Age-related differences in listening effort during degraded speech recognition

Kristi M. Ward, Jing Shen, Pamela E. Souza, Tina M. Greeco-Calub
The Roxelyn & Richard Pepper Department of Communication Sciences & Disorders, Northwestern University, Evanston, IL

Email: kmward@u.northwestern.edu

Abstract

Listening effort refers to the amount of cognitive resources needed by a listener to understand speech. Older listeners are thought to expend more listening effort than younger listeners (Hornsby, 2013). Listening effort is likely dependent on the underlying cognitive function of the listener (Desjardins & Doherty, 2013). Thus, children may have a limited amount of cognitive resources that can be reallocated during highly degraded speech recognition tasks due to their cognitive immaturity. To test the hypothesis, the current study implements a dual-task paradigm to objectively measure listening effort in children (5-12 years old) and young adults (18-25 years old). The dual-task paradigm consists of two simultaneously presented tasks: 1) a primary speech recognition task and a secondary visual task and 2) a secondary visual sequencing task. Performance on the secondary task is used to quantify how listening effort is driven by age-related differences in cognitive function. Preliminary data suggest that children expend more listening effort than adults as the speech signal becomes more degraded. Identifying how listening effort is driven by age-related differences in cognitive function will inform the development of a clinical tool to measure listening effort in populations who are especially prone to chronic mental fatigue, such as those with hearing loss (Hornsby, 2013).

Methods

Participants: Ten children (5 female; 8-12 years old; mean=9.5 years) and ten young adults (8 female; 18-25 years old; mean=20.1 years) participated in the current study. All participants were native English speakers, had normal hearing across a range of frequencies, and had normal or corrected-to-normal vision. All testing procedures were approved by the Institutional Review Board at Northwestern University.

Procedure: Prior to testing, each participant was directed to listen to and repeat five sentences from each of the four conditions of spectral resolution. If participants achieved an accuracy score of 50% or less, they were re-familiarized with 20 additional sentences in the 4-channel condition: 10 with feedback and 10 without feedback.

Testing Protocol (in order of presentation):
1) Speech Recognition (Primary) Task in Isolation — Each of the four test conditions consisted of 20 sentences and were presented in order of presentation. Participants were instructed to repeat each sentence aloud to the best of their ability. Performance in each condition served as a baseline measure of speech perception.
2) Visual Sequencing (Secondary) Task in Isolation — Participants were instructed to perform a secondary task while performing the primary speech recognition task. The secondary task was performed simultaneously in each of the four conditions of spectral resolution. Performance scores were calculated using a selective visual attention measure.
3) Dual-Task Paradigm — The speech recognition (primary) task and visual sequencing (secondary) task were performed simultaneously in each of the four conditions of spectral resolution. Performance scores were calculated using a selective visual attention measure.
4) Visual Sequencing (Secondary) Task in Isolation — Participants repeated the task at the end of the test session to verify that performance was not affected by learning or fatigue.

Stimuli: 180 Bench-Kowal-Bamford (BKB) phonetically-balanced short sentences were presented at 65 dB SPL across four different conditions of varying spectral resolution; 4-, 6-, 8-channel noise-band vocoded, and unprocessed.

Dependent Variables: Accuracy: proportion of keywords correctly identified within each task condition. Accuracy: proportion of trials in which a repeated image eluded a response. Reaction Time (RT): duration of time between the onset of the repeated image and the participant's response.

Test Conditions: Lists of 20 sentences were presented at 65 dB SPL across four different conditions of varying spectral resolution: 4-, 6-, 8-channel noise-band vocoded, and unprocessed.

Overview of Test Measures:

Visual task performance changes systematically with the spectral resolution of speech

Figure 1: There is no difference in speech recognition accuracy between the single (gray bars) and dual (black bars) task for children (A) or adults (B). Percent correct was converted to a z-score for the analysis. A: 4 (age group) x 2 (task) x 4 (condition) mixed ANOVA revealed: 1) No main effect of task [F(1,81)=1.41, p=0.251]. 2) A main effect of condition [F(2,34.6)=313.51, p<0.001] indicating that speech recognition accuracy decreases significantly as spectral resolution decreases; 3) No significant interaction between condition and age group [F(2.34,6)=313.51, p>0.01] demonstrating that children and adults have similar speech recognition accuracy across the degraded conditions.

Figure 2: Performance on the visual task. Reaction time (A) slows and accuracy (B) decreases as the spectral resolution of the speech signal decreases. A 2 (age group) x 2 (condition) repeated measures ANOVA showed: 1) A main effect of condition for both reaction time [F(2,9)=14.12, p<0.003] and accuracy [F(2,8)=7.75, p=0.001]; 2) A main effect of age [F(1,8)=10.58, p<0.004] demonstrating that children (blue bars) are significantly slower and less accurate than adults (red bars) across all conditions.

Figure 3: Listening effort (gray shaded region) as a function of reaction time (A) or accuracy (B) measured during degraded speech recognition. Overall, listening effort increases as spectral resolution decreases. Children exhibit greater listening effort during degraded speech recognition, regardless of the number of spectral channels. In adults, listening effort increases significantly in only the most degraded condition. Children’s, but not adults’, false alarm rate increases with decreasing spectral resolution (C).

Children expend more listening effort than young adults during degraded speech recognition

Summary & Conclusions

1) Listeners’ speech recognition accuracy declined as the spectral resolution of the speech signal decreased. Older listeners performed the same regardless of whether the primary task occurred in isolation or as part of the dual task (Figure 1).
2) Performance on the visual task declined as the speech signal became more degraded. This finding is consistent with increased listening effort during degraded speech perception. Additionally, children were slower and less accurate than the young adults on the visual task across conditions (Figure 2).
3) Listeners expended more effort as the speech became more degraded. Children’s increased listening effort with any degree of degradation, as compared to adults who only showed a significant increase in listening effort in the most severely degraded condition. This is further supported by the fact that children produced a greater proportion of false alarms than adults across conditions (Figure 3).
4) Overall, children are more susceptible to elevated listening effort during degraded speech recognition due to limitations in their underlying cognitive function.

References


Acknowledgments

We would primarily like to thank our participants and their families for their time and diligence. We acknowledge Beverly Wight for providing the visual stimuli for the secondary task. Also, we extend our gratitude to Brook Fergus and Jennifer Romero for their assistance with the technical and analytical aspects of this project. Work supported by NIH grant R01 DC012989 and the School of Communication, Northwestern University.

Footnotes

To generate the noted voiced spectral speech stimuli, each sentence was normalized equally at 65 dB SPL and digitally filtered into the respective number of frequency channels. For the visual sequencing task, each sentence was root-mean-square equalized at 65 dB SPL and had normal or corrected-to-normal vision. The envelope was then multiplied by a broadband noise carrier and passed through the same low-pass filter as the speech signal.