INTELLIGIBILITY, RESPONSE TIME, AND LISTENING EFFICIENCY OF NATIVE AND NON-NATIVE LISTENERS IN VARIOUS ACOUSTICAL CONDITIONS

Alice Lam *[1] **and Murray Hodgson** †[1]

[1] Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

1 Introduction

This study evaluates the speech perception of English as a Second Language (ESL) and native English-speaking students and in various acoustical conditions. Three different metrics for speech intelligibility are compared: the intelligibility score (IS), response time, and listening efficiency (DE).

Speech intelligibility scores are based on the accuracy of participants' responses to listening tests [1]. The response times of participants can be used as a more sensitive measure, to distinguish between conditions that appear equally intelligible [2].

Prodi *et al.* [3] introduced a combined metric, the "direct listening efficiency" or "listening efficiency" (DE), which combines IS and response time to evaluate the amount of effort required from a listener to understand speech. The development and evaluation of this metric has so far been in the Italian language. This project is an initial evaluation of the DE metric in English, for both native and non-native (ESL) speakers.

2 Method

2.1 Test Material

The Diagnostic Rhyme Test (DRT) [1] was used for this study. The test material was split into 16 lists of 6 word pairs, each recorded by one of five native speakers (2 male, 3 female) in an anechoic, low-noise environment (Condition A). These recordings were auralized into conditions B (reverberation dominant) and C (some reverberation plus "babble and activity" noise [4]). T_{mid} = 1.08 s and 0.78 s for conditions B and C, respectively, These conditions corresponded to an STI (Speech Transmission Index) of 0.78 for Condition A, and an STI of 0.575 for Conditions B and C.

2.2 Test Procedure

During the test, participants listened to the recorded word with carrier phrase, and were then asked to select the word that they had heard from one of three options (the two words in the word pair, or "none of the above"), as shown in Figure 1. This process would be repeated for each pair in the list, making up one test. Each participant completed one test from condition A as practice, followed by one test each from conditions A, B and C.

All tests took place in an anechoic environment. Playback of the test material (all monaural) was done using a single KRK Rokit 5 G2 loudspeaker 1.5m away from the seated participant, at head height. Playback was calibrated to 66 dBA, corresponding to "raised" vocal effort [1].

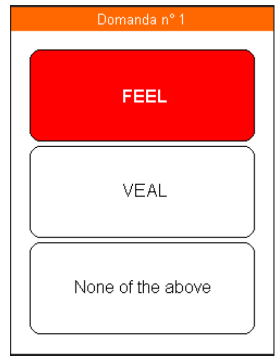

Figure 1: Sample Intelligo test screen.

The Intelligo system [5] was used to facilitate data collection and synchronization with the playback system. Participants used a touchscreen handset to enter responses in real time. Following the tests, participants were asked to fill out a brief questionnaire regarding their proficiency in English, their country of origin, and other languages spoken.

2.3 Participants

All 37 participants were university students aged 18 to 32, with no reported hearing loss. 25 individuals (15 male, 10 female) identified as ESL, and 12 identifying as non-ESL (native speakers of English). We defined an individual as ESL if they were not exposed to the English language until after 3 years of age. The amount of time in which the ESL participants had lived in a primarily-English speaking environment ranged from 6 months to 15 years.

3 Results

The IS and DE were calculated for each test using the method in [3], as follows:

$$Intelligibility\ Score\ (IS) = (average\ raw\ score + 1) * 0.5$$

(raw scores: correct = +1, incorrect = -1, "none of the above" = -.05)

$$Listening\ Efficiency\ (DE, s^{-1}) = \frac{Intelligibility\ Score}{Mean\ Response\ Time\ (s)}$$

We compare the effectiveness of these three metrics in performing two different tasks: differentiating between sound fields (conditions A, B, C) and differentiating language ability (ESL and non-ESL). For the former, the Kruskal-Wallis test was used. Where applicable, the Wilcoxon rank sum test was performed *post hoc* in order to determine which pair of conditions showed significant differences. For the latter, Wilcoxon Rank Sum tests were performed.

* avl@alumni.ubc.ca
† murray.hodgson@ubc.ca

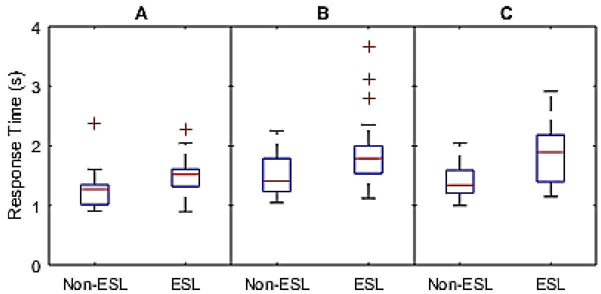

Figure 2: Response times for all cases.

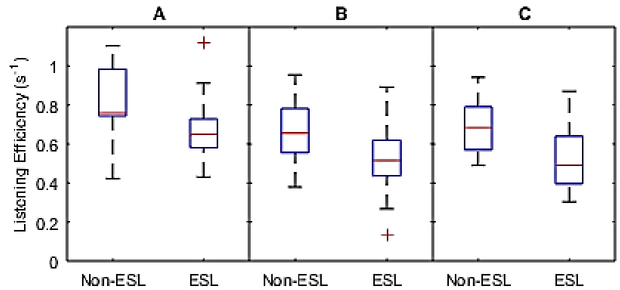

Figure 3: Listening efficiency for all cases.

3.1 Differentiating Sound Fields

Using Intelligibility Score

There was no significant difference between conditions A, B and C in either the ESL or non-ESL case ($p = .3826$ and $p = .1714$, respectively).

Using Response Time

As shown in Figure 2, there was no significant difference between conditions A, B and C for the non-ESL case. In the ESL case, response time was lower (i.e. faster) in A than it was in B or C ($p = .0154$ and $.0127$, respectively).

Using Listening Efficiency

As shown in Figure 3, there was no significant difference between conditions A, B and C for the non-ESL case. In the ESL case, DE was higher in A than it was in B or C ($p = .0101$ and $.0091$, respectively).

3.2 Differentiating Language Ability

Using Intelligibility Score

None of the three conditions showed any statistically significant differences ($p > 0.15$ in all cases) between non-ESL and ESL participants' intelligibility scores.

Using Response Time

As shown in Figure 2, all three conditions showed statistically significant differences between non-ESL and ESL participants' average response times ($p = .0249$, $.0270$ and $.0087$ for conditions A, B and C, respectively).

Using Listening Efficiency

As shown in Figure 3, all three conditions showed statistically significant differences between non-ESL and ESL participants' average DE ($p = .0178$, $.0163$ and $.0079$ for conditions A, B and C, respectively).

4 Discussion

The relative ineffectiveness of using IS alone could be due to the fact that conditions A, B and C were generally similar: in all cases participants could understand what was being said after some thought, and there were no cases where speech was genuinely unintelligible.

When differentiating between sound fields, only the ESL case showed significant differences. This suggests that ESL students have a disadvantage in acoustically non-ideal conditions compared to their non-ESL peers, a result that aligns with [6]. The smaller sample size of non-ESL may also have been a factor.

Comparing non-ESL and ESL results, ESL participants showed a higher average response time and lower average DE. This result aligns with the idea that these two metrics give a picture of "listening effort" as posited in [3].

Though the qualitative results for DE and response time were similar in all cases, in every case the DE results were more significant, showed fewer statistical outliers, and less variance.

5 Conclusion

This study compared the effectiveness of using IS, response time, or DE to differentiate between either acoustical conditions or language ability. It was found in all cases that using DE was the most effective option, with response time being a close second. Since incorporating the intelligibility score in a listening test that already measures response time is relatively trivial, using DE instead of response time alone is the best option to evaluate listening effort in addition to speech intelligibility.

Acknowledgments

We thank Nicola Prodi for providing the Intelligo test system as well as his guidance and advice, and Katie O'Brien for her assistance in organizing and conducting the listening tests.

References

[1] *ANSI S3.2-2009 Method for Measuring the Intelligibility of Speech over Communication Systems*, 2009.

[2] M. H. L. Hecker, K. N. Stevens and C. E. Williams, "Measurements of Reaction Time in Intelligibility Tests," *J. Acoust. Soc. Am.*, vol. 39, no. 6, pp. 1188-1189, 1966.

[3] N. Prodi, C. Visentin and A. Farnetani, "Intelligibility, listening difficulty and listening efficiency in auralized classrooms," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 172-181, 2010.

[4] N. Prodi, C. Visentin and A. Feletti, "On the perception of speech in primary school classrooms: Ranking of noise interference and of age influence," *J. Acoust. Soc. Am.*, vol. 133, no. 1, pp. 255-268, 2013.

[5] N. Prodi, C. Visentin and C. Belletini, "Listening efficiency testing," in *45th AES Conference Proceedings*, Helsinki, 2012.

[6] S. J. van Wijngaarden, H. J. M. Steeneken and T. Houtgast, "Quantifying the intelligibility of speech in noise for non-native listeners," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1906-1916, 2002.