Development of a Method to Assess In-Ear Speech Intelligibility Through Listening Effort

Alexis Pinsonnault-Skvarenina[*1,2], **Philippe Chabot** [†3], **Ajin Tom** [‡3], **and Antoine Bernier** [§3]
[1] Université du Québec (ÉTS), Montréal, Québec, Canada
[2] Centre for Interdisciplinary Research in Music Media and Technology, McGill University, Montréal, Québec, Canada
[3]EERS Global Technologies Inc., Montréal, Québec, Canada

1 Introduction

Active hearing protection devices equipped with an in-ear microphone (IEM) enable in-ear voice pickup which presents better signal-to-noise ratio over ambient microphones in highly noisy conditions. To improve its intelligibility, in-ear speech requires processing which can take many forms, from fixed filtering to spectral domain processing based on machine learning. Objectively comparing these processing strategies can be difficult. In the authors' experience, objective intelligibility assessment techniques have not provided the necessary resolution to compare various processed in-ear speech. They have also failed to capture a concept of listening effort, which intuitively seemed to increase with in-ear speech over reference speech when recorded in silence.

This work presents the development and preliminary validation of a technique that objectively measures intelligibility of in-ear speech material and its associated listening effort. The technique is based on a dual task paradigm, where a primary task measures intelligibility directly. Performance on a secondary task, performed at the same time, provides a measure correlated with listening effort [1]. Dual-task paradigms are commonly used in audiology such as to evaluate the effectiveness of noise reduction algorithms in hearing aids [2, 3].

2 Method

2.1 Participants

Eleven adults participated in this pilot study. All participants self-reported normal hearing sensitivity.

2.2 Procedures

Primary task: word repetition task

Participants were asked to repeat as many words as they could after hearing recordings of 5-word sentences through Sennheiser HD280 Pro headphones. The material of the American English Matrix test [4] was used to create nine lists of ten random sentences of similar syntax. The sentences were recorded for four conditions to be evaluated: noisy reference speech (microphone in front of mouth while subjected to 85 dB(A) and 80 dB(A) noise), IEM speech without any processing, and IEM speech with BWE processing.

*alexis.pinsonnault-skvarenina@etsmtl.ca
†pchabot@eers.ca
‡atom@eers.ca
§abernier@eers.ca

Secondary task: visual-motor task

The chosen visual motor task was a computer game adaptation of the Tower of London test [5]. The goal of the game is to move colored beads on a virtual board to match a given configuration displayed as a target. Colored beads are placed in three stacks on the board, arranged in different patterns and must be moved using a computer mouse.

Subjective speech intelligibility

Subjective speech intelligibility was assessed by asking participants to rate, on a scale of 0 (Bad) to 10 (Excellent), how well they could understand the words.

Procedures

The testing took place in a quiet room. First, after receiving instructions, participants underwent a five sequences familiarization phase for the secondary task. Second, participants were asked to read each of the words of the Matrix. A practice session was then conducted in the dual-task paradigm (word repetition and visual motor concomitantly). Participants were instructed that the primary task (word repetition) should be prioritized. Finally, participants completed the test sessions in single- and dual-task paradigms in random order. After each single-task paradigm involving the word repetition task, subjective speech intelligibility was assessed.

2.3 Analyses

Single-task accuracy was subtracted from dual-task accuracy to obtain two dual-task cost (DTC) scores: DTC Word accuracy and DTC Visual-motor accuracy.

3 Results

3.1 Dual tasks results

Dual-task costs were calculated (Figure 1). No differences in DTC Word were observed amongst conditions. The highest DTC Visual-Motor was observed in the reference speech with 85 dB(A) noise condition ($29\% \pm 17\%$), followed by the 80 dB(A) noise condition ($20\% \pm 13\%$). Both IEM and BWE conditions presented with the similar DTC Visual-Motor (respectively $13\% \pm 12\%$ and $13\% \pm 15\%$).

3.2 Subjective ratings results

Results for the subjective speech intelligibility ratings are presented in Figure 2. Worse subjective intelligibility ratings were obtained with reference speech with 85 dB(A) noise $(2,7 \pm 1,1)$. Reference speech with 80 dB(A) noise and IEM

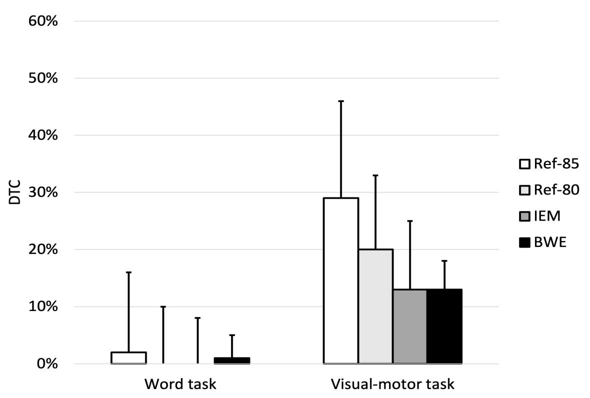

Figure 1: Word and visual-motor task DTC accuracy results for reference speech with 85 dB(A) noise, reference speech with 80 dB(A) noise, IEM, and BWE. Bars represent standard deviations.

speech obtained similar intelligibility ratings ($6,4 \pm 1,4$ and $5,8 \pm 1,3$ respectively). BWE processing obtained the best intelligibility ratings ($7,5 \pm 1,1$).

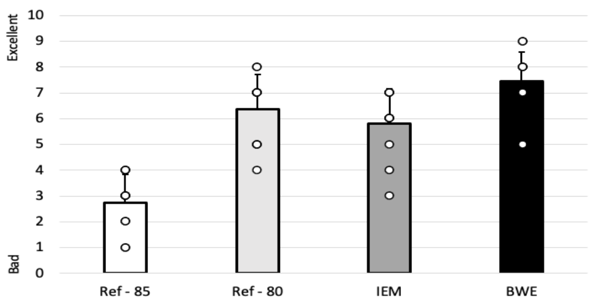

Figure 2: Subjective speech intelligibility ratings for reference speech with 85 dB(A) noise, reference speech with 80 dB(A) noise, IEM , and BWE. White dots represent individual scores. Bars represent standard deviations.

4 Discussion

4.1 Listening effort with in-ear speech

Our preliminary data suggests that a dual-task paradigm can capture nuances an objective intelligibility test alone cannot. According to the DTCs for the visual-motor task, more listening effort is required to understand noisy reference speech than IEM speech (either raw or with BWE processing). This is also supported by participants' subjective evaluation, rating IEM speech with BWE processing to be subjectively the most intelligible over noisy reference speech or raw IEM speech. However, unintuitively, the BWE processing did not improve listening effort over raw IEM speech, despite being subjectively rated as better and having the least performance variability of all the conditions.

Taken together, our results suggest that using a measure of listening effort (i.e., a dual-task paradigm) can be informative when assessing in-ear speech intelligibility. This is supported by previous research that found listening effort to be a better indicator of speech perception than intelligibility accuracy scores [1, 6].

4.2 Limitation and future research

Potential limitation of this preliminary study may have reduced our ability to accurately measure listening effort. First, hearing sensitivity was not objectively measured. It is believed that individuals with hearing loss can present with increased listening effort to understand speech-in-noise [6]. Therefore, if some of our participants did present with a hearing loss, this could have introduced additional variability in our results, which might have affected our ability to measure differences in listening effort across conditions. Second, previous research has suggested that an 80% performance criterion on both word and visual-motor single tasks could improve the potential of a dual-task paradigm to measure listening effort. Since the performances of our participants for the single tasks were very high (nearly 100%), it is unclear how this might have affected the sensitivity of our experimental paradigm. Third, more participants would be needed to achieve statistical significance between the objective scores.

Future development of our test paradigm should seek to refine the secondary task and integrate a time response variable, which might be more suited to evaluate listening effort [1, 6].

5 Conclusions

Our results suggest that a dual-task paradigm to measure listening effort can be a good approach to behaviorally evaluate in-ear speech. Our pilot study supports the use of in-ear speech to improve communication in noisy settings while quantifying its potential for further improvements.

Acknowledgments

The authors would like to acknowledge the financial support received from MITACS IT26677 (SUBV-2021-168) for the ÉTS-EERS Industrial research chair in in ear technologies, sponsored by EERS Global Technologies Inc.

References

[1] Jean-Pierre Gagne, Jana Besser, and Ulrike Lemke. Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in hearing*, 21:2331216516687287, 2017.

[2] Brent Edwards. The future of hearing aid technology. *Trends in amplification*, 11(1):31–45, 2007.

[3] Anastasios Sarampalis, Sridhar Kalluri, Brent Edwards, and Ervin Hafter. Objective measures of listening effort: Effects of background noise and noise reduction. 2009.

[4] B Kreisman, R Carroll, M Zokoll, A Warzybok, P Allen, P Folkeard, K Wagener, and B Kollmeier. Design, optimization and evaluation of an american english matrix sentence test in noise. In *American Academy of Audiology Conference*, 2013.

[5] Timothy Shallice. Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 298(1089):199–209, 1982.

[6] Penny Anderson Gosselin and Jean-Pierre Gagné. Use of a dual-task paradigm to measure listening effort utilisation d'un paradigme de double tâche pour mesurer l'attention auditive. *Revue canadienne d'orthophonie et d'audiologie-Vol*, 34(1), 2010.