Perceived listening effort for a tonal task with contralateral competing signals

William J. Bologna^{a)}

*Department of Hearing and Speech Sciences, University of Maryland, College Park,
Maryland 20742
bolognaw@umd.edu*

Monita Chatterjee

*Auditory Prostheses and Perception Laboratory, Boys Town National Research Hospital,
555 North 30th Street, Omaha, Nebraska 68131
monita.chatterjee@boystown.org*

Judy R. Dubno

*Department of Otolaryngology-Head and Neck Surgery, Medical University of South
Carolina, 135 Rutledge Avenue, MSC 550, Charleston, South Carolina 29425-5500
dubnoj@muscc.edu*

Abstract: Perceived listening effort was assessed for a monaural irregular-rhythm detection task while competing signals were presented to the contralateral ear. When speech was the competing signal, listeners reported greater listening effort compared to either contralateral steady-state noise or no competing signal. Behavioral thresholds for irregular-rhythm detection were unaffected by competing speech, indicating that listeners compensated for this competing signal with effortful listening. These results suggest that perceived listening effort may be associated with suppression of task-irrelevant information, even for conditions where informational masking and competition for linguistic processing resources would not be expected.

© 2013 Acoustical Society of America

PACS numbers: 43.71.Rt, 43.71.Sy, 43.66.Rq [AC]

Date Received: May 31, 2013 **Date Accepted:** August 21, 2013

1. Introduction

“Listening effort” refers to the additional allocation of cognitive resources to support performance of a challenging auditory task (Hicks and Tharpe, 2002). It has been suggested that these resources are attentional in nature (Kahneman, 1973), thus increased listening effort may indicate an increased role of attention in tasks such as listening in a complex environment. In some cases, listening effort can be more sensitive to the processing requirements of a task than behavioral performance. When listeners modulate attention to maintain a consistent level of performance with increasing task demands, this form of compensation can be measured as an increase in listening effort (Mackersie and Cones, 2011).

Listening effort can be assessed with physiologic and behavioral responses to changes in task demands. Objective methods such as pupillometry (e.g., Zekveld *et al.*, 2011; Kuchinsky *et al.*, 2013) and dual-task interference (e.g., Sarampalis *et al.*, 2009) provide means of measuring listening effort that are free from subjective bias, but infer changes in listening effort from a single measurement dimension. A more comprehensive assessment of listening effort may be obtained by evaluating multiple dimensions of listening effort, such as perceived difficulty, and perceived performance, and frustration.

Perceived listening effort is a multidimensional construct, and the relative contributions of each dimension to the subjective demands of a task are likely to vary between

^{a)}Author to whom correspondence should be addressed.

listeners. The NASA Task Load Index (TLX), a self-report measure, was originally designed to measure workload for user-control interfaces (Hart and Staveland, 1988). More recently, the NASA TLX has been adapted as a subjective measure of listening effort for speech perception in background noise (Mackersie and Cones, 2011). The standard questionnaire has six domains, which measure multiple dimensions of perceived workload, such as Mental Demand, Perceived Effort, Frustration, and Performance. Listeners interpret task demands according to their subjective experience, and each interpretation contributes to a multidimensional perspective of perceived listening effort.¹ However, it should be noted that listeners may be reflecting on their experience across several trials and conditions. Thus, differences in listeners' ability to retrospectively evaluate their experience may contribute increased variability to the data (Kuchinsky *et al.*, 2013).

Recent studies of speech perception have shown that listening effort increases when the target speech signal is masked by competing auditory signals (Zekveld *et al.*, 2011; Kuchinsky *et al.*, 2013; Koelewijn *et al.*, 2012a; Koelewijn *et al.*, 2012b). When portions of the speech stimulus are rendered inaudible by a masker, listeners expend greater degrees of listening effort for speech perception (Zekveld *et al.*, 2011; Kuchinsky *et al.*, 2013). When speech is masked by other talkers, informational masking also contributes to listening effort (Koelewijn *et al.*, 2012a; Koelewijn *et al.*, 2012b). It has been suggested that additional listening effort with informational masking could be associated with the perceptual segregation of two voices, or the inhibition of linguistic processing of the masker, or both (Koelewijn *et al.*, 2012a). In one study, individual differences in listening effort with competing speech were correlated with an independent measure of listeners' ability to inhibit irrelevant linguistic information (Koelewijn *et al.*, 2012b). This finding provided evidence that the additional listening effort with informational masking may reflect semantic interference, such that greater effort was expended by listeners to suppress processing of the irrelevant speech masker. However, it remains unclear to what extent this increase in listening effort relates to perceptual segregation of the two voices, and if linguistic competition between two voices is required for the observed increase in listening effort. The present study used a tonal irregular-rhythm detection task to determine whether task-irrelevant competing speech influences perceived listening effort in the absence of voice segregation and linguistic competition.

Irregular-rhythm detection is an objective measure of low-level segregation, in which listeners detect temporal perturbations in an otherwise isochronous sequence of alternating tones (Roberts *et al.*, 2002). This tone-based task does not require linguistic or semantic processing, and so it is unlikely to compete with irrelevant speech for linguistic processing resources. As such, this task is well suited to determine how suppression of irrelevant speech affects perceptual segregation and listening effort without competition for linguistic processing or informational masking.

The goals of the present investigation were to determine the effects of competing speech on perceived listening effort and low-level perceptual segregation, in the absence of informational masking. To this end, listeners completed a monaural irregular-rhythm detection task, and on subsets of trials, competing single-talker speech, steady-state speech-shaped noise, or no competing signal were separately presented to the contralateral ear. Subjective impressions of listening effort were assessed with the NASA TLX, and separate listening effort measures were obtained for each competing signal condition. We hypothesized that competing single-talker speech would be rated as requiring the greatest listening effort, as it is rich in task-irrelevant information and has been shown to require the greatest listening effort in speech perception tasks (e.g., Koelewijn *et al.*, 2012a; Koelewijn *et al.*, 2012b). Detection thresholds for irregular-rhythm were collected to determine if the costs associated with these contralateral competing signals would affect low-level perceptual segregation mechanisms.

2. Methods

2.1 Participants

Twenty participants (8 males, 12 females), 18–29 yr of age (mean age: 24.25 yr), with normal hearing [thresholds ≤ 20 dB hearing level (HL) at octave frequencies from 250

to 8000 Hz], took part in the experiment. The stimulus ear for each participant was chosen in an alternating fashion, and competing signals were always presented contralaterally to the stimulus. Each participant was paid upon completion of this study, as approved by the MUSC Institutional Review Board.

2.2 Irregular-rhythm detection

Irregular-rhythm detection is a two-alternative forced-choice task. The stimulus and task design were based on Roberts *et al.* (2002). The two listening intervals each consist of 12 repetitions of two, 60 ms pure tones (A and B), presented in an alternating sequence at a rate of 10 Hz (ABABAB). Both intervals begin with an isochronous rhythm, but the target interval's rhythm becomes progressively anisochronous (irregular) in the second half of the sequence. The irregular rhythm is created by delaying the position of the B tones, and longer delays create a more pronounced irregular rhythm. When A and B tones differ greatly in frequency, low-level segregation of the tonal sequence disrupts perception of the irregular rhythm, such that both intervals seem to maintain an isochronous rhythm throughout (for a review, see Roberts *et al.*, 2002). Each adaptive run started with the maximum delay of 40 ms, and converged on a delay value (in ms) corresponding to 70.7% for detection of the irregular rhythm (calculated from the last 4 of 6 reversal points). Participants were tested at five frequency separations (A = 1000 Hz; B = 4, 8, 10, 12, or 14 semitones above 1000 Hz); longer delay thresholds were expected for increasing frequency separations.

On subsets of trials, competing signals were presented to the contralateral ear. The competing signals were continuous speech (Audiobook recording of *The History of the Peloponnesian War*), and steady-state speech-shaped noise, shaped to match the long-term average spectrum of the competing speech signal. Participants were also tested on a baseline condition without a competing signal in the contralateral ear. Delay thresholds were obtained for each of the five frequency separations with each of the three competing signals (speech, noise, none), for a total of 15 conditions.

2.3 NASA Task Load Index

A modified version of the NASA TLX was implemented for the current investigation. This modified NASA TLX consisted of five domains (Mental Demand, Perceived Effort, Task Difficulty, Frustration, and Performance), each of which was rated on a 20-point visual-analog scale. The Performance domain prompted listeners to rate their own performance on the task (*"How successful were you in accomplishing what you were asked to do?"*). The domain of Task Difficulty was added to specifically address difficulty associated with the competing signals (*"How difficult was the task with the speech/noise/without any distracter?"*). Individual definitions of workload vary, and subjective assessments of task difficulty have been shown to closely and consistently correlate with objective workload on a variety of tasks (Hart and Staveland, 1988). Thus, the domain of Task Difficulty was added to supplement the other four domains of the NASA TLX and provide a multivariate assessment of perceived listening effort associated with the competing signals.

2.4 Procedures

Participants were seated in a sound attenuating booth in front of a computer monitor with a mouse. Each trial of the irregular-rhythm task consisted of two tonal sequences, one with the delay and one without, presented in random order. Participants were instructed to listen to the rhythm of the tonal sequences and use the mouse to select the sequence with the irregular rhythm (the delay). Feedback was provided on the screen following each trial. Participants were instructed to ignore the competing signals (when present) and to focus only on the tonal sequence task. Each participant completed a single practice session prior to data collection. Practice consisted of one run in each of the 15 conditions (not included in data analysis); the NASA TLX was not administered during practice.

Data collection consisted of three trials in each of the 15 conditions. Testing sessions were split into three blocks, each consisting of five runs (one run per frequency separation) with a single competing signal presented continuously to the contralateral ear. After each block, participants completed a NASA TLX based on their overall impressions of the task (across frequency separations) with that particular competing signal. Participants were told that the NASA TLX was a subjective assessment, and they were encouraged to interpret the prompts according to their own criteria. Three testing sessions were required for each participant to complete the three trials in each condition. The order of conditions was partially counterbalanced based on two separate Latin Squares, one for frequency separation and one for competing signals.

2.5 Apparatus

Tonal sequences for the irregular-rhythm task and contralateral competing signals were output on separate channels from a PC with a Cakewalk™ external sound card. The task was coded in Psycon,² and competing signals were presented in Adobe Audition (Version 1.5). Stimuli and competing signals were presented through left and right Sennheiser HDA200 headphones at 70 dB sound pressure level (SPL).

2.6 Data analysis

Individual subject thresholds for irregular-rhythm detection were calculated as the average of three runs of a given condition. Detection thresholds were analyzed with a two-way repeated-measures analysis of variance (ANOVA) with repeated measures of frequency separation and competing signal. Individual subject responses on the NASA TLX were calculated as the average of three ratings for a given competing signal. Although data from the NASA TLX are subjective and ordinal in nature, analysis with conventional parametric statistics was deemed appropriate, as the dataset was generally non-homogenous (i.e., few identical data points) and relatively free of skewness and kurtosis (e.g., Mackersie and Cones, 2011). Separate one-way ANOVAs were computed for each of the five domains with an alpha level of 0.01 (family-wise error rate of 0.05). Where significant main effects were observed, *post hoc* pairwise comparisons were calculated with a False Discovery Rate (FDR) correction for multiple comparisons.

3. Results

3.1 Detection thresholds

Delay thresholds as a function of frequency separation are displayed in Fig. 1; competing signal is the parameter. Mean thresholds are plotted on the far right. As expected, delay thresholds increased (poorer irregular-rhythm detection) with increasing frequency separation. Repeated-measures ANOVA revealed significant main effects of frequency separation [$F(4,76)=76.209$, $p<0.001$] and competing signal [$F(2,38)=3.994$, $p<0.05$], with no significant interaction [$F(8,152)=1.247$, $p=0.276$]. *Post hoc* pairwise comparisons between competing signals revealed significantly longer delay thresholds (poorer detection) with competing noise compared to no competing signal ($p<0.05$), but no significant difference between competing speech and no competing signal, or between speech and noise competing signals.

3.2 Perceived listening effort

Mean and distribution of responses on the NASA TLX for the three competing signal conditions are displayed in Fig. 2. Larger values indicate greater task demands and poorer performance. Separate ANOVAs revealed a main effect of competing signal for Mental Demand [$F(2,38)=10.571$, $p<0.001$], Perceived Effort [$F(2,38)=13.220$, $p<0.001$], and Task Difficulty [$F(2,38)=55.030$, $p<0.001$]; Frustration was significant, but at the 0.05 level [$F(2,38)=3.545$, $p=0.039$]. *Post hoc* pairwise comparisons between speech and no competing signal were significant for Mental Demand ($p<0.001$), Perceived Effort ($p<0.001$), and Task Difficulty ($p<0.001$). Pairwise

Fig. 1. Delay thresholds plotted as a function of frequency separation. Mean thresholds are plotted on the far right. Error bars represent standard error. The parameter is competing signal condition (none, single-talker speech, speech-shaped noise). Longer delay thresholds at larger frequency separations indicate poorer irregular-rhythm detection.

comparisons between speech and noise were significant for Mental Demand ($p < 0.01$), Perceived Effort³ ($p < 0.01$), and Task Difficulty ($p < 0.001$). Pairwise comparisons between noise and no competing signal were significant only for Task Difficulty ($p < 0.01$). The one-way ANOVA for Performance did not reach significance [$F(2,38) = 1.729$, $p = 0.191$], indicating that participants did not report differences in their performance across the three competing signals, despite reporting greater Mental Demand, Perceived Effort, and Task Difficulty with the competing speech compared to either competing noise or no competing signal.

4. Discussion

The present study minimized many of the factors associated with informational masking to determine whether suppression of task-irrelevant speech contributes to increased

Fig. 2. Subjective ratings of Mental Demand, Perceived Effort, Task Difficulty, Frustration, and Performance for each competing signal condition (none, single-talker speech, speech-shaped noise). Larger values indicate greater task demands and poorer performance. The ends of the whiskers indicate the 10th and 90th percentile, the boundaries of each box indicate the first and third quartile, and the line in the middle of each box indicates the median. Filled circles indicate the mean, and asterisks indicate significant *post hoc* pairwise comparisons.

listening effort. Here, the primary target was a tonal sequence, which was categorically different from the contralateral speech and noise. As such, listeners should have experienced no difficulty identifying the appropriate target for processing and segregating it from the competing signal. Additionally, the irregular-rhythm detection task did not require any semantic or linguistic processing, and so no degree of semantic interference between the primary signal and competing speech was expected. Despite the absence of segregation difficulty and of semantic interference, listeners reported greater perceived listening effort with competing speech compared to either competing noise or no competing signal. Based on these findings, it seems likely that the source of this additional listening effort relates to suppression of meaningful content in the competing speech.

An alternative explanation is that rhythmic fluctuations of the competing speech contributed to listening effort on the irregular-rhythm detection task. Studies by Koelewijn *et al.* (2012a) and Koelewijn *et al.* (2012b) found no difference in either subjective or objective listening effort for sentence recognition with a steady-state compared to fluctuating noise masker. However, in a similar study of speech perception, subjective ratings of effort were greater for modulated compared to steady-state noise (Rudner *et al.*, 2012). Thus, it is unclear to what extent temporal fluctuations in a competing signal affect perceived listening effort. Future investigations of perceived listening effort can test this hypothesis by comparing competing speech to competing noise with a dynamic temporal pattern.

Irregular-rhythm detection was unaffected by competing speech, indicating that this competing signal did not influence low-level segregation of the tonal sequences. However, listeners were poorer at the task with contralateral steady-state noise compared to their performance without a competing signal. It is unclear at this time why performance was reduced in this condition, but listening effort results suggest that the difference in thresholds is not attributable to differences in mental demand or perceived effort between contralateral noise and no competing signal. In addition, no interaction was observed between frequency separation and competing signal, suggesting that the effect of competing noise was independent of low-level segregation of the tonal sequences.

Previous work in the field of human factors has demonstrated that subjective reports by human operators can accurately reflect changes in the objective workload of a task (Moray, 1982; Gopher and Braune, 1984; Hart and Staveland, 1988). The difficulty with subjective assessment is that definitions of abstract concepts like “effort” and “mental demand” will naturally vary between listeners, as will their interpretations of how these factors contributed to the demands of a task (Hart and Staveland, 1988). This complicates the issue of subjectively assessing listening effort in a group of subjects and highlights the importance of using a multidimensional assessment tool. In the present study, competing speech affected perceptions of Mental Demand, Perceived Effort, and Task Difficulty (though not all listeners reported greater values in each domain). Overall, the results indicate that several dimensions of listening effort were affected by competing speech, which provides sufficient evidence to support the conclusion that competing speech increased perceived listening effort. In contrast, competing noise only influenced perceptions of Task Difficulty, and so the evidence that competing noise influenced perceived listening effort is less conclusive. Future investigations are likely to benefit from supplementing subjective measures of listening effort with objective measures such as pupillometry (e.g., Zekveld *et al.*, 2011; Koelewijn *et al.*, 2012a; Koelewijn *et al.*, 2012b; Kuchinsky *et al.*, 2013) or dual task interference (e.g., Sarampalis *et al.*, 2009).

Listening effort can provide unique insight into how listeners cope with the perceptual challenges of real world listening environments. Additional listening effort has been related to fatigue (Hicks and Tharpe, 2002) and memory (Rudner *et al.*, 2012), making listening effort a potentially important topic in the study of aging and hearing loss (Kuchinsky *et al.*, 2013). Recent studies of speech perception have shown

increased listening effort with speech maskers, which possibly reflects the costs of resolving informational masking (Koelewijn *et al.*, 2012a; Koelewijn *et al.*, 2012b). The present study demonstrates that competing speech, even in the absence of informational masking, can increase perceived listening effort, suggesting an inherent cost of suppressing task-irrelevant speech.

Acknowledgments

This work was supported (in part) by research grants P50 DC00422 and R01 DC00184 from NIH/NIDCD and by the South Carolina Clinical and Translational Research (SCTR) Institute, with an academic home at the Medical University of South Carolina, NIH/NCRR Grant number UL1 RR029882. This investigation was conducted in a facility constructed with support from Research Facilities Improvement Program Grant Number C06 RR14516 from the National Center for Research Resources, National Institutes of Health. The authors thank Sandra Gordon-Salant for advice and comments on an earlier draft of this manuscript, and Lois Matthews, Jayne Ahlstrom, and Fu-Shing Lee for advice and assistance.

References and links

- ¹Here, the term “listening effort” refers to the general concept, while Perceived Effort will be used to refer to a specific domain of the NASA TLX.
 - ²Psychoacoustics testing software developed by Bomjun Kwon was obtained from the website: <http://auditorypro.com/download/psycon/psycon.html> (Last viewed 5/20/13).
 - ³In Fig. 2, the mean value for Perceived Effort with competing noise deviates from the median, raising the possibility that use of parametric statistics may not be appropriate. Accordingly, parametric test results using differences between pairs of values from individual subjects were confirmed with a Wilcoxon Signed Rank Test with correction for multiple comparisons. Whereas median values across subjects appear similar for speech and noise, individual subjects consistently reported higher ratings (more effort) for speech compared to noise.
- Gopher, D., and Braune, R. (1984). “On the psychophysics of workload: Why bother with subjective measures?” *Hum. Factors* **26**, 519–532.
- Hart, S. G., and Staveland, L. E. (1988). “Development of NASA-TLX: Results of empirical and theoretical research,” in *Human Mental Workload*, edited by P. A. Hancock, and N. Meshkati (Elsevier, Amsterdam), pp. 239–250.
- Hicks, C. B., and Tharpe, A. M. (2002). “Listening effort and fatigue in school-age children with and without hearing loss,” *J. Speech Lang. Hear. Res.* **45**, 573–584.
- Kahneman, D. (1973). *Attention and Effort* (Prentice-Hall, Englewood Cliffs, NJ).
- Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012a). “Pupil dilation uncovers extra listening effort in the presence of a single-talker masker,” *Ear. Hear.* **33**, 291–300.
- Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., and Kramer, S. E. (2012b). “Processing load induced by informational masking is related to linguistic abilities,” *Int. J. Otolaryngol.* **2012**, 865731.
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., and Eckert, M. A. (2013). “Pupil size varies with word listening and response selection difficulty in older adults with hearing loss,” *Psychophysiology* **50**, 23–34.
- Mackersie, C. L., and Cones, H. (2011). “Subjective and psychophysiological indexes of listening effort in a competing-talker task,” *J. Am. Acad. Audiol.* **22**, 113–122.
- Moray, N. (1982). “Subjective mental workload,” *Hum. Factors* **24**, 25–40.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). “Primitive stream segregation of tone sequences without differences in fundamental frequency or passband,” *J. Acoust. Soc. Am.* **112**, 2074–2085.
- Rudner, M., Lunner, T., Behrens, T., Thoren, E. S., and Rönnberg, J. (2012). “Working memory capacity may influence perceived effort during aided speech recognition in noise,” *J. Am. Acad. Audiol.* **23**, 577–589.
- Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). “Objective measures of listening effort: Effects of background noise and noise reduction,” *J. Speech Lang. Hear. Res.* **52**, 1230–1240.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2011). “Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response,” *Ear. Hear.* **32**, 498–510.