Subjective and psychophysiological indices of listening effort in a competing-talker task

Carol L. Mackersie and **Heather Cones**
School of Speech, Language and Hearing Sciences, San Diego State University

Abstract

Background—The effects of noise and other competing backgrounds on speech recognition performance are well documented. There is less information, however, on listening effort and stress experienced by listeners during a speech recognition task that requires inhibition of competing sounds.

Purpose—The purpose was a) to determine if psychophysiological indices of listening effort were more sensitive than performance measures (percentage correct) obtained near ceiling level during a competing speech task b) to determine the relative sensitivity of four psychophysiological measures to changes in task demand and c) to determine the relationships between changes in psychophysiological measures and changes in subjective ratings of stress and workload.

Research Design—A repeated-measures experimental design was used to examine changes in performance, psychophysiological measures, and subjective ratings in response to increasing task demand.

Study Sample—Fifteen adults with normal hearing participated in the study. The mean age of the participants was 27 (range: 24–54).

Data Collection and Analysis—Psychophysiological recordings of heart rate, skin conductance, skin temperature, and electromyographic activity (EMG) were obtained during listening tasks of varying demand. Materials from the Dichotic Digits Test were used to modulate task demand. The three levels of tasks demand were: single digits presented to one ear (low-demand reference condition), single digits presented simultaneously to both ears (medium demand), and a series of two digits presented simultaneously to both ears (high demand). Participants were asked to repeat all the digits they heard while psychophysiological activity was recorded simultaneously. Subjective ratings of task load were obtained after each condition using the NASA-TLX questionnaire. Repeated-measures analyses of variance were completed for each measure using task demand and session as factors.

Results—Mean performance was higher than 96% for all listening tasks. There was no significant change in performance across listening conditions for any listener. There was, however, a significant increase in mean skin conductance and EMG activity as task demand increased. Heart rate and skin temperature did not change significantly. There was no strong association between subjective and psychophysiological measures, but all participants with mean

normalized effort ratings of greater than 4.5 (i.e. effort increased by a factor of at least 4.5) showed significant changes in skin conductance.

Conclusions—Even in the absence of substantial performance changes, listeners may experience changes in subjective and psychophysiological responses consistent with activation of a stress response. Skin conductance appears to be the most promising measure for evaluating individual changes in psychophysiological responses during listening tasks.

Key Words / terms (MESH)

speech perception; psychophysiology; electromyography; galvanic skin response; skin temperature; heart rate; physiological stress reactivity; listening effort

Numerous studies describe the effects of competing backgrounds on speech recognition performance. Less is known, however, about the non-performance aspects of the listener's experience during speech recognition tasks. It is possible, for example, that people experience communication-related stress and increased listening effort in some situations. In recent years, there has been growing interest in quantifying listening effort experienced during speech communication. In audiology, the measurement of listening effort may be useful in describing differences in the effects of hearing loss across the age span, characterizing the extra-perceptual effects of minimal hearing loss, and quantifying the subtle effects of sensory-aid processing.

Recent approaches aimed at capturing the dimension of listening effort include: 1) subjective ratings 2) measures of processing speed, and 3) cognitive-based paradigms. Subjective measures of listening effort have been used to document benefit from hearing aid processing (Baer et al, 1993; Bentler et al, 2008; Mackersie et al, 2009; Luts et al, 2010), and benefit from adding a second hearing aid (Feurerstein, 1992; Noble and Gatehouse, 2006) or a second cochlear implant (Noble et al, 2008). There is also evidence that the addition of response-time measures may be more sensitive than performance measures alone, (Gatehouse and Gordon, 1990; Baer et al., 1993; Mackersie et al, 1999a; Apoux et al, 2001) but interpretation of individual data may be complicated by time-order effects (Mackersie et al, 1999b; Larsby et al, 2008). For example, Mackersie et al. (1999b) reported systematic changes in response time (gradually increasing or decreasing) across a test session for some participants. Cognitive-based paradigms based on the limited-capacity resource model (Kahneman, 1973) have gained resurgent popularity among audiology and hearing science researchers. Numerous studies from both the cognitive psychology and audiology/hearing science literature have shown that auditory degradation can deplete information processing resources during listening tasks as reflected by performance decrements on a secondary task (e.g., working memory). Deterioration in secondary task performance is often accompanied by little or no change in a primary auditory task (e.g., recognition). Reviews of this work can be found in several recently published sources (Pichora-Fuller and Singh, 2006; Arlinger et al, 2009; Lunner et al, 2009).

It is feasible that a reaction that leads to increased listening effort may also be experienced as stress by the listener. The human stress response includes activation of the autonomic nervous system (ANS) and endocrine system and may or may not be accompanied by an emotional reaction to the stressor (Stokes and Kite, 2001; Staal, 2004). Activation of the autonomic nervous system from exposure to physical, mental, or emotional stressors typically results in increased activity in the sympathetic branch of the ANS accompanied by decreased activity in the parasympathetic branch (Criswell, 1995; Staal, 2004). Arousal of the sympathetic nervous system may result in measureable physiologic changes including changes in respiration rate, cardiac activity (e.g. heart rate, blood pressure), skin conductance, electromyography (muscle tension), skin temperature, and pupil diameter

(Andreassi, 2007). Changes in the balance of sympathetic and parasympathetic activity may appear as changes in heart-rate variability (Berntson et al, 1997).

So far, the study of psychophysiological responses to task difficulty have mostly dealt with the effects of task difficulty during cognitive, linguistic, and computational tasks (e.g.: (Coles, 1974; Cacioppo and Sandman, 1978; Jorna, 1992; Boutcher and Boutcher, 2006)). Changes in cardiac measures and skin conductance are commonly observed when the mental demands of tasks are increased (Kahneman et al, 1969; Clements and Turpin, 1995; Fournier et al, 1999; Miyake, 2001; Wilson and Russell, 2003; Richter et al, 2008). Changes in respiration, electromyography, EEG activity, and skin temperature, however, have been less consistent (Rogers and Elder, 1981; O'Gorman and Lloyd, 1988; Cohen et al, 1992; Hanson et al, 1993; Backs and Seljos, 1994; Veltman and Gaillard, 1998; Fournier et al., 1999; Wilson and Russell, 2003).

Psychophysiological measures of autonomic nervous system activity in response to auditory tasks may provide a means of quantifying the effects of listening effort. A small number of studies conducted using non-speech stimuli have documented cardiac changes during auditory detection and discrimination tasks (Light and Obrist, 1983; Cohen et al., 1992; Van Der Molen et al, 1996). Changes in psychophysiological measures were accompanied by changes in the percentage of correct responses.

One of the few auditory studies using speech stimuli was conducted by O'Gorman and Lloyd (1988). These researchers examined changes in electroencephalographic activity (EEG), skin conductance, and cardiac activity using a dichotic speech recognition task. The primary objective was to determine if group differences in baseline skin conductance would correspond to differences in reactivity and performance during the task. Participants were asked to repeat triplets of word-pairs presented to each ear simultaneously. Listeners with higher baseline skin conductance (categorized as "labiles") had lower accuracy scores and were more reactive during the test (i.e., greater physiological changes were observed) than listeners with lower baseline activity ("stabiles"). This pattern was observed for the skin conductance and cardiac measures, but not for the EEG measures. The authors did not evaluate whether significant changes from baseline were present. Therefore, the correspondence between auditory task difficulty and psychophysiological changes was not established.

More recently, changes in pupil diameter have been measured in responses to speech in noise (Kramer et al, 1997; Zekveld et al, 2010). Predictably, as the noise level increased, performance decreased. The decrease in performance was accompanied by a systematic increase in pupil diameter, suggesting increased listening effort.

Most of the psychophysiological research on task demand has included tasks in which performance varies across conditions. We chose to focus on performance near ceiling level in order to answer the question: Will listeners experience psychophysiological and subjective stress even when their performance scores suggest that they have little difficulty with the task? Our interest was in documenting the potential physiological and emotional costs of maintaining maximum-level performance across listening tasks of varying demand.

The specific objectives of the current study were:

    a. to determine if psychophysiological indicators of autonomic nervous system activity were more sensitive than performance measures (percentage correct) obtained near ceiling level during a competing speech task

    b. to determine the relative sensitivity of four psychophysiological measures (heart rate, skin conductance, skin temperature, EMG) to changes in task demand and

c. to determine the relations between psychophysiological measures and subjective measures of effort and stress.

METHOD

Participants

The required sample size was estimated from a power analysis of the skin conductance changes during listening tasks using preliminary data for eight participants. The power analysis indicated that a sample size of 12 was needed to reach a power of at least 0.80 (actual power estimate = 0.82).

Based on the power analysis, 15 adults were recruited for the experiment (2 men, 13 women). The mean age was 27 years (range: 24–54 years). All participants had pure-tone thresholds of 25 dB HL or lower for frequencies between 250 and 6000 Hz.

Measures

Physiological Measures—The Nexus-10 physiological recording system was used to record heart rate (calculated from a blood volume pulse measure), electromyographic responses (EMG), skin conductance, and skin temperature. Simultaneous multi-channel recordings were obtained throughout the test sessions.

Blood volume pulse was measured using a photoplethysmograph (NX-BVP1A) placed on one index finger (on the hand opposite of the writing hand). This device emits infra-red light through a sensor attached to the finger and measures the amount of infra-red light reflected from the surface of the skin. The peak of blood volume pulse occurs with each heartbeat. In the current study, heart rate was extracted from the blood volume pulse measure and used in the analyses. Therefore, the term "heart rate" (rather than "blood volume pulse") will be used to refer to the measure of interest in this study. Heart rate was calculated in beats per minute based on the pulse rate relative to blood flow through the arteries and blood vessels sampled at 32 samples per second.

Electromyographic activity (EMG) was recorded from the frontalis muscle (forehead area) on both sides of the face using a bipolar recording montage. Three Nexus-10 snap cable surface electrodes (NX-EXG2) were placed on the forehead. The positive and negative electrodes were placed above the right and left eyes and a ground electrode was placed in the middle of the forehead in line with the other electrodes. The digital EMG bandpass filter was set to 20–500 Hz.

Skin conductance measured in micro-Siemens (μS) was recorded from two surface electrodes (NX-GSR1A) placed approximately two inches apart, on the palm of one hand over the thenar and hypothenar muscles. Skin conductance reflects the activation of the eccrine sweat glands, and is a relative measure of the moisture at the level of the skin.

Skin temperature was recorded in degrees Fahrenheit using a temperature sensor (NX-TMP1A) taped to the palm-side middle finger of one hand using medical self-adhesive tape. The temperature sensor is capable of detecting changes of up to $1/1000^{th}$ degree Fahrenheit over a range of 50–104° F. Typically skin temperature is lower than the body's actual internal temperature, by 5–6° F (LenHardt and Sessler, 2006). The skin conductance and skin temperature sensors were also attached to the hand opposite the writing hand.

Speech Recognition Measures—Materials from the Dichotic Digits Test (Musiek, 1983) were chosen for the competing-speech task in order to systematically increase task demand without the influence of masking. The goal was to examine listening effort under

conditions that reflect near-ceiling-level performance. Three conditions were used to vary the auditory task demand. The "low demand" task required repetition of digits (1–10) presented to one ear; this task was used as the reference condition. The "medium-demand" task was a single dichotic digit task in which a different digit was presented to each ear simultaneously. Participants were asked to repeat both digits and two responses were scored on each trial. This task required participants to switch (or divide) attention between the competing speech and retain the information in memory long enough to repeat the digits. The "high-demand" task consisted of pairs of unique digits presented simultaneously to each ear. That is, two digits were presented to one ear while two different digits were presented to the opposite ear and four responses were scored on each trial. All digits were spoken by the same male talker. For each condition, scores were expressed as the percentage of words repeated correctly.

Standard psychophysiological stress protocol—An abbreviated version of the "Long Stress Test" within the Bio Trace+ software (Mind Media, 2006) was administered to determine if a standard psychophysiology stress protocol could be used to predict reactivity during speech recognition tasks. This type of protocol is commonly used as a baseline measure for biofeedback treatment (Schwartz and Andraski, 2003). The protocol consisted of a 2-minute baseline measure, a three-minute visual Stroop Test, a 2-minute rest period, a three-minute math stressor, and a post-test 2-minute rest period. Although the original Long Stress Test protocol also included an expressive stressor in which individuals are asked to talk about a stressful event, it was omitted for this experiment.

During the Stroop Test, the text showing the names of colors was displayed on a 17-inch wide flat-screen computer monitor situated approximately two feet in front of and six inches below the eye-line of the participants. The color of the text differed from the name. Five items were displayed at a time. A new set of five items was displayed every five seconds. Participants were asked to repeat the color of the text for each item, rather than the word itself. For example, the correct response for the word "PURPLE" written in green font would be "green."

During the math stressor segment, participants were asked to count backwards aloud, starting from 1081, in increments of 7. If participants answered incorrectly, they were instructed to start over. Physiologic measures were monitored throughout the stress protocol, but performance (accuracy) was not recorded

Subjective measures – NASA Task Load Index (NASA-TLX)—The NASA-TLX (Hart and Staveland, 1988) is a visual-analog rating scale that assesses the participant's perception of both the task demand and their performance. The rating scales include the following categories and questions: Mental Demand ("How mentally demanding was the task?"), Physical Demand ("How physically demanding was the task?"), Temporal Demand ("How hurried or rushed was the pace of the task?"), Performance ("How successful were you in accomplishing what you were asked to do?"), Effort ("How hard did you have to work to accomplish your level of performance?"), and Frustration ("How insecure, stressed, and annoyed were you ?"). The Frustration rating can be considered a rating of the emotional component of stress. The categorical anchors were "very low" and "very high" at the extremes of the visual-analog scale for all ratings except Performance. The categorical anchors for the Performance scale were "perfect" (low values) and "failure" (high values). We will use the label "Perceived Error" in our presentation of the Performance scale ratings to emphasize that higher values mean higher perceived failure of performance rather than successful performance. Ratings were obtained using the pencil-paper visual-analog version and were later converted to a ten-point scale with "0" indicating the lowest demand and "10"

indicating the highest demand. The dimensions Physical Demand and Temporal Demand were not relevant to the task used and therefore, were not reported.

Procedures

Testing was completed in a quiet room over two experimental sessions. The ambient noise level was 35 dBA. The standard psychophysiological stress profile was administered in the first session before administering the listening tasks. The second session was identical to the first except the standard stress profile was not administered. Participants were seated in a comfortable chair facing away from the tester to minimize visual distractions and reactivity based on visual contact.

An otoscopic examination and hearing screening was performed at the beginning of the first session. The standard psychophysiological stress profile was completed before any experimental auditory testing began.

Digitized speech stimuli were presented from the computer soundcard, routed to the speech channels of an Interacoustics AC40 Clinical Audiometer and delivered to a Telephonics TDH-39P headphone at 50 dB HL (70 dB SPL). Stimuli were presented using custom-written software. Participants were asked to repeat all of the digits that they heard. Recognition scores were expressed as the percentage of word repeated correctly.

Three-minute baseline physiological recordings were obtained before administering the speech recognition tests. During this baseline period, participants were instructed to sit quietly and relax. After each listening condition, the physiologic recordings were paused while the NASA-TLX was administered. Following administration of the NASA-TLX, physiologic measures were recorded for a one-minute recovery period before beginning the next listening condition. During each test session, one 20-item list was administered under each listening condition. Three-minute post-baseline physiologic measures were recorded at the end of each test session. The order of the low-, medium -, and high-demand tasks was counterbalanced among the 15 participants using a Latin-Squares design.

RESULTS

Recognition Scores

The mean recognition scores and standard deviations are shown in Table 1. For the medium- and high-demand (dichotic) tasks, data were collapsed across ears because mean scores for the right and left ears were within 1.6 percentage points.

As expected, mean recognition scores were close to 100%, consistent with performance at or near ceiling levels. Scores for the low-, medium-, and high-demand tasks were within four percentage points of one another. It was not practical to analyze the data using conventional statistical methods because there was no variance in one condition (low-demand, Session 2) and because data were not normally distributed.

Individual data were analyzed using the 95% confidence intervals based on the normal approximation to the binomial distribution. There were no significant differences between any scores for any participant.

Subjective Ratings

Mean ratings of mental demand, effort, frustration, and perceived performance are shown in Figure 1. Mean performance scores, collapsed across the two sessions are shown for reference. Recall that higher numerical ratings correspond to higher perceived mental demand, effort, frustration and perceived error. It can be seen that ratings increased

systematically as the task demand increased with the largest changes observed for "mental demand" and "effort".

Separate repeated-measures analyses of variance were completed for each rating category using session and task demand as factors. The results are summarized in Table 2. There was a significant main effect of task demand for all ratings. Although not shown in Figure 1, there was also a significant main effect of session for three of the four categorical ratings reflecting lower perceived mental demand, effort, and frustration in session 2 than in session 1. There was no effect of session, however, for ratings of perceived error. There was no significant interaction between session and task demand for any rating category, suggesting that the effects of task demand were similar across the two sessions.

Psychophysiological Measures

The effects of task demand on psychophysiological measures are shown in Figure 2. A monotonic increase in mean EMG activity, skin conductance, and heart rate was apparent with increasing task demand. Little change in skin temperature was observed, however. The largest changes for EMG activity and heart rate were observed when the task demand changed from medium to high. In contrast, the largest changes for skin conductance occurred as task demand increased from low to medium.

Separate repeated-measures analyses of variance were completed for each measure using test session and task demand as factors. The results are summarized in Table 3. There was a significant main effect of task demand for EMG and skin conductance measures, but not for skin temperature or heart rate measures. Unlike the subjective ratings, there was no significant effect of session for any measure. Newman-Keuls post-hoc tests confirmed a significant increase in EMG activity when the task demand increased from medium to high, but there was no significant difference between EMG activity for the low- and medium-demand tasks. Post-hoc tests also confirmed a significant increase in skin conductance when the task demand increased from low to medium and from low to high, but there was no significant change between the medium-and high- demand tasks.

For the tasks used in this study, EMG and skin conductance appear to be most sensitive to changes in task demand.

Individual Differences in Physiological Reactivity

The group analyses suggested that at least two physiologic measures were sensitive to changes in task demand for the tasks examined in this study. Examination of individual data suggested that participants varied in their pattern of autonomic nervous system reactivity. That is, some participants showed changes in some measures, but not other measures, with varying patterns among the participants. In order to examine the potential usefulness of these measures, individual z-scores were calculated to determine the number of significant differences for each participant and for each psychophysiological measure. The individual means and standard deviations were calculated within the statistical analysis module of the Bio Trace software. Individual z-scores were then calculated for the medium and high demand tasks using the low-demand task as a reference. Any z-score greater than 2.0 was tallied as a significant change from the low-demand condition.

Mean z-scores collapsed across session and task (medium and high-demand) were significant for 60% of the participants for skin conductance, and 53% of participants for skin temperature. An additional three participants (20%) showed a mean temperature increase rather than decrease. No participant had significant mean changes in EMG or heart rate when averaged across session and task. Based on these individual analyses, skin

conductance appears to hold the greatest promise for detection of changes for individual listeners.

Relationship between skin conductance and subjective measures

The relationship between skin conductance and subjective measures was determined using the individual z-scores from the physiologic measures and the mean normalized change in subjective ratings. Skin conductance was used in the analysis because it appeared to be the most sensitive measure in both the group and individual analyses. The normalized change was calculated by subtracting the rating for the medium- and high-demand tasks from the rating for the low-demand task and dividing this value by the rating for the low-demand task. Pearson Product-Moment correlation coefficients between mean z-scores and normalized ratings (averaged across session and task) were calculated for each physiologic measure and rating dimension.

Figure 3 shows the relationship between normalized changes in perceived effort and skin conductance for all participants (top panel) and with a single outlier removed (bottom panel). The circles denote participants who had mean skin conductance z-scores of greater than 2.0 or less than −2.0 (significant change). The hatched squares indicate participants whose changes in skin conductance were not significant. As shown in the top panel, one participant was a clear outlier in that she showed a highly significant change in skin conductance accompanied by a more modest change (doubling) in perceived effort. It is important to note that the skin conductance standard deviation for this participant was very low, which resulted in a very high z-score for this measure. When the outlier was excluded (bottom), there was a modest association between perceived effort and skin conductance (r $(13) = 0.67$, $p < .05$), accounting for approximately 45% of the variance.

The correlation coefficients with and without the outlier included are shown in Table 4. There were no significant correlations between skin conductance and perceived failure, effort, or frustration.

Based on these findings, the relations between physiologic changes in skin conductance and subjective ratings of task demand appear to be weak, at best. As shown in Figure 3, however, all four participants with mean normalized effort ratings of greater than 4.5 showed significant changes in skin conductance (z-scores > 2.0), whereas only three of eleven participants with normalized effort ratings of 4.5 or lower had significant changes in skin conductance.

Relationship between the psychophysiological stress protocol and listening tasks

A standard psychophysiological stress protocol was included to determine if individual reactivity to the stress protocol corresponded to reactivity on the listening tasks. That is, is it possible to predict who will show significant changes in psychophysiological activity during listening tasks from information obtained from a standard stress protocol?

For each psychophysiological measure, the numbers of individuals with and without significant changes were tallied. The tallies were based on the individual z-scores calculated for the listening tasks and stress protocol. Significant changes that were in the opposite direction than is considered consistent with a stress response (e.g. temperature increase, rather than decrease) were tallied as "no significant change". The tallies were entered into a 2×2 contingency tables and analyzed using the Fisher's Exact Test. The breakdowns included the numbers of participants with: a) no significant change for either test b) significant changes for 1–2 conditions of both tests (e.g. math or Stroop test and medium or high-demand listening) c) significant changes for 1–2 conditions of the stress protocol, but not the listening tasks and d) significant changes for 1–2 conditions of the listening tasks,

but not the stress protocol. The frequency tables and Fisher's Exact Test results are summarized in Table 5. A significant number of participants (12/15) who showed significant changes in skin conductance during the stress protocol also showed significant changes in skin conductance during the listening tasks. There were no other significant associations.

DISCUSSION

A significant increase in group mean EMG activity and skin conductance was observed with increasing task demand, although there was little to no effect on performance. These findings are consistent with systematic arousal of the sympathetic nervous system corresponding to increased listening effort and task engagement. Significant changes in mean heart rate and temperature were not observed. Although a number of individual participants showed significant changes in skin temperature during listening tasks, these changes were not always consistent with the expected decrease associated with vasoconstriction and increased stress. Therefore, it is not clear whether skin temperature would be a reliable indicator of sympathetic nervous system arousal during listening tasks.

Psychophysiological changes were accompanied by systematic changes in subjective ratings of mental demand, effort, perceived error in performance, and frustration/annoyance/stress. The systematic increase in these measures is consistent with increased subjective stress. Increased ratings of subjective stress are consistent with a recent study in which increased state anxiety was reported during a dichotic listening task (Roup and Chiasson, 2010).

Mean ratings of mental demand, effort, and frustration were lower in session two than in session one, suggesting that the emotional reactions to the tasks may have decreased. There were, however no significant changes in psychophysiological measures across the two sessions or in perceived error in performance. The decrease in subjective indicators may be attributable to increased comfort in the experimental setting, decreased uncertainty (and perhaps decreased anxiety) regarding what was expected of them, and/or increased familiarity with the task. It is also possible that the self-awareness of the costs of task demands decreased across sessions, but the true physiologic costs in terms of the physiologic stress response did not change. The stability in the perceived error rating and psychophysiological measures suggests that the tasks remained demanding despite the changes in the participants' subjective reactions.

The significant changes in mean skin conductance and EMG with increased task demand provide evidence that psychophysiological measures may be useful indicators of listening effort and possibly, communication-related stress. In clinical settings, however, the primary interest is in changes within an individual rather than changes in mean data. Overall, 60 % of participants showed significant changes in skin conductance as task demand increased. In contrast, despite significant changes in mean EMG with increasing task demand, there were no individuals who showed statistically significant changes, primarily because of the variability of the measure within individuals. Based on these findings, skin conductance appears to be more sensitive than EMG measures to physiologic changes in reactivity that occur with changes in listening task demand.

The magnitude of skin conductance changes with increasing task demand were not strongly associated with changes in subjective ratings. A general disassociation between subjective and physiologic indicators of stress has been reported by other investigators (Miyake, 2001; Wilson and Sasse, 2001) and suggests that subjective and physiologic measures provide independent sources of information. Despite the generally weak association between skin conductance and subjective measures, all participants in the current study with mean normalized effort ratings greater than 4.5 also showed significant mean changes in skin

conductance. Further research is needed to determine if effort ratings greater than 4.5 can be used to predict who will show significant psychophysiological changes during other types of listening tasks.

Reactivity on a standard stress protocol used in biofeedback assessment (visual Stroop Test and math stressor) was associated with reactivity on the listening tasks, but only for the skin conductance measure. This finding suggests that a standard psychophysiological pre-test may useful in identifying people who are most likely to show sympathetic nervous system arousal during listening tasks.

The task and stimuli used in the current study were chosen to target ceiling-level performance to enable us to examine psychophysiological reactivity with little or no change in performance. The competing speech task used in the present study was ideally suited for this purpose because it minimized the effects of energetic masking while imposing a moderate load on attentional processes and working memory.

In contrast to the relatively simple acoustic demands of the current study, the auditory scene of real-world environments is often exquisitely complex with moment-to-moment, often unpredictable, changes in the acoustic signal. Moreover, in many environments, background noise is present. Further work is needed to determine if physiologic changes observed in the present study would be observed during speech recognition tasks that involve greater acoustic and linguistic complexities.

Significant psychophysiological reactivity and subjective stress with increasing task demand in the face of near-perfect performance indicates that focused listening may have physiological, cognitive, and emotional costs that are not reflected in performance scores. This stress reactivity varies substantially from individual to individual. Given the negative long-term effects of sustained or repeated stress on health (Lovallo, 2005), further consideration of stress reactivity in our clinical populations is warranted.

Acknowledgments

Abbreviations

EMG	electromyography
ANOVA	analysis of variance
NASA-TLX	National Aeronautics and Space Administration Task Load Index

References

Andreassi, JL. Psychophysiology: Human behavior and physiological response. 5th edition. Mahwah, NJ: Lawrence Erlbaum; 2007.

Apoux F, Crouzet O, Lorenzi C. Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: effects on identification performance and response times. Hear Res. 2001; 153:123–131. [PubMed: 11223303]

Arlinger S, Lunner T, Lyxell B, Pichora-Fuller MK. The emergence of cognitive hearing science. Scand J Psychol. 2009; 50:371–384. [PubMed: 19778385]

Backs RW, Seljos KA. Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. International Journal of Psychophysiology. 1994; 16:57–68. [PubMed: 8206805]

Baer T, Moore BCJ, Gatehouse S. Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairement: effects on intelligibility quality, and response time. J Rehabil Res Dev. 1993; 30:49–72. [PubMed: 8263829]

Bentler R, Wu Y-H, Kettel J, Hurtig R. Digital noise reduction: outcomes from laboratory and field studies. Int J Audiol. 2008; 47:447–460. [PubMed: 18698521]

Berntson GG, Bigger JT Jr, Eckberg DL, Grossman P, Kaufmann PG, Malik M, Nagaraja HN, Porges SW, Saul JP, Stone PH, van der Molen MW. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology. 1997; 34:623–648. [PubMed: 9401419]

Boutcher YN, Boutcher SH. Cardiovascular response to Stroop: Effect of verbal response and task difficulty. Biological Psychology. 2006; 73:235–241. [PubMed: 16730405]

Cacioppo JT, Sandman CA. Physiological differentiation of sensory and cognitive tasks as a function of warning, processing demands, and reported unpleasantness. Biological Psychology. 1978; 6:181–192. [PubMed: 667241]

Clements K, Turpin G. Effects of feedback and task difficulty on electrodermal activity and heart rate: An examination of Fowles' three arousal model. Journal of Psychophysiology. 1995; 9:231–242.

Cohen BH, Davidson RJ, Senulis JA, Saron CD, Weisman DR. Muscle tension patterns during auditory attention. Biological Psychology. 1992; 33:133–156. [PubMed: 1525291]

Coles MG. Physiological activity and detection: The effects of attentional requirements and the prediction of performance. Biological Psychology. 1974; 2:113–125. [PubMed: 4447813]

Criswell, E. Biofeedback and Somatics. Novato, CA: Freeperson Press; 1995.

Feurerstein JF. Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort. Ear Hear. 1992; 13:80–86. [PubMed: 1601196]

Fournier LR, Wilson GF, Swain CR. Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: Manipulations of task difficulty and training. International Journal of Psychophysiology. 1999; 31:129–145. [PubMed: 9987059]

Gatehouse S, Gordon J. Response times to speech stimuli as measures of benefit from amplification. British Journal of Audiology. 1990; 24:63–68. [PubMed: 2317602]

Hanson EKS, Schellekens JMH, Veldman JBP, Mulder LJM. Psychomotor and cardiovascular consequences of mental effort and noise. Human Movement Science. 1993; 12:607–626.

Hart, S.; Staveland, L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.; Meshkat, N., editors. Human Mental Workload. North-Holland Elsevier Science; 1988. p. 139-183.

Jorna PG. Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. Biological Psychology. 1992; 34:237–257. [PubMed: 1467395]

Kahneman, D. Attention and Effort. Englewood Cliffs NJ: Prentice Hall; 1973.

Kahneman D, Tursky B, Shapiro D, Crider A. Pupillary, heart rate, and skin resistance changes during a mental task. Journal of Experimental Psychology. 1969; 79:164–167. [PubMed: 5785627]

Kramer SE, Kapteyn TS, Festen JM, Kuik DJ. Assessing aspect of hearing handicap by means of pupil dilation. Audiology. 1997; 36:155–164. [PubMed: 9193733]

Larsby B, Hällgren M, Lyxell B. The interference of different background noises on speech processing in elderly hearing impaired subjects. Int J Audiol. 2008; 47:S83–S90. [PubMed: 19012115]

LenHardt R, Sessler D. Estimation of mean body temperature from mean skin and core temperature. Anesthesiology. 2006; 105:1117–1121. [PubMed: 17122574]

Light KC, Obrist PA. Task difficulty, heart rate reactivity, and cardiovascular responses to an appetitive reaction time task. Psychophysiology. 1983; 20:301–312. [PubMed: 6867223]

Lovallo, WR. Stress & Health. 2nd Edition. Thousand Oaks: Sage Publications; 2005.

Lunner T, Rudner M, Ronnberg J. Cognition and hearing aids. Scand J Psychol. 2009; 50:395–403. [PubMed: 19778387]

Luts H, Eneman K, Wouters J, Schulte M, Vormann M, Buechler M, Dillier N, Houben R, Dreschler WA, Froehlich M, Puder H, Grimm G, Hohmann V, Leijon A, Lombard A, Mauler D, Spriet A.

Multicenter evaluation of signal enhancement algorithms for hearing aids. J Acoust Soc Am. 2010; 127:1491–1505. [PubMed: 20329849]

Mackersie CL, Neuman AC, Levitt H. A comparison of response time and word recognition meausre using a word-monitoring and closed-set identification task. Ear and Hearing. 1999a; 20:140–148. [PubMed: 10229515]

Mackersie CL, Neuman AC, Levitt H. Response time and word recognition measures using a word-monitoring task: List equivalency and time-order effects. Ear Hear. 1999b; 20:515–520. [PubMed: 10613389]

Mackersie CL, Qi Y, Boothroyd A, Conrad N. Evaluation of cellular phone technology with digital hearing aid features: effects of encoding and individualized amplification. J Am Acad Audiol. 2009; 20:109–118. [PubMed: 19927674]

Miyake S. Multivariate workload evaluation combining physiological and subjective measures. International Journal of Psychophysiology. 2001; 40:233–238. [PubMed: 11228350]

Musiek F. Assessment of central auditory dysfunction: The dichotic digit test revisited. Ear Hear. 1983; 4:79–83. [PubMed: 6840415]

Noble W, Gatehouse S. Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ). Int J Audiol. 2006; 45:172–181. [PubMed: 16579492]

Noble W, Tyler R, Dunn C, Bhullar N. Unilateral and bilateral cochlear implants and the implant-plus-hearing-aid profile: comparing self-assessed and measured abilities. Int J Audiol. 2008; 47:505–514. [PubMed: 18608531]

O'Gorman J, Lloyd JE. Electrodermal lability and dichotic listening. Psychophysiology. 1988; 25:538–546. [PubMed: 3186882]

Pichora-Fuller MK, Singh G. Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. Trends Amplif. 2006; 10:29–59. [PubMed: 16528429]

Richter M, Friedrich A, Gendolla GHE. Task difficulty effects on cardiac activity. Psychophysiology. 2008; 45:869–875. [PubMed: 18665860]

Rogers RL, Elder ST. Immediate effects of repeated and non-repeated instructions and task difficulty on task, cardiovascular, and respiratory performance. Psychophysiology. 1981; 18:534–539. [PubMed: 7280151]

Roup CM, Chiasson KE. Effect of dichotic listening on self-reported state anxiety. Int J Audiol. 2010; 49:88–94. [PubMed: 20151882]

Schwartz, MA.; Andraski, F. Biofeedback: A Practioner's Guide. New York: Guilford Press; 2003.

Staal, MA. Stress, cognition, and human performance: A literature review and conceptual framework. Moffett Field, CA: NASA Ames Research Center; 2004.

Stokes, AF.; Kite, K. On grasping a nettle and becoming emotional. In: Hancock, PA.; Desmond, PA., editors. Stress, workload, and fatigue. Mahwah, NJ: L. Erlbaum; 2001.

Van Der Molen MW, Somson R, Jennings JR. Does the heart know what the ear hears? Psychophysiology. 1996; 33:547–554. [PubMed: 8854742]

Veltman JA, Gaillard AWK. Physiological workload reactions to increasing levels of task difficulty. Ergonomics. 1998; 41:656–669. [PubMed: 9613226]

Wilson GF, Russell CA. Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. Human Factors. 2003; 45:635–643. [PubMed: 15055460]

Wilson, GM.; Sasse, MA. Straight from the Heart - Using Physiological Measurements in the Evaluation of Media Quality. Proceedings of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB) Convention 2001, Symposium on Emotion, Cognition and Affective Computing; York, UK. 2001. p. 63-73.

Zekveld AA, Kramer SE, Festen JM. Pupil response as an indication of effortful listening:the influence of sentence intelligibility. Ear Hear. 2010; 31:480–490. [PubMed: 20588118]

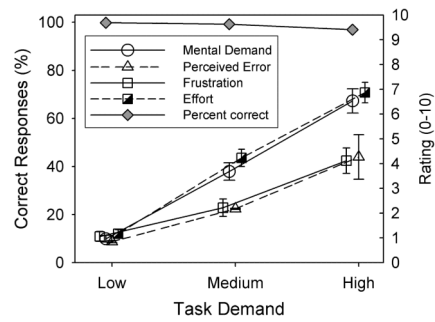**Figure 1.**
Mean NASA-TLX ratings of mental demand, effort, perceived error, and frustration/stress for the low-, medium-, and high-demand tasks. Recognition scores are also shown. The error bars denote ± 1 standard error.
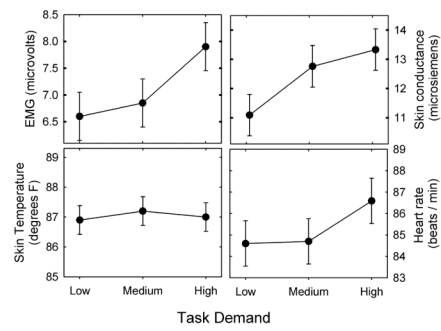
Figure 2.
Mean EMG, skin conductance, skin temperature, and heart rate measures for the low-, medium-, and high-demand tasks. The error bars denote ± 1 standard error.

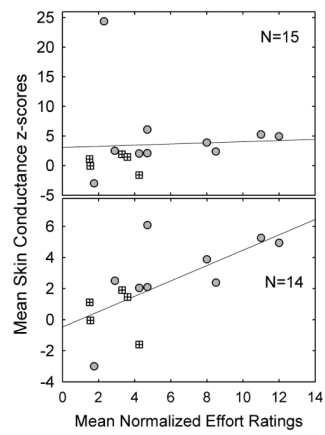**Figure 3.**
Relationship between mean normalized effort ratings and skin conductance for all participants (top panel) and for 14 of the 15 participants (outlier excluded). The circles denote participants who had mean skin conductance z-scores of greater than 2.0 or less than −2.0 (significant change). The hatched squares indicate participants whose changes in skin conductance were not significant.

Table 1

Group mean word recognition scores with standard deviations for three levels of task demand.

	Demand	Mean	SD
Session 1	Low	99.7	1.3
	Medium	99.3	2.0
	High	96.5	3.8
Session 2	Low	100	0.0
	Medium	99.3	1.5
	High	97.8	2.1

Table 2

Repeated-measures analyses-of-variance results for subjective measures of task load (NASA-TLX). The significance levels of statistically significant factors are shown in bold.

Mental Demand	df	F	p	η^2_p
Session	(1,14)	4.67	**0.048**	0.25
Task Demand	(2,28)	96.52	**< .0001**	0.87
Session × Task Dem	(2,28)	1.77	0.19	0.11
Effort	df	F	p	η^2_p
Session	(1,14)	5.45	**0.03**	0.28
Task Demand	(2,28)	125.24	**< .0001**	0.90
Session × Task Dem	(2,28)	2.49	0.10	0.15
Perceived Performance	df	F	p	η^2_p
Session	(1,14)	3.2	0.09	0.19
Task Demand	(2,28)	47.63	**< .0001**	0.77
Session × Task Dem	(2,28)	0.50	0.61	0.03
Frustration	df	F	p	η^2_p
Session	(1,14)	7.09	**0.02**	0.34
Task Demand	(2,28)	25.97	**< .0001**	0.65
Session × Task Dem	(2,28)	2.29	0.12	0.14

Table 3

Repeated-measures analyses-of-variance results for psychophysiological measures. The significance levels of statistically significant factors are shown in bold.

EMG	df	F	p	η^2_p
Session	(1,14)	0.29	0.59	0.02
Task Demand	(2,28)	4.75	**0.02**	0.25
Session × Task Dem	(2,28)	0.01	0.91	0.01
Skin conductance	df	F	p	η^2_p
Session	(1,14)	0.40	0.54	0.03
Task Demand	(2,28)	5.40	**0.01**	0.28
Session × Task Dem	(2,28)	0.93	0.41	0.06
Skin temperature	df	F	p	η^2_p
Session	(1,14)	3.78	0.07	0.21
Task Demand	(2,28)	0.14	0.87	0.01
Session × Task Dem	(2,28)	1.74	0.19	0.11
Heart rate	df	F	p	η^2_p
Session	(1,14)	0.93	0.35	0.06
Task Demand	(2,28)	2.13	0.14	0.14
Session × Task Dem	(2,28)	1.64	0.12	0.2

Table 4

Correlation coefficients for associations between mean changes in skin conductance (z-scores) and NASA TLX task ratings: Mental Demand, Perceived failure (Performance), Effort, and Frustration. The second and third columns show correlations coefficients with and without the outlier included.

	All (n=15)	Excluding outlier (S05)
Mental demand	−0.15	0.15
Perceived failure	−0.25	0.05
Effort	0.05	**0.67**[*]
Frustration	−0.33	0.29

The asterisk and bold font indicate a statistically significant finding.

Table 5

Frequency tables and summary of the Fisher's Exact Test (FET) analyses of the number individuals with and without significant changes in at least one condition for the standard stress protocol or the listening tasks. The row/column labels "Listening 1–2" and "Stress Protocol 1–2" are the number of individuals who showed one or two significant changes; the labels "Stress Protocol 0" and "Listening 0" denote the number of individuals who did not have any significant changes.

	Stress Protocol 0	Stress Protocol 1–2	FET *p*
Skin Conductance			
Listening - 0	2	0	< .03*
Listening 1–2	1	12	
Heart Rate			
Listening - 0	6	6	0.23
Listening 1–2	0	3	
Temperature			
Listening - 0	0	2	0.14
Listening 1–2	9	4	
EMG			
Listening - 0	2	11	0.74
Listening 1–2	0	2	