Evaluating Two Measures of Listening Effort

Paul Pendergraft, B.A.

Robyn M. Cox, Ph.D., Research Advisor

Au.D. Research Project 2013
Introduction

• Conventional audiometric tests do not tell us everything about an individual’s “real world” listening experience.

• For instance, although digital noise reduction does not affect speech perception performance, it is thought that it may reduce listening effort in noisy listening conditions.

• A test to assess listening effort could provide a way to measure benefit from hearing aid features that does not show up on conventional tests.

Sarampalis, Kalluri, Edwards (2009)
Different Approaches to Assessing Listening Effort

• Subjective Rating
  o Subjects are asked to rate how much effort they used to successfully understand a speech signal

• Dual-Task Assessment
  o Subjects are asked to perform a secondary task while simultaneously performing a speech perception task

• Physiological Measures
  o Objective measures are made while subjects perform a speech perception task
Subjective Rating of Effort

• Individuals may perform about equally well on a speech perception test in more and less difficult listening conditions.

• Although the speech performance score may only vary slightly, more effort is required to achieve that score in the more difficult condition.

• Asking a subject to rate listening effort is the most direct way to assess listening effort, and has high face validity.
Hypothetical Results from Subjective Rating Assessment

Speech Perception Performance

- Easy Speech Task
- Difficult Speech Task
- Very Difficult Speech Task

Speech Task Effort Rating

- Easy Speech Task
- Difficult Speech Task
- Very Difficult Speech Task
Dual-Task Assessment

- Individuals have finite cognitive resources
- More difficult tasks require more cognitive resources
- When multi-tasking, individuals allocate those resources according to the priority of the tasks they are trying to accomplish
- In a dual-task paradigm, a subject is assigned primary and secondary tasks
  - As difficulty of the primary task increases, more cognitive resources are required to successfully complete the primary task, leaving fewer available for the secondary task
  - With fewer cognitive resources available for secondary task, performance on secondary task decreases

Fraser, Gagné, Alepins, Dubois (2010)
Hypothetical Results from Dual-Task
Purpose

• To evaluate the validity and sensitivity of one objective assessment, and one subjective assessment of listening effort

• The goal was to determine which method was more appropriate for measuring listening effort and could be included in future studies of hearing aid benefit
Methods

- The Revised Speech Perception in Noise (R-SPIN) test was used for speech material, presented at a conversational level.
- The R-SPIN contains 16 lists of 25 sentences.
- Participants are asked to repeat the last word of each sentence.
- 8 of the lists are High Predictability (HP) lists in which the last word of the sentence might be deduced by using contextual clues.
  - E.g. “The plow was pulled by an **OX**”
- 8 of the lists are Low Predictability (LP) lists in which the last word of the sentence cannot be deduced by using contextual clues.
  - E.g. “Bill might discuss the **FOAM**”

Methods

Two assessments of listening effort were explored
• An objective dual-task assessment of listening effort
• A subjective self-report measure of listening effort

The 16 R-SPIN lists were presented at -4, -2, 0, and 2 dB SNR
• For each listening effort assessment (dual-task or rating), the subjects were presented with one list at each SNR level, in each contextual condition (HP/LP)
• We expected the 2 dB SNR, High Predictability condition to require the least listening effort
• And the -4 dB SNR, Low Predictability condition to require the most listening effort.
Dual-Task Assessment

• The dual-task assessment used speech perception as the primary task, and word recall as the secondary task
  o After 5 presentations, testing was paused and subjects were asked to recall the last 5 responses they had given. Subjects could repeat the words in any order.
  o We expected subjects to remember fewer words in more difficult listening conditions in which more cognitive resources would be required to successfully understand speech for the primary task
Subjective Rating of Effort

- Subjects were asked to rate the effort required to understand speech halfway through, and at the end of each listening condition
  - We expected subjects to rate conditions with lower speech scores as requiring more effort

- Participants were asked to use the 7-point scale shown here.

Listening Effort Scale

1. No effort
2. Very little effort
3. Little effort
4. Moderate effort
5. Considerable effort
6. Much effort
7. Extreme effort

This scale was provided to subjects on 8.5” x 11” (letter size) paper.
Methods

• Speech materials were presented in the sound field from a speaker at 0° azimuth
• The noise (multi-talker babble) was presented at a constant 67 dB SPL
  o The amplitude of the speech signal was varied to achieve the various SNR conditions
• Presentations alternated between HP and LP lists
• Context order (HP or LP first), SNR order, and Task order (dual-task or rating) were counter-balanced across subjects
Subjects

- 22 participants (6 male)
- Aged 23 – 56 years
- All had English as first language
- None reported more than mild hearing difficulty
Methods

• The data was first checked for outliers
• Skew and kurtosis were assessed numerically
• The effect of SNR and contextual condition were analyzed using a 4x2 repeated measures ANOVA
Research Questions:

1. o Is the rating task a valid assessment of listening effort?
   o Is it a sensitive assessment of listening effort?

2. o Is the dual-task a valid assessment of listening effort?
   o Is it a sensitive assessment of listening effort?

3. Does the dual-task give us the same information as the rating task?

4. For HP lists, does the measure of listening effort affect speech perception performance?

5. For LP lists, does the measure of listening effort affect speech perception performance?
Before we investigated these research questions, we wanted to be sure that the different listening conditions had been substantially different in difficulty. If they were, we expected that:

- Change in SNR would affect primary task (speech perception) performance
- Change in contextual condition would affect primary task (speech perception) performance
Results
Did speech perception performance vary as expected with SNR and contextual condition?
Context was found to have a significant effect on mean speech perception score ($P<.001$). SNR was found to have a significant effect on mean speech perception score ($P<.001$). Within HP lists, paired comparisons at each incremental increase of SNR (i.e. -4 dB SNR vs -2 dB SNR, 0 dB SNR vs 2 dB SNR) were all found to be significant at the $\alpha = .01$ level of significance. This was also true for LP lists.
Did speech perception performance vary as expected with SNR?
  o More difficult SNR conditions yielded poorer speech perception scores
  o For each contextual condition, a change in SNR produced a significant change in mean speech perception score

Did speech perception performance vary as expected between HP and LP lists?
  o There was a significant difference in mean speech performance scores between HP and LP lists at every SNR

These results indicate that the various listening conditions are substantially different in difficulty and, therefore, we expect that they required different amounts of effort for best listening performance.
Is the rating task a valid and sensitive assessment of listening effort?

- If it is, effort ratings should follow a trend similar to speech performance scores
- Higher SNR presentations should be rated as easier than lower SNR presentations
- High Predictability lists should be rated as easier than Low Predictability lists
Is the rating task a valid measure of listening effort?

- The main effects of contextual condition and SNR were both found to be statistically significant (P<.001)
- Paired comparisons at each incremental increase of SNR (-4 dB SNR vs -2 dB SNR, -2 dB SNR vs 0 dB SNR, 0 dB SNR vs 2 dB SNR) were all found to be significant at the $\alpha = .01$ level of significance
- At each SNR, HP lists were rated as significantly easier than LP lists, at the $\alpha = .05$ level of significance

Is the rating task a sensitive measure of listening effort?

- For HP lists, the difference in mean speech perception score between 0 to 2 dB SNR conditions was about 1 word
- The difference on the rating task was close to a whole increment on the 7-point rating scale
There was a significant interaction between SNR and context. At 0 dB SNR the P value was <.001, but at -4 dB SNR, the P value was .025. Contextual condition had a bigger impact in easier SNR conditions, perhaps because subjects were less able to hear the contextual clues in harder SNR conditions. At -4 dB SNR a floor effect was observed: HP and LP lists were both given ratings approaching the maximum difficulty rating on the listening effort scale.
Is the dual-task a valid and sensitive assessment of listening effort?

- If it is, word recall scores should follow a trend similar to speech performance scores.
- Subjects should remember more words in easier SNR presentations, compared to harder SNR presentations.
- Subjects should remember more words in High Predictability lists, compared to Low Predictability lists.
Word Recall Scores

Speech Perception Scores

# of Words Correctly Remembered

-4 dB SNR  -2 dB SNR  0 dB SNR  2 dB SNR

High Predictability Lists

Low Predictability Lists

# of Words Correctly Remembered

-4 dB SNR  -2 dB SNR  0 dB SNR  2 dB SNR
Is the dual-task a valid measure of listening effort?

- SNR was found to have a significant effect on mean word recall score (P<.001). Context was also found to have a significant effect on mean memory score (P=.002).
- Despite this statistical significance, the differences in mean number of words recalled across listening conditions was comparatively small.
- None of the paired comparisons at incremental increases of SNR were found to be significant at the α = .01 level of significance.
- At a given SNR, mean word recall scores for HP lists were not significantly greater than for LP lists at the α = .05 level of significance.

Is this dual-task a sensitive measure of listening effort?

- The mean difference in speech perception score between -4 dB SNR LP and 2 dB SNR HP conditions was 16.8 words.
- The mean difference in word recall scores between those conditions was 2.4 words.
  - Although the measure is statistically valid, its sensitivity to differences in difficulty is poor.
Does the dual-task give us the same information as the rating task?

- If the dual-task and rating task are both valid and sensitive measures of listening effort, we expect to see a significant relationship between rating scores and word recall scores.
- To assess this relationship we plotted the mean rating score against the mean word recall score for each listening condition.
- If there is a relationship between the two assessments we expect to see word recall scores consistently increasing as rating scores decrease.
Individual Ratings and Word Recall Scores in the 8 Listening Conditions

Word Recall Scores
(# of Words Correctly Remembered)

Listening Effort Rating

Hard

Easy

HP, -4 dB SNR
HP, -2 dB SNR
HP, 0 dB SNR
HP, 2 dB SNR
LP, -4 dB SNR
LP, -2 dB SNR
LP, 0 dB SNR
LP, 2 dB SNR
Does the method of listening effort measurement affect speech perception performance?

- If it does, we expect to see higher speech perception scores during either rating task or dual-task.
- There would be a consistent trend in which speech perception scores are higher for one task at all SNRs.
When speech perception scores at all SNRs were combined, the effect of task was significant (P=.030). When looking at speech scores at each SNR, mean speech perception scores were significantly lower during the dual-task only at -2 dB SNR (P=.041).
The measure of listening effort was not found to have significant effect on speech perception scores for LP lists (P=.480).
Conclusion

- The various presentation conditions used in this study represented more and less difficult listening conditions, as demonstrated by statistically significant differences in speech perception scores.

- The rating task was found to be a valid and sensitive measure of listening effort.

- Although the dual-task was a statistically valid measure of listening effort, it was not a sufficiently sensitive measure of listening effort.
Possible limitations of this dual-task design

- Participants were required to respond on the speech perception task even when they had not heard the target word.
- It was observed that participants were easily able to remember the “guess” words (which were sometimes fabrications).
- The difficulty of the word recall task varied considerably between subjects.
  - Some subjects always performed well, even in difficult listening conditions.
  - Some subjects always performed poorly, even in easy listening conditions.
- Our results indicated that the dual-task approach used in this study is not valid for speech tests where the speech perception score is low.
References


Acknowledgements

Special thanks to Dr. Jingjing Xu for his assistance with instrumentation set-up, and statistical analysis