HF-HRV during the baseline periods. The correlations between these two HRV indices during the baseline were higher, ranging from 0.38 for the baseline preceding the difficult observed condition to 0.55 for the baseline preceding the difficult alone condition (average across all four baselines = 0.47). Secondly, the blood pressure measures only correlated with a measure of PNS activity, RMSSD in a single condition (easy alone condition, correlation coefficients ranging from 0.32 to 0.42). Otherwise, the relationships between HRV and blood pressure measures were weak. Finally, the correlations between PEP, our measure of cardiac SNS activity, and the blood pressure measures were moderate in all conditions (average correlation coefficients ranging from -0.49 to -0.51). These observations, as explored further in the discussion section (section 4.3), increased our uncertainty when interpreting the ANS origins of the cardiovascular response demonstrated in this study.

**Table 4**
Pearson correlation coefficients for the association of cardiovascular variables

| **Alone difficult** | RMSSD | HF HRV | PEP | HR | SBP | DBP | MAP |
|---|---|---|---|---|---|---|---|
| RMSSD | - | **.33** | -.16 | .26 | -.06 | -.18 | -.04 |
| HF HRV | | - | -.08 | -.07 | -.02 | .15 | .21 |
| PEP | | | - | -.16 | **-.40** | **-.38** | **-.57** |
| HR | | | | - | **.40** | **.38** | .25 |
| SBP | | | | | - | **.57** | **.82** |
| DBP | | | | | | - | **.69** |
| MAP | | | | | | | - |
| | | | | | | | |
| **Alone easy** | RMSSD | HF HRV | PEP | HR | SBP | DBP | MAP |
| RMSSD | - | -.06 | -.27 | -.02 | **.37** | **.42** | **.32** |
| HF HRV | | - | **-.43** | -.26 | .09 | .10 | .18 |
| PEP | | | - | .19 | **-.48** | **-.41** | **-.55** |
| HR | | | | - | -.12 | -.04 | .05 |
| SBP | | | | | - | **.57** | **.77** |
| DBP | | | | | | - | **.84** |
| MAP | | | | | | | - |
| | | | | | | | |
| **Observed difficult** | RMSSD | HF HRV | PEP | HR | SBP | DBP | MAP |
| RMSSD | - | .06 | .16 | .01 | -.06 | -.21 | -.02 |
| HF HRV | | - | -.27 | .19 | .02 | .18 | .26 |
| PEP | | | - | **-.42** | **-.53** | **-.59** | **-.46** |
| HR | | | | - | **.44** | .21 | .21 |
| SBP | | | | | - | **.68** | **.86** |
| DBP | | | | | | - | **.82** |
| MAP | | | | | | | - |
| | | | | | | | |
| **Observed easy** | RMSSD | HF HRV | PEP | HR | SBP | DBP | MAP |
| RMSSD | | -0.10 | 0.12 | 0.12 | -0.05 | 0.14 | 0.09 |
| HF HRV | | | -0.27 | -0.18 | 0.10 | -0.08 | 0.10 |
| PEP | | | | **-0.32** | -0.19 | **-0.31** | **-0.48** |
| HR | | | | | **0.35** | **0.60** | **0.59** |
| SBP | | | | | | **0.48** | **0.74** |
| DBP | | | | | | | **0.88** |
| MAP | | | | | | | - |
| **Average correlation coefficients** | RMSSD | HF HRV | PEP | HR | SBP | DBP | MAP |
| RMSSD | - | 0.06 | -0.04 | 0.10 | 0.06 | 0.05 | 0.09 |
| HF HRV | | - | -0.27 | -0.08 | 0.05 | 0.09 | 0.19 |
| PEP | | | - | -0.19 | **-0.41** | **-0.43** | **-0.52** |
| HR | | | | - | 0.28 | **0.31** | 0.29 |
| SBP | | | | | - | **0.58** | **0.80** |
| DBP | | | | | | - | **0.82** |
| MAP | | | | | | | - |

Values higher than 0.3 are presented in bold

**Table 5**
Means (and SEMs) of subjective rating scales

| | **Social observation condition** | | | |
|---|---|---|---|---|
| | **Alone** | | **Observed** | |
| **Intelligibility condition** | **Difficult** | **Easy** | **Difficult** | **Easy** |
| Effort | 6.60 (0.38) | 4.60 (0.46) | 6.67 (0.33) | 4.42 (0.40) |
| Preference for change | 7.45 (0.49) | 5.42 (0.58) | 7.32 (0.40) | 5.09 (0.65) |
| Giving up | 3.90 (0.54) | 2.13 (0.33) | 4.07 (0.56) | 1.81 (0.33) |
| Stress | 5.01 (0.47) | 3.39 (0.41) | 5.59 (0.48) | 3.50 (0.41) |

**Table 6**
Results of univariate tests on subjective ratings

| **Effect** | **Subjective rating scale** | **$F_{[1,21]}$*** | **p value** | **Partial $\eta^2$** |
|---|---|---|---|---|
| Intelligibility | Effort | 25.78 | **<.01** | .55 |
| | Preference for change | 31.01 | **<.01** | .60 |
| | Giving up | 17.26 | **<.01** | .45 |
| | Stress | 37.01 | **<.01** | .64 |

Corrected p values presented. Those demonstrating significance (<0.05) are presented in bold. Degrees of freedom are 21 for all as one participant missed a question.

### *Subjective ratings*

Means and standard errors for subjectively rated effort, preference to improve the scenario, tendency to give up and stress can be found in Table 5. The multivariate analysis revealed a significant effect of intelligibility on the subjective rating scales (Wilks' lambda = 0.28, $F_{[4,18]} = 11.43$, p = <0.01, $\eta p^2 = 0.72$), such that at the difficult condition, the subjective ratings increased compared to at the easy condition. No significant effect of social state on the subjective rating scales was found (Wilks' lambda = 0.87, $F_{[4,18]} = 0.65$, p = 0.63, $\eta p^2 = 0.13$) and no interaction effect was demonstrated (Wilks' lambda = 0.95, $F_{[4,18]} = 0.22$, p = 0.93, $\eta p^2 = 0.05$). Results of follow up univariate analyses conducted on the subjective rating data can be seen in Table 6. p values were corrected to account for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; Martínez-Cagigal, 2021). For each of the subjective rating scales a significant effect of intelligibility was demonstrated ($p < 0.01$). This revealed that at the difficult intelligibility condition, participants rated their effort, preference to improve the scenario, tendency to give up and their stress level to be higher compared to the easy condition.

## Discussion

The main aim of this study was to assess the effects of social observation and intelligibility on hearing-impaired individuals' listening effort investment during a speech reception task. To index effort, we implemented subjective rating scales (self-perceived effort investment, preference to improve the listening situation, stress and tendency to give up), in addition to a range of cardiovascular measures including two HRV parameters (RMSSD and HF-HRV), PEP, heart rate and three blood pressure measures (SBP, DBP and MAP). We anticipated that the combination of these physiological measures would provide a comprehensive picture of the cardiovascular response associated with listening effort. More specifically, in keeping with motivational intensity theory (Brehm & Self, 1989), we expected that effort (as shown by the subjective ratings and cardiovascular reactivity) would vary as a function of both intelligibility (task demand) and social observation state (success importance). We hypothesized that there would be no strong impact of social observation state on effort in the easy condition, but at the difficult condition we expected observation to result in an increase of effort investment.

Contrary to our expectations, the most prominent finding of this study was that observation during a speech-in-noise task resulted in an increase in general cardiovascular reactivity, regardless of the difficulty of the task. These observation-related changes in cardiovascular reactivity, driven by increased blood pressure, were not accompanied by significant changes in the participants' subjective ratings. Instead, participants' subjective ratings demonstrated a significant effect of intelligibility only; an effect that was missing in the cardiovascular data. Indeed, disparities between physiological measures, speech recognition performance and subjective ratings are not uncommon in the field (McGarrigle et al., 2014; Strand et al., 2018; Zekveld et al., 2010; Zekveld et al., 2011). Several authors have highlighted the need to include multiple different measures of listening effort, as it is likely that different measures reflect different aspects of listening effort (Alhanbali et al., 2019; Francis & Love, 2020; Strand et al., 2021). Below we will discuss the potential mechanisms and constructs demonstrated by our findings.

### *Social observation*

Our results demonstrated that social observation during a listening task resulted in an increase in blood pressure. This may reflect increased effort investment related to heightened success importance (Gendolla & Richter, 2006). The presence of the observers may have motivated the participants to succeed at the task, thus prompting them (the participants) to invest more effort than when they were alone. Although an intelligibility-moderated relationship was anticipated, it is possible that the participants were preoccupied by the presence of the observers and therefore less sensitive to changes in intelligibility (Richter, 2010). This is supported by a study which demonstrated that the context of a memory or visual scanning task impacted whether participants' PEP reactivity scores were sensitive to task demand or reward level (Richter, 2010). Participants who were questioned before the task about reward were sensitive to the reward manipulation, whereas those who were asked questions about task demand were sensitive to the task demand manipulation. In the context of our study, participants may have been focused on the observation aspect of the experiment and less on the intelligibility manipulation, resulting in an increased effort investment when observed, regardless of intelligibility level. This explanation, however, is not supported by the subjective ratings of effort, which showed no effect of observation. Other interpretations may therefore be warranted.

Beyond increased effort investment, another construct reported to elicit changes in blood pressure is task engagement (Fairclough et al., 2009). Task engagement is a broad, multidimensional concept, encompassing effort investment, motivation and affective changes (Fairclough et al., 2009). In one experiment, participants undertook a working memory task, while SBP was measured. Participants received pre-arranged feedback on their performance: a positive feedback group was shown that their performance improved over time, whereas a negative feedback group was shown the opposite. The participants who received positive feedback exhibited an increased SBP compared to those who received negative feedback. The authors interpreted this to be a result of increased task engagement (Fairclough et al., 2009). Unfortunately, in our study, as the observers were absent during the baseline periods, it is not possible to determine whether the observation-based changes were specific to the task, and thus the level of task engagement, or simply associated with the observers' presence in the room. Interestingly, our participant subjective rating of "giving up", which might be considered the inverse of task engagement, did not reveal any effect of social condition. Outside of the context of task engagement, changes in affect have also been related to cardiovascular reactivity (including blood pressure) (Maier et al., 2003; Neumann & Waldstein, 2001), however these are beyond the scope of the current discussion.

Instead of attributing the present findings to changes in effort investment or task engagement, the demonstrated increase in blood pressure during observation could be solely due to an increase in general physiological stress associated with the presence of, and potential evaluation from, others (Baron, 1986; Maier et al., 2003; Woody et al., 2018). Such a response to the presence of the observers could occur irrespective of the nature of the current (listening) task. Similar findings to ours were demonstrated in a study by Woody et al., (2018), where participants performed a speech either in the presence of a two-member audience (social-evaluative threat condition) or to a video camera (non-social-evaluative condition). In addition to the social manipulation, cognitive load was manipulated (load or non-load conditions). Heart rate and blood pressure increased in response to the social-evaluative threat manipulation, but no impact of cognitive load (task demand) was demonstrated. The authors interpreted this to reflect a social-evaluative threat response to the audience as stressors.

In the present study, had the changes in blood pressure while observed been related to social-evaluative threat, one might expect this to be reflected in the subjective stress rating, which it was not. Previous work has demonstrated a relationship between a task involving the presence of an audience, physiological markers of stress and subjective report. For instance, Hellhammer and Schubert's (2012) participants performed the Trier Social Stress Test, which involves giving an interview and performing a mental arithmetic task aloud in front of a panel of judges. Before, during and after the test, participants rated their stress perception, anxiety, and emotional insecurity on a visual analogue scale. All three subjective ratings were significantly increased during and after the task compared to baseline, showing that participants were more stressed, anxious, and emotionally insecure. In addition, the subjective stress ratings taken during the task were able to predict heart rate parameters. It is possible that the Trier Social Stress Test elicits more stress than our test paradigm, explaining the difference in our results.

In light of the above discussion, we believe that the most plausible explanation is that the presence of observers resulted in increased stress associated with potential evaluation. Though this was not supported by an observation effect on the subjective rating scales (one of which referred specifically to stress), we suspect that when selecting their subjective ratings, the participants considered only the listening task rather than the whole situation, omitting contextual factors such as the observers' presence. The subjective rating scales also specifically referred to speech understanding and the listening task. Interestingly, a similar result (subjective ratings showing an effect of intelligibility not co-presence) was demonstrated in another study that introduced a second participant to a listening task (Pielage et al., 2021). The authors attributed this finding to inherent weaknesses of the subjective rating scales employed. Some authors have hypothesised that instead of rating their effort, participants substitute different constructs that are easier to assess, such as performance or task difficulty (Moore & Picou, 2018; Picou et al., 2017; Picou & Ricketts, 2018). Future work of this type may benefit from providing participants with more explicit instructions when completing the subjective rating scales about considering the whole situation, not only the listening-related aspects. Future work may also benefit from revising the challenge and threat subjective rating that was implemented in this study. Critique of this model often rejects the view that challenge and threat are opposing constructs on a unidimensional continuum (Uphill et al., 2019). The difficulty experienced by our participants in understanding this rating scale adds weight to this criticism.

Stress whilst performing a task under observation has been shown to result in both an increase in SNS activity (measured by PEP) and a decrease in PNS activity (measured by HRV) (Bosch et al., 2009; Weissman & Mendes, 2021). The lack of a sensitive response to observation in both our SNS (PEP) and PNS (HRV) measures was surprising as other studies have demonstrated an effect on these measures of including even a single audience member (Bosch et al., 2009). However, this disparity in results may be attributed to the notable difference between the studies in terms of the participant demographics and study designs. Weissman and Mendes (2021) and Bosch et al., (2009) included young participants, presumably with normal-hearing, whereas our participants were older, hearing-impaired individuals. There is some evidence to suggest that the magnitude of PEP reactivity decreases with age, and specifically in participants with heart disease (Bertel et al., 1980; Gurel et al., 2019). Though we believed our participants were free of cardiovascular disease, it is possible that some had undetected or undiagnosed conditions (Tan et al., 2018). Interestingly, a significant effect of social condition was also absent in a similar study involving hearing-impaired individuals and RMSSD (Mackersie & Kearney, 2017).

### Intelligibility

We can be confident that our manipulation of intelligibility was successful by inspecting the performance data and subjective ratings. Subjective ratings were sensitive to intelligibility, demonstrating that at the more difficult intelligibility condition, participants rated their effort, stress, preference to improve audibility and tendency to give up as higher, compared to at the easier condition. Similar findings have been demonstrated in other studies (Mackersie & Cones, 2011; Moore & Picou, 2018; Pielage et al., 2021; Plain et al., 2020; Zekveld et al., 2010).

A plausible explanation for the lack of an overall effect of intelligibility on the cardiovascular data may relate to the specific intelligibility conditions presented in the present study. Previous work has implemented four or more SNRs (with different auditory material and maskers), spanning a wider range of the psychometric curve (Mackersie et al., 2015; Plain et al., 2020; Seeman & Sims, 2015). Whereas the present study applied two individually adapted SNRs targeting relatively close performance levels, situated at the middle to the right-hand side of the psychometric function (50 and 80% correct). A broader range of more distinct listening demand levels may have allowed us to detect an effect of intelligibility with greater sensitivity. This point is perhaps even more salient because we included hearing-impaired participants in our study. It is likely that in this population, physiological differences between the easy and difficult listening conditions were not substantial. Indeed, previous work measuring pupil dilation has shown that hearing-impaired participants have a similar peak pupil dilation (indicative of effort), across a wide range of SNRs, whereas listeners with normal hearing show a reduced peak pupil dilation at higher compared to lower SNRs (Ohlenforst et al., 2017). This is probably due to a compensatory effort investment required to perform even an easy task with a hearing impairment (Hockey, 1997).

Another explanation for the differing results between our study and others that implemented cardiovascular measures during listening tasks may be the type of masking noise presented during the task. For instance, the present study applied a four-talker babble masker, whereas Plain et al., (2020) applied a single-interfering talker masker. Single-talker maskers have a higher informational masking component than other types of masker (Brungart, 2001). It is possible that PEP is more sensitive to single-talker masker for this reason.

### Interpreting the ANS origins of these findings

The ANS origins of the cardiovascular responses detected in the present study are not clear cut. The significant multivariate effect of observation was driven by the blood pressure measures, changes which result from contributions of both the SNS and PNS. We included two measures of cardiac PNS activity: RMSSD and HF-HRV, neither of which correlated with the blood pressure measures, nor revealed any significant changes related to the experimental manipulations. More importantly, these two measures were poorly correlated to one another, despite both being measures of PNS activity (Kleiger et al., 1991). RMSSD and HF-HRV have been shown to be highly correlated, yet the relationship between these measures was weak in the present study (see Table 4). The discrepancy between these measures may be in part due to respiration rate, which was not monitored during the experiment. It is possible that participants' breathing was not confined purely to the 0.15 – 0.4Hz frequency band during the speech reception task, which may have added noise to the HF-HRV measure. Previous work has demonstrated that HF-HRV is susceptible to the effects of respiration whereas RMSSD is not (Penttilä et al., 2001). This might explain the lack of correlation between these two indices. A task related interference in the measures is supported by the stronger correlation found between the two indices during the baseline, when the participant was not required to speak.

Interestingly though, the blood pressure measurements were highly correlated with the SNS-measure, PEP (Table 4). Although PEP itself did not demonstrate a significant effect of either experimental manipulation, the correlational analysis clearly demonstrates a strong relationship between PEP and the blood pressure measures (see Fig. 6), which were sensitive to observation. This may suggest that the demonstrated changes in blood pressure are a result of SNS activity, rather than PNS activity. Based upon the above reasoning, there is some evidence to suggest that SNS activation may have contributed to the cardiovascular response demonstrated in the present study. Without a main effect in PEP, RMSSD or HF-HRV, we are not able to conclusively determine from this dataset which branch of the ANS dominated during the experiment.

### Limitations of the work and future directions

Several limitations existed in the present work. Firstly, due to unexpected issues with cardiovascular data quality in our test population, the study included three fewer participants than the 26 recommended by our power calculation. We do not anticipate that this has had a large impact on the presented results because our power calculation was deliberately conservative in nature. The power calculation was conducted based upon detecting a multivariate effect. It is possible that for some variables, such as PEP, HF-HRV, RMSSD and heart rate, the univariate analyses may have been under-powered and comprising of too much noise to detect an effect at a significant level.

As discussed above in section 4.1, another methodological limitation was the timing of the presence of the observers. We acknowledge that this limits the ability to interpret the physiological changes presented here as representing a task related construct. We instead interpret our findings as physiological stress associated with potential evaluation from the observers, irrespective of the task. Future work would benefit from the addition of a baseline that has observers present and an alternative, non-auditory task, such that a distinction might be made between a purely observation effect and a task-related observation effect.

Finally, the balancing of the order of conditions across participants was not equal. 17 participants began with the observed conditions followed by the alone conditions, whereas 12 participants began with the alone conditions followed by the observed conditions. Imperfect balancing may have inadvertently biased the results towards a social observation effect.

## Conclusions

This study measured the cardiovascular reactivity of hearing-impaired participants during a speech reception task at two intelligibility levels and two social observation states. We demonstrated that while observed, cardiovascular reactivity was increased, compared to when the task was performed alone. Of the cardiovascular measures employed, all three measures of blood pressure demonstrated this effect of observation. The same effect was absent from the subjective rating data, which instead demonstrated a significant effect of intelligibility. To our knowledge, these data are the first to demonstrate social observation-mediated blood pressure changes during a speech reception task in hearing-impaired participants. Our final aim was to learn more about the relative contributions of the two branches of the ANS during listening. However, based upon our results, we were not able to conclusively determine whether the SNS or PNS, or a combination of both, were driving the cardiovascular response demonstrated.

## Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Sklodowska-Curie grant agreement No 765329.

## Declarations of interest

None

## Author statements

Bethany Plain: writing – original draft preparation, conducting experiment

Hidde Pielage: conducting experiment, reviewing manuscript

Michael Richter, Tanveer A. Bhuiyan, Thomas Lunner, Sophia E. Kramer & Adriana A. Zekveld: conceptualization, supervision, methodology, reviewing and editing manuscript

## Acknowledgements

The authors are very grateful to the research clinicians, Hjørdis Huginsdóttir Eide, Renskje Hietkamp, Franziska Behnen and Christine Winther for their input into the experimental design, organizing participant visits and conducting the test sessions in Danish. Also, we'd like to extend our gratitude to Antoine Favre-Félix and Lorenz Fiedler for their assistance in laboratory set-up.

## Data sharing statement

The authors are not able to share the data due to GDPR regulations and the relatively small pool of test participants at Eriksholm Research Centre, which makes full anonymisation of the data very difficult.

## References

Ahmed, S.S., Levinson, G.E., Schwartz, C.J., Ettinger, P.O., 1972. Systolic time intervals as measures of the contractile state of the left ventricular myocardium in man. Circulation 46 (3), 559–571. doi:10.1161/01.CIR.46.3.559.

Alhanbali, S., Dawes, P., Lloyd, S., Munro, K.J., 2017. Self-reported listening-related effort and fatigue in hearing-impaired adults. Ear Hear. 38 (1), e39–e48. doi:10.1097/AUD.0000000000000361.

Alhanbali, S., Dawes, P., Millman, R.E., Munro, K.J., 2019. Measures of listening effort are multidimensional. Ear Hear. 40 (5), 1084–1097. doi:10.1097/AUD.0000000000000697.

Baron, R.S., 1986. Distraction-conflict theory: Progress and problems. Adv. Exp. Soc. Psychol. 19 (C), 1–40. doi:10.1016/S0065-2601(08)60211-7.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc.: Series B (Methodological) 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Berntson, G.G., Cacioppo, J.T., Quigley, K.S., 1991. Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. Psychol. Rev. 98 (4), 459–487. doi:10.1037/0033-295X.98.4.459.

Berntson, G.G., Thomas Bigger, J., Eckberg, D.L., Grossman, P., Kaufmann, P.G., Malik, M., Nagaraja, H.N., Porges, S.W., Saul, J.P., Stone, P.H., Van Der Molen, M.W., 1997. Heart rate variability: Origins methods, and interpretive caveats. In *Psychophysiology* (Vol. 34, Issue 6. Blackwell Publishing Inc 623–648. doi:10.1111/j.1469-8986.1997.tb02140.x.

Bertel, O., Bühler, F.R., Kiowski, W., Lütold, B.E., 1980. Decreased beta-adrenoreceptor responsiveness as related to age, blood pressure, and plasma catecholamines in patients with essential hypertension. Hypertension 2 (2), 130–138. doi:10.1161/01.HYP.2.2.130.

Bess, F.H., Hornsby, B.W.Y., 2014. Commentary: Listening can be exhausting-fatigue in children and adults with hearing loss. In *Ear and Hearing*. Lippincott Williams and Wilkins 35 (6), 592–599. doi:10.1097/AUD.0000000000000099.

Bird, K.D., Hadzi-Pavlovic, D., 2014. Controlling the maximum familywise Type I error rate in analyses of multivariate experiments. Psychol. Methods 19 (2), 265–280. doi:10.1037/a0033806.

Blascovich, J., Tomaka, J., 1996. The Biopsychosocial Model of Arousal Regulation. Adv. Exp. Soc. Psychol. 28 (C), 1–51. doi:10.1016/S0065-2601(08)60235-X.

Bosch, J.A., De Geus, E.J.C., Carroll, D., Goedhart, A.D., Anane, L.A., Veldhuizen Van Zanten, J.J., Helmerhorst, E.J., Edwards, K.M., 2009. A general enhancement of autonomic and cortisol responses during social evaluative threat. Psychosom. Med. 71 (8), 877–885. doi:10.1097/PSY.0b013e3181baef05.

Brehm, J.W., Self, E.A., 1989. The intensity of motivation. Annu. Rev. Psychol. 40 (1), 109–131. doi:10.1146/annurev.ps.40.020189.000545.

Brungart, D.S., 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. J. Acoust. Soc. Am. 109 (3), 1101–1109. doi:10.1121/1.1345696.

Byrd, D.L., Reuther, E.T., McNamara, J.P.H., DeLucca, T.L., Berg, W.K., 2015. Age differences in high frequency phasic heart rate variability and performance response to increased executive function load in three executive function tasks. Front. Psychol. 5, 1470. doi:10.3389/fpsyg.2014.01470, NOV.

Christensen, J., Saunders, G., Porsbo, M., & Pontoppidan, N. (2020). *The everyday acoustic environment and its association with human heart rate: evidence from real-world data logging with hearing aids and wearables*. https://doi.org/10.1101/2020.07.08.180869

Cvijanović, N., Kechichian, P., Janse, K., Kohlrausch, A., 2017. Effects of noise on arousal in a speech communication setting. Speech Commun. 88, 127–136. doi:10.1016/j.specom.2017.02.001.

Dorman, M.F., Spahr, A., Gifford, R.H., Cook, S., Zhang, T., Loiselle, L., Yost, W., Cardy, L., Whittingham, J.A., Schramm, D., 2012. Current research with cochlear implants at Arizona State University. J. Am. Acad. Audiol. 23 (6), 385–395. doi:10.3766/jaaa.23.6.2.

Evans, J., 1996. Straightforward statistics for the behavioral sciences https://psycnet.apa.org/record/1995-98499-000.

Fairclough, S.H., Ewing, K.C., Roberts, J., 2009. Measuring task engagement as an input to physiological computing. In: Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. ACII doi:10.1109/ACII.2009.5349483, *2009*.

Francis, A.L., Bent, T., Schumaker, J., Love, J., Silbert, N., 2021. Listener characteristics differentially affect self-reported and physiological measures of effort associated with two challenging listening conditions. Attent. Percept. Psychophys. doi:10.3758/s13414-020-02195-9.

Francis, A.L., Love, J., 2020. Listening effort: Are we measuring cognition or affect, or both?. In: Wiley Interdisciplinary Reviews: Cognitive Science, 11. Wiley-Blackwell, p. e1514. doi:10.1002/wcs.1514.

Francis, A.L., MacPherson, M.K., Chandrasekaran, B., Alvar, A.M., 2016. Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. Front. Psychol. 263. doi:10.3389/fpsyg.2016.00263, *7*(MAR).

Gendolla, G.H.E., Richter, M., 2006. Cardiovascular reactivity during performance under social observation: The moderating role of task difficulty. In: International Journal of Psychophysiology, 62. Elsevier Science, pp. 185–192. doi:10.1016/j.ijpsycho.2006.04.002.

Gendolla, G.H.E., Wright, R.A., Richter, M., Gendolla, G.H.E., Wright, R.A., Richter, M., 2019. Advancing issues in motivation intensity research. In: The Oxford Handbook of Human Motivation, pp. 372–392. doi:10.1093/oxfordhb/9780190666453.013.21.

Giuliano, R.J., Gatzke-Kopp, L.M., Roos, L.E., Skowron, E.A., 2017. Resting sympathetic arousal moderates the association between parasympathetic reactivity and working memory performance in adults reporting high levels of life stress. Psychophysiology 54 (8), 1195–1208. doi:10.1111/psyp.12872.

Gordan, R., Gwathmey, J.K., Xie, L.-H., 2015. Autonomic and endocrine control of cardiovascular function. World J. Cardiol. 7 (4), 204. doi:10.4330/wjc.v7.i4.204.

Gurel, N.Z., Carek, A.M., Inan, O.T., Levantsevych, O., Abdelhadi, N., Hammadah, M., O'Neal, W.T., Kelli, H., Wilmot, K., Ward, L., Rhodes, S., Pearce, B.D., Mehta, P.K., Kutner, M., Garcia, E., Quyyumi, A., Vaccarino, V., Raggi, P., Douglas Bremner, J., Shah, A.J., 2019. Comparison of autonomic stress reactivity in young healthy versus aging subjects with heart disease. PLoS One (5) 14. doi:10.1371/journal.pone.0216278.

Hällgren, M., Larsby, B., Lyxell, B., Arlinger, S., 2001. Evaluation of a cognitive test battery in young and elderly normal-hearing and hearing-impaired persons. J. Am. Acad. Audiol. 12 (7), 357–370. https://pubmed.ncbi.nlm.nih.gov/11500010/.

Harris, W.S., Schoenfeld, C.D., Weissler, A.M., 1967. Effects of adrenergic receptor activation and blockade on the systolic preejection period, heart rate, and arterial pressure in man. J. Clin. Invest. 46 (11), 1704–1714. doi:10.1172/JCI105661.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. 52 (C), 139–183. doi:10.1016/S0166-4115(08)62386-9.

Hasson, D., Theorell, T., Wallén, M.B., Leineweber, C., Canlon, B., 2011. Stress and Prevalence of Hearing Problems in the Swedish Working Population doi:10.1186/1471-2458-11-130.

Hellhammer, J., Schubert, M., 2012. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 37 (1), 119–124. doi:10.1016/j.psyneuen.2011.05.012.

Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol. 45 (1–3), 73–93. doi:10.1016/S0301-0511(96)05223-4.

Holman, J.A., Drummond, A., Hughes, S.E., Naylor, G., 2019. International Journal of Audiology Hearing impairment and daily-life fatigue: a qualitative study Hearing impairment and daily-life fatigue: a qualitative study. doi:10.1080/14992027.2019.1597284.

Hopstaken, J.F., van der Linden, D., Bakker, A.B., Kompier, M.A.J., 2015. The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. Biol. Psychol. 110, 100–106. doi:10.1016/j.biopsycho.2015.06.013.

Kahneman, D., 1973. Attention and Effort Englewood Cliffs. New Jersey.

Kleiger, R.E., Bigger, J.T., Bosner, M.S., Chung, M.K., Cook, J.R., Rolnitzky, L.M., Steinman, R., Fleiss, J.L., 1991. Stability over time of variables measuring heart rate variability in normal subjects. Am. J. Cardiol. 68 (6), 626–630. doi:10.1016/0002-9149(91)90355-O.

Koelewijn, T., Zekveld, A.A., Lunner, T., Kramer, S.E., 2018. The effect of reward on listening effort as reflected by the pupil dilation response. Hear. Res. 367, 106–112. doi:10.1016/j.heares.2018.07.011.

Levenson, R.W., 2014. The autonomic nervous system and emotion. Emot. Rev. 6 (2), 100–112. doi:10.1177/1754073913512003.

Loewenfeld, I.E., Lowenstein, O., 1999. The Pupil : Anatomy, Physiology, and Clinical Applications. Butterworth-Heinemann.

Lovallo, W.R., 2005. Stress & health: biological and psychological interactions. Stress and Health: Biological and Psychological Interactions doi:10.4135/9781452233543.

Mackersie, C.L., Calderon-Moultrie, N., 2016. Autonomic nervous system reactivity during speech repetition tasks: heart rate variability and skin conductance. Ear Hear. 37, 118S–125S. doi:10.1097/AUD.0000000000000305.

Mackersie, C.L., Cones, H., 2011. Subjective and psychophysiological indexes of listening effort in a competing-talker task. J. Am. Acad. Audiol. 22 (2), 113–122. doi:10.3766/jaaa.22.2.6.

Mackersie, C.L., Kearney, L., 2017. Autonomic nervous system responses to hearing-related demand and evaluative threat. In: American Journal of Audiology, 26. American Speech-Language-Hearing Association, pp. 373–377. doi:10.1044/2017_AJA-16-0133 Issue 3S.

Mackersie, C.L., Macphee, I.X., Heldt, E.W., 2015. Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. Ear Hear. 36 (1), 145–154. doi:10.1097/AUD.0000000000000091.

Maier, K.J., Waldstein, S.R., Synowski, S.J., 2003. Relation of cognitive appraisal to cardiovascular reactivity, affect, and task engagement. Ann. Behav. Med. 26 (1), 32–41. doi:10.1207/S15324796ABM2601_05.

Malik, M., Camm, A.J., Bigger, J.T., Breithardt, G., Cerutti, S., Cohen, R.J., Coumel, P., Fallen, E.L., Kennedy, H.L., Kleiger, R.E., Lombardi, F., Malliani, A., Moss, A.J., Rottman, J.N., Schmidt, G., Schwartz, P.J., Singer, D.H., 1996. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Eur. Heart J. 17 (3), 354–381. doi:10.1093/oxfordjournals.eurheartj.a014868.

Martínez-Cagigal, V., 2021. *Multiple Testing Toolbox* (Retrieved May 13, 2021). MATLAB Central File Exchange https://www.mathworks.com/matlabcentral/fileexchange/70604-multiple-testing-toolbox.

McArdle, W., Katch, F., & Katch, V. (2010). *Exercise physiology: nutrition, energy, and human performance*. https://www.google.com/books?hl=en&lr=&id=XOyjZX0Wxw4C&oi=fnd&pg=PR17&dq=Exercise+physiology:+Nutrition,+energy,+and+human+performance+(7th+ed.)&ots=MwtgS5kw4l&sig=PJmUd7R6HC0M8F0L7q8xy6cS_hk

McCorry, L.K., 2007. Physiology of the autonomic nervous system. Am. J. Pharm. Educ. (4) 71. doi:10.5688/aj710478.

McCoy, S.L., Tun, P.A., Cox, L.C., Colangelo, M., Stewart, R.A., Wingfield, A., 2005. Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. Q. J. Exp. Psychol. Sect. A: Human Exp. Psychol. 58 (1), 22–33. doi:10.1080/02724980443000151.

McGarrigle, R., Dawes, P., Stewart, A.J., Kuchinsky, S.E., Munro, K.J., 2017. Pupillometry reveals changes in physiological arousal during a sustained listening task. Psychophysiology 54 (2), 193–203. doi:10.1111/psyp.12772.

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. In: International Journal of Audiology, 53. Informa Healthcare., pp. 433–440. doi:10.3109/14992027.2014.890296.

Melis, C., van Boxtel, A., 2007. Autonomic physiological response patterns related to intelligence. Intelligence 35 (5), 471–487. doi:10.1016/j.intell.2006.09.007.

Moore, T.M., Picou, E.M., 2018. A potential bias in subjective ratings of mental effort. J. Speech Lang. Hear. Res. 61 (9), 2405–2421. doi:10.1044/2018_JSLHR-H-17-0451.

Nachtegaal, J., Kuik, D.J., Anema, J.R., Goverts, S.T., Festen, J.M., Kramer, S.E., 2009. Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an internet-based national survey on hearing. Int. J. Audiol. 48 (10), 684–691. doi:10.1080/14992020902962421.

Neumann, S.A., Waldstein, S.R., 2001. Similar patterns of cardiovascular response during emotional activation as a function of affective valence and arousal and gender. J. Psychosom. Res. 50 (5), 245–253. doi:10.1016/S0022-3999(01)00198-2.

Newlin, D.B., Levenson, R.W., 1979. Pre-ejection period: measuring beta-adrenergic influences upon the heart. Psychophysiology 16 (6), 546–552. doi:10.1111/j.1469-8986.1979.tb01519.x.

Nielsen, J.B., Dau, T., 2011. The Danish hearing in noise test. International Journal of Audiology 50 (3), 202–208. doi:10.3109/14992027.2010.524254.

Nolte, R.N., Wright, R.A., Turner, C., Contrada, R.J., 2008. Reported fatigue, difficulty, and cardiovascular response to a memory challenge. Int. J. Psychophysiol. 69 (1), 1–8. doi:10.1016/j.ijpsycho.2008.02.004.

Ohlenforst, B., Wendt, D., Kramer, S.E., Naylor, G., Zekveld, A.A., Lunner, T., 2018. Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. Hear. Res. 365. doi:10.1016/j.heares.2018.05.003.

Ohlenforst, B., Zekveld, A.A., Jansma, E.P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., Kramer, S.E., 2017. Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. In: Ear and Hearing, 38. Lippincott Williams and Wilkins, pp. 267–281. doi:10.1097/AUD.0000000000000396 Issue 3.

Ohlenforst, B., Zekveld, A.A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N.J., Kramer, S.E., 2017. Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. Hear. Res. 351, 68–79. doi:10.1016/j.heares.2017.05.012.

Penttilä, J., Helminen, A., Jartti, T., Kuusela, T., Huikuri, H.V., Tulppo, M.P., Coffeng, R., Scheinin, H., 2001. Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: Effects of various respiratory patterns. Clin. Physiol. 21 (3), 365–376. doi:10.1046/j.1365-2281.2001.00337.x.

Pichora-Fuller, M.K., Kramer, S.E., Eckert, M.A., Edwards, B., Hornsby, B.W.Y., Humes, L.E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C.L., Naylor, G., Phillips, N.A., Richter, M., Rudner, M., Sommers, M.S., Tremblay, K.L., Wingfield, A., 2016. Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). Ear Hear. 37. doi:10.1097/AUD.0000000000000312, 5S-27S.

Picou, E.M., Bean, B.N., Marcrum, S.C., Hornsby, B.W., Ricketts, T.A., 2019. Moderate reverberation does not increase subjective fatigue, subjective listening effort, or behavioral listening effort in school-aged children. Front. Psychol. 10, 1749. doi:10.3389/fpsyg.2019.01749, JULY.

Picou, E.M., Moore, T.M., Ricketts, T.A., 2017. The effects of directional processing on objective and subjective listening effort. J. Speech Lang. Hear. Res. 60 (1), 199–211. doi:10.1044/2016_JSLHR-H-15-0416.

Picou, E.M., Ricketts, T.A., 2018. The relationship between speech recognition, behavioural listening effort, and subjective ratings. Int. J. Audiol. 57 (6), 457–467. doi:10.1080/14992027.2018.1431696.

Pielage, H., Zekveld, A.A., Saunders, G.H., Versfeld, N.J., Lunner, T., Kramer, S.E., 2021. The presence of another individual influences listening effort, but not performance. Ear Hear. Publish Ah. doi:10.1097/aud.0000000000001046.

Plain, B., Richter, M., Zekveld, A.A., Lunner, T., Bhuiyan, T., Kramer, S.E., 2020. Investigating the influences of task demand and reward on cardiac pre-ejection period reactivity during a speech-in-noise task. *Ear Hear., Publish Ah.* doi:10.1097/aud.0000000000000971.

Richter, M., 2010. Pay attention to your manipulation checks! Reward impact on cardiac reactivity is moderated by task context. Biol. Psychol. 84 (2), 279–289. doi:10.1016/j.biopsycho.2010.02.014.

Richter, M., 2016. The moderating effect of success importance on the relationship between listening demand and listening effort. Ear Hear. 37, 111S–117S. doi:10.1097/AUD.0000000000000295.

Richter, M., Friedrich, A., Gendolla, G.H.E., 2008. Task difficulty effects on cardiac activity. Psychophysiology 45 (5), 869–875. doi:10.1111/j.1469-8986.2008.00688.x.

Richter, M., Gendolla, G.H.E., 2009. The heart contracts to reward: Monetary incentives and preejection period. Psychophysiology 46 (3), 451–457. doi:10.1111/j.1469-8986.2009.00795.x.

Richter, M., Gendolla, G.H.E., Wright, R.A., 2016. In: Chapter Five - Three Decades of Research on Motivational Intensity Theory: What We Have Learned About Effort and What We Still Don't Know, 3. Elsevier, pp. 149–186. doi:10.1016/bs.adms.2016.02.001 https://doi.org/ (A. J. B. T.-A. in M. S. Elliot (ed.).

Rönnberg, J., Rudner, M., Foo, C., Lunner, T., 2008. Cognition counts: A working memory system for ease of language understanding (ELU). Int. J. Audiol. 47 (SUPPL. 2). doi:10.1080/14992020802301167.

Seeman, S., Sims, R., 2015. Comparison of psychophysiological and dual-task measures of listening effort. J. Speech Lang. Hear. Res. 58 (6), 1781–1792. doi:10.1044/2015_JSLHR-H-14-0180.

Shaffer, F., Ginsberg, J.P., 2017. An overview of heart rate variability metrics and norms. Front. Public Health 5. doi:10.3389/fpubh.2017.00258.

Shaffer, F., Venner, J., 2013. heart rate variability anatomy and physiology. Biofeedback 41 (1), 13–25. doi:10.5298/1081-5937-41.1.05.

Sherwood, A., Allen, M.T., Fahrenberg, J., Kelsey, R.M., Lovallo, W.R., van Doornen, L.J.P., 1990. Methodological guidelines for impedance cardiography. Psychophysiology 27 (1), 1–23. doi:10.1111/j.1469-8986.1990.tb02171.x.

Sherwood, A., Allen, M.T., Obrist, P.A., Langer, A.W., 1986. Evaluation of Beta-Adrenergic influences on cardiovascular and metabolic adjustments to physical and psychological stress. Psychophysiology 23 (1), 89–104. doi:10.1111/j.1469-8986.1986.tb00602.x.

Shoushtarian, M., Weder, S., Innes-Brown, H., McKay, C.M., 2019. Assessing hearing by measuring heartbeat: the effect of sound level. PLoS One 14 (2), e0212940. doi:10.1371/journal.pone.0212940.

Shukla, A., Harper, M., Pedersen, E., Goman, A., Suen, J.J., Price, C., Applebaum, J., Hoyer, M., Lin, F.R., Reed, N.S., 2020. Hearing loss, loneliness, and social isolation: a systematic review. Otolaryngol.-Head Neck Surg. (5) 622–633. doi:10.1177/0194599820910377.

Steinhauer, S.R., Siegle, G.J., Condray, R., Pless, M., 2004. Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. Int. J. Psychophysiol. 52 (1), 77–86. doi:10.1016/j.ijpsycho.2003.12.005.

Strand, J.F., Brown, V.A., Merchant, M.B., Brown, H.E., Smith, J., 2018. Measuring listening effort: convergent validity, sensitivity, and links with cognitive and personality measures. J. Speech Lang. Hear. Res. 61 (6), 1463–1486. doi:10.1044/2018_JSLHR-H-17-0257.

Strand, J.F., Ray, L., Dillman-Hasso, N.H., Villanueva, J., Brown, V.A., 2021. Understanding speech amid the jingle and jangle: recommendations for improving measurement practices in listening effort research. Audit. Percept. Cognit. 1–20. doi:10.1080/25742442.2021.1903293.

Tan, H.E., Lan, N.S.R., Knuiman, M.W., Divitini, M.L., Swanepoel, D.W., Hunter, M., Brennan-Jones, C.G., Hung, J., Eikelboom, R.H., Santa Maria, P.L., 2018. Associations between cardiovascular disease and its risk factors with hearing loss-A cross-sectional analysis. Clin. Otolaryngol. 43 (1), 172–181. doi:10.1111/coa.12936.

Tarvainen, M.P., Niskanen, J.P., Lipponen, J.A., Ranta-aho, P.O., Karjalainen, P.A., 2014. Kubios HRV - Heart rate variability analysis software. Comput. Methods Programs Biomed. 113 (1), 210–220. doi:10.1016/j.cmpb.2013.07.024.

Uphill, M.A., Rossato, C.J.L., Swain, J., O'Driscoll, J, 2019. Challenge and threat: a critical review of the literature and an alternative conceptualization. In: Frontiers in Psychology, 10. Frontiers Media S.A., p. 1255. doi:10.3389/fpsyg.2019.01255 Issue JUL.

Wang, Y., Naylor, G., Kramer, S.E., Zekveld, A.A., Wendt, D., Ohlenforst, B., Lunner, T., 2018. Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. Ear Hear. 39 (3), 573–582. doi:10.1097/AUD.0000000000000512.

Weissman, D.G., Mendes, W.B., 2021. Correlation of sympathetic and parasympathetic nervous system activity during rest and acute stress tasks. Int. J. Psychophysiol. 162, 60–68. doi:10.1016/J.IJPSYCHO.2021.01.015.

Winn, M.B., Wendt, D., Koelewijn, T., Kuchinsky, S.E., 2018. Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. Trends in Hearing. SAGE Publications Inc Vol. 22, p. 2331216518800869 doi:10.1177/2331216518800869.

Woody, A., Hooker, E.D., Zoccola, P.M., Dickerson, S.S., 2018. Social-evaluative threat, cognitive load, and the cortisol and cardiovascular stress response. Psychoneuroendocrinology 97, 149–155. doi:10.1016/j.psyneuen.2018.07.009.

Wright, R.A., 1996. Brehm's theory of motivation as a model of effort and cardiovascular response. In: The Psychology of Action: Linking Cognition and Motivation to Behavior, pp. 424–453 January *1996*.

Wu, Y.H., Stangl, E., Zhang, X., Perkins, J., Eilers, E., 2016. Psychometric functions of dual-task paradigms for measuring listening effort. Ear Hear. 37 (6), 660–670. doi:10.1097/AUD.0000000000000335.

Zekveld, A.A., Koelewijn, T., Kramer, S.E., 2018. The pupil dilation response to auditory stimuli: current state of knowledge. Trends in Hearing, 22. SAGE Publications Inc doi:10.1177/2331216518777174.

Zekveld, A.A., Kramer, S.E., 2014. Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. Psychophysiology 51 (3), 277–284. doi:10.1111/psyp.12151.

Zekveld, A.A., Kramer, S.E., Festen, J.M., 2010. Pupil response as an indication of effortful listening: the influence of sentence intelligibility. Ear Hear. 31 (4), 480–490. doi:10.1097/AUD.0b013e3181d4f251.

Zekveld, A.A., Kramer, S.E., Festen, J.M., 2011. Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. In: Ear and Hearing, 32. Ear Hear, pp. 498–510. doi:10.1097/AUD.0b013e31820512bb.

Zekveld, A.A., van Scheepen, J.A.M., Versfeld, N.J., Veerman, E.C.I., Kramer, S.E., 2019. Please try harder! The influence of hearing status and evaluative feedback during listening on the pupil dilation response, saliva-cortisol and saliva alpha-amylase levels. Hear. Res. 381, 107768. doi:10.1016/j.heares.2019.07.005.

Ziegler, M.G., 2012. Psychological stress and the autonomic nervous system. In: Primer on the Autonomic Nervous System. Elsevier Inc, pp. 291–293. doi:10.1016/B978-0-12-386525-0.00061-5.