Independent Studies and Capstones

Program in Audiology and Communication Sciences

2015

Evaluation of the American English Matrix Test with cochlear implant recipients

Alison M. Kiolbasa
Washington University School of Medicine in St. Louis

Follow this and additional works at: http://digitalcommons.wustl.edu/pacs_capstones

EVALUATION OF THE AMERICAN ENGLISH MATRIX TEST WITH COCHLEAR IMPLANT RECIPIENTS

by

Alison M. Kiolbasa

A Capstone Project
submitted in partial fulfillment of the
requirements for the degree of:

Doctor of Audiology

Washington University School of Medicine
Program in Audiology and Communication Sciences

May 20, 2016

Approved by:
Lisa G. Potts, Ph.D., Capstone Project Advisor
Jill B. Firszt, Ph.D., Second Reader

Abstract: The American English Matrix Test (AMT) is a recently developed adaptive speech-in-noise test that determines a speech reception threshold. The aim of the proposed study is to evaluate speech recognition in cochlear implant (CI) recipients using the AMT, determine test-retest reliability, and compare it to current clinical speech recognition tests.

Acknowledgements

I would like to thank Lisa Potts for her mentorship throughout the past year. Lisa provided me with an immense amount of support and guidance throughout this project, and served as a true role model for my time at Wash U. Thank you to Jill Firszt, my second reader, for welcoming me into her lab and helping me through this research experience. I would also like to give special thanks to Ruth Reeder for her help with the statistics for this project as well as the various other contributions to its development and completion.   I would like to thank the other members of the Firszt Lab and Adult Cochlear Implant team for their help in recruiting participants, their support, and kind words of wisdom: Noel Dwyer, Laura Holden, Tim Holden, Brenda Gotter, Karen Mispagel, Sarah Unfried, and Sallie Vanderhoof. Without this incredible team of people to support and motivate me, I would not have been able to complete this project. Finally, I would like to thank my participants for their time and generosity in completing this study.

Table of Contents

List of Tables and Figures

INTRODUCTION

Speech recognition testing is used to determine candidacy for cochlear implantation and monitor performance of cochlear implant (CI) recipients. Assessment with speech materials provides an objective measure of speech understanding, which suggests how an individual may perform in everyday environments and allows for the evaluation of equipment functioning. Measures are given at multiple input levels in quiet and noise to effectively assess individuals with different levels of performance and avoid floor or ceiling effects.

A minimum speech test battery designed by Luxford (2001) suggests that postlingually deafened CI patients be evaluated over time utilizing Hearing in Noise Test (HINT) sentences (Nilsson, Soli, & Sullivan, 1994; MSTB, 2011) and the Consonant Nucleus Consonant (CNC) monosyllabic word test (Peterson & Lehiste, 1962). In the last few years, the AzBio Sentence test (Spahr et al., 2012) has been incorporated into CI evaluations, often in addition to or in place of the HINT sentences (Fabry, Firszt, Gifford, Holden, & Koch, 2009). Fabry et al (2009) completed a survey of clinical practices and found most clinicians utilized the CNC word test, but a variety of sentence tests were used. A reason cited for this was to avoid ceiling effects as CI recipients' performance improved (Fabry et al., 2009). As the patient adapts to electrical hearing and grows more accustomed to the sound, they are typically able to perform more difficult sentence testing. Therefore, to be clinically useful, a speech recognition test should be efficient to administer and allow for the evaluation of multiple skill levels, in order to avoid floor or ceiling effects.

The CNC test is a commonly utilized monosyllabic word test. It was developed by Peterson and Lehiste in 1962 and is designed to have equal phonemic distribution throughout the English language. There are a total of 10 lists each containing 50 words with approximately the

same set of phonemes distributed throughout. It is considered a contextually low test because it presents a single monosyllabic word, with only a carrier word (ready). A low context word test is commonly utilized in conjunction with a sentence test to provide maximum information about the recipient's performance (Fabry et al., 2009). Similarly, a sentence test is used to determine traditional CI candidacy per the FDA guidelines (Luxford, 2001).

Test materials most widely recognized and currently utilized are everyday sentences. These sentences are designed to sound like speech that the listener may hear in everyday environments, for example the HINT sentences (Nilsson et al., 1994). These sentences mimic common speech using short sentences (6 to 8 words) at a first grade reading level presented by a single male speaker with slow speaking rate; this makes comprehension and repetition of the vocabulary used in the sentences easy for adult CI recipients (Nilsson et al., 1994). They are scored based on key word recognition. The HINT sentences were adapted from the Bamford-Knowl-Bench (BKB) Sentences, which utilize familiar nouns and verbs from the British language. The HINT consists of 250 sentences placed in twenty-five phonemically balanced sentence lists with ten sentences per list (Nilsson et al, 1994). The HINT can be given as a fixed or adaptive test. In fixed noise testing, the signal-to-noise ratio (SNR) is static with the speech and noise at certain input levels. A common SNR for testing CI recipients is +8 dB SNR, but often ranges from +5 to +10 dB SNR. An adaptive noise test allows for the clinician to obtain the SNR required to recognize the sentence a certain percent of the time, typically 50% (Nilsson et al, 1994). The adaptive condition of the test adjusts the SNR during testing, which allows more flexibility prior to reaching a ceiling. Normal hearing individuals often perform at a negative SNR indicating that the noise can be louder than the signal (Nilsson et al., 1994), where CI recipients have been reported scoring above 10 dB SNR on average (Plyler, Bahng, & von

Hapsburg, 2008).   When testing is performed in noise, "HINT noise" is typically used, which is a speech spectrum shaped noise developed specifically for use with this test in the adaptive condition (Nilsson et al., 1994). Testing in the fixed condition can be performed using the HINT spectrum noise, or multi-talker babble noise. However, even in the adaptive condition many CI recipients find this material to be simple due to the use of a single speaker, slow sentence rate, and simple words. Another issue with this type of test material is that there is a limited number of sentence lists available. Therefore, a patient returning for multiple visits may be evaluated on a highly redundant set of test material (Hochmouth et al. 2012).

Another everyday sentence test is the AzBio sentence test (Spahr et al. 2012). This is more challenging than the HINT, allowing for a greater test difficulty, but still poses the issue of a limited number of test lists available. The repetition of sentence lists may allow for memorization, or a ceiling effect resulting in an unrealistic measure of actual performance of the CI recipient. The AzBio sentence test was developed for use in research testing, but was quickly transitioned into clinical use. There are 250 sentences in 33 lists. Sentences are spoken by four different speakers—two male and two female—with a natural rate of speech, and a sentence length varies between 3 to 12 words (Spahr et al, 2012). This test is performed in fixed noise and provides a percent correct score. Four-talker babble (4TB) noise was used in the development and validation of the test and is recommended for the AzBio Sentences clinically to represent everyday conversational noise (Spahr et al, 2012). In a review of the literature performed by Gifford et al in 2008, the authors found that only 1/137 participants reached a ceiling in quiet (Gifford, Shallop, & Peterson, 2008). The AzBio test has been shown to be a valid and reliable tool for clinical and research audiologists (Schafer, Pogue, & Milrany, 2012). However, there are still a limited number of lists available and there is no manner in which to

advance the difficulty of the test, after adding noise.

To address this issue, another type of sentence test has been developed and is widely utilized in Europe. A matrix test allows for testing using a fixed syntactical structure of a sentence and limited contextual cues. In 1982, Hagerman developed the first matrix sentence test in Sweden, from which current versions are based. The formula selects a word from a category (name, verb, number, object and noun) to create a five-word sentence. A word from each category is chosen at random to create the sentence. A list of 10-20 sentences is compiled from these words, ensuring that no sentence is repeated twice (Hagerman, 1982; Houben et al., 2014). The benefits of using a matrix test over the everyday sentence lists are that the sentences have low redundancy and a large sample of sentences is available reducing repetition of sentences to the listener. The matrix test requires training time prior to the start of testing because the sentences are not common. Studies have shown that without ample training, the final score may be negatively impacted indicting worse performance than the patient's true performance level (Hagerman 1982; Hey, Hocke, Hedderich, & Muller-Deile, 2014; Hochmuth et al., 2012; Wagener, Josvassen, & Ardenkjaer, 2003).

Since the implementation of the Hagerman and Oldenburg sentences, various other matrix sentence tests have been utilized in other languages. The DANTALE II is for Danish speakers, developed by Wagener et al. in 2003. This test was based off the Oldenburg sentences, but can be tested with various types of noise. A mean SNR of -8.43 dB SNR was reported for normal hearing individuals by Wagener and colleagues (Wagener et al., 2003), which is better than the SNR reported for normal hearing individuals on the HINT sentences (Nilsson et al., 1994). Matrix tests have also been developed and cited for Dutch, Polish, Spanish, Finnish, and French speakers (Dietz et al., 2014; Houben et al., 2014; Hochmuth et al., 2012; Jansen et al.,

2012; Ozimek, Warzybok, & Kutzner, 2010; Theelen-van den Hoek, Houben, & Dreschler, 2014).

Hey and colleagues (2014) tested the Oldenburg sentences with cochlear implant recipients to determine the test-retest reliability, SNR function, and establish criteria for performance outliers. Participants were given 30 training sentences in quiet at 65 dB SPL. An adaptive noise procedure was used with Oldenburg speech-spectrum shaped noise, starting at different SNRs (+5, +3, +2, +1, 0, -1, and -2) to create discrimination functions of the various input levels. The results showed good test-retest reliability. The average SNR was -0.54 for CI recipients (Hey et al., 2014), which is better than reported SNR levels on the HINT sentences.

Recently, the HörTech™ group has released matrix sentence tests in various languages, including American English (HörTech, 2014). The American English Matrix Test (AMT) follows the basic matrix structure that consists of a name, verb, number, adjective, and noun. It has 36 lists, each containing twenty sentences. The AMT can be given in quiet and in noise, and can also be given as a closed set test for patients who have limited speech understanding (Hochmuth et al., 2012). This allows for patients at various levels of performance to be tested utilizing the same test battery.

The purposes of this study were to 1) evaluate speech recognition in adult CI recipients using the AMT, 2) determine test-retest reliability of the AMT, 3) evaluate the effects of noise type and noise condition, and 4) compare AMT results with result obtained using more common speech recognition measures.

METHODS

Participants

Seventeen CI recipients (7 males; 10 females) enrolled in the study. The mean age was 64 years (range 47.5-78.2) and mean CI experience was 8.5 years (range 1-26); all participants had at least one year of CI experience. Participants had received either an Advanced Bionics (n = 7) or a Cochlear America's (n = 10) device. Tables 1 and 2 provide individual demographic information. Inclusion criteria for this study required patients to have open set word recognition in quiet defined as a CNC score of greater than 30% on a single list. Other criteria were speakers of American English and the ability to repeat sentences verbally. Patients with bilateral devices (n=5) were evaluated while wearing one device such that all participants were tested unilaterally. Two participants (P12 and P14) were unable to complete the testing in noise; therefore their data were excluded from the analysis.

The study protocol was approved by the Human Research Protection Office at Washington University School of Medicine (IRB # 201411083). Participants were recruited from the Washington University School of Medicine Cochlear Implant Program and were reimbursed for their time and travel.

Procedures and Test Materials

Participants completed two test sessions, approximately 2-4 weeks apart. Testing procedures included: sound field thresholds, CNC words, AzBio sentences in quiet and noise, and AMT sentences in quiet and noise. All testing was completed with the participants' using their preferred everyday program, volume, and sensitivity levels. Prior to speech recognition testing, sound field thresholds were obtained from .25-6 kHz using warble tones to verify CI function.

Two lists of CNC words were presented in Quiet at 60 dB SPL. Two lists of AzBio sentences were presented in two conditions: 1) Quiet at 60 dB SPL, and 2) Fixed SNR (sentences at 60 dB SPL at +8 dB SNR) with 4TB Noise. Three lists of AMT sentences were presented in a variety of conditions including: 1) Quiet at 60 dB SPL, 2) Fixed SNR (sentences at 60 dB SPL at +8 SNR) with 4TB Noise, 3) Fixed SNR (sentences at 60 dB SPL at +8 SNR) with AMT Noise, 4) Adaptive SNR (noise at 60 dB SPL) with 4TB Noise, and 5) Adaptive SNR (noise at 60 dB SPL) with AMT Noise. AMT sentences were from the HörTech™ International Matrix Test (HörTech, 2014).

All tests (CNC, AzBio, or AMT) were presented in a randomized order. First, test order was randomized, then test conditions (quiet, noise, noise type) within each test were randomized. For example, the AMT had Fixed SNR and Adaptive SNR test conditions, as well as the different noise types in each condition. Therefore, the order of testing for AMT was assigned as Fixed SNR or Adaptive SNR and then the order of noise was randomly assigned.

All testing was completed in a 10' by 10' IAC double walled sound booth. Sentence materials were presented using a Dell 530 computer with a Lynx Studio Technology Inc. L22 soundcard installed with the use of an ART SLA1 amplifier. The sentences were played through a Urei 809 speaker at $0°$ azimuth from the listener. Calibration was completed prior to the start of the study. Calibration measurements were made using a Bruel and Kjaer 2230 sound level meter with a 1/3rd octave band filter attachment type 1625. An RMS detector method was used with slow time weighting. The selected frequency weighting was a band-pass filter from 20 Hz − 20 kHz.

Data Analysis

All comparisons were determined *a priori*. Repeated measures analysis of variance (ANOVA) was used to analyze differences among sessions and lists. There were three lists presented for each of the five AMT conditions during each session. There were two lists presented for each AzBio condition and the CNC words for each session. A repeated measures ANOVA was also performed to compare the two tests (AMT and AzBio) that were administered in the same two noise conditions (Quiet and with 4TB Noise). Results were assessed and corrected for sphericity using Mauchly's Test. Post hoc pair-wise analyses were adjusted for multiple comparisons using Bonferonni corrections. Bonferonni adjusted, two-tailed, paired t-tests analyzed differences between the 4TB Noise and AMT Noise for both the Fixed SNR and Adaptive SNR test conditions and for session one and two CNC scores. Significance was set at 0.05 for all analyses.

RESULTS

Figure 1 shows group mean soundfield thresholds ($\pm$ 1 SD) for tested frequencies averaged across Sessions 1 and 2. Group mean thresholds were below 30 dB HL across the frequency range. Thresholds were comparable between sessions and verified audibility and proper equipment function prior to speech recognition testing.

A two (Session) by three (List Order) repeated measures ANOVA was completed for each AMT condition. As shown in Figure 2, a significant learning effect was identified for AMT sentences presented at a Fixed SNR with 4TB Noise, had a significant list order effect [$F_{(2, 28)}=17.35$, $p<0.001$]. Bonferroni adjusted post-hoc pairwise comparisons indicated results for List 1 were significantly poorer than List 2 ($p<0.05$) and List 3 ($p<0.001$); however, Lists 2 and 3 were not significantly different from each other ($p>0.05$). The means for Lists 1 (purple bar), 2

(teal bar), and 3 (green bar) were 78.9% (SD=10.9), 83.4% (SD=8.6), and 85.5% (SD=8.9), respectively. No other AMT conditions showed a significant learning effect for list order.

Three AMT conditions—Fixed SNR with AMT Noise, Adaptive SNR with 4TB Noise, and Adaptive SNR with AMT Noise—had a significant session effect. Figure 3 shows the mean group scores for the AMT in Quiet and the two Fixed SNR conditions (4TB Noise and AMT Noise) collapsed across lists for Session 1 (blue bars) and Session 2 (red bars). The mean score for the AMT in Quiet for Session 1 was 93.3% (SD=4.5) and for Session 2 94.8% (SD=4.8). The mean scores for the Fixed SNR with 4TB Noise for Session 1 was 81.1% (SD=10.8) and for Session 2 was 84.2% (SD=10.1). The scores for AMT in Quiet and the Fixed SNR with 4TB Noise were not significantly different across sessions (p>0.05). The mean scores for the Fixed SNR in AMT Noise for Session 1 was 88.6% (SD=6.8) and for Session 2 was 91.8% (SD=6.1). There was a significant session effect for the Fixed SNR in AMT Noise [$F(1, 14)=6.95$, $p<0.05$], showing Session 2 had better scores than Session 1.

The mean scores for Session 1 (blue bars) and Session 2 (red bars) for the AMT Adaptive Conditions—Adaptive SNR with 4TB Noise and Adaptive SNR with AMT Noise—are shown in Figure 4. In the Adaptive Conditions, a lower score indicates better performance. The mean scores for AMT Adaptive SNR with 4TB Noise collapsed across lists for Session 1 was 3.7 dB SNR (SD=2.0) and for Session 2 was 2.5 dB SNR (SD=2.3). The mean scores for AMT Adaptive SNR with AMT Noise collapsed across lists for Session 1 was -0.1 dB SNR (SD=1.9) and for Session 2 was -1.4 dB SNR (SD=2.1). There was a significant session effect for the both the Adaptive SNR in 4TB Noise [$F(1, 14)=7.08$, $p<0.05$] and Adaptive SNR in AMT Noise [$F(1, 14)=9.37$, $p<0.05$]. In these conditions, participants' scores improved from Session 1 to Session 2.

A two (Session) by three (List Order) repeated measures ANOVA was completed for the AzBio sentences in Quiet and with 4TB Noise. Figure 5 shows the mean scores for the AzBio sentences in Quiet and with 4TB Noise for Session 1 (blue bars) and Session 2 (red bars). No significant list order effect was identified in either the Quiet or 4TB Noise ($p > 0.05$). There was a significant session effect for AzBio sentences in Quiet [$F(1, 14) = 6.88$, $p < 0.05$]. The mean scores for AzBio in Quiet collapsed across lists for Session 1 was 78.2% (SD=13.9) and Session 2 was 82.1% (SD=10.6). Scores improved from Session 1 to Session 2 in Quiet. The means for AzBio with 4TB Noise collapsed across lists for Session 1 was 29.9% (SD=13.0) and for Session 2 was 33.6% (SD=14.2). There was no significant session effect for AzBio with 4TB Noise ($p > 0.05$).

Figure 5 also shows the mean scores for the CNC words collapsed across lists for Session 1 (blue bars) and Session 2 (red bars). A paired t-test comparing CNC results between sessions was completed. An analysis for list effect was not completed as the CNC lists were presented as paired lists based on the findings of Skinner et al., 2006. The mean CNC score for Session 1 was 65.6% (SD=13.5) and for Session 2 was 65.0% (SD=12.3). These scores were not significantly different between sessions ($p > 0.05$).

A two-tailed, paired t-test comparing the two different noise types used in the AMT test (4TB Noise and AMT Noise) in both the Fixed SNR and Adaptive SNR conditions was completed. Figure 6 shows the difference in scores for the AMT Fixed SNR with 4TB Noise (blue dotted bar) and AMT Noise (purple striped bar). Scores are reported as means averaged across lists and sessions. The mean for the AMT Fixed SNR with 4TB Noise was 82.6% (SD=9.2) and AMT Fixed SNR with AMT Noise was 90.2% (SD=6.0). Scores were significantly better with AMT Noise compared to 4TB Noise [$t(14) = -4.96$, $p < 0.001$]. Figure 7 shows the difference between scores for the AMT Adaptive SNR with 4TB Noise (blue dotted

bar) and AMT Noise (purple stripped bar). Again, scores are reported as means averaged across lists and sessions. The mean for the AMT Adaptive SNR with 4TB Noise was 3.1 dB SNR (SD=2.0) and AMT Adaptive SNR with AMT Noise was -0.8 dB SNR (SD=1.8). Scores were significantly better with AMT Noise compared to 4TB Noise [t(14)=20.78, p<0.001]. 4TB Noise produced poorer scores in both the Fixed SNR and Adaptive SNR conditions.

A two (Test) by two (Noise Condition) repeated measures ANOVA compared AMT to AzBio sentences scores in Quiet and AMT to AzBio sentence scores with Fixed SNR 4TB Noise. Figure 8 shows the mean scores for the AMT in Quiet (green bar) and with 4TB Noise (green dotted bar) and the AzBio sentences in Quiet (maroon bar) and with 4TB Noise (maroon dotted bar). Scores are reported as means averaged across all sessions and lists. The mean for the AMT in Quiet was 94.1% (SD=4.2) and AzBio sentences in Quiet was 80.2% (SD=12.0). The mean for the AMT Fixed SNR with 4TB Noise was 82.6% (SD=9.2) and AzBio sentences with 4TB Noise was 31.8% (SD=13.0). There was a significant difference between AMT and AzBio sentences [F(1, 14)=240.22, p<0.001]. Scores in noise were also significantly poorer than those in quiet on both tests [F(1, 14)=359.60, p<0.001]. There was also a significant interaction between the sentence test and noise condition [F(1,14)=291.86, p<0.001]. Specifically, 4TB Noise had a greater impact on AzBio sentences than on the AMT sentences.

The following moderators were examined for correlations with test measures—age, years of hearing loss, years of severe-profound hearing loss, years of hearing aid use prior to implantation, and years of CI use. No correlations were significant (p>0.05).

DISCUSSION

Speech recognition testing is utilized with CI recipients to determine candidacy and monitor performance. Different types of speech tests are typically used for these evaluations, including word and sentence tests presented in quiet and noise. The use of a battery of measures provides a comprehensive evaluation to better understand real world performance.

Word tests provide a good measure of speech recognition, but are a low context and usually given in quiet. Sentence tests evaluate speech recognition with contextual cues and are often presented in quiet and in background noise. The addition of background noise makes the test more difficult. The AzBio sentence test was recently developed and designed to be more challenging than previous sentence tests (Fabry et al., 2009; Spahr et al, 2012). It is considered an everyday sentence test, because the content mimics everyday speech and language.

A matrix test is another type of speech recognition test. Sentences do not mimic natural speech and it uses a fixed syntactical structure where each sentence is comprised of a name, verb, number, object and noun. It does not have the same contextual cues as everyday sentences because it would be difficult to predict the next word based on the previous word, giving the test a low internal redundancy (Hagerman, 1982; Houben et al., 2014). The matrix selection formulates a large number of sentence lists without repetition. Studies have shown learning effects with this test format. Therefore, without proper training on the test material, a participant's score may be negatively impacted (Hagerman 1982; Hey et al., 2014; Hochmuth et al., 2012; Wagener et al., 2003).

The American English Matrix Test (AMT) is a recently developed matrix test. It was developed as the English version to other available matrix tests which are commonly used in Europe. The current study used the AMT to evaluate experienced CI recipients. The AMT was

12

given in five different test conditions, including quiet, with two different types of noise, and two different types of test formats. Three test lists were given in each of the AMT conditions, and scores were analyzed for learning effects and list differences. Prior to testing with AMT, two practice lists were provided to each participant in the AMT Adaptive SNR with AMT Noise to control for effects of training previously mentioned. List differences were found for only one condition, the AMT Fixed SNR with 4TB Noise, where List 2 and List 3 scores were significantly higher than List 1, but List 2 and List 3 scores were not different from each other. Therefore, a learning effect was present for the AMT Fixed SNR with 4TB Noise, which suggests that the AMT with a Fixed SNR required additional practice. Similar training effects were found by Hagerman (1982), Hey et al. (2014), Hochmuth et al. (2012), and Wagener et al. (2003) in various studies of Matrix Tests in other languages. No learning effects were found for the other AMT conditions, indicating that the two training lists provided at the beginning of the testing were enough to achieve maximum score. In addition, there was no learning effect for the AzBio sentences as there was no significant difference between lists. This is consistent with what Spahr et al. (2012) report in their findings, validating the list equivalency of the AzBio.

All tests were given at two different sessions, approximately 2-4 weeks apart to examine test-retest reliability. Small but significant session differences were found for the AMT in the Fixed SNR with AMT Noise condition and the AzBio in Quiet condition. Scores improved from Session 1 to Session 2 for these two tests. The AMT in Fixed SNR with AMT Noise had a test-retest difference of 3.2 percentage points. The AzBio sentences in Quiet had a test-retest difference of 3.9 percentage points. However, Spahr et al. (2012) noted for the AzBio sentences that when two lists were given, a change of greater than 10 percentage points was significant

based on 95% confidence interval. This indicates that the test difference shown in this study was not clinically meaningful.

Significant session differences were also found for the two adaptive AMT conditions—4TB Noise and AMT Noise. In the adaptive condition, the noise stayed at a constant level, but the speech level was adjusted so the SNR constantly changed to achieve a 50% threshold level. The test-retest difference between sessions were similar for both types of noise, with a difference of 1.2 dB SNR for the AMT Adaptive SNR with 4TB Noise and a difference of 1.3 dB SNR for the AMT Adaptive SNR with AMT Noise. Hey et al. (2014) reported test-retest difference of 1.1 dB SNR for CI recipients with the Oldenburg Matrix test in Oldenburg noise. The findings are almost identical to the current study, for both types of noise. This may indicate that even though a difference between sessions was found, there is still good test-retest reliability for the AMT Adaptive SNR with both types of noise.

There were no list or session effects reported for the AMT in Quiet, AMT in Fixed SNR with AMT Noise, or CNC words. The CNC test was challenging for participants, with an average score of 65.3%, probably due to the low context. The current study's CNC scores were similar to the 61.5% reported by Holden et al. (2013) for CI recipients on CNC words (Holden, et al., 2013).

Previous research on matrix tests has primarily utilized the test in the adaptive format. The German version of the Matrix Test, the Oldenburg Sentence Test utilized with Oldenburg Spectrum Noise showed an average SNR of -0.54 dB SNR for CI recipients (Hey et al., 2014). This is similar to the SNR of -0.8 dB SNR found in the current study for the Adaptive SNR with AMT Noise. There was a significant difference in SNRs between noise types. The Adaptive SNR with 4TB Noise average was 3.12 dB SNR. Even in a Fixed SNR, there was a difference of

7.6 percentage points between the noise types, with 4TB Noise showing poorer scores. Therefore, the 4TB Noise was significantly more difficult than the AMT Noise. This suggests that the 4TB Noise is more difficult and perhaps a more effective masker.

The AMT in quiet and fixed noise was compared to common speech recognition measures used in the United States. Results indicated that scores on the AMT Fixed SNR were significantly higher than the AzBio sentences for both Quiet and 4TB Noise conditions. Mean scores in Quiet for the AMT in Quiet were 13.9 percentage points better compared to the AzBio sentences, indicating that the AMT test was significantly easier. The same was seen for the AMT in Fixed SNR with 4TB Noise with mean scores 50.8 percentage points better than AzBio sentences with 4TB Noise. Noise significantly degraded performance on both tests, but the difference between quiet and 4TB Noise scores were larger for the AzBio sentences (48.4 percentage points) than for the AMT (11.5 percentage points). This was seen as a significant interaction between Noise and sentence test, indicating that 4TB had a greater impact on AzBio sentences than on the AMT. Again, this shows that the AzBio is a more difficult test, particularly when utilized in 4TB background noise.

CONCLUSIONS

The AMT manual recommends two lists be given for training and then presentation of three lists (60 sentences) to avoid learning effects (HörTech, 2014). For one AMT condition, participants achieved their maximum score on the third list. The practice lists and three test lists makes the AMT more time consuming to give, and may not be as efficient as other measures for clinical use.

The type of noise significantly affected performance for both the fixed and adaptive conditions and should, therefore, be considered when administering and/or comparing

performance on the AMT. The 4TB Noise increased the difficulty of the AMT compared to the AMT noise, which was designed to be used with the AMT test.

The adaptive AMT scores obtained in the current study agree well with previous research on matrix tests in other languages. This would allow comparison of research among international research facilities. The AMT Fixed SNR in Quiet and with 4TB Noise were found to be significantly easier compared to AzBio sentences. The AMT Fixed SNR is likely approaching a ceiling effect. Individual participants were able to reach scores of 100% in quiet and with both types of noise. The AzBio sentences, on the other hand, provide room for improvement, especially in noise. Therefore, the use of AMT in quiet and fixed noise may not allow for evaluation of better performing CI recipients.

FUTURE RESEARCH

Specific list differences should be evaluated, to ensure that all 36 sentence lists are equivalent. It is possible that the AMT, specifically with Adaptive SNR will be beneficial for clinical use in a patient population that does not perform well on speech recognition, such as pre-lingually deafened adults or early CI recipients (less than 1 year experience). Studies should also look at the utilization of this test with children, to assess if it may be appropriate for use with pediatric populations.

REFERENCES

Dietz, A. & Buschermohle, M. & Aarnisalo, A.A. & VanHanen, A. & Hyyrynen, T. & Aaltonen, O., et al. (2014, February 10). The development and evaluation of the Finnish Matrix Sentence Test for speech intelligibility assessment. *Acta Oto-Laryngologica, 134,* 728-737.

Fabry, D. & Firszt, J.B. & Gifford, R.H. & Holden, L.K., & Koch, D. (2009, May/June). Evaluating speech perception benefit in adult cochlear implant recipients. *AudiologyToday, 21(3),* 36-43.

Gifford, R.H. & Shallop, J.K., & Peterson, A.M. (2008, January 22). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiology & Neurotology, 13,* 193-205.

Hagerman, B. (1982, n.d.). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology, 11,* 79-87.

Hey. M. & Hocke, T. & Hedderich, J. & Muller-Deile, J. (2014, June 22). Investigation of a matrix sentence test in noise: reproducibility and discrimination function in cochlear implant patients. *International Journal of Audiology, 53(12),* 895-902.

Hochmuth, S. & Brand, T. & Zokoll, M.A. & Castro, F.Z. & Wardenga, N., & Kollmeier, B. (2012, February 24). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology, 51,* 536-544.

Holden, L.K. & Finley, C.C. & Firszt, J.B. & Holden, T.A. & Brenner, C. & Potts, L.G., et al. (2013, n.d.). *Ear & Earing, 34(3),* 342-360.

Houben, R. & Koopman, J. & Luts, H. & Wagener, K.C. & van Wieringen, A. & Verschuure, H., et al. (2014, April 26). Development of a Dutch Matrix sentence test to assess speech intelligibility in noise. *International Journal of Audiology, 53(10),* 760-763.

HörTech gGmbH. International Matrix Test, reliable speech audiometry in noise. (2014, n.d.).

Jansen, S. & Luts, H. & Wagener, K.C. & Kollmeier, B. & Rio, M.D. & Dauman, R., et al. (2012, n.d.). Comparison of three types of French speech-in-noise tests: a multi-center study. *International Journal of Audiology, 51*, 164-173.

Luxford, W.M. (2001, n.d.). Minimum speech test battery for postlingually deafened adult cochlear implant patients. *Otolaryngology-Head and Neck Surgery, 124,* 125-126.

Minimum speech test battery (MSTB, June n.d.) for adult cochlear implant users. (2011).

Nilsson, M. & Soli, S.D. & Sullivan, J.A. (1994, February n.d.). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95(2),* 1085-1099.

Ozimek, E. & Warzybok, A. & Kutzner, D. (2010, February 2). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology, 49,* 444-454.

Peterson, G.E. & Lehiste, I. (1962, February n.d.). Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders, 27(1),* 62-70.

Plyler, P.N. & Bahng, J. & von Hapsburg, D. (2008, April n.d.). The acceptance of background noise in adult cochlear implant users. *Journal of Speech, Language, and Hearing Research, 51,* 502-515.

Schafer, E. & Pogue, J. & Milrany, T. (2012, November 7). List equivalency of the AzBio

    sentence test in noise for listeners with normal-hearing sensitivity or cochlear implants.

    Journal of the American Academy of Audiology, 23, 501-509.

Spahr, A.J. & Dorman, M.F. & Litvak, L.M. & Wie, S.V. & Gifford, R.H. & Loizou, P.C., et al.

    (2012, n.d.). Development and validation of the AzBio sentence lists. *Ear & Hearing,*

    33(1), 112-117.

Theelen-van den Hoek & F.L., Houben, R. & Dreschler, W.A. (2014, May 4). Investigation into

    the applicability and optimization of the matrix sentence test for use with cochlear

    implant users. *International Journal of Audiology, 53(11),* 817-828.

Wagener, K. & Josvassen, J.L. & Ardenkjaer, R. (2003, n.d.). Design, optimization and

    evaluation of a Danish sentence test in noise. *International Journal of Audiology, 42,* 10-

    17.

Participant	Gender	Age (years)	Length of HL (years)	Length of SP Loss (years)	Length of HA use (years)
P1	F	62.4	49	10	3
P2	F	64.5	40	29	40
P3	F	56.1	20	1	17
P4	F	57.8	41	5	40
P5	M	62.7	50	40	2
P6	M	47.5	36	35	35
P7	F	60.2	21	21	0
P8	F	68.9	60	38	53
P9	F	74.8	26	26	26
P10	M	65.8	27	1	20
P11	F	56.1	55	5	4
P12	M	55.3	39	1	39
P13	M	58.6	38	10	29
P14	F	76.3	46	8	10
P15	F	72.3	4	10	4
P16	M	75.8	19	15	30
P17	M	78.2	51	24	0
Mean		**64.3**	**36.6**	**16.4**	**20.7**

HL-hearing loss, SP-severe-to-profound, HA-hearing aid

Table 1. Demographic information for each participant . P12 and 14 (highlighted in red) did not complete Session 2 and their test data were not utilized in the analysis.

Participant	Ear Tested	Everyday Modality	Length of CI use (years)	Device Company	Processor
P1	L	Bimodal	3.6	Cochlear Nucleus	CP810
P2	L	Bilateral	11	Cochlear Nucleus	CP910
P3	R	Bilateral	6.9	Cochlear Nucleus	Freedom
P4	L	Bilateral	10	Cochlear Nucleus	CP810
P5	R	Unilateral	2.3	Advanced Bionics	Naida Q70
P6	L	Bimodal	1.9	Advanced Bionics	Naida Q70
P7	L	Bilateral	8	Advanced Bionics	Naida Q70
P8	L	Bimodal	7.8	Advanced Bionics	Harmony
P9	R	Unilateral	10.1	Advanced Bionics	Harmony
P10	L	Bimodal	3.4	Advanced Bionics	Harmony
P11	L	Unilateral	13	Cochlear Nucleus	CP910
P12	L	Bimodal	10.5	Cochlear Nucleus	Freedom
P13	L	Bilateral	3.8	Cochlear Nucleus	CP810
P14	R	Unilateral	18	Cochlear Nucleus	Freedom
P15	R	Unilateral	26.3	Cochlear Nucleus	Freedom
P16	L	Bimodal	7.8	Cochlear Nucleus	CP910
P17	L	Bimodal	7.3	Advanced Bionics	Naida Q70
Mean			8.9		

CI-Cochlear Implant

Table 2. Cochlear implant (CI) information for each participant. P12 and 14 (highlighted in red) did not complete Session 2 and their test data were not included in the analysis.
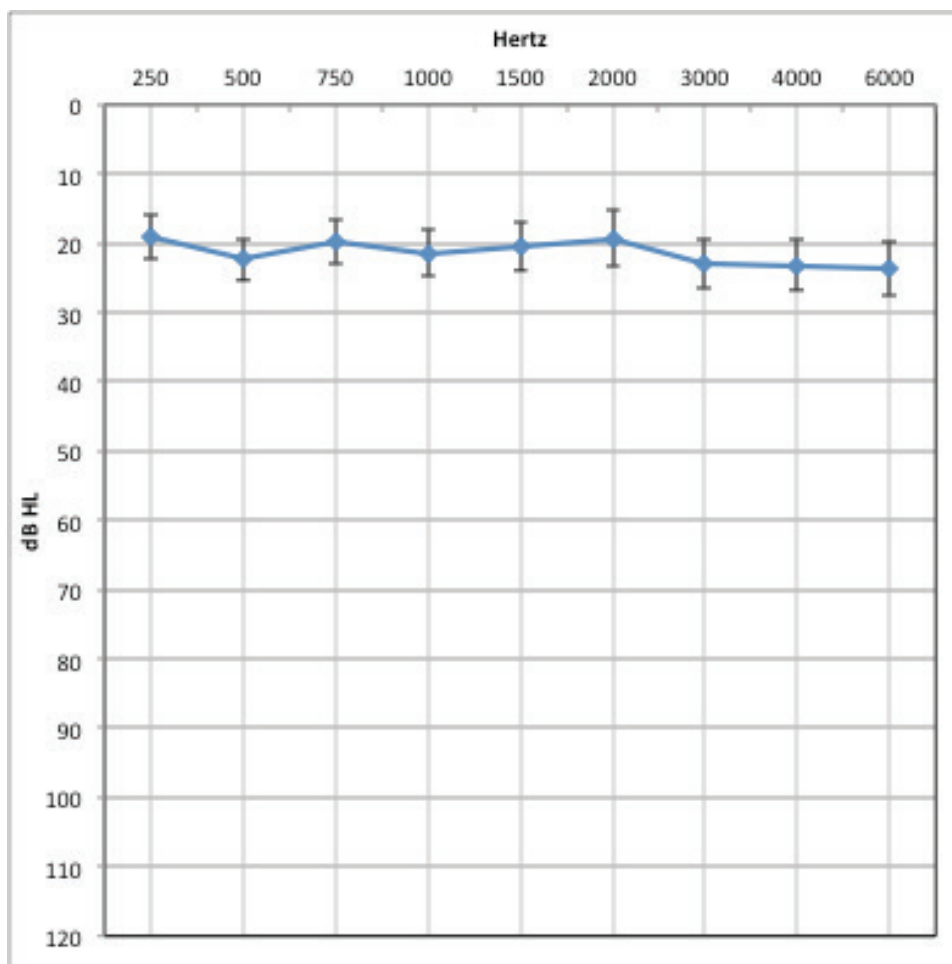
Figure 1. Mean aided soundfield thresholds for the CI ear tested are shown for 250-6000 Hz collapsed across sessions. Error bars are ± 1 SD.
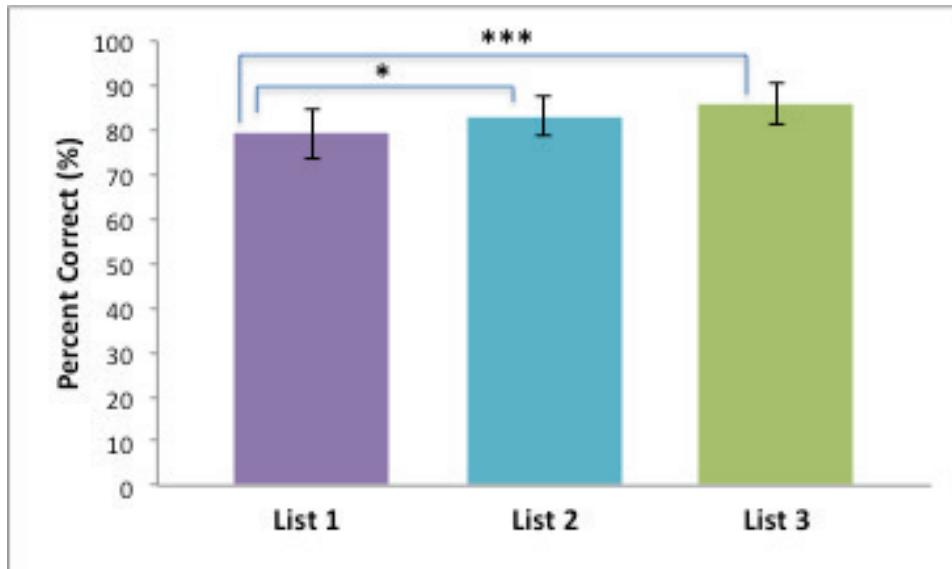
Figure 2. Mean scores for AMT Fixed SNR with 4TB Noise for List 1 (purple bar), List 2 (teal bar) and List 3 (green bar) collapsed across session are shown. Error bars are ± 1 SD, *** $p<0.001$, *$p<0.05$.
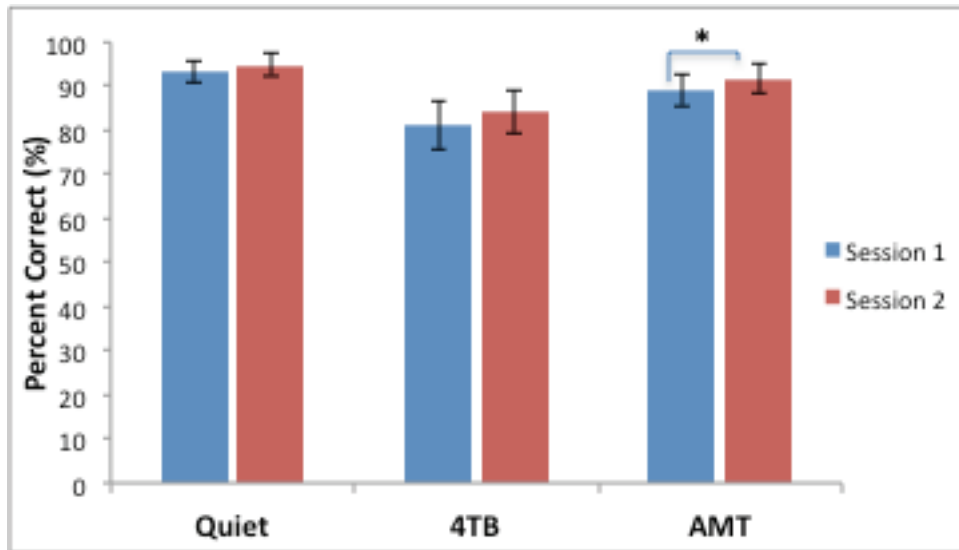
Figure 3. Mean scores for AMT in Quiet, Fixed SNR with 4TB Noise and Fixed SNR with AMT Noise collapsed across lists are shown for Session 1 (blue bars) and Session 2 (red bars). Error bars are ± 1 SD, *p<0.05.
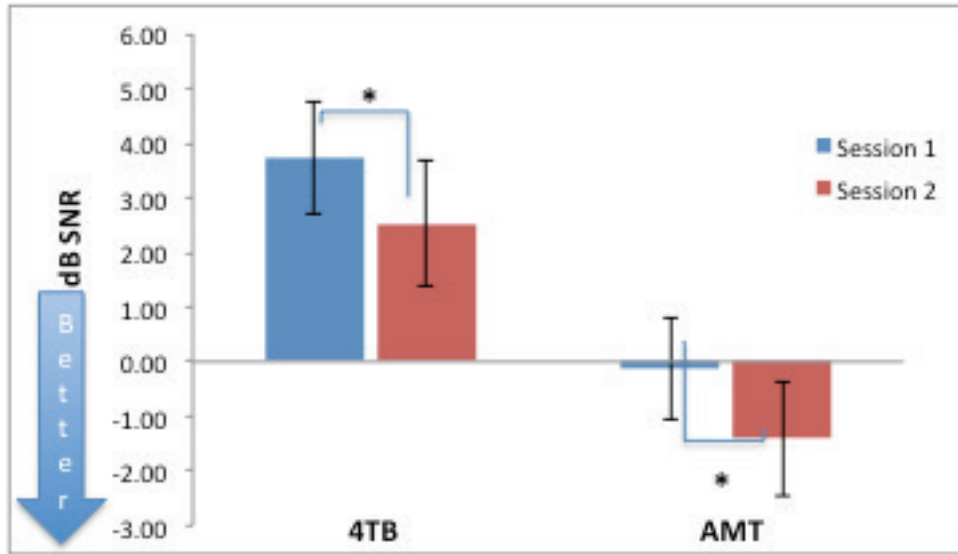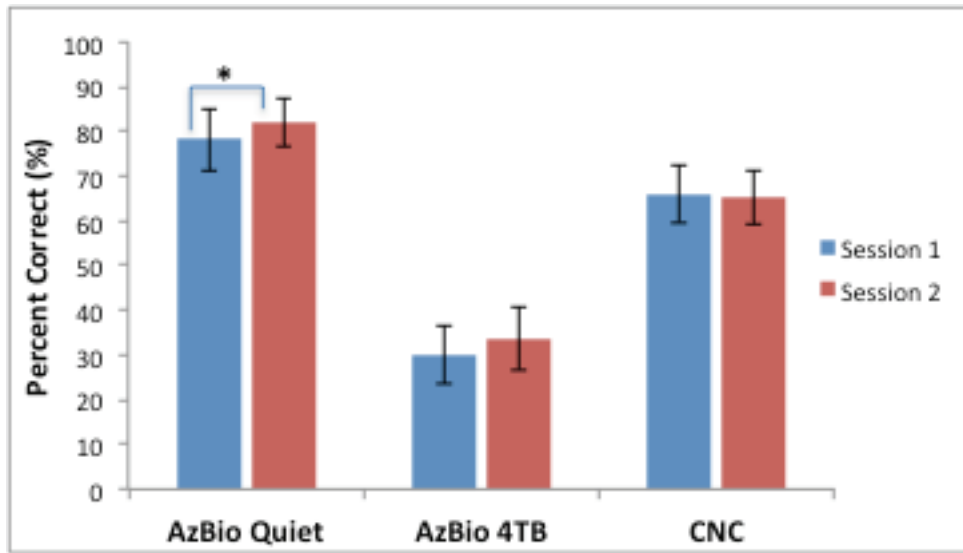
Figure 4. Mean scores for AMT Adaptive SNR with 4TB Noise and Adaptive SNR with AMT Noise collapsed across lists are shown for Session 1 (blue bars) and Session 2 (red bars). Error bars are ± 1 SD, *p<0.05.

Figure 5. Mean scores for AzBio sentences in Quiet and with 4TB Noise and CNC words collapsed across lists are shown for Session 1 (blue bars) and Session 2 (red bars). Error bars are ± 1 SD, *p<0.05.
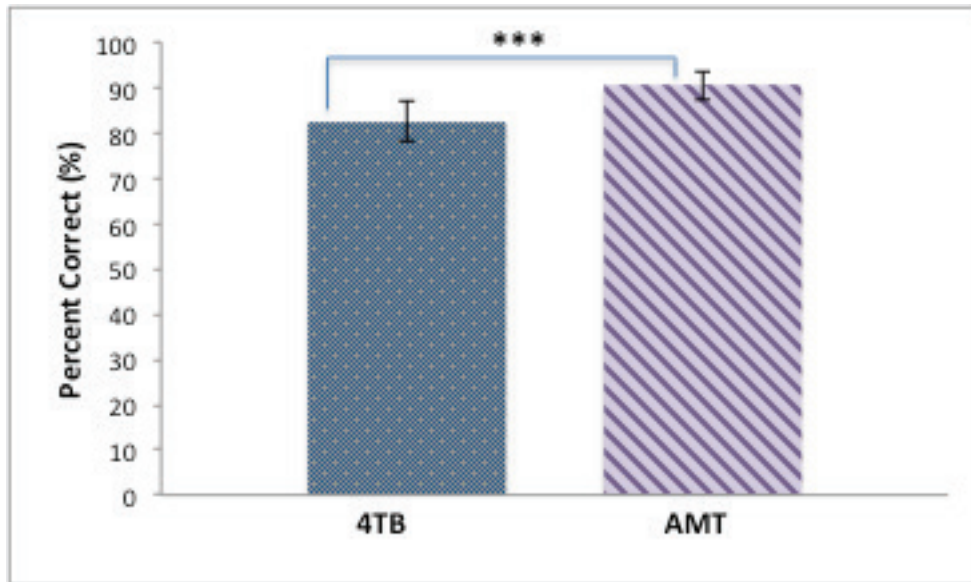
Figure 6. Mean scores for AMT Fixed SNR with both 4TB Noise (blue dotted bar) and AMT Noise (purple striped bar) averaged across lists and sessions are shown. Error bars are ± 1 SD, ***p<0.001.
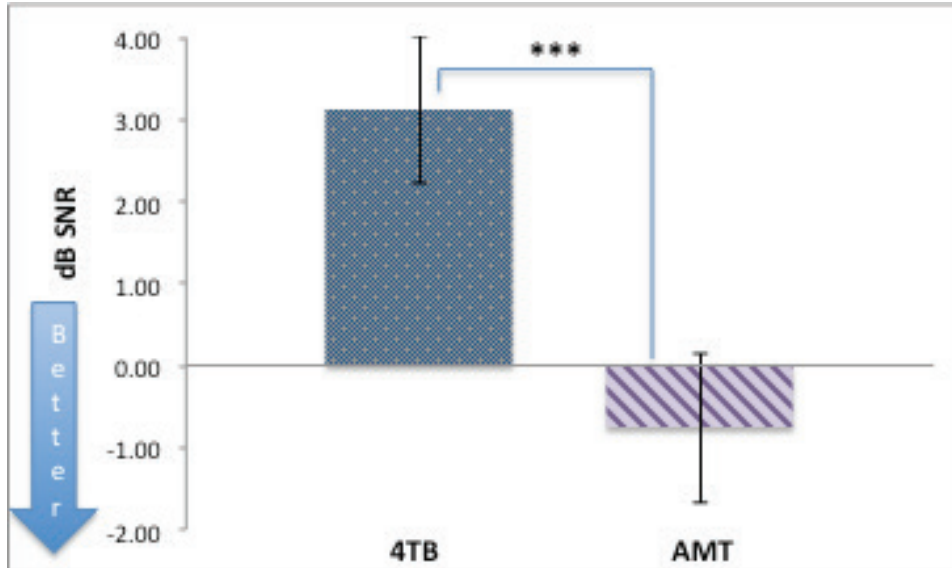
Figure 7. Mean scores for the AMT Adaptive SNR with 4TB Noise (blue dotted bar) and AMT Noise (purple striped bar) averaged across lists and sessions are shown. Error bars are ± 1 SD, *** p<0.001.
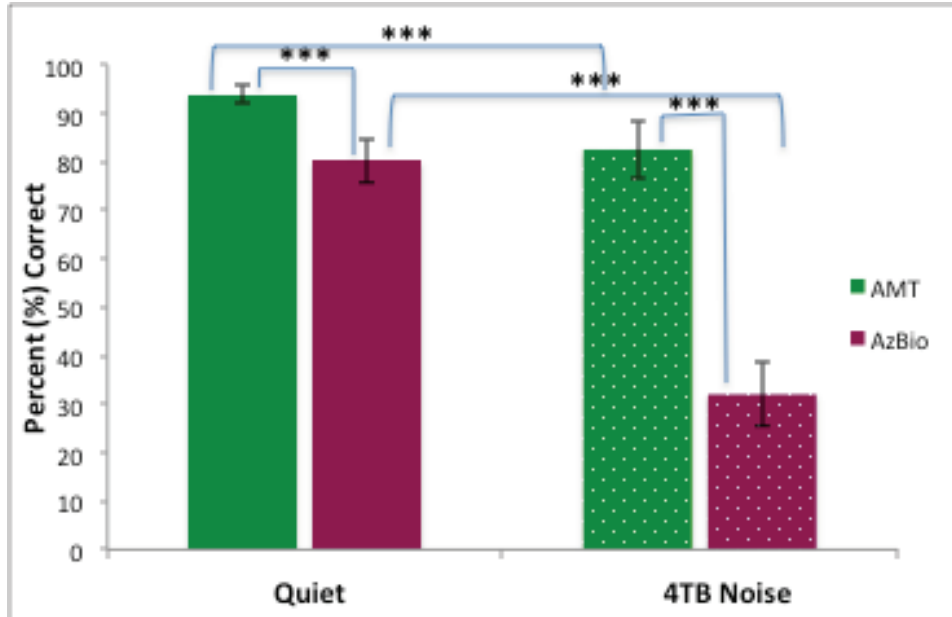
Figure 8. Mean scores for AMT Fixed SNR in Quiet (green bar) and with 4TB Noise (striped green dotted bar) and the AzBio sentences in Quiet (maroon bar) and with 4TB Noise (maroon dotted bar) averaged across all sessions are shown. Error bars are ± 1 SD, ***p<0.001.