

**CONTEXTUAL MOMENTARY ASSESSMENT OF SPEECH-IN-NOISE
LISTENING SITUATIONS AMONG HEARING AID USERS: VALIDITY AND
RELIABILITY**

by

LISE GILLEN

H.BSc., The University of Toronto, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Audiology and Speech Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2015

© Lise Gillen, 2015

Abstract

Currently, all hearing aid benefit outcome measures rely on retrospective self-report, which can often be inaccurate due to memory decay, recollection biases, and the use of cognitive heuristics. Contextual momentary assessment (CMA) involves repeated collection of real-time data on an individual's experience in their natural environment; CMAs circumvent the error and bias related to retrospective assessments, making them more ecologically valid for capturing day-to-day variations in experiences.

The purpose of the present paper was to answer three research questions: (a) Is CMA capable of facilitating valid and reliable evaluations of subjective listening experiences in lab-controlled acoustic conditions?; (b) Is CMA validity and reliability altered significantly by the timing of the CMA relative to the listening event (Experiment I)?; (c) Is CMA validity and reliability altered by the presence of, or focus on a secondary task (Experiment II)?

To address these research questions, this study employed a block-randomized, within-subject design where 12 participants with sensorineural hearing loss were fitted with hearing aid(s), and completed CMA ratings based on listening situations where they performed a sentence repetition task. The study was comprised of two experiments involving three independent variables: (a) speech level; (b) signal-to-noise ratio (SNR); (c) CMA timing (Experiment I), or task focus (Experiment II). CMAs were composed of four rating dimensions: intelligibility, noisiness, listening effort, and loudness.

For the listening situations employed in this lab study, the reliability, construct validity, and criterion validity results were as follows: (a) intelligibility ratings were

reliable, demonstrated construct validity, and had the strongest correlation with intelligibility scores when the CMA was completed after listening situations where there was no secondary task; (b) noisiness ratings were reliable, demonstrated construct validity, and correlated the strongest with measured background noise intensities when rated while experiencing the listening situation; (c) listening effort ratings were unreliable and had questionable construct validity; (d) loudness ratings were reliable, demonstrated construct validity, and correlated the strongest with measured speech intensities when rated while experiencing the listening situation. Based on these results, CMA ratings of intelligibility, loudness, and noisiness, but not listening effort, show potential to be useful for measuring hearing aid benefit.

Preface

This thesis was based on research conducted in the Amplification Research Lab at the University of British Columbia. The study design was created in conjunction with L. Jenstad and G. Singh, and the author L. Gillen was primarily responsible for set-up, participant recruitment, data collection, and data analysis.

This study was reviewed and approved by the Behavioural Research Ethics Board of the University of British Columbia. The certificate number of the ethics certificate obtained is H13-02869.

Table of Contents

Abstract	ii
Preface.....	iv
Table of Contents	v
List of Tables	ix
List of Figures	xi
List of Abbreviations	xvii
Acknowledgements.....	xviii
Dedication	xix
Chapter 1: Literature Review.....	1
1.1 Introduction.....	1
1.2 Audiological outcome measures	2
1.3 Potential issues with currently used outcome measures	2
1.4 Contextual momentary assessment (CMA)	4
1.5 Challenges to CMA usage	5
1.6 CMA validity and reliability.....	5
1.6.1 CMA response timing.....	6
1.6.2 Presence of a secondary task	7
1.7 Purpose.....	8
1.8 Research questions.....	9
Chapter 2: Methodology	10
2.1 Introduction.....	10
2.2 Participants.....	11

2.3 Apparatus and stimuli	14
2.3.1 Auditory stimuli	14
2.3.2 Contextual momentary assessment questionnaire	14
2.3.3 Hearing aids	15
2.3.4 Secondary task	16
2.3.5 Equipment	17
2.3.5 Questionnaires	18
2.4 Procedure	19
2.4.1 Session I (screening)	19
2.4.2 Session II	21
2.4.3 Practice session (Experiment I)	21
2.4.4 Experiment I—timing of ratings	21
2.4.5 Practice session (Experiment II)	23
2.4.6 Experiment II—task focus	23
2.5 Data analysis	25
2.5.1 Overview	25
2.5.2 Criterion validity	25
2.5.3 Construct validity	26
2.5.3 Reliability	27
2.5.4 Secondary task (connect-the-dots) performance	28
Chapter 3: Results	29
3.1 Overview	29
3.2 Experiment I—timing of ratings	31
3.2.1 Intelligibility	31
3.2.2 Noisiness	38
3.2.3 Listening effort	44
3.2.4 Loudness	48
3.3 Experiment II—task focus	52

3.3.1 Connect-the-dots performance.....	52
3.3.2 Intelligibility	55
3.3.3 Noisiness.....	60
3.3.4 Listening effort	65
3.3.5 Loudness	71
Chapter 4: Discussion and Conclusion	77
4.1 Introduction.....	77
4.2 Intelligibility	78
4.3. Noisiness	81
4.4 Listening effort	83
4.5 Loudness	86
4.6 Implications	89
4.7 Strengths and limitations	89
4.8 Conclusion	90
References.....	92
Appendices.....	106
Appendix A Participant audiograms listed by code.....	106
Appendix B Hearing thresholds across participants	118
Appendix C CMA rating instructions.....	119
Appendix D Hearing aid fitting distance from NAL-NL1 target gain.....	120
Appendix E Screenshots of a sample connect-the-dots trial	121
Appendix F Questionnaire package	122
Appendix G Informed consent form	136
Appendix H Hearing history form	139

Appendix I	MoCA test form	140
Appendix J	MoCA administration instructions	141
Appendix K	Loudness discomfort levels methodology.....	146
Appendix L	Sentence repetition keywords correct (%) by participant age.....	147
Appendix M	CMA rating means across participants.....	148

List of Tables

Table 2.1. Participant characteristics listed by increasing age.....	13
Table 3.1. Participant characteristics (age and gender), and their corresponding means and standard deviations for connect-the-dots performance (baseline, focus on sentence repetition, focus on connect-the- dots).....	54
Table 4.1. Summary of results for the intelligibility rating dimension.....	80
Table 4.2. Summary of results for the noisiness rating dimension	82
Table 4.3. Summary of results for the listening effort rating dimension	85
Table 4.4. Summary of results for the loudness rating dimension.....	88
Table L.1. Mean sentence repetition performance and standard deviations for each participant listed by age	147
Table M.1. CMA rating means and 95% confidence intervals of the mean across participants for Experiment I	148

Table M.2. CMA rating means and 95% confidence intervals of the mean
across participants for Experiment II.....149

List of Figures

Figure 3.1. Correlation between intelligibility ratings and sentence repetition percent keywords correct for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B)	32
Figure 3.2. Mean ratings for intelligibility made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL)	35
Figure 3.3. Comparison of mean intraclass correlation coefficients (ICCs) for each rating dimension (intelligibility, noisiness, listening effort, and loudness) made while (W), or after (A) listening to the sentences, and with focus on either the sentence repetition task (SR), or with focus on the secondary connect-the-dots task (CTD)	37
Figure 3.4. Correlation between noisiness ratings and corresponding calibrated background noise intensities for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B)	39
Figure 3.5. Mean ratings for noisiness made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL)	42

Figure 3.6. Correlation between listening effort ratings and sentence repetition percent keywords correct for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B)45

Figure 3.7. Mean ratings for listening effort made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL).....47

Figure 3.8. Correlation between loudness ratings and corresponding calibrated speech intensities for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B)49

Figure 3.9. Mean ratings for loudness made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL).....51

Figure 3.10. Correlation between intelligibility ratings and sentence repetition percent keywords correct for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B)56

Figure 3.11. Comparison of mean ratings for intelligibility made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL).....59

Figure 3.12. Correlation between noisiness ratings and corresponding calibrated background noise intensities for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B).62

Figure 3.13. Mean ratings for noisiness made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL).....64

Figure 3.14. Correlation between listening effort ratings and sentence repetition performance (percent keywords correct) for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B)67

Figure 3.15. Correlation between listening effort ratings and secondary connect-the-dots performance (dots/min) for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B)68

Figure 3.16. Mean ratings for listening effort made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL)70

Figure 3.17. Correlation between loudness ratings and corresponding calibrated speech intensities for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B)73

Figure 3.18. Mean ratings for loudness made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL).....75

Figure A.1. Participant 109’s audiogram.....106

Figure A.2. Participant 155’s audiogram.....107

Figure A.3. Participant 164's audiogram.....	108
Figure A.4. Participant 171's audiogram.....	109
Figure A.5. Participant 223's audiogram.....	110
Figure A.6. Participant 314's audiogram.....	111
Figure A.7. Participant 449's audiogram.....	112
Figure A.8. Participant 505's audiogram.....	113
Figure A.9. Participant 687's audiogram.....	114
Figure A.10. Participant 733's audiogram.....	115
Figure A.11. Participant 881's audiogram.....	116
Figure A.12. Participant 906's audiogram.....	117
Figure B.1. Mean hearing thresholds for the left and right ear across participants.....	118

Figure D.1. Mean distance of hearing aid fittings from NAL-NL1 gain targets across participants	120
Figure E.1. Connect-the-dots task instructions screen.....	121
Figure E.2. Connect-the-dots task testing screen.....	121

List of Abbreviations

APHAB	Abbreviated Profile of Hearing Aid Benefit
BFI	Big Five Inventory
BKB	Bamford-Kowal-Bench
CMA	Contextual Momentary Assessment
COSI	Client Oriented Scale of Improvement
EMA	Ecological Momentary Assessment
GHABP	Glasgow Hearing Aid Benefit Profile
HHIE/A	Hearing Handicap Inventory for the Elderly or for Adults
HL	Hearing level
ICC	Intraclass correlation coefficient
IEEE	Institute of Electrical and Electronic Engineers
IOI-HA	International Outcome Inventory for Hearing Aids
jnd	Just noticeable difference
LDL	Loudness discomfort level
MoCA	Montreal Cognitive Assessment
PTA	Pure-tone average
rms	root mean square
SADL	Satisfaction with Amplification in Daily Life
SEM	Standard error of measurement
SNR	Signal-to-noise ratio
SPL	Sound pressure level

Acknowledgements

Funding for this research project was provided by Mitacs Inc. with Unitron Hearing Ltd as the industry partner.

I would like to first thank my supervisor, Dr. Lorianne Jenstad for her expertise, encouragement, and unwavering support. Her guidance through the process of planning, executing and writing this thesis project was integral to its success. Lorianne, it has been a pleasure to work with you and learn from you. Thank you so much for giving me this opportunity.

Second, I thank my committee members Dr. Gurjit Singh and Dr. Anita DeLongis for their insight and expertise. Their input was incredibly helpful throughout the research process.

Third, I thank my research assistants Flora Pang and Stephanie Kore for contributing so much of their time and skillful attention to detail towards helping me with this thesis project.

Fourth, I am thankful for the support of my family, friends, my fellow St. John's College members, and my fellow AMPLab members Foong Yen Chong and Myron Huen.

Lastly, I would like to thank my audiology classmates: Chelsea, Dorothy, Gill, Hope, Judy, Nicole, Sukaina, and Yiwan. I am so grateful to have met and studied alongside you ladies! You are all such smart, supportive, and wonderful women. I will always cherish the time we spent together at UBC.

Dedication

To all who suffer from hearing loss and those who support them.

Chapter 1: Literature Review

1.1 Introduction

Hearing loss is associated with social isolation, poor cognitive functioning (Lin et al., 2011), incident dementia (Lin et al., 2011), and overall morbidity and mortality (Cacioppo, Hawkley, Norman, & Berntson, 2011), particularly among older adults. Despite the rapid advancement of hearing aid technology, three out of four people with a mild hearing loss, and six out of ten with moderate-to-severe hearing loss do not use hearing aids (Kochkin, 2009). Additionally, only 55% of 3000 hearing aid users surveyed were either “satisfied” or “very satisfied” with their hearing aids (Kochkin, 2010). Whether individuals with hearing loss are satisfied or dissatisfied with their hearing aids is related to the number of personally important listening situations where the user perceives a significant hearing aid benefit, allowing them to reclaim activities that were previously difficult due to their hearing loss (Kochkin, 2010). Accordingly, increasing the probability for user-perceived hearing aid benefit and satisfaction is the primary goal of the hearing aid fitting and fine-tuning process, and demands the provision of a systematic approach, supported by evidence (Valente et al., 2006). The hearing aid fitting and fine-tuning process is typically performed in a controlled, quiet setting, which does not accurately predict the benefit of hearing aid use in everyday situations (Cox, Alexander, & Rivera, 1991; Boymans & Dreschler, 2000). In order to capture hearing aid benefit in everyday situations, outcome measure questionnaires are used to assess and document the difference between the wearer’s auditory performance or experience after several weeks or months of hearing aid use and their auditory performance or experience before hearing

aid fitting (Humes, 2001). Due to their importance in evaluating hearing aid satisfaction, benefit and usage, outcome measures are of interest to clinicians, hearing aid providers, researchers, hearing aid manufacturers, third-party payers, and hearing aid wearers (Humes, 2003).

1.2 Audiological outcome measures

Outcome measures are typically performed objectively and subjectively as a part of the hearing aid fitting process. Commonly used objective outcome measures focus on measuring an increase in speech audibility or intelligibility in everyday listening situations (Valente et al., 2006). There are many subjective self-report outcome measures available for use in clinic, addressing areas such as auditory disability (activity limitation) with and without hearing aids (e.g. Abbreviated Profile of Hearing Aid Benefit [APHAB] [Cox & Alexander, 1995]), auditory handicap (participation restriction) with and without hearing aids (e.g. the Hearing Handicap Inventory for the Elderly or for Adults [HHIE/A] [Newman, Weinstein, Jacobson, & Hug, 1990; Ventry & Weinstein, 1982]), patient satisfaction with hearing aids (e.g. Satisfaction with Amplification in Daily Life [SADL] [Cox & Alexander, 1999]), or several dimensions related to hearing aid use (e.g. the International Outcome Inventory for Hearing Aids [IOI-HA] [Cox & Alexander, 2002]). There are also questionnaires addressing the evaluation of patient-specific goals (e.g. the Client Oriented Scale of Improvement [COSI] [Dillon, James, & Ginis, 1997], and the Glasgow Hearing Aid Benefit Profile [GHABP] [Gatehouse, 1999]).

1.3 Potential issues with currently used outcome measures

Perceived satisfaction and benefit, as evaluated by subjective outcome measures,

vary based on multiple factors, including the patient's hearing profile (Gatehouse, 1994), amplification variables (Souza & Tremblay, 2006), personality and perceived needs of the patient (Hutchinson, Duffy, & Kelly, 2001; Schum, 1999), hearing aid provider comments (Bentler, Neibuhr, Johnson, & Flamme, 2000), whether amplification was delivered in a privately or publicly funded clinic (Cox, Alexander, & Gray, 2005), and even the format and wording of questionnaire items (Schwarz, 1999). Another potential source of outcome assessment variability is the retrospective recall required when performing any of the available self-reported outcome measures. A large body of literature indicates that retrospective assessment of events and experiences can often be inaccurate, due to memory decay, recollection biases, and the use of cognitive heuristics to fill the gaps in their memory so as to "make a good story" (Clark & Teasdale, 1982; Gorin & Stone, 2004, Shiffman, Stone, & Hufford, 2008). Furthermore, memory for auditory stimuli has been long considered relatively weaker than memory for visual or tactile stimuli (Bigelow & Poremba, 2014; Cohen, Horowitz, & Wolfe, 2009). An individual's memory for certain auditory stimuli characteristics can last for just a few seconds, or up to as long as 24 hours, depending on what characteristics of the sound(s) are being recalled (Deutsch, 1975, Winkler & Cowan, 2005; Yarmey & Matthys, 1992). The inaccuracy of retrospective assessments is increasingly likely for average first-time hearing aid users, who are 66-70 years old (Kochkin, 2009); studies indicate that a rapid decline in episodic memory performance begins at about 60-80 years of age (Burke & Light, 1981; Nilsson, 2003). Additionally, current outcome measures are performed at a single time point, and are unable to accurately capture changes in an individual's experiences or behaviour over time, and across situations (Carney, Tennen, Affleck, del Boca, & Kranzler, 1998). These

findings suggest that retrospective self-reported outcome measures of hearing aid benefit may be inadequate for measuring perceived hearing aid benefit, especially if the goal of the outcome measure is a contextual association such as measuring benefit during one of the wearer's varying everyday listening situations.

1.4 Contextual momentary assessment (CMA)

The growing widespread use of smartphones in North America (Barbour, 2013; Smith, 2013), allows for the use of a different methodological approach to collect subjective outcome measures of perceived hearing aid benefit. This methodological approach is known as contextual momentary assessment (CMA) or ecological momentary assessment (EMA) and involves repeated collection of real-time data on an individual's experience in their natural environment (Shiffman et al., 2008). CMA circumvents the error and bias related to retrospective assessments, is more ecologically valid than retrospective assessments, and is highly suited for capturing day-to-day variations in mood, experiences and behaviours (Shiffman et al., 2008). CMA yields qualitatively different information in comparison to traditional retrospective self-report measures; CMA is more valid and better-suited for capturing experiences in a specific moment whereas retrospective measures are better suited for obtaining an overall impression or prediction of future behavior (Shiffman et al., 2008). Based on this evidence, it is possible that CMA could be a better method for measuring hearing aid benefit, especially for specific everyday listening situations. Although CMA employs various technologies, targets of assessment, and schedules of data collection, it is always focused on obtaining repeated measures in real-time from individuals in their natural environment (Stone,

Shiffman, Atienza, & Nebeling, 2007).

1.5 Challenges to CMA usage

Challenges for the application of CMA to clinical treatment generally include participant burden, sample bias, poor compliance, increased reactivity (i.e., the potential for behavior or experience to be affected by the act of assessing it), and the user's technological ability (Palmier-Claus, 2011). Recent studies on employing CMA to assess hearing aid experiences have found that CMA is feasible for evaluating subjective experiences by hearing aid users (Hasan, Chipara, Wu, & Aksan, 2014; Hasan, Lai, Chipara, & Wu, 2013). Galvez et al. (2012) found that CMA was feasible among elderly hearing aid users; the potential limitations of participant burden and poor compliance were minimized with a maximum of four alerts per day and three minutes to complete each questionnaire. Additionally, Galvez et al. (2012) reported that the hearing aid users did not demonstrate reactivity based on pre- and post- HHIE scores. As for the limitation of the user's technological ability, a study in 2013 found that 56% of adults in the U.S. are smartphone owners¹ (Smith, 2013). The growing ownership of smartphones suggests that the CMA limitation of user technological ability with smartphones will likely become negligible in the near future.

1.6 CMA validity and reliability

Before the CMA methodology can be applied to collect outcome measures for any clinical treatment, it is important that the validity and reliability be established (Thiele,

¹ Smartphone ownership decreases with age; 55% of individuals between 45 and 54, 39% between 55 and 64, and 18% over 65 own a smartphone (Smith, 2013).

Laireiter, & Baumann, 2002). There are two possible threats to CMA validity and reliability as a hearing aid outcome measure: (a) response timing relative to the event, (b) presence of a secondary task.

1.6.1 CMA response timing

CMA response timing (i.e., the length of time between the CMA prompt and the completion of the CMA) relative to a particular listening event has the potential to affect the validity and reliability of CMA ratings. The suspension of a CMA response is necessary in the field at times when CMA is inconvenient (e.g., when driving, during a movie, during a meeting). CMA response timing issues have been previously researched, with evidence showing that suspending a CMA prompt until after an event, or frequently suspending a CMA prompt may threaten the ecological validity of the data (Scollon, Kim-Prieto, & Diener, 2003; Tennen, Affleck, Coyne, Larsen, & DeLongis, 2006). An individual's memory for sounds could threaten the ecological validity of CMAs targeting the perceived benefit of hearing aids in a particular listening situation. The ability to identify voices can vary from 2 minutes (Winkler & Cowan, 2005) to 24 hours (Yarmey & Matthys, 1992), and the ability to remember sound qualities such as pitch can decay within a few seconds (Deutsch, 1975). Additionally, annoyance ratings performed while listening to highly disruptive speech and completing a secondary task (visual memory) have been shown to be much higher in comparison to post-task annoyance ratings (Zimmer, Ghani, & Ellermeier, 2008). Based on these findings, it is important to determine whether the validity and reliability of CMA responses are altered when ratings are made immediately after a listening event as opposed to ratings made while

experiencing the particular listening event. Future studies will determine an acceptable length of time after the listening event before CMA ratings could potentially become invalid and/or unreliable.

1.6.2 Presence of a secondary task

Since multi-tasking occurs regularly in daily life, the presence of a secondary task at the time of completing the CMA could potentially affect the validity and reliability of CMA ratings. Currently, many people engage in media-multitasking (Foehr, 2006), such as watching TV while using a smartphone. It is possible that the presence of a secondary task while the individual is trying to attend to speech or music, or simply the presence of having to complete the CMA during a listening situation may affect CMA ratings. Fraser, Gagne, Alepins, and Dubois (2010) demonstrated that a secondary task has an effect on listening effort; they found that subjective ratings of listening effort were significantly higher when participants were performing a dual-task paradigm involving speech recognition and tactile pattern recognition in comparison to effort ratings when performing the single task of speech recognition.

When performing two tasks simultaneously (e.g., communicating while performing a secondary task), the prioritization and allocation of attentional resources may affect how accurately individuals perceive and recall a listening event. Studies on dual-task paradigms indicate that the division of attention based on task prioritization can have an effect on task performance (Fisher, 1975; Janssen, Brumby, & Garnett, 2012; Wickens, 2003). Additionally, memory encoding is challenged when an individual's attention is divided between two tasks (Naveh-Benjamin, Craik, Guez, & Krueger, 2005;

Nyberg, Nilsson, Olofsson, & Backman, 1997). Based on previous research regarding the effects of multi-tasking on task performance and memory encoding, it would be useful to determine whether the presence and relative prioritization of a secondary task affects the rating validity and reliability for sound and speech characteristics.

1.7 Purpose

CMA has the potential to be a useful approach for collecting an accurate measure of hearing aid benefit during specific day-to-day listening situations, providing a more complete description of the hearing aid users' everyday experiences during their treatment than retrospective questionnaires. Outcome measures collected using the CMA methodology can be used in several other ways: (a) to verify and fine-tune the functioning of automatic program switching², (b) to guide the adjustment of the hearing aid settings for particular listening situations, (c) to allow for the provision of a more individually-tailored counseling strategy regarding the individual's lifestyle and related listening situations. The use of CMAs to collect data regarding hearing aid users' experiences has already been deemed feasible among hearing aid users (Galvez et al., 2012; Hasan et al., 2013; Hasan et al., 2014), but before outcome measures collected using the CMA methodology can be used for optimizing hearing aid benefit, we must establish whether

² Automatic program switching is a hearing aid feature that responds to a wearer's day-to-day variation in their listening environment by classifying the acoustic scene, and based on the classification, applies changes to microphone directionality, turns features such as noise reduction on or off, and adjusts the frequency response of the hearing aid to best suit the listening environment (Edwards, 2007). Automatic program selection can sometimes be inconsistent with the wishes of individual wearers, particularly in speech-in-noise situations; further analysis of the acoustic parameters and subjective preferences of the individual wearer may help to optimize and fine-tune the automatic program switching feature (Buechler, 2001).

CMA is capable of facilitating valid and reliable subjective measures of listening experiences.

1.8 Research questions

The overall primary objective of this study was to determine whether CMA is capable of facilitating valid and reliable evaluations of subjective listening experiences in lab-controlled acoustic conditions. Secondary and tertiary objectives of this study were to determine whether CMA validity and reliability are altered significantly by the timing of the CMA relative to the listening event (Experiment I), or by the presence of a secondary task, with focus on either the primary or the secondary task (Experiment II).

Chapter 2: Methodology

2.1 Introduction

This study employed a block-randomized, within-subject factorial design involving three factors: (a) 3 levels of speech intensity for sentences [50, 65, and 80 dB SPL]; (b) 3 levels of signal-to-noise ratio (SNR) [0, +5, +10 dB]; (c) 2 levels of CMA timing (Experiment I) [rated while or after listening to the sentences], or task attention focus (Experiment II) [sentence repetition focus or connect-the-dots focus].

Participants completed the study over one or more sessions in the Amplification Research Lab at the University of British Columbia. During the initial screening, which lasted up to 1 hour, participants completed the following tasks: (a) informed consent; (b) hearing history and general health questionnaire; (c) cognitive screening test; (d) hearing test; (e) near vision screening test; (f) hearing aid fitting; and (g) questionnaires regarding demographics, hearing status, hearing aid satisfaction, and personality.

During the data collection phase, which was completed either immediately after the screening or over one or more additional sessions, participants performed Experiment I and Experiment II, beginning with a practice session for each experiment. In both Experiment I and Experiment II, a CMA was completed for each listening trial. For each and every listening trial in both experiments, participants repeated back 4 sentences (with a per-trial overall minimum of 16 keywords) and completed a CMA regarding the 4 sentences. In Experiment I, participants completed the CMA either while listening to the four sentences, or immediately after they were done listening to the four sentences. In Experiment II, participants performed a secondary visuo-motor task while listening to the

sentences simultaneously, then completed the CMA immediately after they were done listening to the sentences. In Experiment II, the participants' attention was focused on either the primary sentence repetition task or on the secondary connect-the-dots task. Experiment II also incorporated eight equally-distributed baseline trials, with four involving only sentence repetition in quiet, and four involving only connect-the-dots. For both experiments, participants experienced 3 trials for each of the 18 experimental conditions (54 total trials per experiment). Overall, every participant performed 116 listening trials in total ([Experiment I = 54] + [Experiment II = 54 + 8 baseline trials = 62] = 116) across both experiments. The trial listening conditions were presented in a random order for each participant using a Latin square randomization to minimize sequence effects. Sequence effects were further minimized by assigning half of the participants to begin with Experiment I, and the other half to begin with Experiment II. The time to complete all the trials ranged from 1.5 to 3 hours.

2.2 Participants

Participants included 12 adults, 3 males and 9 females, who were 21 to 76 years of age (mean age, 49.1 years). All participants were experienced hearing aid users (at least 6 months of recent, regular hearing aid use) with a mild to severe sensorineural (i.e., lesion or disease of the inner ear and/or the auditory nerve) hearing loss, based on a three-frequency (0.5, 1, and 2 kHz) pure-tone average (PTA) ranging from 31 to 64 dB HL (mean PTA, 50 dB HL). Audiograms for each individual participant are presented in order by their code number in Appendix A. Mean, maximum, minimum, 25th, and 75th percentile pure-tone air conduction thresholds for both ears in dB HL at 0.25, 0.5, 1, 2, 3,

4, 6, and 8 kHz across participants are shown in Appendix B. Participant characteristics (age, gender, hearing aid experience, right and left ear pure-tone averages, cognitive screening, and health-related QoL) are shown in Table 2.1. One participant (505) had a unilateral profound loss, and was fitted with only one hearing aid on their aidable ear for the study; their unaidable (left) ear threshold was not included in the data used for Appendix B. All other participants were fitted binaurally.

Participants were recruited through the distribution of pamphlets advertising the study in community centres, senior centres, and hearing aid clinics in the Vancouver area, and through email advertisement to past study participants with hearing aid experience. All participants were initially screened through phone and email contact to meet the following inclusion criteria: over 18 years of age, fluent in English, able to transport themselves to and from the lab, and have at least 6 months of current hearing aid experience. At the first session, participants were screened to meet the following additional inclusion criteria: completed informed consent; passed (scored ≥ 26) the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005); near visual acuity equal to or better than 20/100; self-reported minimum of “5” on a 10-point scale of health-related quality of life; sensorineural hearing loss (average air-bone gap no more than 15 dB across 0.5, 1, and 2 kHz), and have a three-frequency PTA (0.5, 1, and 2 kHz) between 25 dB HL and 70 dB HL binaurally, or in one ear if the other ear’s PTA was over 70 dB HL.

Table 2.1. Participant characteristics listed by increasing age

Participant code #	Age	Gender	Hearing aid experience (years)	Pure-tone average (dB HL)		MoCA score	Health-related QoL score
				Right	Left		
109	21	M	15	60	57	28	7
505	24	F	20	37	107	29	9
171	30	F	15	45	47	29	10
906	35	F	30	52	33	28	8
449	37	M	32	57	57	30	10
223	42	F	35	60	55	30	10
881	57	M	7	27	28	28	10
155	61	F	26	57	45	27	5
687	66	F	30	58	55	29	9
164	67	F	16	50	50	29	8
314	73	F	15	48	43	30	10
733	76	F	10	47	55	27	9

Note. Pure-tone averages (PTA) are based on 3 frequencies: 0.5, 1, and 2 kHz. dB HL = Hearing level (American National Standards Institute, 2010); MoCA = Montreal Cognitive Assessment; health-related QoL = health-related quality of life.

2.3 Apparatus and stimuli

2.3.1 Auditory stimuli

The auditory stimuli were 536 female-voiced sentences with 3 to 5 keywords each, from the Institute of Electrical and Electronic Engineers (IEEE) (Brouwer, Van Engen, Calandruccio, & Bradlow, 2012; IEEE, 1969), and Bamford-Kowal-Bench (BKB) (Bench, Kowal, & Bamford, 1979; Calandruccio et al., 2010; Van Engen, 2010) sentence lists, in a background of four-talker babble noise (Auditec of St. Louis, 1971). All sentences were adjusted to have the same root-mean-square (rms) level using PRAAT (version 5.3.77). The sound intensity levels chosen for the sentences (50, 65, and 80 dB SPL), and the noise (0, +5, +10 dB SNR) are considered to represent the range of speech levels and SNR in various environments (Pearsons, Bennett, & Fidell, 1977; Smeds, Wolters, & Rung, 2015).

The noise used for all trials was a four-talker babble taken from a commercially available recording composed of two male and two female talkers reading independent passages (Auditec of St. Louis, 1971). The 11-minute-long noise track was played continuously throughout each trial, beginning from the point that the noise track was paused at the end of the previous trial. The noise and speech levels were calibrated weekly throughout data collection, and biological listening checks were performed daily prior to testing.

2.3.2 Contextual momentary assessment questionnaire

The CMA questionnaire was composed of four sound quality or speech intelligibility ratings and was completed using the touch screen of an android smartphone

(Motorola Moto G), which required a near visual acuity of 20/125 to complete. Four dimensions were rated using an 11-point Likert-type scale: (a) intelligibility, (b) noisiness, (c) listening effort, (d) loudness. The four rating dimensions, end-point categories, and instructions to be used for each question were taken from Preminger and Van Tasell (1995a, 1995b). CMA rating instructions and end-point categories used are shown in Appendix C.

Preminger and Van Tasell (1995a, 1995b) employed a 100-point rating scale, and based on Miller (1956), assumed that a realistic confidence interval for responses was approximately 20 points. Instead of a 100-point scale, this study used an 11-point Likert-type scale with end point categories for the two extremes of the scale. The 11-point Likert-type (0 to 10) scale was chosen based on the following factors: (a) the rating scale had to fit and be easily visible on a smartphone screen; (b) the number of scale intervals used was sufficient for describing the range of subjective categories for each rating dimension; (c) test-retest reliability tends to be maximal at 7 response categories, and decreases minimally for scales with more than 10 response categories (Preston & Coleman, 2000; Weijters, Cabooter & Schillewaert, 2010); (d) it is a standard scale used for outcome measures in audiology.

2.3.3 Hearing aids

All participants were fitted with a pair of Phonak Ambra microM behind-the-ear (BTE) hearing aids. The aids were fitted using SlimTubes, closed domes, and the NAL-NL1 fitting prescription through Noah 4 and Phonak Target™ 3.2 software. Real-ear verification of the fitting was completed using the FONIX® 7000, to ensure that the

frequency response to the speech-weighted spectrum at 50, 65, and 80 dB SPL input levels was within +/- 8 dB of the NAL-NL1 targets prescribed for 500, 1000, and 2000 Hz whenever possible (per the criteria established by Polonenko et al., 2010). Mean, range, 25th, and 75th percentiles for dB distance from NAL-NL1 prescription gain targets at 0.5, 1 and 2 kHz for both ears in response to a 65 dB SPL speech-weighted spectrum are shown in Appendix D.

The hearing aids were set with SoundRecover³ off, and a single, fixed hearing aid program called “Speech in Noise”. “Speech in Noise” is one of the frequency response settings from Phonak’s automatic program-switching system “SoundFlow”. The program “Speech in Noise” was chosen because it would be triggered in the field under acoustic situations similar to all the lab-controlled listening conditions. The hearing aid settings were then saved to Noah 4 when the fitting was complete.

2.3.4 Secondary task

Participants completed the secondary connect-the-dots task on a touch screen using the SANZEN “Trail Making” program (Keller, 2014), which was developed for creating and administering cognitive tests for neuropsychological assessment. The program allowed for greater ease of administering the secondary connect-the-dots task, and was capable of timing participants’ responses, automating the scoring, and providing

³ SoundRecover’s primary objective is to restore audibility for high frequency inputs through to 10 kHz. SoundRecover compresses the signal above a specified cut off frequency whereas frequencies below the cut off frequency will be amplified normally (Phonakpro.com, 2014). Due to variable clinical outcomes (Glista, Scollie, Bagatto, Seewald, Parsa, & Johnson, 2009; O’Brien, Yeend, Hartley, Keidser, Nyffeler, 2010; Simpson, 2009) and lack of evidence-based research to guide the selection of frequency compression parameters, the SoundRecover feature was not enabled for this study.

accurate results, but was not used as a cognitive test in this context. The reasons for using connect-the-dots as the secondary task were as follows: (a) the task was sufficiently challenging and distracting, but not overly challenging that participants would require extensive practice beforehand; (b) most participants over 18 years of age would be physically able to complete the task while simultaneously performing the listening task; (c) the task was non-linguistic and visuo-motor, so it would minimally interfere with the auditory and language processing involved in the listening task (Baddeley & Logie, 1999; Pashler, 1994).

The secondary connect-the-dots task required participants to touch circled numbers on a screen starting at 1 (labeled “START”) and connect the numbered circles from 1 to 40 in increments of 1 as quickly and as accurately as possible. The placement coordinates of the numbered dots were randomly chosen using a MATLAB randomization function for each connect-the-dots trial. In total, 58 connect-the-dot trials with randomized patterns were created so that no pattern was repeated for each participant. For screenshots of a sample connect-the-dots trial, see Appendix E.

2.3.5 Equipment

Participants were seated in a double-walled sound-treated booth facing a speaker (Behringer TRUTH B2031A) 2 ft directly in front at 0° azimuth in the horizontal plane. A touch screen display (Elo Touch Solutions 1915L 19” Desktop Touchmonitor) was placed on a table to the side of their handedness for the secondary connect-the-dots task. A microphone (Crown Sound Grabber II Pressure Zone Microphone) was mounted on the table in front of the participant and connected to a digital voice recorder (Olympus DS-

2300) to record their spoken responses on the sentence repetition task for later verification, if needed. The participant was provided with a written copy of the CMA rating dimension instructions on the table in front of them. Appendix C shows the instructions for rating each CMA dimension.

To run the experiment, the experimenter used a custom-written (© Gillen) MATLAB (version R2013b) function that automated the following: randomization of the listening trial conditions, presentation of the four-talker noise at the correct SNR for each listening trial, presentation of randomly-chosen sentence tracks at the appropriate speech level, collection of the number of keywords repeated back correctly for each sentence, as well as prompting the experimenter when to display the connect-the-dots task, and when to alert the participant to begin filling out the CMA on the handheld device. The MATLAB program commands were viewed on the experimenter's primary display monitor. A secondary display (which could be controlled by the experimenter) was split using an analogue monitor splitter to the touch screen monitor (Elo Touch Solutions 1915L 19" Desktop Touchmonitor) in the sound booth for the participant to view and use to complete the connect-the-dots task. The participant's touch screen display was controlled by the experimenter through a data transfer switch, and was used to alert the participant when to complete the CMAs for Experiment I (timing of rating).

2.3.5 Questionnaires

Each participant filled out questionnaires regarding demographics, hearing status, hearing aid satisfaction, and personality (see Appendix F for questionnaire package). The questionnaire on demographics gathered information regarding age, gender, living

arrangement, education level, employment status, income, and how recently they (or their family) had immigrated to Canada. The questionnaires on hearing status included the Hearing Handicap Inventory for the Elderly or for Adults (HHIE/A) (Newman et al., 1990; Ventry & Weinstein, 1982), and the Abbreviated Profile of Hearing Aid Benefit (APHAB) (Cox & Alexander, 1995). The questionnaires on hearing aid satisfaction included the Satisfaction with Amplification in Daily Life (SADL) (Cox & Alexander, 1999), and the International Outcome Inventory for Hearing Aids (IOI-HA) (Cox & Alexander, 2002); these questionnaires were filled out regarding the participants' own hearing aids. The personality questionnaire used was the Big Five Inventory (BFI) (John, Donahue, & Kentle, 1991; John, Naumann, & Soto, 2008). Questionnaire data were collected as possible co-variates for later analyses, but will not be reported further in this thesis.

2.4 Procedure

2.4.1 Session I (screening)

Participants initially completed informed consent (see Appendix G), and filled out a questionnaire on their hearing history and general health (see Appendix H). If the participant scored less than “5” on a 10-point scale of health-related quality of life, they were asked if they felt well enough to continue the session, or if they would like to reschedule their appointment. Following the consent and hearing health forms, the experimenter administered a 10-minute cognitive screening test, the Montreal Cognitive

Assessment⁴ (MoCA) (Nasreddine et al., 2005). See Appendix I for the MoCA test form, and Appendix J for MoCA administration instructions. The MoCA was chosen over the Mini-Mental State Examination (MMSE) due to research evidence suggesting that the MoCA was more sensitive to mild cognitive impairment (Freitas, Simões, Alves, & Santana, 2013).

Participants' hearing thresholds were then obtained using insert earphones (ER-3A) while seated in a double-walled, sound-attenuating room with ambient noise levels complying with American national standards (American National Standards Institute, 1999). The hearing evaluation included otoscopy, pure-tone air- (at 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz) and bone-conduction (at 0.5, 1, 2, 3 and 4 kHz) audiometry, and loudness discomfort levels (LDLs) for 500 Hz and 3000 Hz (see Appendix K for LDL methodology used). Pure-tone audiometry was performed by the experimenter using a 5 dB step-size bracketing method (2 steps down, one step up, threshold at the tone intensity with 50% response rate) with a GSI-61 audiometer calibrated in accordance with American national standards for threshold measurement (American National Standards Institute, 2010). A near vision screening was also performed using a Reading Text chart at 16 cm from the participant's eyes, to screen for at least 20/100 corrected near vision. The hearing aid fitting was real-ear verified using the FONIX® 7000 to meet NAL-NL1 prescription gain targets, and then saved to Noah 4. Lastly, participants were given a questionnaire package on demographics, hearing status, hearing aid satisfaction, and personality (see Appendix F) to complete at the lab or at home.

⁴ One participant opted to receive the MoCA task instructions in written form so they could be sure that they understood the task correctly.

2.4.2 Session II

Session II lasted up to 3 hours and was performed either immediately after the screening session or over multiple separate sessions, depending on the participant's preference⁵. Session II consisted of Experiments I and II; each participant was randomly assigned to start with either Experiment I or II, and, for Experiment II, to start with either their attention focused on sentence repetition or connecting-the-dots. Before each experiment, the participant performed up to two practice trials for familiarization with the task. Each participant was randomly assigned a particular sequence of listening conditions for each experiment in order to reduce sequence effects.

2.4.3 Practice session (Experiment I)

In the practice session for Experiment I, participants were told to read the instructions for the four CMA rating dimensions, and were provided a written copy of the instructions if they needed to remind themselves of the instructions at any time. The participants were then instructed on how to use the smartphone to fill out the CMA. When participants reported that they were comfortable using the handheld device, two practice trials were performed to ensure that the participant understood the instructions for the experimental task.

2.4.4 Experiment I—timing of ratings

In Experiment I, CMAs were performed either while, or immediately after listening to the sentences in one of the nine different listening conditions. For every

⁵ Only one participant (out of twelve) was tested across two separate sessions. All other participants completed the testing in one session.

randomly presented listening condition, participants performed a sentence repetition task where they repeated back four randomly chosen sentences, for a total of 16-18 keywords per trial. The presentation timing of each sentence as well as the scoring of each sentence was controlled by the experimenter who was able to use visual cues to aid with scoring (they were able to see the participant through the sound booth window). An audio recording of the participant's responses was collected for later verification, if needed.

For each listening trial, participants were alerted to begin completing the CMA on the smartphone via a message presented on the touch screen display. When the participant was to perform the CMA while listening to the sentences, the alert always occurred immediately after the experimenter entered the number of keywords correct score for the first sentence. The rating responses were logged by the smartphone once completed. Prior to testing, participants were instructed on the sentence repetition task and were told to complete the CMA ratings as soon as they saw the alert on the touch screen. The verbatim instructions were as follows:

You will hear a woman say sentences with people talking in the background. Please repeat back the sentences that the woman says. Repeat as much of the sentence as you can, even if you only could hear one word. It is ok to guess if you are unsure. You will be alerted to complete an evaluation of your listening situation on the computer screen. Please complete your ratings on the device as soon as you are alerted. You may take a break at any time. Any questions?

Through the course of Experiment I, all nine listening conditions were experienced three times, each with a CMA performed either while or immediately after listening to the sentences for a total of fifty-four listening trials (3 speech levels * 3 SNRs * 2 rating timing levels * 3 repetitions).

2.4.5 Practice session (Experiment II)

In the practice session for Experiment II, if the participants had not done so already, they were familiarized with the four CMA rating dimension instructions, and how to use the smartphone to complete the CMA. Participants were then familiarized with the connect-the-dots task, and performed 5 timed (30 second) trials of the connect-the-dots task. Before starting the test trials, two practice trials were completed to ensure that the participant was comfortable with the instructions for the experiment.

2.4.6 Experiment II—task focus

In Experiment II, CMA was always performed immediately after listening to the sentences, with attention primarily focused on either the sentence repetition task or a secondary non-linguistic, visuo-motor task of connect-the-dots. Each listening trial involved a sentence repetition task and a simultaneous connect-the-dots task. The sentence presentation began as soon as the participant pressed the circled number “1” (START) on the touch-screen. When the keywords correct score for the fourth sentence track was entered, the touch screen for the secondary task was shut off, and the participant completed a CMA. Equally distributed throughout Experiment II (every 7 trials), there were 8 baseline trials. When focus was directed at sentence repetition, the baseline trials involved sentence repetition of four sentences in quiet at average intensity (65 dB SPL), with no secondary task. When focus was directed at the connect-the-dots task, the baseline trials involved only the connect-the-dots task. The scores on the connect-the-dots task during these baseline trials were collected in order to help calculate the participants’ baseline performance for the connect-the-dots task.

Participants were instructed to regard either the sentence repetition task or the connect-the-dots (secondary) task as the more important of the two, using instructions adapted from those shown to be effective in altering primary attention (Fisher, 1975). The verbatim instructions for the primary task (sentence repetition) focus were as follows:

You will work on two tasks simultaneously. Repeat back the sentences that the woman says, and use the touch screen to connect the dots in order starting from 1 as accurately and as quickly as possible. We are asking that you regard the **sentence repetition** task as the main and most important of the two. At the end of the session, please complete the evaluation of your listening situation on the handheld device. Let me know if you need to rest between test sessions. Remember, the **sentence repetition** task is the main and most important of the two.

The verbatim instructions for the secondary task (connect-the-dots) focus were as follows:

You will work on two tasks simultaneously. Repeat back the sentences that the woman says, and use the touch screen to connect the dots in order starting from 1 as accurately and as quickly as possible. We are asking that you regard the **connect-the-dots** task as the main and most important of the two. At the end of the session, please complete the evaluation of your listening situation on the handheld device. Let me know if you need to rest between test sessions. Remember, the **connect-the-dots** task is the main and most important of the two.

All participants then performed 3 repetitions of the 9 listening conditions with 4 equally distributed rest trials (31 trials total). Participants were then instructed to change their attention focus using the instructions framework taken from Fisher (1975), before performing the same number of trials again (another 31 listening trials) with the new task focus.

In summary, participants experienced all nine listening conditions three times for each task focus for a total of 54 trials (3 speech levels * 3 SNRs * 2 task focus levels * 3 repetitions) plus 8 baseline trials for a total of 62 trials in Experiment II.

2.5 Data analysis

2.5.1 Overview

Two important traits of an outcome measure are validity and reliability. Validity is the ability of an instrument to measure what it intends to measure, and reliability is the ability of the instrument produce the same value each time it is used if no change has occurred (Devon et al., 2007). The validity for each of the four rating dimensions was analyzed by examining criterion validity and construct validity. Criterion validity refers to the strength of the relationship between the subjective ratings and their corresponding objective measures, while construct validity is the degree to which an instrument measures what it is intended to measure (Devon et al., 2007). Reliability was evaluated using intra-rater reliability for each rating dimension based on the repeated measurements participants performed in the same lab-controlled listening situations.

2.5.2 Criterion validity

The criterion validity was evaluated for the intelligibility, noisiness, listening effort, and loudness ratings by evaluating how strongly they correlated with related objective measures. The general guidelines used for interpreting the correlation coefficients were as follows: 0 = Zero, .1–.3 = Weak, .4–.6 = Moderate, .7–.9 = Strong, 1 = Perfect (Dancey & Reidy, 2004). Intelligibility ratings for each completed CMA were correlated to the percent total keywords the participant repeated back correctly (scored by the experimenter) in the corresponding listening trial. Noisiness and loudness ratings were correlated to the known, calibrated intensity levels that were presented for the background noise and the speech level during the corresponding listening trial.

Criterion validity analysis of the listening effort rating dimension was not as straightforward as intelligibility, noisiness and loudness. Analyzing listening effort is problematic because subjective ratings are often not correlated with objective measures of listening effort (Fraser et al., 2010; McGarrigle et al., 2014; Sarampalis, Kalluri, Edwards, & Hafter, 2009). However, listening effort has some relation to intelligibility in speech-in-noise situations and is often measured objectively in research using recall, or reaction-based secondary task paradigms (Picou, Ricketts, Hornsby, 2013). To examine listening effort validity, this study compared listening effort ratings to intelligibility (percent keywords correct) for Experiments I and II, and to performance on the secondary connect-the-dots task in Experiment II.

The effects of rating timing (Experiment I), task focus (Experiment II), and secondary task presence on the criterion validity of the ratings were analyzed by determining whether the correlation of subjective ratings to their corresponding objective measures were significantly stronger under one condition compared to the other. These dependent correlations were compared by hand using Steiger's (1980) modification of Dunn and Clark's z with a backtransformed average z procedure (Silver, Hittner, and May, 2004, Equation 5), which was shown to have a minimal Type I error rate (Silver et al., 2004).

2.5.3 Construct validity

The construct validity was evaluated for the four rating dimensions by comparing the change in ratings to the following a priori hypotheses: (a) intelligibility ratings will increase (improve) with increasing SNR, (b) noisiness ratings will increase (be perceived

as louder) with increasing background noise⁶, (c) listening effort will increase (worsen) with decreasing SNR, (d) loudness ratings will increase with increasing speech level. The CMA ratings (dependent variables) for each of the 4 dimensions (intelligibility, noisiness, listening effort, loudness) were analyzed for each of the two experiments with a three-way repeated-measures analysis of variance (ANOVA) using the IBM SPSS Statistics program (IBM Corp., 2013) (Experiment I independent variables—CMA timing [2 levels], speech level [3 levels], signal-to-noise ratio [3 levels]; Experiment II independent variables—task focus [2 levels], speech level [3 levels], signal-to-noise ratio [3 levels]).

The effect of secondary task presence was analyzed by comparing CMA ratings made with and without the secondary task: focus on the sentence repetition task in Experiment II compared to ratings made after the sentences in Experiment I using a three-way repeated-measures ANOVA (independent variables--number of tasks [2 levels], speech level [3 levels], signal-to-noise ratio [3 levels]).

2.5.3 Reliability

Reliability, defined in this study as the consistency of ratings within a specific listening condition within each participant (intra-rater reliability), was evaluated using intraclass correlation coefficients (ICC). ICCs were calculated for the four rating dimensions of intelligibility, noisiness, listening effort, and loudness for each timing condition for Experiment I (timing of ratings), and for each task focus condition for

⁶ Increasing background noise level is shorthand for “decrease in SNR within each speech level”. The background noise levels employed in this study ranged from 40 dB to 80 dB by 5 dB increments.

Experiment II (task focus). The ICCs and their 99% confidence intervals⁷ were determined by hand using a method appropriate for randomized block designed experiments where participants perform multiple measurements for each condition (Gwet, 2008, Equation 10; Balakrishnan & Gwet, 2014). The general guidelines used for interpreting the intra-rater reliability ICCs were as follows: $< .40$ = “Poor”, $.40-.74$ = “Fair to good”, $.75-1.00$ = “Excellent” (Cicchetti & Sparrow 1981).

2.5.4 Secondary task (connect-the-dots) performance

Performance on the secondary connect-the-dots task was analyzed to examine whether participants as a whole did indeed focus on either the primary or secondary task as instructed. Baseline performance on the connect-the-dots task was calculated using the dots per minute scores for the final 2⁸ practice connect-the-dot trials and the 4 equally distributed baseline trials that were performed throughout the testing block in Experiment II where participants were focused on connect-the-dots. Using paired t-tests, connect-the-dots performance was compared for the following: (a) the focus on connect-the-dots condition versus the focus on sentence repetition condition, (b) the baseline performance versus focus on sentence repetition, and (c) the baseline performance versus focus on connect-the-dots.

⁷The comparison of ICCs in this study was a multiple inference, and requires adjusting confidence intervals to 99% in order to reduce Type I error (Hochberg & Tamhane, 1978).

⁸ Pilot studies with 3 normal-hearing young adults showed that the last 2 of 5 practice connect-the-dots trials reached an asymptote where dots per minute scores were within +/- 2 of each other.

Chapter 3: Results

3.1 Overview

Criterion validity

Ratings for intelligibility, noisiness, and loudness were strongly correlated (Dancey & Reidy, 2004) with their corresponding objective measures (percent keywords correct, background noise intensity, and speech level) regardless of rating timing, secondary task presence, or task focus. Intelligibility had the strongest correlation with percent keywords correct when rated after the listening event, in the absence of a secondary task. Noisiness and loudness had the strongest correlation with background noise intensity and speech intensity when ratings were made while the participant was listening to the sentences.

Ratings for listening effort were moderately correlated with the objective measure of intelligibility (percent keywords correct) in the single task condition, and rating timing had no effect on the strength of correlation. When a secondary task was present, listening effort ratings correlated weakly with percent keywords correct, and task focus had no effect on the strength of correlation. Listening effort ratings were also moderately correlated with secondary connect-the-dots task performance (correct dots per minute), and was not significantly different due to task focus.

Construct validity

Results for Experiment I and Experiment II revealed that ratings for three of the four dimensions, intelligibility, noisiness, and loudness, changed significantly in the

patterns predicted: (a) intelligibility ratings increased (improved) with increasing signal-to-noise ratio (SNR) (i.e., speech increased in intensity compared to the background noise); (b) noisiness ratings increased (worsened) with increasing speech level and with decreasing SNR under each of the three speech levels; (c) loudness ratings increased with increasing speech level. CMA timing (Experiment I) had a significant effect on noisiness at 80 dB SPL, and on loudness at 65 dB SPL and 80 dB SPL. Both noisiness and loudness ratings made after the sentences were lower than ratings made while listening to the sentences. The presence of a secondary task, and whether the task focus was on the sentence repetition task or on the secondary task (connect-the-dots) did not have any effect on ratings for all any of the 4 dimensions across all 9 listening conditions.

Results for the fourth rating dimension, listening effort, partly agree with the hypothesis that effort increases with decreasing SNR. In Experiment I, an interaction between SNR and speech level was evident. Increasing SNR resulted in a significant decrease in listening effort ratings, but only for 65 dB SPL and 80 dB SPL speech levels. Listening effort ratings showed ceiling effects (maximum listening effort) for 50 dB SPL. In Experiment II, listening effort ratings were unaffected by changes in SNR when a secondary task was present. Listening effort data for Experiment II show that there were ceiling effects (maximum listening effort) for many participants.

Reliability

Intra-rater reliability was fair to good (Cicchetti & Sparrow 1981) for intelligibility ratings in both Experiments I and II, and fair to good for noisiness and loudness ratings made after listening to the sentences and when a secondary task was present. Intra-rater

reliability was excellent for noisiness and loudness ratings made while listening to the sentences. Listening effort intra-rater reliability was poor across both Experiments I and II.

3.2 Experiment I—timing of ratings

3.2.1 Intelligibility

Criterion validity

Scatter plots with regression lines for percent keywords correct and intelligibility ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B) are shown in Figure 3.1. In both conditions, the participants' intelligibility ratings correlated strongly (Dancey & Reidy, 2004) with the percent keywords they repeated back correctly⁹ (while: $r = .75$, 95% CI [.70, .80], $p < .001$; after: $r = .83$, 95% CI [.80, .86], $p < .001$). There was an effect of rating timing on the correlation between intelligibility ratings and percent keywords correct; intelligibility ratings made after listening to the sentences correlated more strongly with their corresponding percent keywords correct than intelligibility ratings made while listening to the sentences, $z = 3.64$, $p < .001$.

⁹ Sentence repetition performance (percent correct keywords) means and standard deviations for each participant, and means and standard errors across participants in each of the 9 listening conditions are shown in Appendix L.

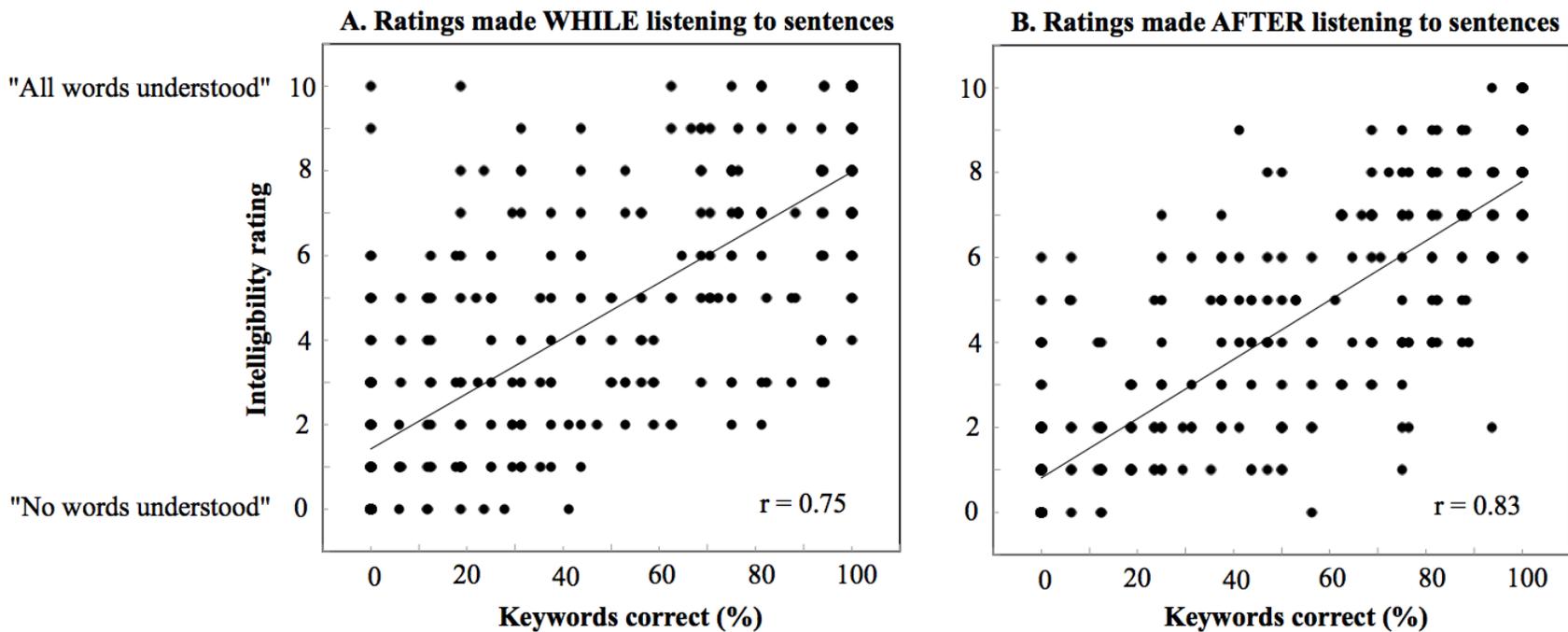


Figure 3.1. Correlation between intelligibility ratings and sentence repetition percent keywords correct for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B). Pearson's correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment I intelligibility ratings are shown in Figure 3.2, and listed in Table M.1 in Appendix M. A three-way analysis of variance showed that timing of rating (CMA completion either while listening to or immediately after listening to the sentences) had no effect on intelligibility ratings, $F(1, 11) = 1.42, p = .259$. There was an interaction between SNR and speech level on intelligibility ratings, $F(4, 44) = 9.12, p < .001$. The results revealed a general pattern of increasing (improved) intelligibility ratings with increasing SNR, as well as with increasing speech level.

At the 50 dB SPL speech level, increases in SNR from 0 to +10, $t(11) = 12.38, p = .005$, and from +5 to +10 $t(11) = 12.38, p = .005$, led to significantly higher intelligibility ratings; however, an increase in SNR from 0 to +5 had no effect, $t(11) = 3.93, p = .073$. At 65 dB SPL, increases in SNR from 0 to +5, $t(11) = 36.80, p < .001$, from +5 to +10, $t(11) = 20.71, p < .001$, and from 0 to +10, $t(11) = 51.20, p < .001$, led to significantly higher intelligibility ratings. At 80 dB SPL, increases in SNR from 0 to +5, $t(11) = 66.18, p < .001$, from +5 to +10, $t(11) = 17.83, p < .001$, and from 0 to +10, $t(11) = 73.38, p < .001$, led to significantly higher intelligibility ratings.

At 0 SNR, increases in speech level from 50 to 65 dB SPL, $t(11) = 13.04, p = .004$, and from 50 to 80 dB SPL, $t(11) = 28.39, p < .001$, led to significantly higher intelligibility ratings. However, an increase in speech level from 65 to 80 dB SPL at 0 SNR had no effect, $t(11) = 0.50, p = .495$. At +5 SNR, increases in speech level from 50 to 65 dB SPL, $t(11) = 29.13, p < .001$, and from 50 and 80 dB SPL, $t(11) = 42.42, p < .001$, led to significantly higher intelligibility ratings (improved intelligibility); however,

an increase in speech level from 65 to 80 dB SPL had no effect on intelligibility ratings, $t(11) = 1.74, p = .214$. At an SNR of +10, increases in speech level from 50 to 65 dB SPL, $t(11) = 24.38, p < .001$, and from 50 to 80 dB SPL, $t(11) = 27.15, p < .001$, led to significantly higher intelligibility ratings. However, an increase in speech level from 65 to 80 dB SPL at +10 SNR had no effect on intelligibility ratings, $t(11) = 3.06, p = .108$.

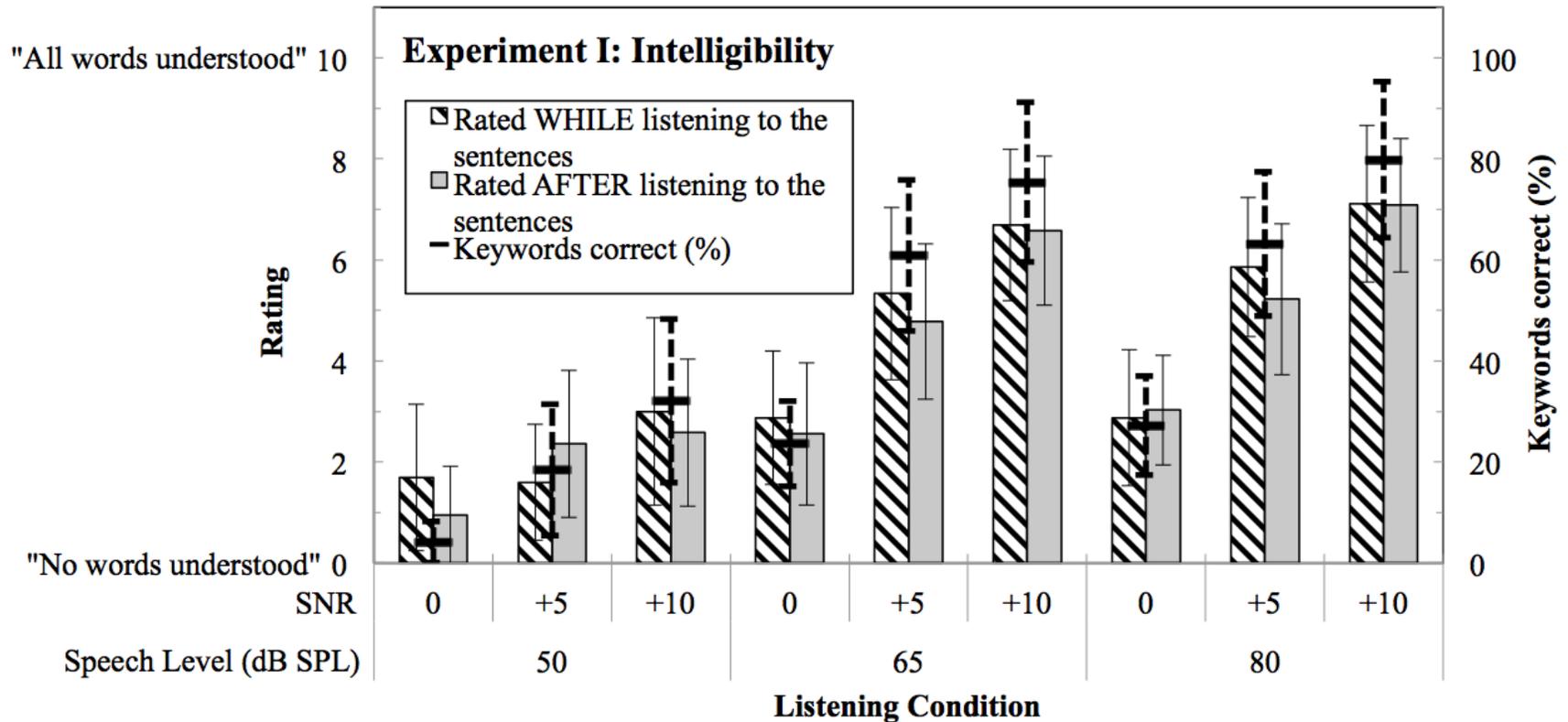


Figure 3.2. Mean ratings for intelligibility made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). Mean percent keywords correct for each listening condition in Experiment I across participants are shown for comparison. Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for intelligibility ratings and their 99% confidence intervals are shown in Figure 3.3. The intelligibility intra-rater reliability was fair to good (Cicchetti & Sparrow 1981) regardless of CMA timing. The ICC for intelligibility ratings made while listening to the sentences was .53, 99% CI [.32, .90], and .56 for intelligibility ratings made after listening to the sentences, 99% CI [.35, .91].

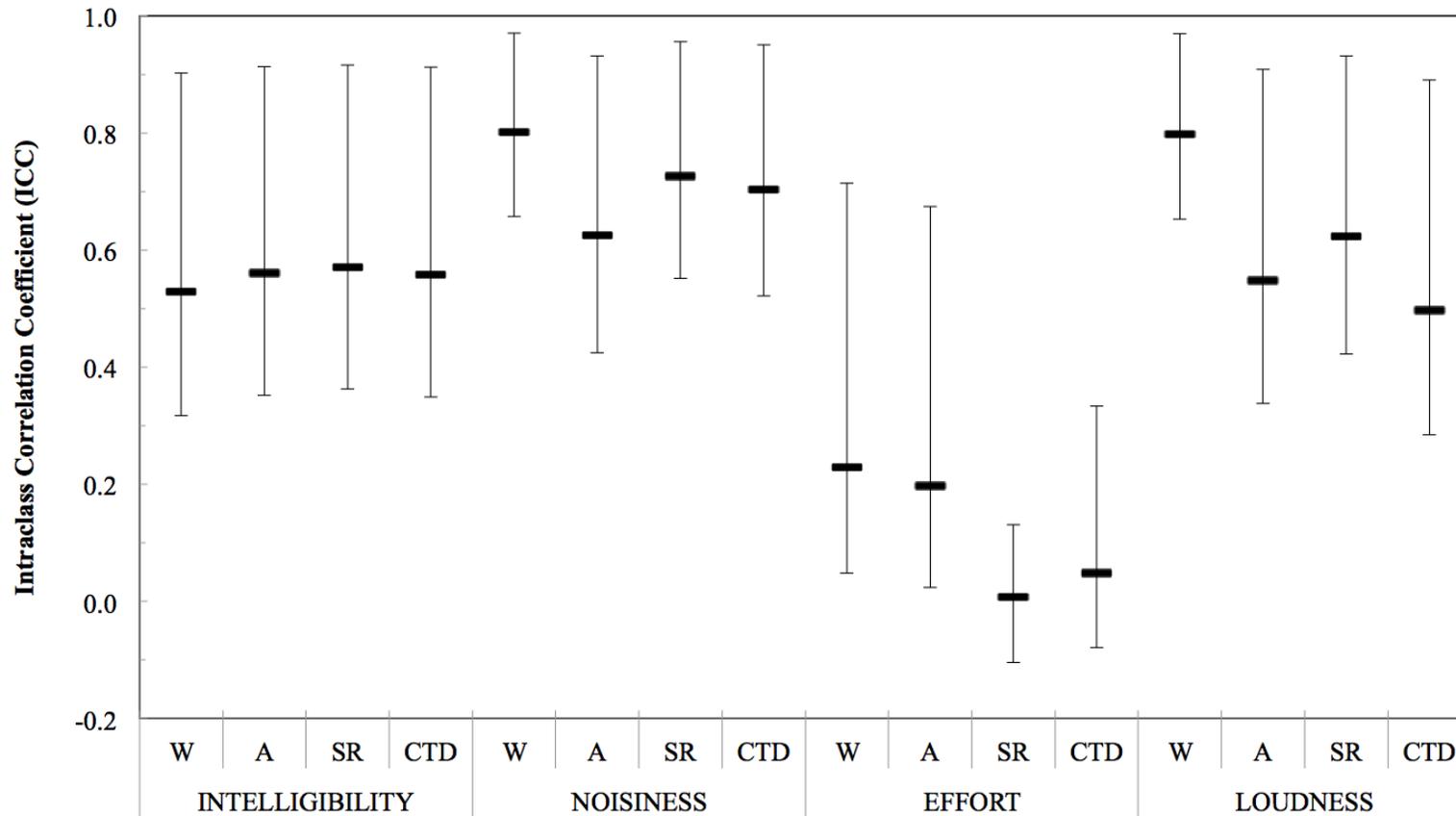


Figure 3.3. Comparison of mean intraclass correlation coefficients (ICCs) for each rating dimension (intelligibility, noisiness, listening effort, and loudness) made while (W), or after (A) listening to the sentences, and with focus on either the sentence repetition task (SR), or with focus on the secondary connect-the-dots task (CTD). Error bars indicate 99% confidence intervals for the mean.

3.2.2 Noisiness

Criterion validity

Scatter plots with regression lines for background noise intensity and noisiness ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B) are shown in Figure 3.4. When participants rated noisiness while listening to the sentences, their ratings correlated strongly (Dancey & Reidy, 2004) with the range of background noise intensities, $r = .77$, 95% CI [.73, .81], $p < .001$. When participants rated noisiness after listening to the sentences, their ratings correlated moderately with the range of background noise intensities $r = .66$, 95% CI [.60, .72], $p < .001$. There was an effect of rating timing on the correlation between noisiness ratings and background noise intensity; noisiness ratings made while listening to the sentences correlated more strongly with their corresponding background noise intensities than noisiness ratings made after listening to the sentences, $z = 4.35$, $p < .001$.

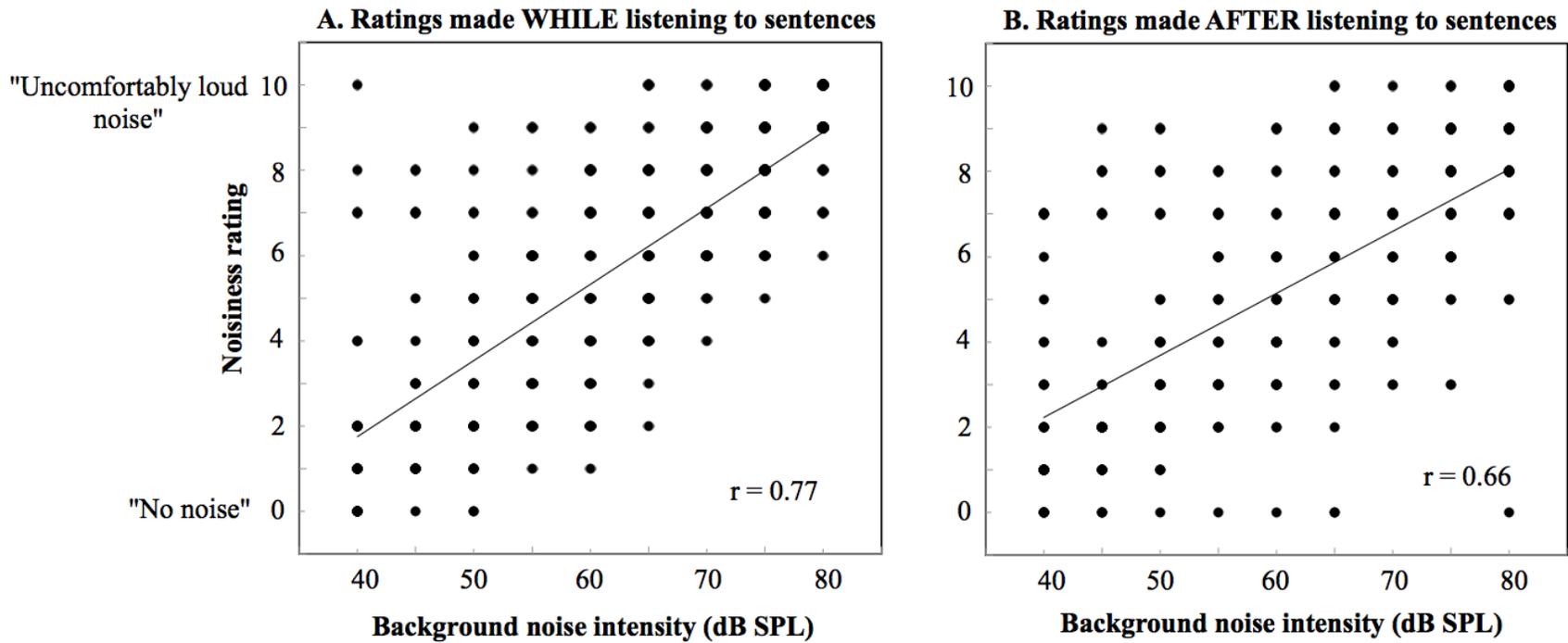


Figure 3.4. Correlation between noisiness ratings and corresponding calibrated background noise intensities for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B). Pearson's correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment I noisiness ratings are shown in Figure 3.5, and listed in Table M.1 in Appendix M. A three-way analysis of variance showed that there was an interaction between timing of rating (CMA completion either while listening to or immediately after listening to the sentences) and speech level on noisiness ratings, $F(2, 22) = 5.95, p = .021$; noisiness ratings made after listening to the sentences were significantly lower than noisiness ratings made while listening to the sentences, but only when the speech level was 80 dB SPL. There was no interaction between speech level and SNR, $F(4, 44) = 0.55, p = .700$, but there was a main effect of speech level, $F(2, 22) = 85.88, p < .001$, and a main effect of SNR, $F(2, 22) = 57.36, p < .001$. The results revealed a pattern of increasing (worse) noisiness ratings (the noise was perceived as louder) with increasing speech level, and with decreasing SNR.

Changes in SNR had a main effect on noisiness ratings, $F(2, 22) = 57.36, p < .001$. Increasing SNR from 0 to +5, $t(11) = 117.30, p < .001$, from +5 to +10 dB SNR, $t(11) = 14.19, p = .003$, and from 0 to +10 SNR, $t(11) = 85.51, p < .001$, led to significantly lower (better) noisiness ratings (the noise was perceived as quieter).

When ratings were made while listening to the sentences, an increase in speech level from 50 to 65 dB SPL, $t(11) = 67.57, p < .001$, from 65 to 80 dB SPL $t(11) = 35.57, p < .001$, and from 50 to 80 dB SPL, $t(11) = 197.50, p < .001$, led to significantly higher noisiness ratings (the noise was perceived as louder). When ratings were made immediately after listening to the sentences, an increase in speech level from 50 to 65 dB SPL, $t(11) = 27.03, p < .001$, from 65 to 80 dB SPL, $t(11) = 32.15, p < .001$, and from 50 to 80 dB SPL, $t(11) = 99.02, p < .001$, also led to significantly higher noisiness ratings. At

80 dB SPL, an interaction between rating timing and speech level occurred such that noisiness ratings made after the sentences were significantly lower (the noise was perceived as quieter) than noisiness ratings made while listening to the sentences, $t(11) = 12.74, p = .004$. There was no effect of timing on noisiness ratings at 50 dB SPL, $t(11) = 1.15, p = .307$, or at a 65 dB SPL speech level, $t(11) = 0.94, p = .352$.

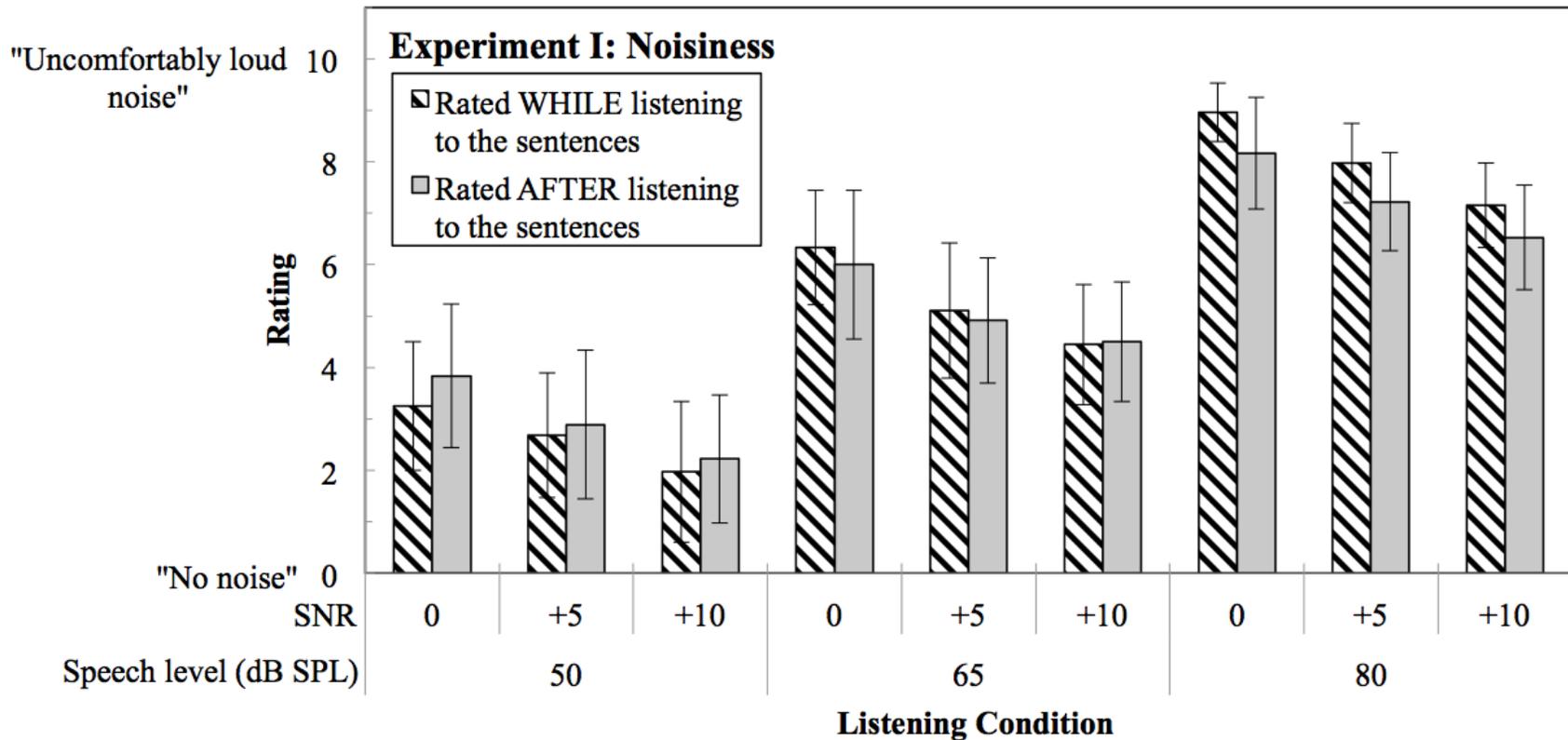


Figure 3.5. Mean ratings for noisiness made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). The background noise intensities (dB SPL) for each listening condition are listed along the bottom of the figure for comparison. Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for noisiness ratings and their 99% confidence intervals are shown in Figure 3.3. The noisiness intra-rater reliability was fair to good (Cicchetti & Sparrow 1981) when ratings were made after listening to the sentences, and excellent when noisiness was rated while listening to the sentences. The ICC for noisiness ratings made while listening to and repeating back the sentences was .80, 99% CI [.66, .97], and .63 for noisiness ratings made after listening to and repeating back the sentences, 99% CI [.43, .93].

3.2.3 Listening effort

Criterion validity

Scatter plots with regression lines for percent keywords correct and their corresponding listening effort ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B) are shown in Figure 3.6. When participants rated listening effort while listening to the sentences, their ratings correlated moderately (Dancey & Reidy, 2004) with the percent keywords correct, $r = -.46$, 95% CI $[-.37, -.53]$, $p < .001$. When participants rated listening effort after listening to the sentences, their ratings also correlated moderately with the percent keywords correct, $r = -.43$, 95% CI $[-.34, -.52]$, $p < .001$. There was no effect of rating timing on the correlation between listening effort ratings and percent keywords correct, $z = 0.36$, $p = .716$.

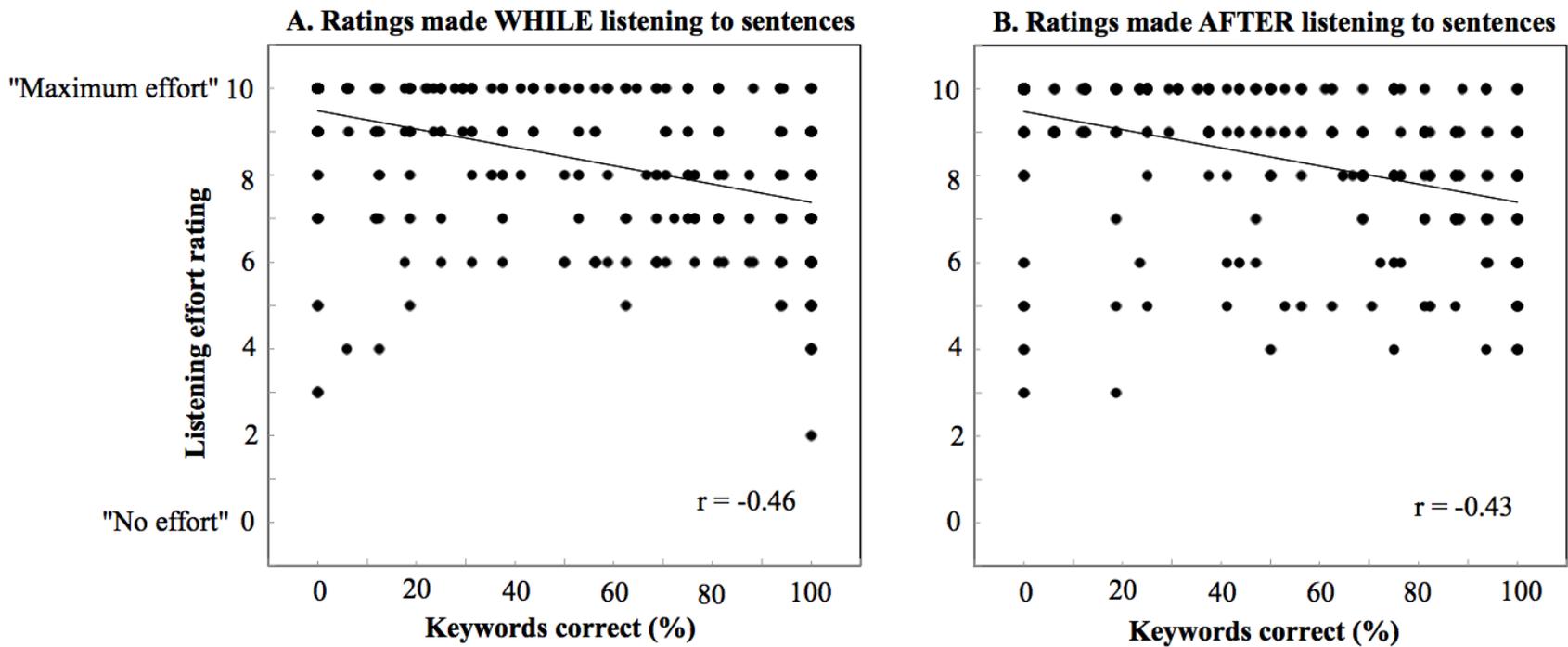


Figure 3.6. Correlation between listening effort ratings and sentence repetition percent keywords correct for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B). Pearson's correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment I listening effort ratings are shown in Figure 3.7, and listed in Table M.1 in Appendix M. A three-way analysis of variance showed that timing of rating (CMA completion either while listening to or immediately after listening to the sentences) had no effect on listening effort ratings, $F(1, 11) = 0.38, p = .549$. There was an interaction between speech level and SNR, $F(4, 44) = 4.47, p = .028$, a main effect of SNR, $F(2, 22) = 18.31, p < .001$, but no main effect of speech level, $F(2, 22) = 2.64, p = .094$ on listening effort ratings. The results suggested a pattern of decreasing listening effort (less effort was needed) with increasing SNR.

At a 50 dB SPL speech level, changes in SNR had no effect on listening effort, $F(2, 22) = 1.75, p = .198$. At 65 dB SPL, an increase in SNR from 0 to +5, $t(11) = 12.21, p = .005$, from +5 to +10 SNR, $t(11) = 19.25, p < .001$, and from 0 to +10 SNR, $t(11) = 6.77, p = .025$, led to significantly lower listening effort ratings. At 80 dB SPL, an increase in SNR from 0 to +5, $t(11) = 5.47, p = .039$, from +5 to +10 SNR, $t(11) = 16.14, p = .002$, and from 0 to +10, $t(11) = 11.21, p = .006$, led to significantly lower listening effort ratings.

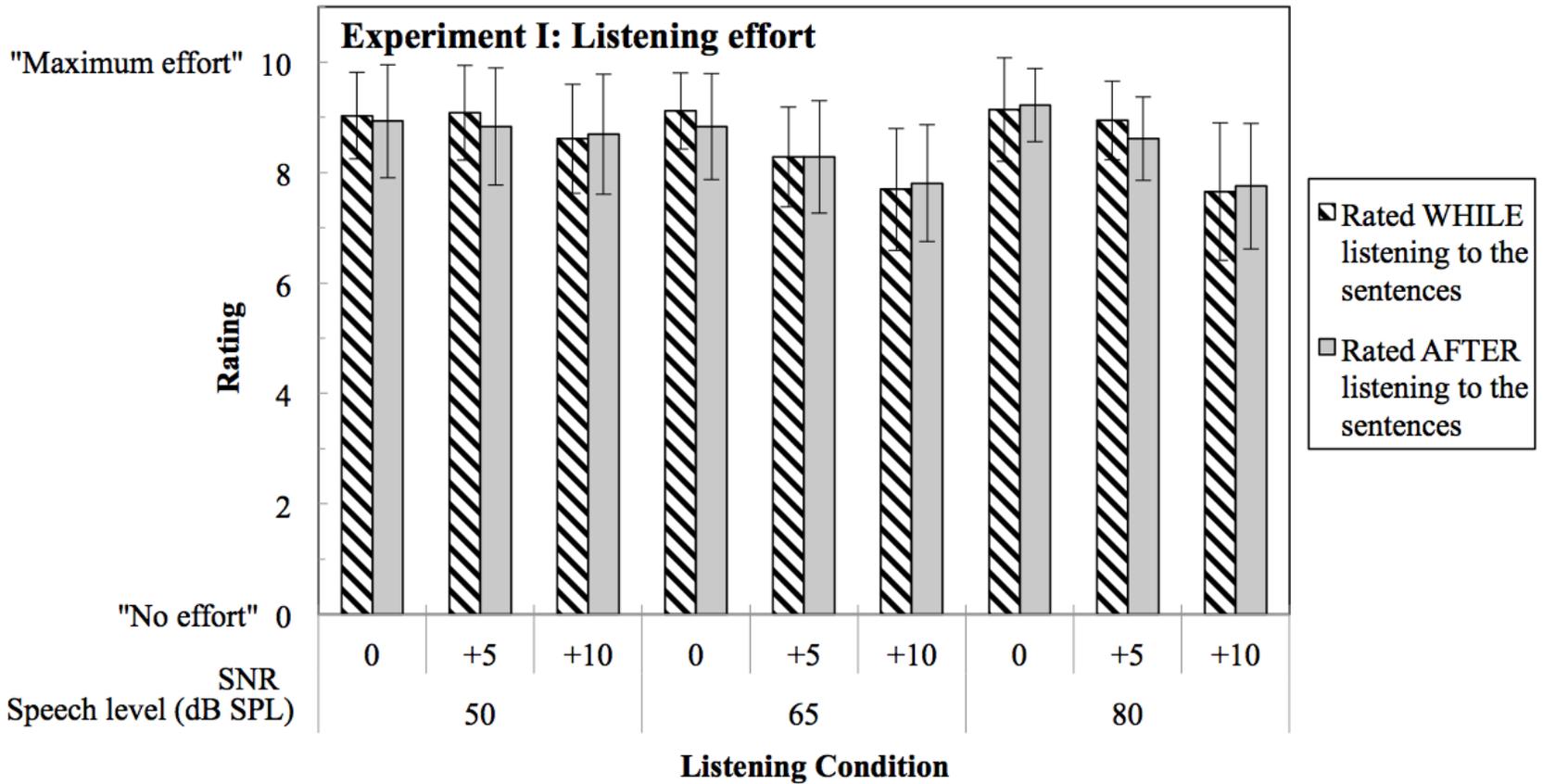


Figure 3.7. Mean ratings for listening effort made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for listening effort ratings and their 99% confidence intervals are shown in Figure 3.3. The listening effort intra-rater reliability was poor (Cicchetti & Sparrow 1981) across Experiment I. The ICC for listening effort ratings made while listening to the sentences was .23, 99% CI [.05, .71], and .20 for listening effort ratings made after listening to the sentences, 99% CI [.02, .68].

3.2.4 Loudness

Criterion validity

Scatter plots with regression lines for speech intensity and their corresponding loudness ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B) are shown in Figure 3.8. When participants rated loudness while listening to the sentences, their ratings correlated strongly (Dancey & Reidy, 2004) with the range of speech intensities, $r = .74$, 95% CI [.69, .79], $p < .001$. When participants rated loudness after listening to the sentences, their ratings also correlated strongly with the range of speech intensities, $r = .65$, 95% CI [.59, .71], $p < .001$. There was an effect of rating timing on the correlation between loudness ratings and speech intensity; loudness ratings made while listening to the sentences correlated more strongly with their corresponding speech intensities than loudness ratings made after listening to the sentences, $z = 3.508$, $p < .001$.

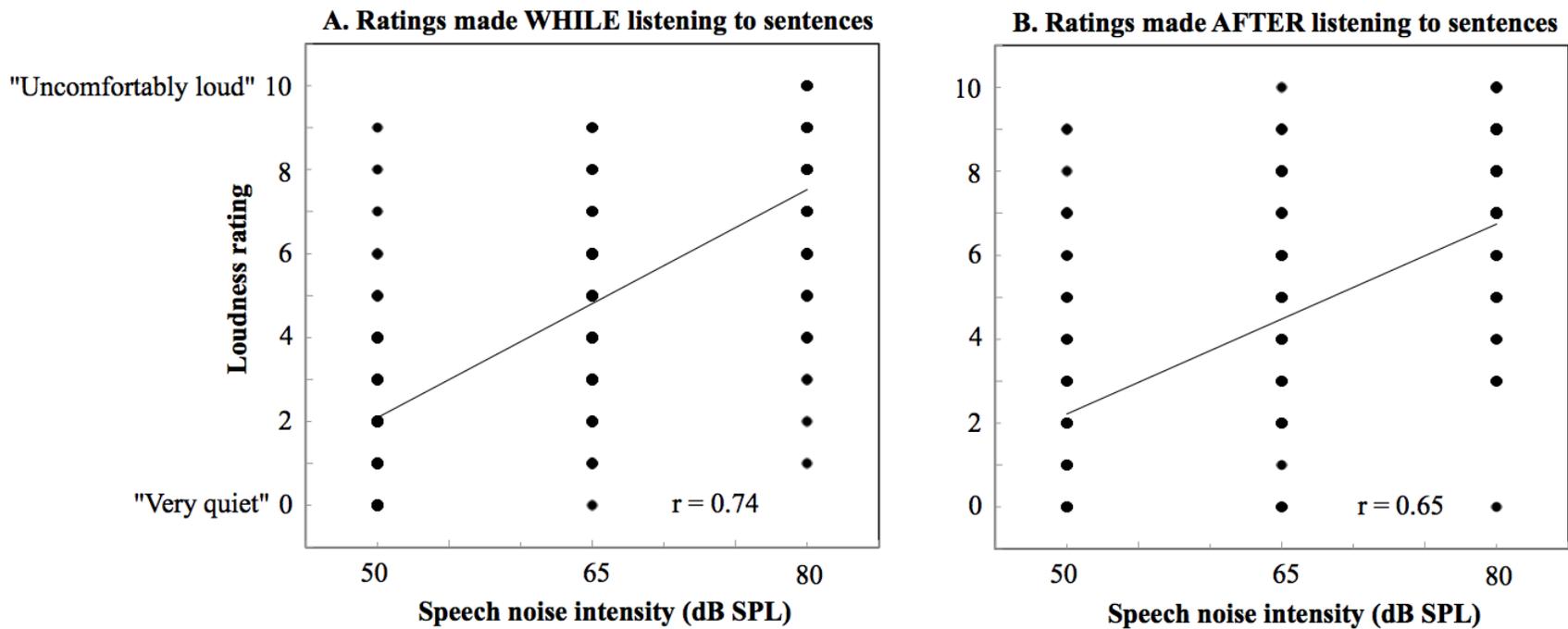


Figure 3.8. Correlation between loudness ratings and corresponding calibrated speech intensities for ratings made while listening to the sentences (panel A), and after listening to the sentences (panel B). Pearson's correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment I loudness ratings are shown in Figure 3.9, and listed in Table M.1 in Appendix M. A three-way analysis of variance showed that there was an interaction between CMA timing (CMA completion either while listening to or immediately after listening to the sentences) and speech level on loudness ratings, $F(2, 22) = 4.03, p = .032$; loudness ratings made after the sentences were significantly lower (perceived as quieter) than loudness ratings made while listening to the sentences, but only when the speech level was at 65 or 80 dB SPL. There was no interaction between SNR and speech level, $F(4, 44) = 0.64, p = .639$, and no main effect of SNR on loudness ratings, $F(2, 22) = 1.14, p = .322$. The results revealed a pattern of increasing loudness ratings with increasing speech level.

When loudness ratings were made while listening to the sentences, an increase in speech level from 50 to 65 dB SPL, $t(11) = 41.37, p < .001$, from 65 to 80 dB SPL, $t(11) = 23.29, p < .001$, and from 50 to 80 dB SPL, $t(11) = 142.1, p < .001$, led to significantly higher loudness ratings. When loudness ratings were made after the sentences, an increase from 50 to 65 dB SPL, $t(11) = 43.84, p < .001$, from 65 to 80 dB SPL, $t(11) = 27.36, p < .001$, and from 50 to 80 dB SPL, $t(11) = 152.00, p < .001$, led to significantly higher loudness ratings. Loudness ratings made after the sentences were significantly lower than loudness ratings made while listening to the sentences at 65 dB SPL, $t(11) = 8.02, p = .016$, and at 80 dB SPL, $t(11) = 9.38, p = .011$. There was no effect of CMA timing on loudness ratings at 50 dB SPL, $t(11) = 0.57, p = .468$.

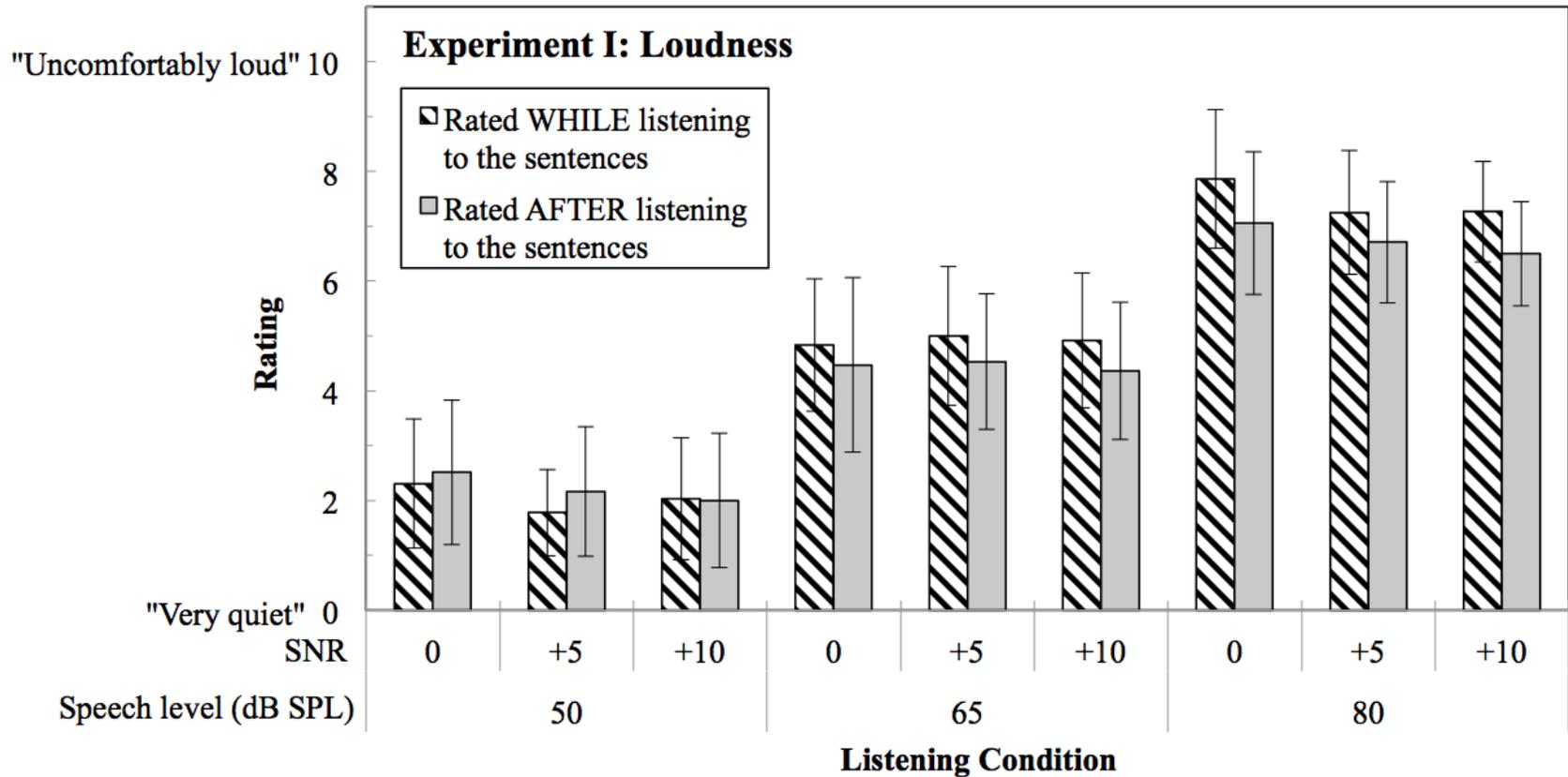


Figure 3.9. Mean ratings for loudness made while or immediately after listening to four sentences in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for loudness ratings and their 99% confidence intervals are shown in Figure 3.3. The loudness intra-rater reliability was fair to good (Cicchetti & Sparrow 1981) when ratings were made after listening and repeating back the sentences in Experiment I, and excellent when rated while listening to the sentences. The ICC for loudness ratings made while listening to the sentences was .80, 99% CI [.65, .97], and .55 for loudness ratings made after listening to the sentences, 99% CI [.34, .91].

3.3 Experiment II—task focus

3.3.1 Connect-the-dots performance

Connect-the-dots performance data from Experiment II were collected and analyzed to determine whether participants appeared to prioritize one of the two tasks as instructed. Means and standard deviations for each participant, and means and standard errors across participants for connect-the-dots performance (correct dots per minute) are shown in Table 3.1. Across participants, the connect-the-dots performance was significantly higher (increased correct dots per minute) when participants were instructed to focus on the secondary connect-the-dots task than when participants were instructed focus was on the primary sentence repetition task, $t(323) = 9.32, p < .001$. The average performance when participants were focused on the connect-the-dots task was also significantly higher than the average baseline performance, $t(59) = 2.77, p = .007$. There was no significant difference between average baseline performance and average performance when participants were focused on the sentence repetition task, $t(59) = 0.70,$

$p = .489.$

Table 3.1. Participant characteristics (age and gender), and their corresponding means and standard deviations for connect-the-dots performance (baseline, focus on sentence repetition, focus on connect-the-dots). Overall mean and standard errors across participants are listed at the bottom of the table.

Participants		Connect-the-dots performance (dots per minute)		
Age	Gender	Baseline	SR Focus	CTD Focus
21	M	21.3±3.5	26.7±3.4	26.3±4.9
24	F	22.5±5.0	20.4±4.6	21.7±4.9
30	F	29.2±2.6	35.5±7.0	39.1±4.9
35	F	27.1±2.4	24.7±5.5	30.1±6.0
37	M	25.9±4.9	18.6±6.9	27.6±4.8
42	F	28.0±6.1	8.1±5.7	26.9±3.6
57	M	18.8±3.0	11.6±3.0	18.4±2.8
61	F	14.1±5.9	15.0±4.5	15.2±4.6
66	F	13.6±2.9	18.1±4.2	18.5±7.1
67	F	19.3±4.1	18.4±5.1	19.6±5.9
73	F	14.6±5.9	10.8±3.1	14.1±4.9
76	F	20.7±1.7	21.7±5.0	25.8±7.9
Overall mean and standard error		21.2±0.8	19.1±0.5	23.6±0.5

3.3.2 Intelligibility

Criterion validity

Scatter plots with regression lines for percent keywords correct and intelligibility ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B) are shown in Figure 3.10. When participants rated intelligibility while focused on the primary sentence repetition task, their ratings correlated strongly (Dancey & Reidy, 2004) with the percent keywords they repeated back correctly, $r = .76$, 95% CI [.71, .81], $p < .001$. When participants rated intelligibility while focused on the secondary connect-the-dots task, their ratings also correlated strongly with the percent keywords they repeated back correctly, $r = .73$, 95% CI [.68, .78], $p < .001$. There was no effect of task focus on the level of correlation between intelligibility ratings and percent keywords correct, $z = 1.23$, $p = .221$.

There was a significant effect of secondary task presence on intelligibility ratings; the correlation for intelligibility ratings to percent keywords correct made after listening to the sentences in Experiment I was stronger than the correlation for intelligibility ratings made when participants were focused on the primary sentence repetition task in Experiment II, $z = 3.14$, $p < .001$.

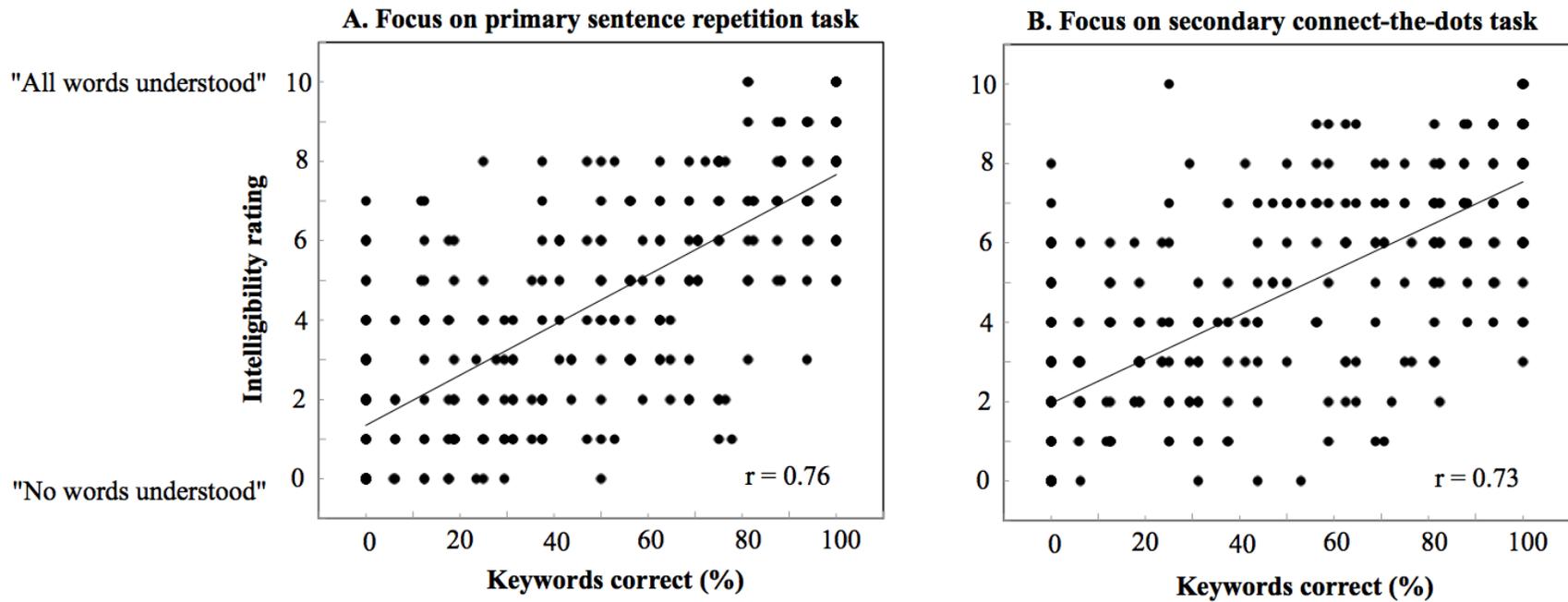


Figure 3.10. Correlation between intelligibility ratings and sentence repetition percent keywords correct for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B). Pearson's correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment II intelligibility ratings are shown in Figure 3.11, and listed in Table M.2 in Appendix M. A three-way analysis of variance showed that the presence of a secondary task had no effect on ratings for intelligibility, $F(1,11) = 2.24, p = .163$. Task focus (whether the participant was instructed to focus on the primary sentence repetition task or the secondary connect-the-dots task) also had no effect on intelligibility ratings, $F(1,11) = 1.39, p = .263$. There was an interaction between SNR and speech level on intelligibility ratings, $F(4,44) = 6.72, p < .001$. The results revealed a pattern of increasing CMA intelligibility ratings (improved intelligibility) with increasing SNR, as well as with increasing speech level.

At the 50 dB SPL speech level, an increase in SNR from 0 to +5, $t(11) = 6.24, p = .030$, from +5 to +10, $t(11) = 19.22, p < .001$, and from 0 to +10, $t(11) = 24.05, p < .001$, led to significantly higher intelligibility ratings. At 65 dB SPL, an increase in SNR from 0 to +5, $t(11) = 22.57, p < .001$, from +5 to +10, $t(11) = 11.06, p = .007$, and from 0 to +10, $t(11) = 55.68, p < .001$, led to significantly higher intelligibility ratings. At 80 dB SPL, an increase in SNR from 0 to +5, $t(11) = 37.13, p < .001$, from +5 to +10, $t(11) = 34.37, p < .001$, and from 0 to +10, $t(11) = 65.53, p < .001$, led to significantly higher intelligibility ratings.

At 0 SNR, an increase in speech level from 50 to 65 dB SPL, $t(11) = 18.80, p < .001$, and from 50 to 80 dB SPL, $t(11) = 10.67, p = .007$, led to significantly higher intelligibility ratings. However, there was no effect on intelligibility ratings, $t(11) = 0.21, p = .676$, when the speech level was increased from 65 to 80 dB SPL. At +5 SNR, an increase in speech level from 50 to 65 dB SPL, $t(11) = 30.70, p < .001$, and from 50 to 80

dB SPL, $t(11) = 21.83$, $p < .001$, led to significantly higher intelligibility ratings.

However, there was no effect on intelligibility ratings, $t(11) = 0.18$, $p = 0.657$, when the speech level was increased from 65 to 80 dB SPL. At +10 SNR, an increase in speech level from 50 to 65 dB SPL, $t(11) = 13.04$, $p = .004$, from 50 to 80 dB SPL, $t(11) = 28.39$, $p < .001$, and from 65 to 80 dB SPL led to significantly higher intelligibility ratings.

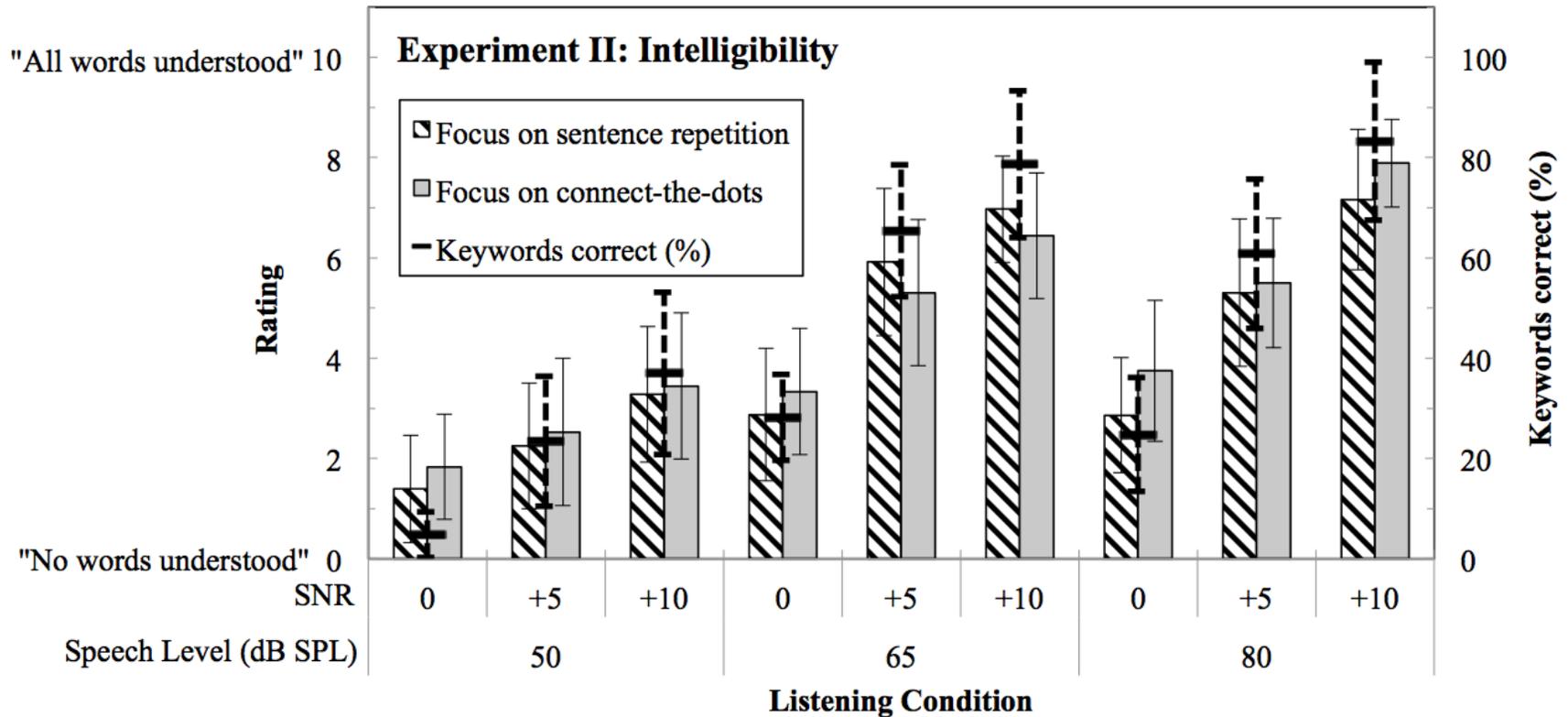


Figure 3.11. Comparison of mean ratings for intelligibility made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). Average percent keywords correct for each listening condition in Experiment II across participants are shown for comparison. Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for intelligibility ratings and their 99% confidence intervals are shown in Figure 3.3. The intelligibility intra-rater reliability was fair to good (Cicchetti & Sparrow 1981) across Experiment II regardless of secondary task presence or task focus. The ICC for intelligibility ratings made while the participants' focus was on sentence repetition was .57, 99% CI [.36, .92], and .56 for intelligibility ratings made while the participants' focus was on the connect-the-dots task, 99% CI [.35, .91].

3.3.3 Noisiness

Criterion validity

Scatter plots with regression lines for background noise intensity and their corresponding noisiness ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B) are shown in Figure 3.12. When participants rated noisiness while focused on the primary sentence repetition, their ratings correlated moderately (Dancey & Reidy, 2004) with the range of background noise intensities, $r = .66$, 95% CI [.60, .72], $p < .001$. When participants rated noisiness while focused on the secondary connect-the-dots task, their ratings also correlated moderately with the range of background noise intensities, $r = .67$, 95% CI [.61, .73], $p < .001$. There was no effect of task focus on the level of correlation between noisiness ratings and the range of background noise intensities, $z = .39$, $p = .698$.

Secondary task presence had no effect on noisiness ratings; there was no significant difference between the correlation for noisiness ratings made after listening to

the sentences in Experiment I to the correlation for noisiness ratings made when participants were focused on the primary sentence repetition task in Experiment II, $z = 0.05$, $p = .957$.

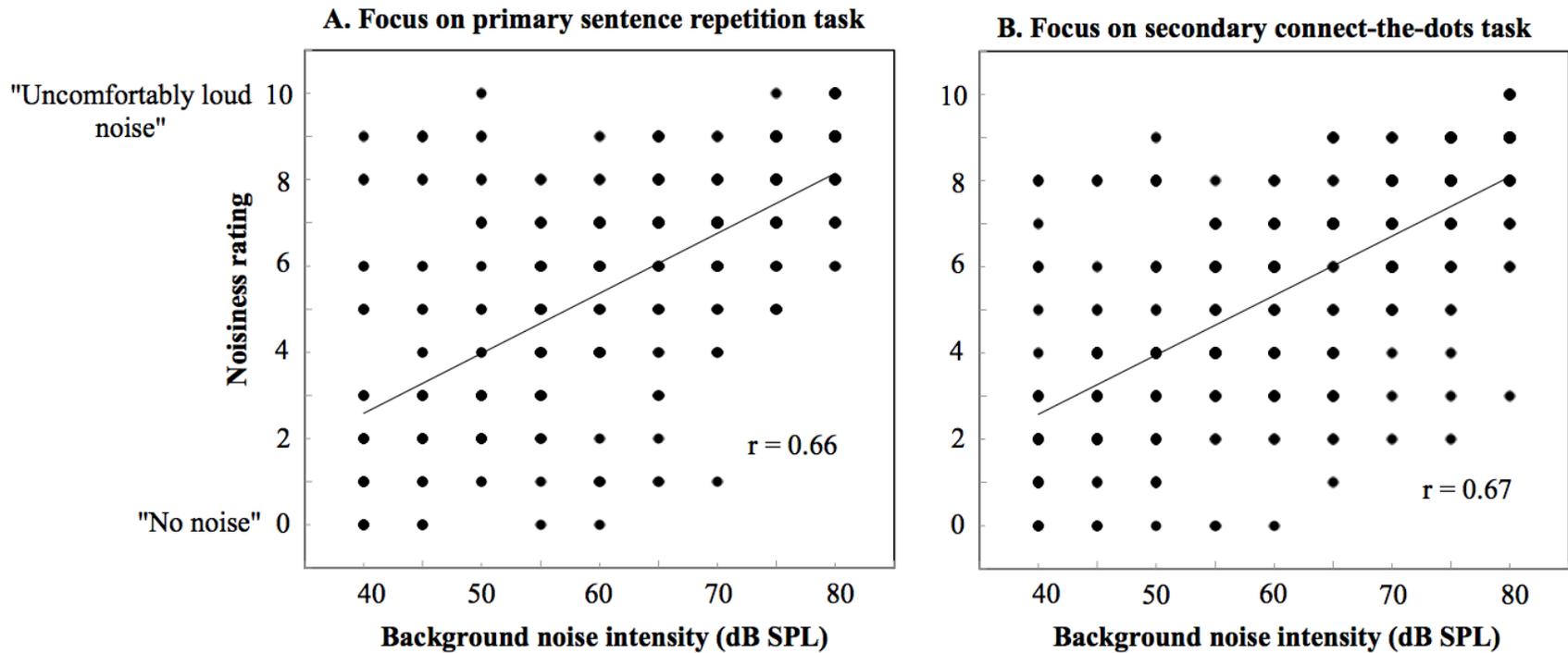


Figure 3.12. Correlation between noisiness ratings and corresponding calibrated background noise intensities for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B). Pearson’s correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment II noisiness ratings are shown in Figure 3.13, and listed in Table M.2 in Appendix M. The presence of a secondary task had no effect on noisiness ratings, $F(1, 11) = 0.68, p = .426$, and there was no effect of task focus on noisiness ratings, $F(1, 11) = 0.03, p = .874$. There was a main effect of speech level, $F(2, 22) = 42.75, p < .001$, and a main effect of SNR, $F(2, 22) = 97.00, p < .001$, such that noisiness ratings increased (the noise was perceived as louder) with increasing speech level, as well as with decreasing SNR. There was no significant interaction between speech level and SNR, $F(4, 44) = 0.51, p = .215$. The results revealed a pattern of increasing noisiness ratings with increasing speech level, and with decreasing SNR.

An increase in speech level from 50 to 65 dB SPL, $t(11) = 19.55, p < .001$, from 65 to 80 dB SPL, $t(11) = 37.82, p < .001$, and from 50 to 80 dB SPL, $t(11) = 58.64, p < .001$, led to significantly higher noisiness ratings. An increase in SNR from 0 to +5, $t(11) = 55.70, p < .001$, from +5 to +10 dB SNR, $t(11) = 67.05, p < .001$, and from 0 to +10 SNR, $t(11) = 147.1, p < .001$, led to significantly lower noisiness ratings (the noise was perceived as quieter).

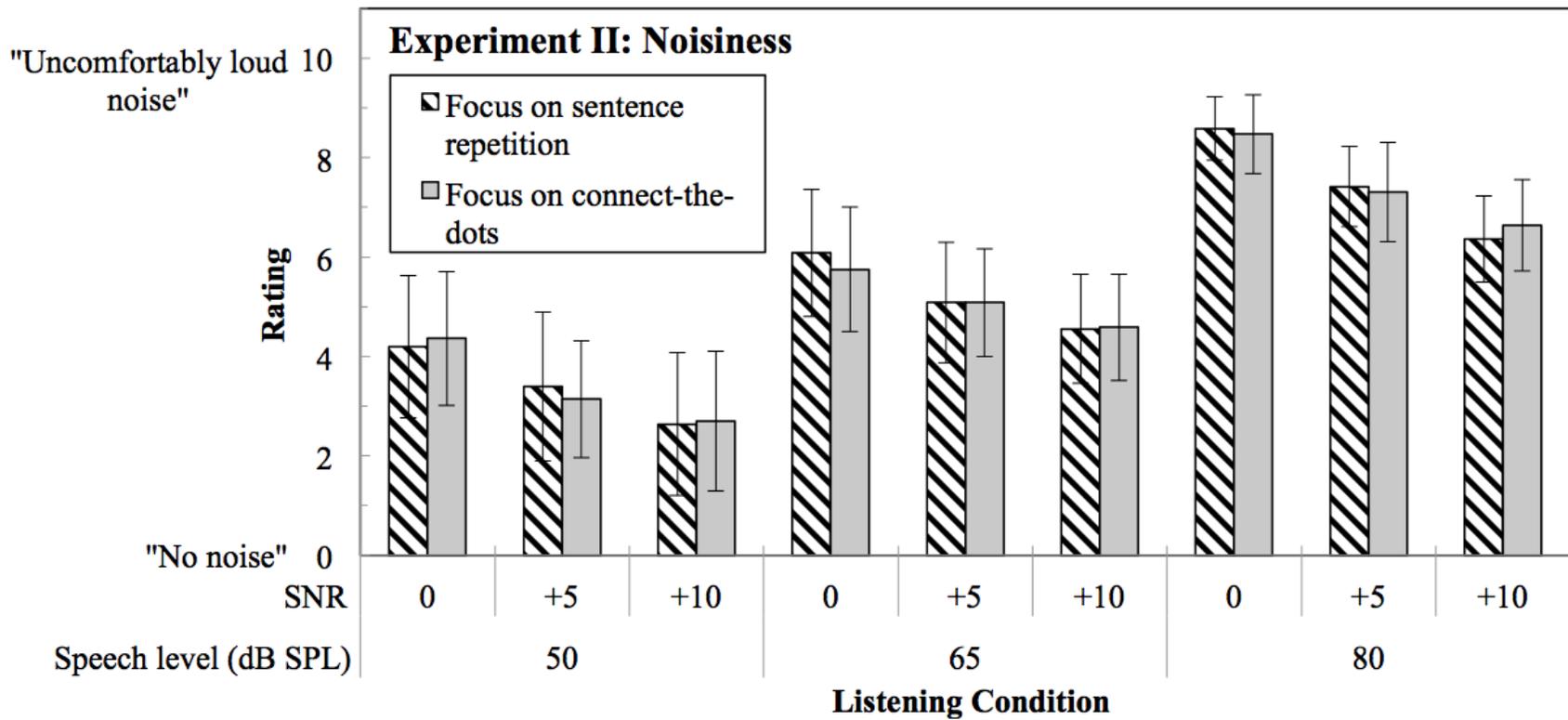


Figure 3.13. Mean ratings for noisiness made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). The background noise intensities (dB SPL) for each listening condition are listed along the bottom of the figure for comparison. Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for noisiness ratings and their 99% confidence intervals are shown in Figure 3.3. The noisiness intra-rater reliability was fair to good (Cicchetti & Sparrow 1981) when ratings were made with a secondary task present, regardless of the task focus. The ICC for noisiness ratings made while the participants' focus was on sentence repetition was .73, 99% CI [.55, .96], and .70 for noisiness ratings made while the participants' focus was on the connect-the-dots task, 99% CI [.52, .95].

3.3.4 Listening effort

Criterion validity

Scatter plots with regression lines for percent keywords correct and their corresponding listening effort ratings made with focus on the primary sentence repetition (panel A), and with focus on the secondary connect-the-dots task (panel B) are shown in Figure 3.14. When participants rated listening effort while focused on the sentence repetition task, their ratings correlated weakly (Dancey & Reidy, 2004) with the percent keywords correct, $r = -.32$, 95% CI [-.42, -.22], $p < .001$. When participants rated listening effort while focused on the connect-the-dots task, their ratings also correlated weakly with the percent keywords correct, $r = -.11$, 95% CI [-.22, .00], $p < .001$. There was no effect of task focus on the correlation between listening effort ratings and percent keywords correct, $z < 0.01$, $p = .999$. There was also no significant effect of secondary task presence; the correlation for listening effort ratings made after listening to the sentences in Experiment I to the correlation for listening effort ratings made when participants were

focused on the primary sentence repetition task in Experiment II were not significantly different, $z = 0.02$, $p = .983$.

Scatter plots with regression lines for correct dots per minute (secondary connect-the-dots task performance) and their corresponding listening effort ratings made with focus on the primary sentence repetition (panel A), and with focus on the secondary connect-the-dots task (panel B) are shown in Figure 3.15. There was no effect of task focus on the level of correlation between listening effort ratings and correct dots per minute, $z = 0.60$, $p = .548$. When participants rated listening effort while focused on the sentence repetition task, their ratings correlated moderately with their dots per minute score, $r = -.47$, 95% CI [-.56, -.38], $p < .001$. When participants rated listening effort while focused on the connect-the-dots task, their ratings also correlated moderately with their dots per minute score, $r = -.46$, 95% CI [-.55, -.37], $p < .001$.

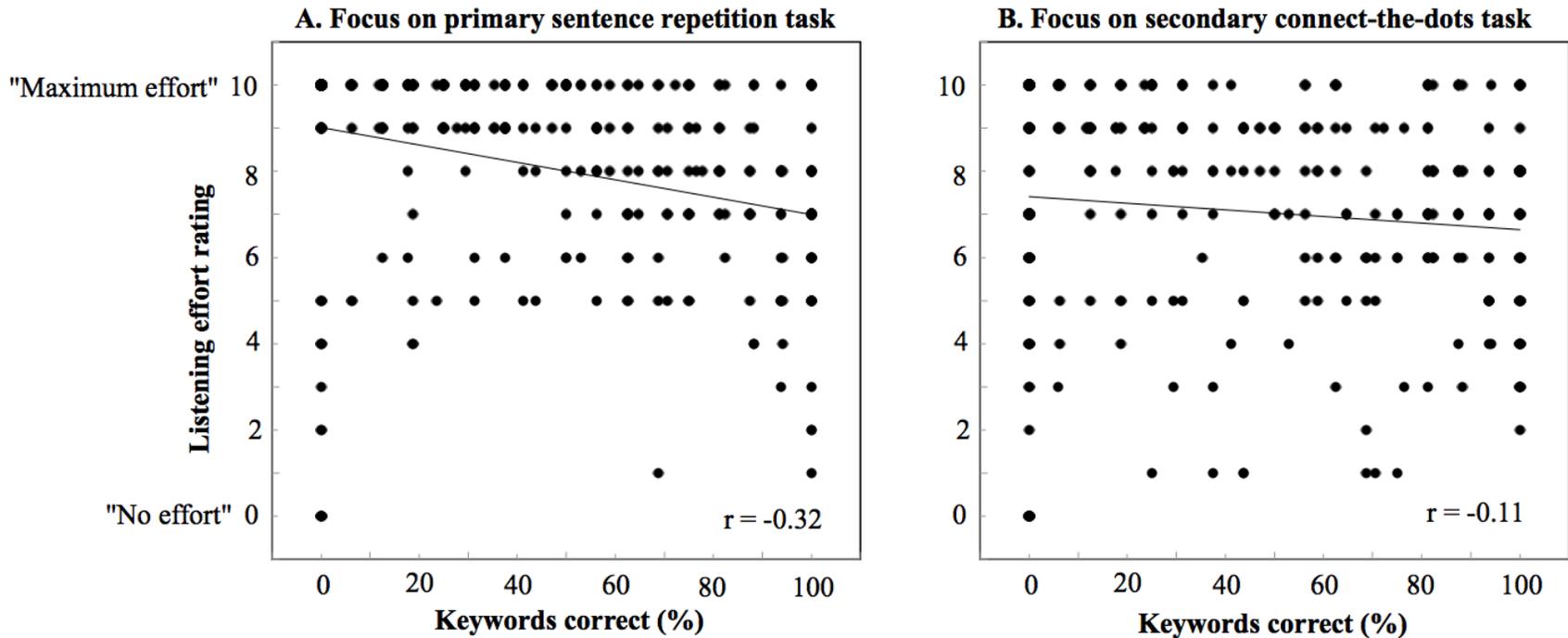


Figure 3.14. Correlation between listening effort ratings and sentence repetition performance (percent keywords correct) for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B). Pearson's correlation coefficients are shown inside the graphs.

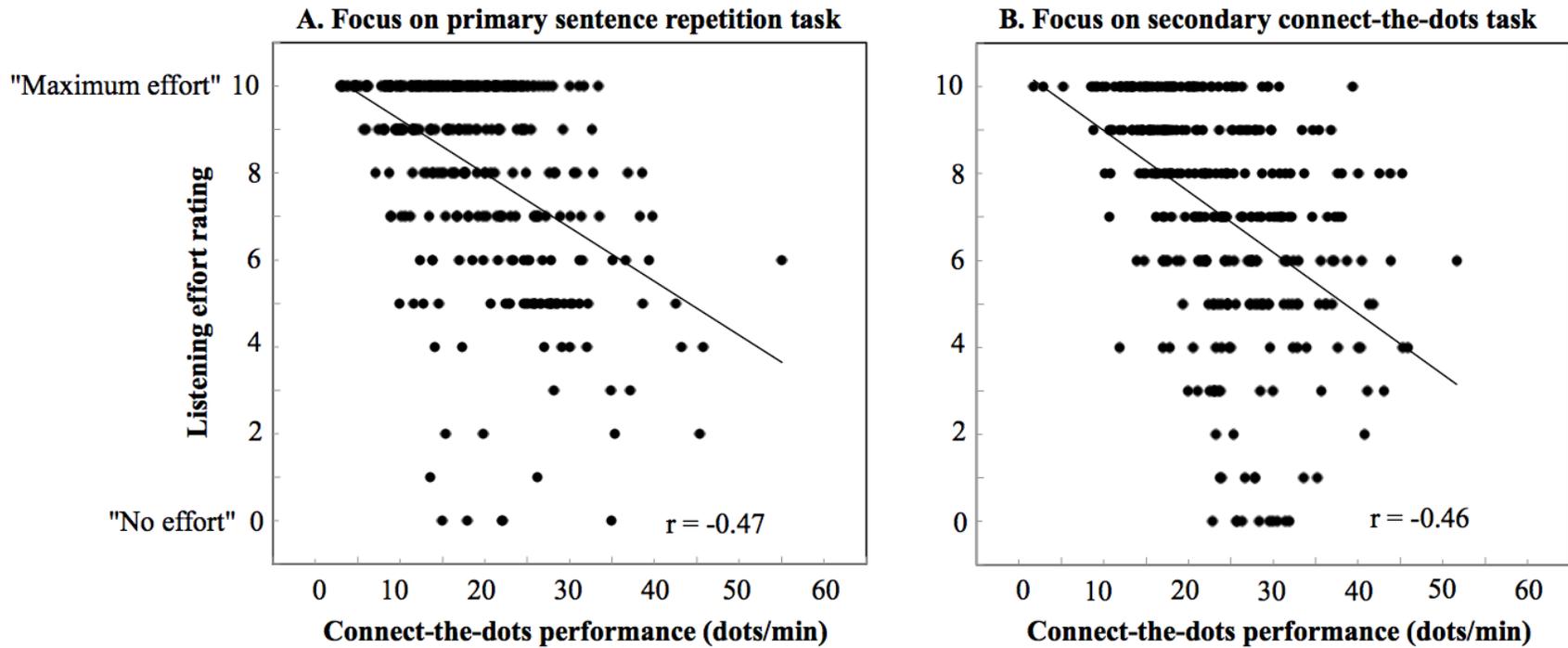


Figure 3.15. Correlation between listening effort ratings and secondary connect-the-dots performance (dots/min) for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B). Pearson's correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment II listening effort ratings are shown in Figure 3.16, and listed in Table M.2 in Appendix M. The results revealed no significant differences among mean listening effort ratings for task focus, $F(1, 11) = 2.98$, $p = .112$, SNR, $F(2, 22) = 1.91$, $p = .173$, or speech level, $F(2, 22) = 0.48$, $p = .627$. There was also no interaction between SNR and speech level on listening effort ratings, $F(4, 44) = 1.62$, $p = .186$. The presence of a secondary task had no effect on ratings for listening effort, $F(1, 11) = 2.32$, $p = .156$.

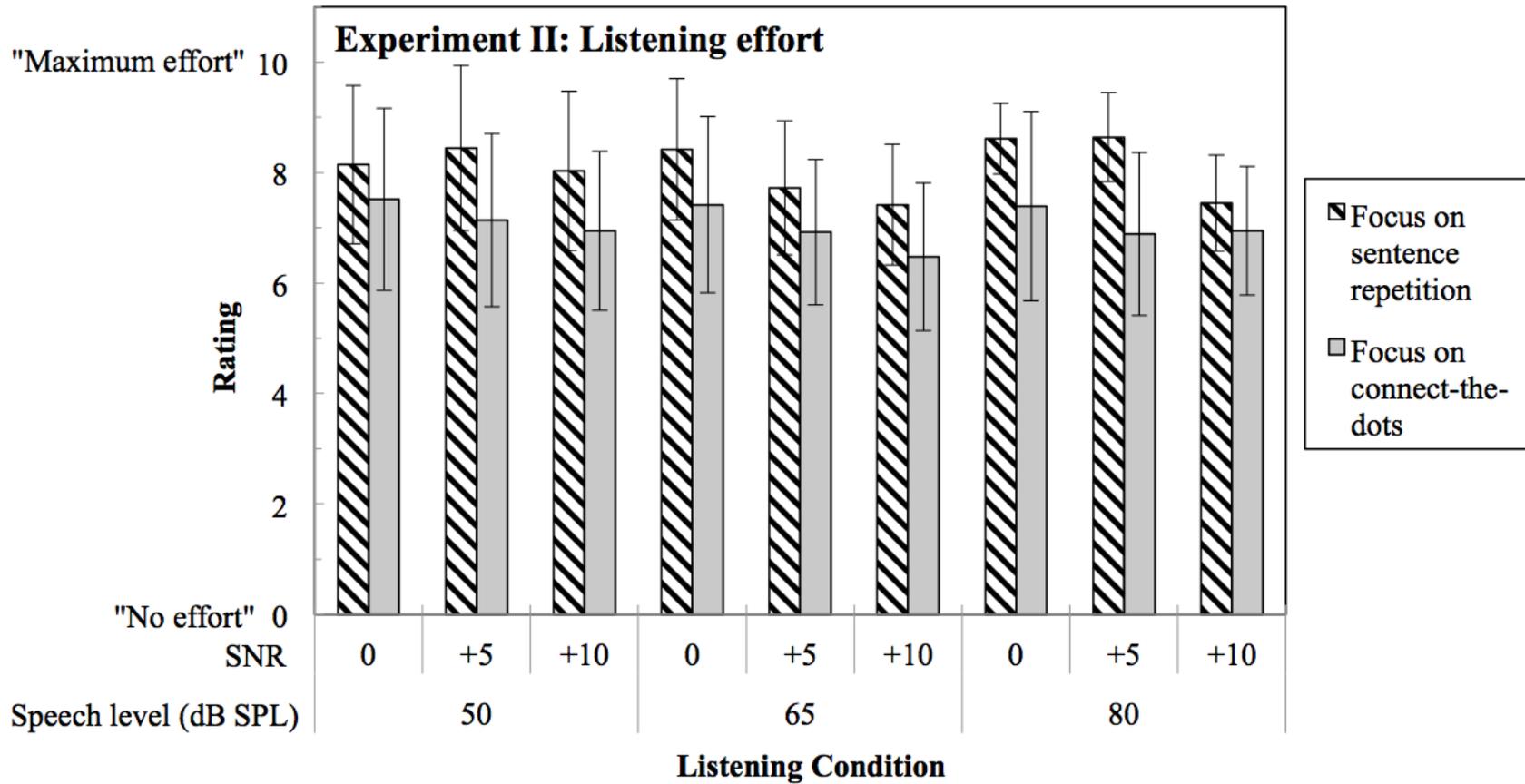


Figure 3.16. Mean ratings for listening effort made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for listening effort ratings and their 99% confidence intervals are shown in Figure 3.3. The listening effort intra-rater reliability was poor (Cicchetti & Sparrow 1981) across Experiment II regardless of secondary task presence or task focus. The ICC for listening effort ratings made while the participants' focus was on sentence repetition was .01, 99% CI [-.11, .13], and .05 for listening effort ratings made while the participants' focus was on the connect-the-dots task, 99% CI [-.08, .33].

3.3.5 Loudness

Criterion validity

Scatter plots with regression lines for speech intensity and their corresponding loudness ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B) are shown in Figure 3.17. There was no effect of secondary task presence on loudness ratings; the correlation for loudness ratings made after listening to the sentences in Experiment I was not significantly different than the correlation for loudness ratings made when participants were focused on the primary sentence repetition task in Experiment II, $z = 0.60, p = .273$. There was also no effect of task focus on the level of correlation between loudness ratings and the range of speech intensities, $z = 0.52, p = .302$. When participants rated loudness while focused on the primary sentence repetition, their ratings correlated strongly (Dancey & Reidy, 2004) with the range of speech intensities, $r = .62, 95\% \text{ CI } [.55, .69], p < .001$. When participants rated loudness while focused on the secondary connect-the-dots task,

their ratings also correlated strongly with the range of speech intensities, $r = .64$, 95% CI [.58, .70], $p < .001$.

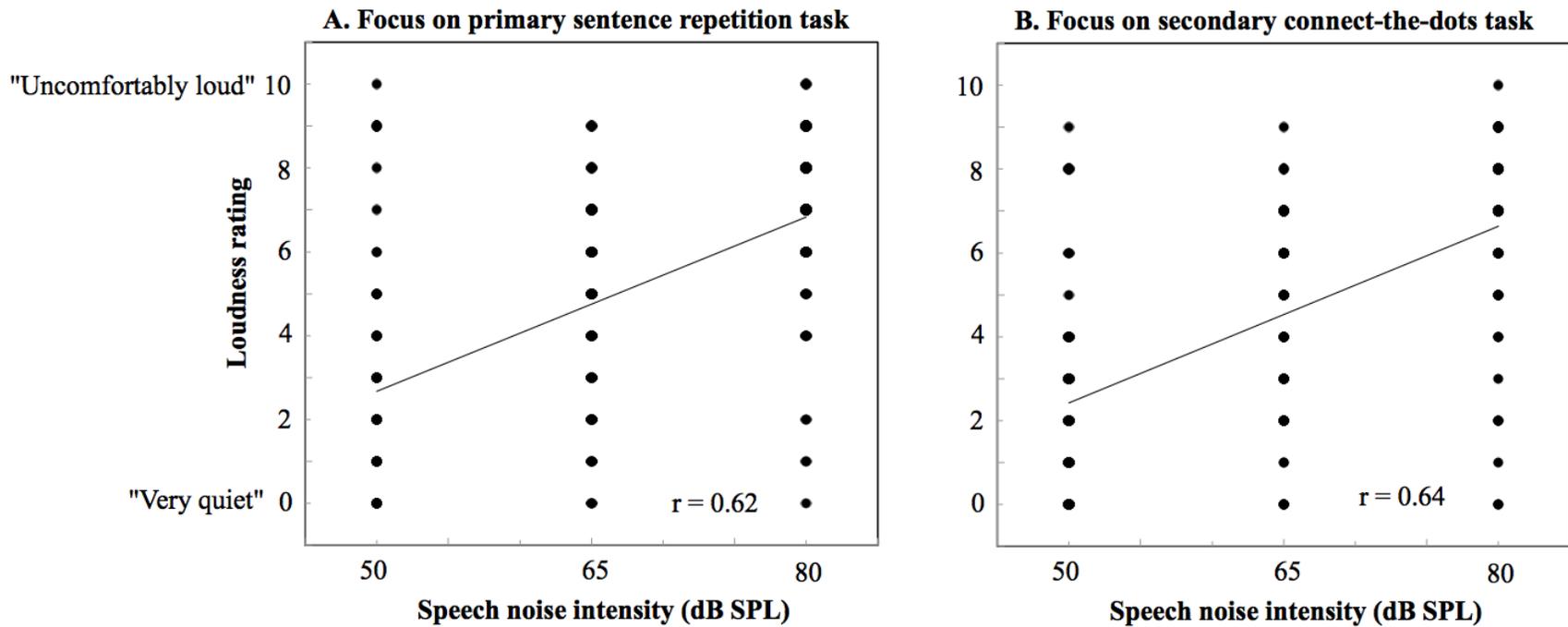


Figure 3.17. Correlation between loudness ratings and corresponding calibrated speech intensities for ratings made while focused on the primary sentence repetition task (panel A), and while focused on the secondary connect-the-dots task (panel B). Pearson's correlation coefficients are shown inside the graphs.

Construct validity

Means and 95% confidence intervals for Experiment II loudness ratings are shown in Figure 3.18, and listed in Table M.2 in Appendix M. There was no effect of secondary task presence, $F(1, 11) = 1.06, p = 0.326$, task focus, $F(1, 11) = 1.26, p = .283$, SNR, $F(2, 22) = 0.26, p = .639$, and no interaction between SNR and speech level, $F(4, 44) = 1.16, p = .343$, on loudness ratings. There was only a— main effect of speech level, $F(2, 22) = 52.30, p < .001$, on loudness ratings. The results revealed a pattern of increasing loudness ratings with increasing speech level.

An increase in speech level from 50 to 65 dB SPL, $t(11) = 28.03, p < .001$, from 65 to 80 dB SPL, $t(11) = 26.04, p < .001$, and from 50 to 80 dB SPL, $t(11) = 45.84, p < .001$, led to significantly higher loudness ratings.

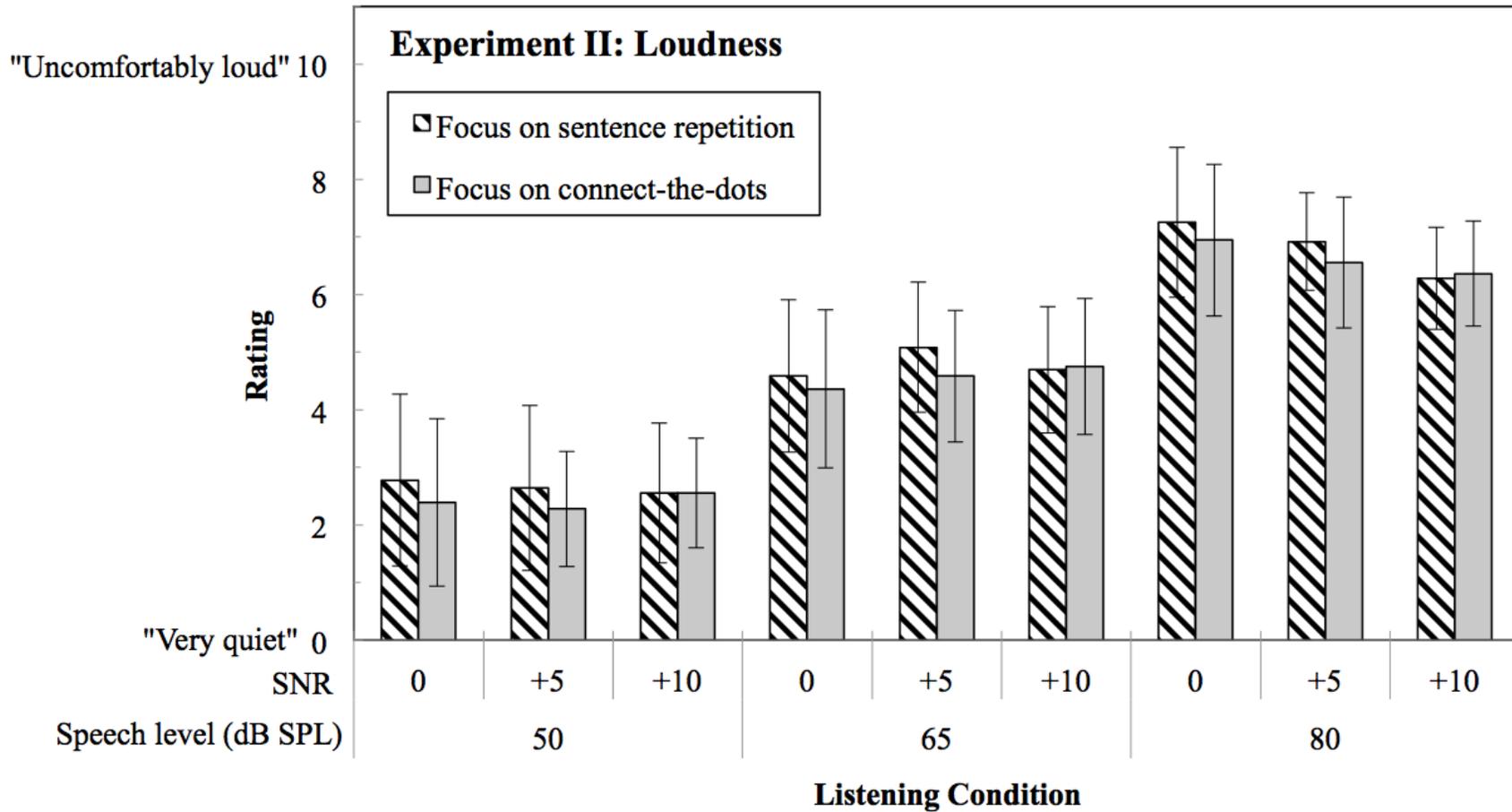


Figure 3.18. Mean ratings for loudness made with instructed focus on either the primary sentence repetition task or the secondary connect-the-dots task in each of the 9 listening conditions that varied by signal-to-noise-ratio (SNR) (0, +5, +10) and speech level (50, 65, 80 dB SPL). Error bars indicate 95% confidence intervals for the mean.

Reliability

The intraclass correlation coefficients (ICCs) for loudness ratings and their 99% confidence intervals are shown in Figure 3.3. The loudness intra-rater reliability was fair to good (Cicchetti & Sparrow 1981) regardless of the task focus. When loudness was rated while listening to the sentences, the intra-rater reliability was excellent. The ICC for loudness ratings made while the participants' focus was on sentence repetition was .62, 99% CI [.42, .93], and .50 for loudness ratings made while the participants' focus was on the connect-the-dots task, 99% CI = [.28, .89].

Chapter 4: Discussion and Conclusion

4.1 Introduction

Due to the recent proliferation of smartphones and the many hearing aid technological advancements over the last decade, an efficient and accurate method for collecting listening-situation-specific hearing aid benefit outcome measures is now possible; this methodology is called contextual momentary assessment (CMA), and involves the repeated measurement of experiences on a day-to-day basis (Shiffman et al., 2008). CMA can provide real-time subjective hearing aid benefit data in relation to any of an individual's varied listening situations, thereby circumventing the biases and memory requirements inherent in retrospective outcome measures. Recent research indicates that using smartphones for collecting data on hearing aid benefit is feasible among adult hearing aid users (Galvez et al., 2012). However, no studies have been done to establish the validity and reliability of CMA ratings when evaluating auditory experiences, or whether factors such as rating timing relative to the listening event, secondary task presence, and relative focus on the primary or secondary task threaten the validity and reliability of CMA ratings. The current study was performed to address these issues.

Validity and reliability are important psychometric attributes for any outcome measure instrument. This study evaluated the criterion validity, construct validity, and intra-rater reliability of four speech and sound quality rating dimensions: intelligibility, noisiness, listening effort, and loudness. Criterion validity was evaluated by determining the significance and strength of correlations between subjective ratings and their corresponding objective measures. Construct validity was evaluated by comparing the

resulting pattern to a predicted pattern of ratings for each dimension. The intra-rater reliability of ratings was evaluated through intraclass correlation coefficients (ICCs) based on the within-subject repeated measures for each of the nine lab-controlled speech-in-noise listening conditions. The current study's validity and reliability results for the four speech and sound quality rating dimensions are discussed below along with their implications.

4.2 Intelligibility

A summary of the intelligibility rating dimension results for validity and reliability are shown in Table 4.1. The method used in this study for defining and rating intelligibility was valid and reliable for the range of speech-in-noise listening situations presented. Intelligibility ratings were most correlated ($r = .83$) to objective measures of intelligibility (percent keywords correct) when rated immediately after a listening event where the participant was not performing a secondary task. This correlation is similar to those found in previous studies examining the relationship between objective and subjective ratings of intelligibility in hearing impaired individuals; Cox, Alexander, and Rivera (1998) reported a correlation of $r = .92$, while Cienkowski and Speaks (2000) reported a correlation of $r = .86$ between intelligibility ratings and objective intelligibility scores. The presence of a secondary task significantly reduced the correlation strength between subjective and objective measures of intelligibility, which may be due to the detrimental effect of divided attention on the encoding of memories (Naveh-Benjamin et al., 2005; Nyberg et al., 1997). Timing of the intelligibility rating with respect to the listening event also made a difference in correlation strength; intelligibility ratings were

significantly more correlated to the objective percent keywords correct when rated after the listening event was over. In fact, during data collection, several study participants mentioned that their confidence regarding the accuracy of their intelligibility judgments was best when they made ratings immediately after the listening event.

The construct used in this study for defining and rating intelligibility was found to be valid across Experiment I and Experiment II. Participants perceived an increase in intelligibility with increasing SNR as hypothesized, as well as with increasing speech level. The 11-point Likert-type rating scale method was appropriate for characterizing subjective intelligibility based on the range of listening conditions employed in the study, which are considered to represent the range of speech levels and SNR in various real-world environments (Pearsons et al., 1977; Smeds et al., 2015).

The intra-rater reliability for intelligibility ratings across Experiment I and Experiment II was fair to good (Cicchetti & Sparrow 1981), indicating that when presented with the same lab-controlled listening condition, participants had somewhat similar intelligibility ratings. The variance of intra-rater ratings among repeated measures in the same listening condition could be attributed to varied familiarity with the presented sentences (Sakamoto, Iwaoka, Suzuki, Amano, & Kondo, 2004), or varying levels of fatigue and attention during each listening trial.

Table 4.1. Summary of results for the intelligibility rating dimension

Intelligibility	Experiment I - timing of ratings			Experiment II - task focus			Secondary task effect?
	WHILE	AFTER	Timing effect?	SR	CTD	Focus effect?	
Criterion validity (r) ^a	.75*	.83*	.83 > .75*	.76*	.73*	.76 = .73	.83 > .76*
Construct validity ^b	$p < .001$	$p < .001$	$p = .259$	$p < .001$	$p < .001$	$p = .263$	$p = .163$
Reliability (ICC) ^c	.53	.56	-	.57	.56	-	-

Note. SR = sentence repetition; CTD = connect-the-dots; r = Pearson's correlation coefficient.

^aIntelligibility ratings were correlated with the participants' corresponding objective intelligibility scores (percent keywords correct).

^bBased on hypothesis test outcome for: H₀ = Intelligibility ratings do not change significantly; H₁ = Intelligibility ratings increase (improve) with decreasing SNR (background noise becomes more intrusive).

^cIntraclass correlation coefficient (ICC) < 0.40 = Poor (Cicchetti & Sparrow, 1981).

* $p < .01$.

4.3. Noisiness

A summary of the noisiness rating dimension results for validity and reliability are shown in Table 4.2. The method used in this study for defining and rating noisiness was valid and reliable for the range of speech-in-noise listening situations that were presented. Noisiness ratings were most correlated to the range of background babble noise intensities when ratings were made while participants were involved in the listening event and were not simultaneously performing a secondary task. When analyzing the construct validity of noisiness ratings, it was evident that noisiness ratings made after a particularly noisy (more intense background noise) listening event tended to be lower than ratings made while the listening event was occurring. Participants appeared to perceive the louder listening situations as less extreme when they made ratings afterwards. In general, noisiness ratings appeared to have construct validity; participants perceived an increase in noisiness with increasing background noise as hypothesized.

The reliability for noisiness ratings was fair to good (Cicchetti & Sparrow 1981), for all conditions except for ratings made while listening to the sentences in Experiment I, which had excellent intra-rater reliability. The noisiness rating results also suggest that participants were able to reliably detect the 5 dB changes in background noise level employed in this study. These results were hypothesized, because the human just noticeable difference (jnd) for changes in sound intensity ranges from 0.25 to 2.5 dB SPL (Valente, Fernandez, & Monroe, 2010).

Table 4.2. Summary of results for the noisiness rating dimension

Noisiness	Experiment I - timing of ratings			Experiment II - task focus			Secondary task effect?
	WHILE	AFTER	Timing effect?	SR	CTD	Focus effect?	
Criterion validity (r) ^a	.77*	.66*	.77 > .66*	.66*	.67*	.66 = .67	.66=.66
Construct validity ^b	$p < .001$	$p < .001$	$p = .021^c$	$p < .001$	$p < .001$	$p = .874$	$p = .426$
Reliability (ICC) ^d	.80	.63	-	.73	.70	-	-

Note. SR = sentence repetition; CTD = connect-the-dots; r = Pearson's correlation coefficient.

^aNoisiness ratings were correlated with the corresponding background noise intensities.

^bBased on hypothesis test outcome for: H₀ = Noisiness ratings do not change significantly; H₁ = Noisiness ratings increase (perceived as louder) with increasing speech level and with decreasing SNR (background noise becomes more intrusive) at each speech level.

^cRating timing interacted with speech level such that at the 80 dB SPL speech level, noisiness ratings made after listening to the sentences were significantly lower than noisiness ratings made while listening to the sentences.

^dIntraclass correlation coefficient (ICC) < 0.40 = Poor (Cicchetti & Sparrow, 1981).

* $p < .01$.

4.4 Listening effort

A summary of the listening effort rating dimension results for validity and reliability are shown in Table 4.3. Listening effort ratings for the speech-in-noise situations used in this study had poor reliability, but show signs of validity based on moderate-strength correlations with objective scores for intelligibility (percent keywords correct) in Experiment I and secondary task performance in Experiment II. Listening effort ratings were only slightly correlated to intelligibility scores (percent keywords correct) in Experiment II. The significant, but only moderate correlations with both intelligibility scores and secondary task performance indicate that listening effort can be explained to an extent by each of these objective measures, but not entirely. Indeed, there are many studies on listening effort that report a lack of correlation between subjective and objective measures of listening effort (Fraser et al., 2010; McGarrigle et al., 2014; Sarampalis et al., 2009).

Listening effort ratings made regarding the speech-in-noise conditions employed in this study revealed ceiling effects. Listening effort ratings also demonstrated questionable construct validity; listening effort ratings decreased (improved) with increasing SNR as expected in Experiment I, but did not significantly change with increasing SNR in Experiment II. Raters often reported they were exerting maximum effort, with some experiencing maximum effort when the SNR was at the maximum level employed in this study (+10 dB), indicating that the SNR levels used were very challenging for many of the participants. Research has shown that when the SNR drops below 15 dB, hearing impaired individuals have to spend more effort on listening,

whereas normal hearing individuals tend to spend more effort on listening when the SNR drops below 5 dB (Valente, Hosford-Dunn, & Roeser, 2008). Schulte et. al. (2009) showed that listening effort is at a maximum until SNR increases up to a level (that varies from person-to-person) where listening effort begins to rapidly decrease; consequently, there can be different ratings of listening effort for the same objective intelligibility score.

The intra-rater reliability for listening effort was poor across both experiments. The variability in listening effort ratings in this lab study could be due to individual factors such as working memory capacity (Larsby, Hällgren, Lyxell, 2008), age (Tomporowski, 2003), and fatigue (Janisse, 1977). In the field, however, contextual factors such as familiarity with the speaker (Van Engen & Peelle), ability to lip read (Picou et al., 2011), and type of background noise (Schulte et al., 2009) may also influence an individual's listening effort.

Table 4.3. Summary of results for the listening effort rating dimension

Listening effort	Experiment I - timing of ratings			Experiment II - task focus			Secondary task effect?
	WHILE	AFTER	Timing effect?	SR	CTD	Focus effect?	
Criterion validity (r) ^a	-.46**	-.43**	-.46 = -.43	-.32** (-.47**)	-.11* (-.46**)	-.32 = -.11	-.43 = -.32
Construct validity ^b	$p = .001$	$p = .001$	$p = .549$	$p = .173$	$p = .173$	$p = .112$	$p = .156$
Reliability (ICC) ^c	.23	.20	-	.01	.05	-	-

Note. SR = sentence repetition; CTD = connect-the-dots; r = Pearson's correlation coefficient.

^aListening effort ratings were correlated with both intelligibility scores (percent keywords correct), and secondary connect-the-dots task performance (correct dots per minute) shown in brackets.

^bBased on hypothesis test outcome for: H_0 = Listening effort ratings do not change significantly; H_1 = Listening effort ratings decrease (improve) with increasing SNR (background noise becomes less intrusive).

^cIntraclass correlation coefficient (ICC) < 0.40 = Poor (Cicchetti & Sparrow, 1981).

* $p < .05$. ** $p < .01$.

4.5 Loudness

A summary of the loudness rating dimension results for validity and reliability are shown in Table 4.4. The method used in this study for defining and rating loudness was valid and reliable for the range of speech-in-noise listening situations that were presented. Loudness ratings were most correlated with the range of speech intensities when ratings were made while participants were involved in the listening event and were not simultaneously performing a secondary task. When analyzing the construct validity of loudness ratings, it was evident that loudness ratings made after a particularly loud listening event tended to be lower than ratings made while the listening event was occurring. Similar to our findings with noisiness ratings, participants appeared to perceive the louder listening situations as less extreme when they made ratings afterwards. In general, loudness ratings appeared to demonstrate construct validity; participants perceived an increase in loudness with increasing speech level as hypothesized.

The reliability for loudness ratings was fair to good (Cicchetti & Sparrow 1981), for all conditions except for ratings made while listening to the sentences in Experiment I, which had excellent intra-rater reliability. As with the noisiness rating dimension, loudness rating validity and reliability was best while the listening situation was occurring, because the speech source was immediately available for evaluation.

The pattern of change in loudness ratings showed that participants were able to reliably detect the 15 dB changes in speech level employed in this study; this finding was hypothesized, and as with the noisiness rating dimension, converges with previous research findings on the human jnd for changes in sound intensity. The human jnd for sound intensity ranges from 0.25 to 2.5 dB SPL (Valente et al., 2010), which is well

below the 15 dB step-size that was used for the speech stimuli, and the 5 dB step-size that was used for the background noise stimuli in this study. The loudness rating scale is likely capable of detecting smaller than 15 dB changes in speech intensity in the field.

Table 4.4. Summary of results for the loudness rating dimension

Loudness	Experiment I - timing of ratings			Experiment II - task focus			Secondary task effect?
	WHILE	AFTER	Timing effect?	SR	CTD	Focus effect?	
Criterion validity (r) ^a	.74*	.65*	.74 > .65*	.62*	.64*	.62 = .64	.65=.62
Construct validity ^b	$p < .001$	$p < .001$	$p = .032^c$	$p < .001$	$p < .001$	$p = .283$	$p = .326$
Reliability (ICC) ^d	.80	.55	-	.62	.50	-	-

Note. SR = sentence repetition; CTD = connect-the-dots; r = Pearson's correlation coefficient.

^aLoudness ratings were correlated with the corresponding intensity of speech (dB SPL).

^bBased on hypothesis test outcome for: H₀ = Loudness ratings do not change significantly; H₁ = Loudness ratings increase with increasing speech level.

^cRating timing interacted with speech level such that at the 65 and 80 dB SPL speech level, loudness ratings made after listening to the sentences were significantly lower than loudness ratings made while listening to the sentences.

^dIntraclass correlation coefficient (ICC) < 0.40 = Poor (Cicchetti & Sparrow, 1981).

* $p < .01$.

4.6 Implications

For listening situations similar to those employed in this lab study, intelligibility, noisiness, and loudness ratings have at least fair to good validity and reliability when rated either while or immediately after the listening situation regardless of the presence of a secondary task. Although there were small significant differences in validity for the intelligibility, noisiness, and loudness dimensions due to timing, these effects may not be clinically relevant. Listening effort ratings, on the other hand, appear to be highly unreliable for speech-in-noise situations (from 0 to +10 dB SNR) among hearing aid users who have a sensorineural hearing loss. Despite poor reliability, listening effort ratings may still be practical as an outcome measure by using the dimension to detect when an individual experiences a significant decrease (improvement) in listening effort. A sudden decrease in perceived listening effort has been shown to occur at a particular SNR point (Schulte et al., 2009), which varies from person to person due to both individual and contextual factors. To control some of this variability, listening effort ratings should be accompanied by data regarding contextual and/or individual factors in the field, such as whether the hearing aid user was able to lip read, how familiar they were with the speaker, and what type of background noise was present. It is also possible that the method used in this study for measuring listening effort was not adequate for accurately capturing the intended meaning of “listening effort”.

4.7 Strengths and limitations

A primary strength of this study was the lab-controlled factorial design that included the random presentation of listening conditions for each participant. The level of

evidence for CMA's ability to facilitate valid and reliable evaluations of intelligibility, noisiness, listening effort, and loudness among experienced hearing aid users with a mild to severe sensorineural hearing loss is high. However, the results of this study should be considered in light of three limitations. One limitation is that there were only 12 participants used in this study. Another limitation of this study was that participants knew they were going to perform ratings of their listening situation, and may have been more attentive and accurate when completing the ratings than in a field setting. Additionally, the results of this study were measured in a laboratory setting and cannot be easily generalized beyond the limited number of listening conditions used; this limitation was necessary in order to be able to make in-lab assumptions based on controlled acoustic conditions, which is a fundamental strength of this study.

4.8 Conclusion

From these findings, we concluded that CMA ratings for intelligibility, noisiness, and loudness among hearing aid users with a mild to severe sensorineural hearing loss were valid and reliable for the lab-controlled speech-in-noise listening conditions used in this study. Noisiness and loudness ratings were most valid and reliable when rated while the listening situation of interest was occurring, and when the rater was not performing more than one task simultaneously. Intelligibility was best evaluated immediately after the listening situation. Listening effort however, had poor reliability and showed ceiling effects among hearing aid users for the range of speech-in-noise situations used in this study. Listening effort may still be practical for use as an outcome measure for easier listening situations where the SNR is better than +10 dB. Further work examining the

utility of the CMA methodology for collecting outcome measures on hearing aid benefit in field-based speech-in-noise listening situations is warranted.

References

- American National Standards Institute. (1999). *Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms*. ANSI/ASA S3.1-1999 (R2013). New York, NY: American National Standards Institute.
- American National Standards Institute. (2010). *Specification for Audiometers*. ANSI/ASA S3.6-2010. New York, NY: American National Standards Institute.
- Auditec of St. Louis. (1971). Four-talker babble [Recording]. St. Louis, MO: Auditec of St. Louis.
- Baddeley, A. D., and Logie, R. H. (1999). Working memory: The multiple-component model. In: Miyake, A., and Shah, P., editors. *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control*. Cambridge: Cambridge University Press, 28–61. doi:10.1017/CBO9781139174909.005
- Barbour, M. B. (2013). *Close to half of Canadians now own a Smartphone*. Ipsos marketing research. Retrieved May 2, 2014, from <http://www.ipsos-na.com/news-polls/pressrelease.aspx?id=6005>
- Bench, J., Kowal, A., and Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, 13, 108–112.
- Bentler, R. A., Neibuhr, D. P., Johnson, T. A. and Flamme, G. A. (2000). Impact of digital labeling on outcome measures. *Ear and Hearing*, 24, 215–224. doi:10.1097/01.AUD.0000069228.46916.92
- Bigelow, J., and Poremba, A. (2014). Achilles' ear? Inferior human short-term and

recognition memory in the auditory modality. *PLoS One*, 9(2):e89914.

doi:10.1371/journal.pone.0089914

Boymans, M., and Dreschler, W. A. (2000). Field trials using a digital hearing aid with active noise reduction and dual-microphone directionality. *Audiology*, 39(5), 260–268.

Brouwer, S., Van Engen, K., Calandruccio, L., and Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *Journal of the Acoustical Society of America*, 131(2), 1449-1464. doi: 10.1121/1.3675943

Buechler, M. (2001). How good are automatic program selection features? A look at the usefulness and acceptance of an automatic program selection mode. *The Hearing Review*, 9, 50–54.

Burke, D. M., and Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin*, 90(3), 513–546. doi:10.1037/0033-2909.90.3.513

Cacioppo, J. T., Hawkley, L. C., Norman, G. J., and Berntson, G. G. (2011). Social isolation. *Annals of the New York Academy of Sciences*, 1231, 17–22.
doi:10.1111/j.1749-6632.2011.06028.x

Calandruccio, L., Van Engen, K., Dhar, S., and Bradlow, A. R. (2010). The effectiveness of clear speech as a masker. *Journal of Speech, Language, and Hearing Research*, 53, 1458-1471. doi: 10.1044/1092-4388(2010/09-0210)

Carney, M. A., Tennen, H., Affleck, G., del Boca, F. K., and Kranzler, H. R. (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time “electronic interviews.” *Journal of Studies on Alcohol*, 59(4),

447–454.

- Cicchetti, D. V., and Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*(2), 127–137.
- Cienkowski, K. M., and Speaks, C. (2000). Subjective vs. objective intelligibility of sentences in listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 43*, 1205–1210. doi: 10.1044/jslhr.4305.1205
- Clark, D. M., and Teasdale, J. D. (1982). Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. *Journal of Abnormal Psychology, 91*(2), 87–95. doi:10.1037/0021-843X.91.2.87
- Cohen, M. A., Horowitz, T. S., and Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences USA, 106*(14), 6008–6010. doi:10.1073/pnas.0811884106
- Cox, R. M., and Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing, 16*(2), 176–186. doi:10.1097/00003446-199504000-00005
- Cox, R. M., and Alexander, G. C. (1999). Measuring satisfaction with amplification in daily life: The SADL scale. *Ear and Hearing, 20*, 306–320.
doi:10.1097/00003446-199908000-00004
- Cox, R. M., and Alexander, G. C. (2002). The International Outcome Inventory for Hearing Aids (IOI-HA): psychometric properties of the English version. *International Journal of Audiology, 41*(1), 30–35.
- Cox, R. M., Alexander, G. C., and Gray, G. A. (2005). Hearing aid patients in private practice and public health (Veterans Affairs) clinics: are they different? *Ear and*

Hearing, 26(6), 513–528.

Cox, R. M., Alexander, G. C., and Rivera, I. M. (1991). Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners.

Journal of Speech and Hearing Research, 34(4), 904-915.

Dancey, C., and Reidy, J. (2004). *Statistics without Maths for Psychology: using SPSS for Windows*, London: Prentice Hall.

Deutsch, D. (1975). Auditory memory. *Canadian Journal of Psychology*, 29(2), 87–105.

Devon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J.,

. . . Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39(2), 155–164. doi:10.1111/j.1547-5069.2007.00161.x

Dillon, H., James, A., and Ginis, J. (1997). The Client Oriented Scale of Improvement (COSI) and its relationship to several other measures of benefit and satisfaction provided by hearing aids. *Journal of the American Academy of Audiology*, 8(1), 27–43.

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, 11(1), 31–45.

Fisher, S. (1975). The microstructure of dual-task interaction. 2. The effect of task instructions on attentional allocation and a model of attention-switching.

Perception, 4(4), 459–474. doi:10.1068/p040459

Foehr, U. G. (2006). *Media multi-tasking among American youth: Prevalence, predictors and pairings: Report*. Menlo Park, CA: Kaiser Family Foundation. Retrieved May 2, 2014, from <http://files.eric.ed.gov/fulltext/ED527858.pdf>

- Fraser, S., Gagné, J.-P., Alepins, M., and Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research, 53*(1), 18–33. doi:10.1044/1092-4388(2009/08-0140)
- Freitas, S., Simões, M. R., Alves, L., and Santana, I. (2013). Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease. *Alzheimer Disease and Associated Disorders, 27*(1), 37–43. doi:10.1097/WAD.0b013e3182420bfe
- Galvez, G., Turbin, M. B., Thielman, E. J., Istvan, J. A., Andrews, J. A., and Henry, J. A. (2012). Feasibility of ecological momentary assessment of hearing difficulties encountered by hearing aid users. *Ear and Hearing, 33*(4), 497–507. doi:10.1097/AUD.0b013e3182498c41
- Gatehouse, S. (1994). Components and determinants of hearing aid benefit. *Ear and Hearing, 15*(1), 30–49. doi:10.1097/00003446-199402000-00005
- Gatehouse, S. (1999). Glasgow Hearing Aid Benefit Profile: Derivation and validation of a client-centred outcome measure for hearing aid services. *Journal of the American Academy of Audiology, 10*(2), 80–103.
- Glista, D., Scollie, S., Bagatto, M., Seewald, R., Parsa, V., and Johnson, A. (2009). Evaluation of nonlinear frequency compression: Clinical outcomes. *International Journal of Audiology, 48*(9), 632–644. doi:10.1080/14992020902971349
- Gorin, A. A., and Stone, A. A. (2004). Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In: Baum, A., Revenson, T., and Singer, J., editors. *Handbook of Health Psychology*. Mahwah, NJ: Lawrence

- Erlbaum Associates Inc., 405–413.
- Gwet, K. L. (2008). Intrarater reliability. Wiley Encyclopedia of Clinical Trials. John Wiley & Sons, Inc.
- Gwet, K. L. (2014). Intrarater reliability. In: Balakrishnan, N., editor. *Methods and Applications of Statistics in Clinical Trials: Planning, Analysis, and Inferential Methods*, Volume 2. Hoboken, NJ: John Wiley & Sons Inc.
doi:10.1002/9781118596333.ch19
- Hasan, S. S., Chipara, O., Wu, Y-H., and Aksan, N. (2014). Evaluating auditory context and their impacts on hearing aid outcomes with mobile phones. Proceedings from PervasiveHealth '14: The 8th International Conference on Pervasive Computing Technologies for Healthcare. doi:10.4108/icst.pervasivehealth.2014.254952
- Hasan, S. S., Lai, F., Chipara, O., and Wu, Y-H. (2013). AudioSense: Enabling real-time evaluation of hearing aid technology in-situ. Proceedings from CBMS 2013 Porto: IEEE 26th International Symposium on Computer-Based Medical Systems.
doi:10.1109/CBMS.2013.6627783
- Hochberg, Y., and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. New York: John Wiley & Sons Inc.
- Humes, L. E., Garner, C. B., Wilson, D. L., and Barlow, N. N. (2001). Hearing-aid outcome measures following one month of hearing aid use by the elderly. *Journal of Speech, Language, and Hearing Research*, 44, 469–486. doi:10.1044/1092-4388(2001/037)
- Humes, L. E. (2003). Modeling and predicting hearing aid outcome. *Trends in Amplification*, 7, 41–75. doi:10.1177/108471380300700202

- Hutchinson, K. M., Duffy, T. L. and Kelly, L. J. (2005). How personality types correlate with hearing aid outcome measures. *Hearing Journal*, 58(7), 28–34.
doi:10.1097/01.HJ.0000286418.68332.c1
- Institute of Electrical and Electronics Engineers. (1969). IEEE Recommended Practice for Speech Quality Measurements. *Audio and Electroacoustics, IEEE Transactions*, 17(3), 225–246. doi:10.1109/TAU.1969.1162058
- Janisse, M. P. (1977). Pupillometry. Washington, DC: Hemispheric.
- Janssen, C. P., Brumby, D. P., and Garnett, R. (2012). Natural break points: The influence of priorities and cognitive and motor cues on dual-task interleaving. *Journal of Cognitive Engineering and Decision Making*, 6(1), 5–29. doi: 10.1177/1555343411432339
- John, O. P., Donahue, E. M., and Kentle, R. L. (1991). *The big five inventory: Versions 4a and 54*. Berkeley, CA: University of California, Institute of Personality and Social Research.
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In: John, O. P., Robins, R. W., and Pervin, L. A. editors. *Handbook of personality: Theory and research*. New York, NY: Guilford Press, 114–158.
- Keller, F. R. (2014). Trails Test (Version 1.0) [Software]. Available from: <https://neuropsychological-assessment-tests.com/>
- Kochkin, S. (2009). MarkeTrak VIII: 25-year trends in the hearing health market. *Hearing Review*, 62(10), 12–31.
- Kochkin, S. (2010). MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly

increasing. *The Hearing Journal*, 63(1), 19–32.

doi:10.1097/01.HJ.0000366912.40173.76

Larsby, B., Hällgren, M., and Lyxell, B. (2008). The interference of different background noises on speech processing in elderly hearing-impaired subjects. *International Journal of Audiology*, 47(2), S83–90. doi: 10.1080/14992020802301159

Lin, F. R., Ferrucci, L., Metter, E. J., An, Y., Zonderman, A. B., and Resnick, S. M. (2011). Hearing loss and cognition in the Baltimore Longitudinal Study of Aging. *Neuropsychology*, 25(6), 763–770. doi:10.1037/a0024238

Lin, F. R., Metter, E. J., O'Brien, R. J., Resnick, S. M., Zonderman, A. B., and Ferrucci, L. (2011). Hearing loss and incident dementia. *Archives of Neurology*, 68(2), 214–220. doi:10.1001/archneurol.2010.362

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., and Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group ‘white paper’. *International Journal of Audiology, Early Online*, 1–13.
doi:10.3109/14992027.2014.890296

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The psychological review*, 63, 81–97.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society*, 53(4), 695–699. doi:10.1111/j.1532-5415.2005.53221.x

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., and Krueger, S. (2005). Divided attention

- in younger and older adults: Effects of strategy and relatedness on memory performance and secondary task costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 520–537. doi:10.1037/0278-7393.31.3.520
- Newman, C., Weinstein, B., Jacobson, G., and Hug, G. (1990). The Hearing Handicap Inventory for Adults: Psychometric adequacy and audiometric correlates. *Ear and Hearing*, *11*, 430–433. doi:10.1097/00003446-199012000-0004
- Nilsson, L. G. (2003). Memory function in normal aging. *Acta Neurologica Scandinavica*, *107*(S179), 7–13. doi:10.1034/j.1600-0404.107.s179.5.x
- Nyberg, L., Nilsson, L-G., Olofsson, U., and Bäckman, L. (1997). Effects of division of attention during encoding and retrieval on age differences in episodic memory. *Experimental Aging Research*, *23*(2), 137–143. doi:10.1080/03610739708254029
- O'Brien, A., Yeend, I., Hartley, L., Keidser, G., and Nyffeler, M. (2010). Evaluation of frequency compression and high-frequency directionality. *The Hearing Journal*, *63*(8), 32–37. doi:10.1097/01.HJ.0000387928.98098.66
- Palmier-Claus, J. (2011). The clinical uses of momentary assessment. *Acta Psychiatrica Scandinavica*, *124*(4), 241–241. doi:10.1111/j.1600-0447.2011.01761.x
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220-244. doi:10.1037/0033-2909.116.2.220
- Pearsons, K. S., Bennett, R., and Fidell, S. (1977). Speech Levels in Various Noise Environments. EPA-600/1-77-025. Washington, D.C.: United States Environmental Noise Protection Agency.
- PhonakPro, (2014). SoundRecover. Retrieved November 22, 2014 from:
<http://www.phonakpro.com/com/b2b/en/evidence/topics/soundrecover.html>

- Picou, E. M., Ricketts, T. A., and Hornsby, B. W. Y. (2011). Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research*, 54, 1416–1430. doi:10.1044/1092-4388(2011/10-0154)
- Polonenko, M. J., Scollie, S. D., Moodie, S., Seewald, R. C., Lurnagaray, D., Shantz, J., and Richards, A. (2010). Fit to targets, preferred listening levels, and self-reported outcomes for the DSL v5.0a hearing aid prescription for adults. *International Journal of Audiology, Early Online*, 1–11. doi:10.3109/14992021003713122
- Preminger, J. E., and Van Tasell, D. J. (1995a). Quantifying the relation between speech quality and speech intelligibility. *Journal of Speech and Hearing Research*, 38, 714–725.
- Preminger, J. E., and Van Tasell, D. J. (1995b). Measurement of speech quality as a tool to optimize the fitting of a hearing aid. *Journal of Speech and Hearing Research*, 38, 726–736.
- Preston, C.C., and Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15. doi:10.1016/S0001-6918(99)00050-5
- Sakamoto, S., Iwaoka, N., Suzuki, Y., Amano, S., and Kondo, T. (2004). Complementary relationship between familiarity and SNR in word intelligibility test. *Acoustical Science and Technology*, 25(4), 290–292. doi:10.1250/ast.25.290
- Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, 52, 1230–1240

- Saunders, G. H., and Cienkowski, K. M. (2002). A test to measure subjective and objective speech intelligibility. *Journal of the American Academy of Audiology*, *13*, 38–49.
- Schulte, M., Vormann, M., Wagener, K., Büchler, M., Dillier, N., Dreschler, W., . . . Wouters, J. (2009). Proceedings from Workshop Hearing Screening and Technology: *Listening effort scaling and preference rating for hearing aid evaluation*. Brussels. Retrieved on December 2, from:
http://hearcom.eu/lenya/hearcom/authoring/about/DisseminationandExploitation/Workshop/S2B-3_Michael-Schulte_Hearing-Aid-Scaling-Rating.pdf
- Schum, D. J. (1999). Perceived hearing aid benefit in relation to perceived needs. *Journal of the American Academy of Audiology*, *10*, 40–45.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105. doi:10.1037/0003-066X.54.2.93
- Scollon, C. N., Kim-Prieto, C., and Diener, E. (2003). Experience sampling: promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, *4*, 5–34.
doi:10.1023/A:1023605205115
- Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32.
doi:10.1146/annurev.clinpsy.3.022806.091415
- Silver, N. C., Hittner, J. B., and May, K. (2004). Testing dependent correlations with nonoverlapping variables: A Monte Carlo simulation. *Journal of Experimental Education*, *73*, 53–69. doi:10.3200/JEXE.71.1.53-70
- Simpson, A. (2009). Frequency-lowering devices for managing high-frequency hearing

loss: A review. *Trends in Amplification*, 13(2), 87–106.

doi:10.1177/1084713809336421

- Smeds, K., Wolters, F., and Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, 26(2), 183–196(14). doi: 10.3766/jaaa.26.2.7
- Smith, A. (2013). *Smartphone ownership 2013*. PewResearch Internet Project. Retrieved May 2, 2014, from <http://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251.
- Souza, P. E., and Tremblay, K. L. (2006). New perspectives on assessing amplification effects. *Trends in Amplification*, 10(3), 119–143. doi:10.1177/1084713806292648
- Stone, A. A., Shiffman, S., Atienza, A. A., and Nebeling, L. (2007). *The science of real-time data capture: self-reports in health research*. Oxford: Oxford University Press.
- Tennen, H., Affleck, G., Coyne, J. C., Larsen, R. J., and DeLongis, A. (2006). Paper and plastic in daily diary research: comment on Green, Rafaeli, Bolger, Shrout, and Reis. *Psychological Methods*, 11, 112–118. doi:10.1037/1082-989X.11.1.112
- Thiele, C. Laireiter, A. R., and Baumann, U. (2002). Diaries in clinical psychology and psychotherapy: a selective review. *Clinical Psychology and Psychotherapy*, 9, 1–37. doi:10.1002/cpp.302
- Tomporowski, P. D. (2003). Performance and perceptions of workload among young and older adults: Effects of practice during cognitively demanding tasks. *Educational*

Gerontology, 29, 447–466. doi:10.1080/713844359

Valente, M., Abrams, H., Benson, D., Chisholm, T., Citron, D., Hampton, D., . . .

Sweetow, R. (2006). Guidelines for the audiologic management of adult hearing impairment. *AudiologyOnline*. Retrieved May 2, 2014, from <http://www.audiologyonline.com/articles/guideline-for-audiologic-management-adult-966>

Valente M., Fernandez, E., and Monroe B. (2010). *Audiology Answers for Otolaryngologists*. New York: Thieme Medical Publishers.

Valente, M., Hosford-Dunn, H., and Roeser, R. J. (2008). *Room Acoustics. Audiology: Treatment*. New York: Thieme Publishing Group.

Van Engen, K. J., and Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, 577. doi:10.3389/fnhum.2014.00577

Ventry, I. M., and Weinstein, B. E. (1982). The Hearing Handicap Inventory for the Elderly: A new tool. *Ear and Hearing*, 3(3), 40–46. doi:10.3342/kjorl-hns.2011.54.12.828

Weijters, B., Cabooter, E., and Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *Working Papers of Faculty of Economics and Business Administration*, Ghent University, Belgium.

Wickens, C. D., and Gosney, J. L. (2003). Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting: *Redundancy, modality, and priority in dual task interference*. 1590–1594. doi: 10.1177/154193120304701301

Winkler, I., and Cowan, N. (2005). From sensory to long-term memory. *Experimental*

Psychology, 52(1), 3–20. doi:10.1027/1618-3169.52.1.3

Yanz, J. (2006). The future of wireless devices in hearing care: A technology that promises to transform the hearing industry. *The Hearing Review*, 13(1), 18–20.

Yarmey, A. D., and Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology*, 6, 367–377.

Zimmer, K., Ghani, J., and Ellermeier, W. (2008). The role of task interference and exposure duration in judging noise annoyance. *Journal of Sound and Vibration*, 311(3), 1039–1051. doi:10.1016/j.jsv.2007.10.002

Appendices

Appendix A Participant audiograms listed by code

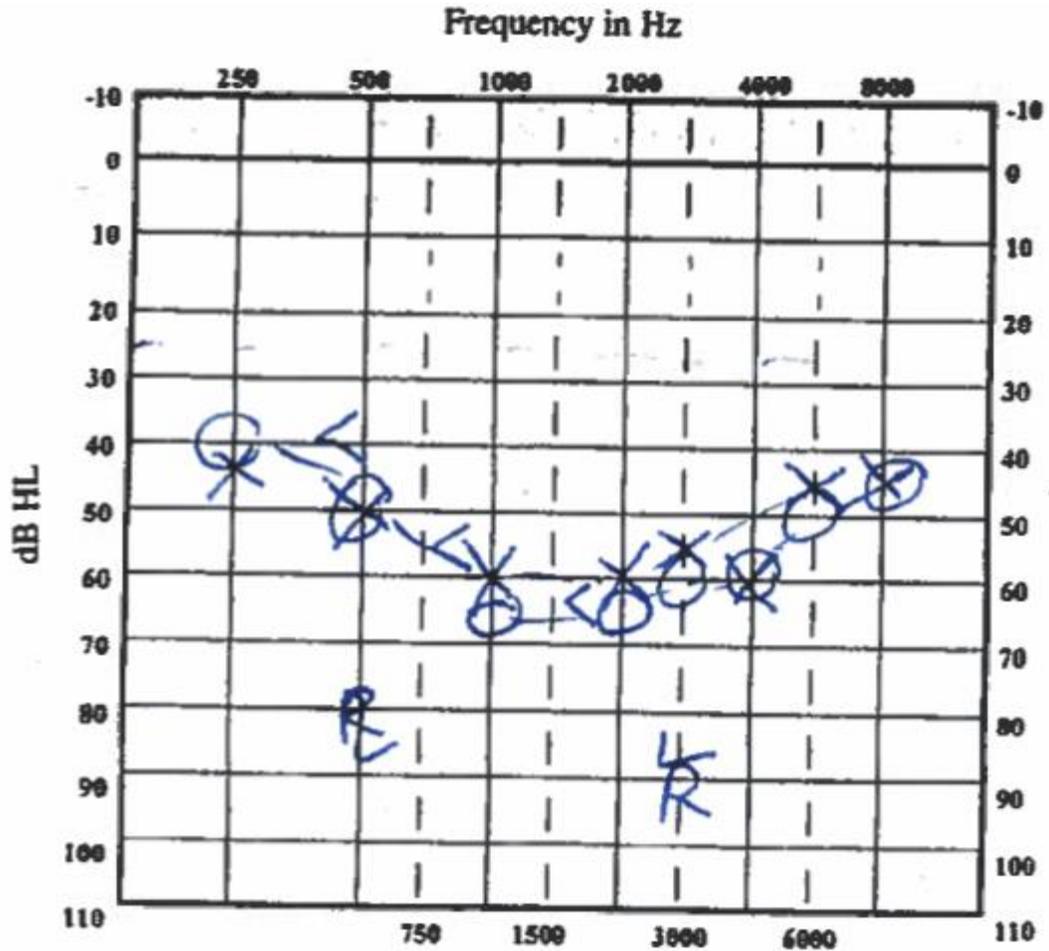


Figure A.1. Participant 109's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

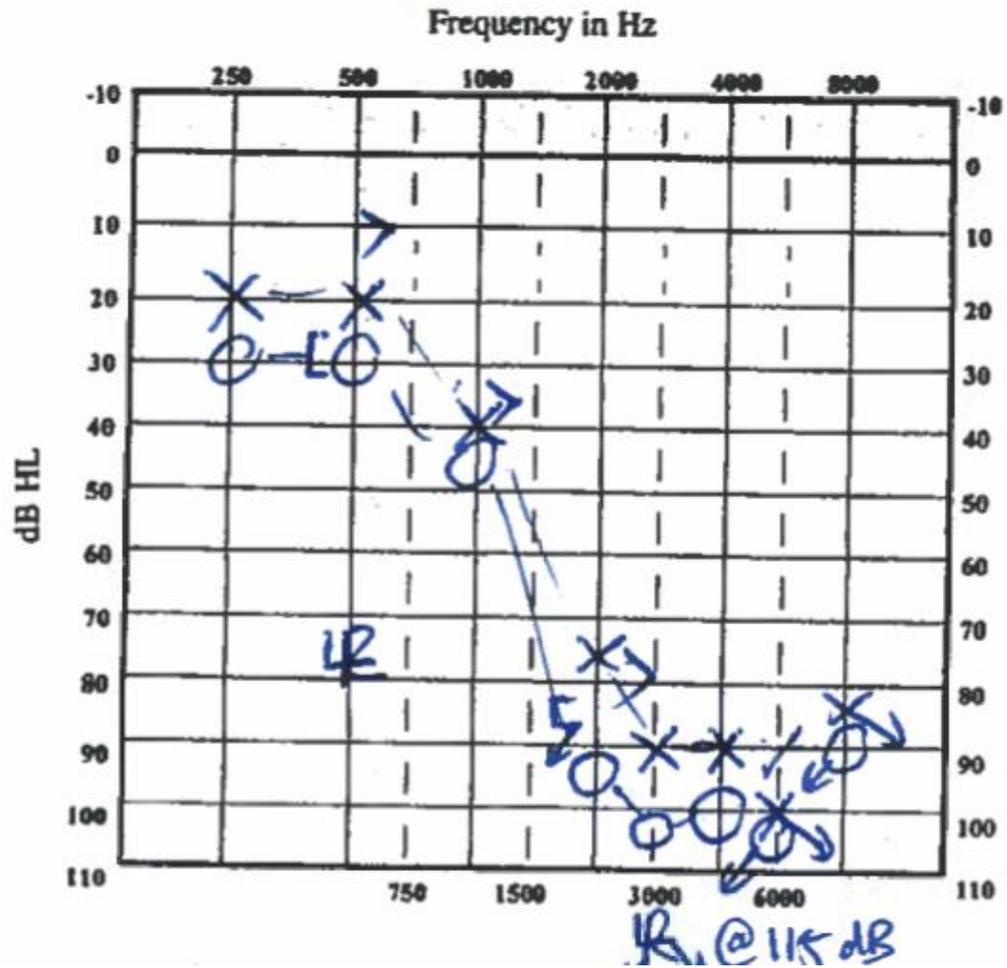


Figure A.2. Participant 155's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

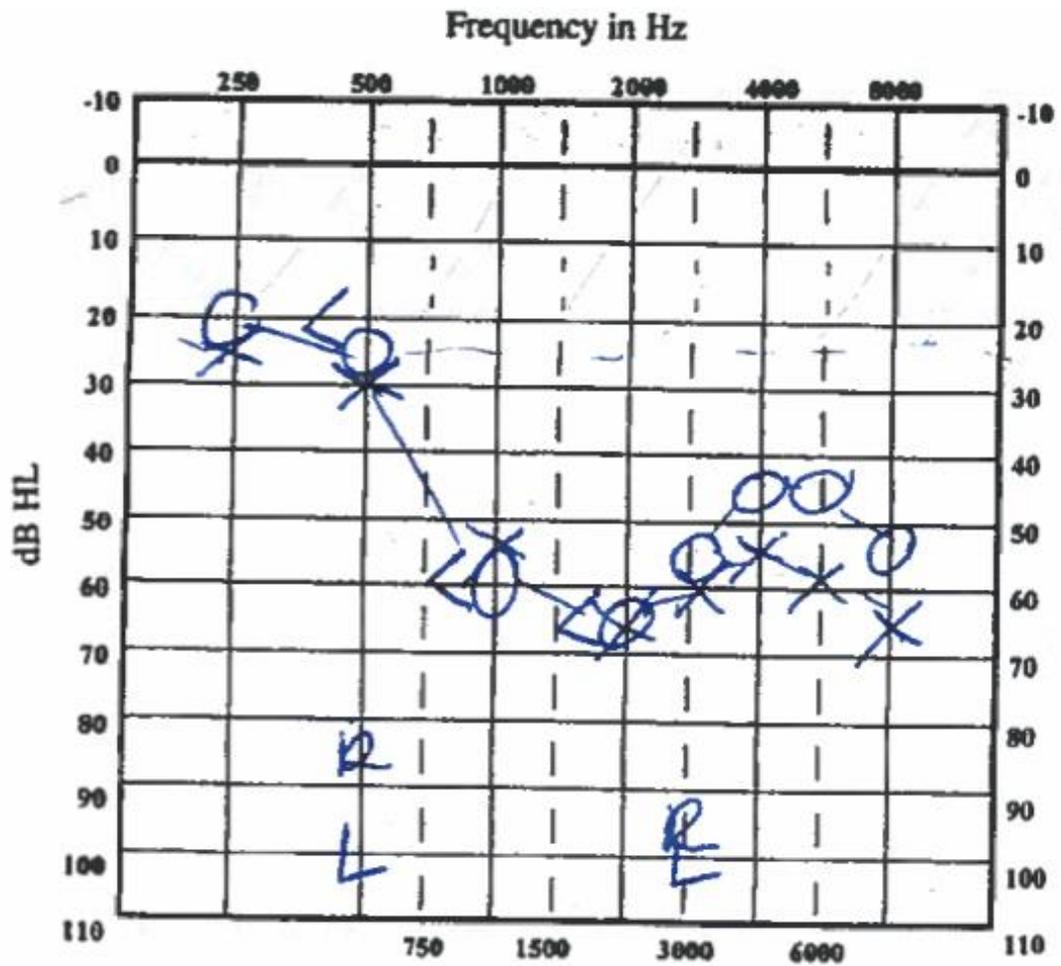


Figure A.3. Participant 164's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

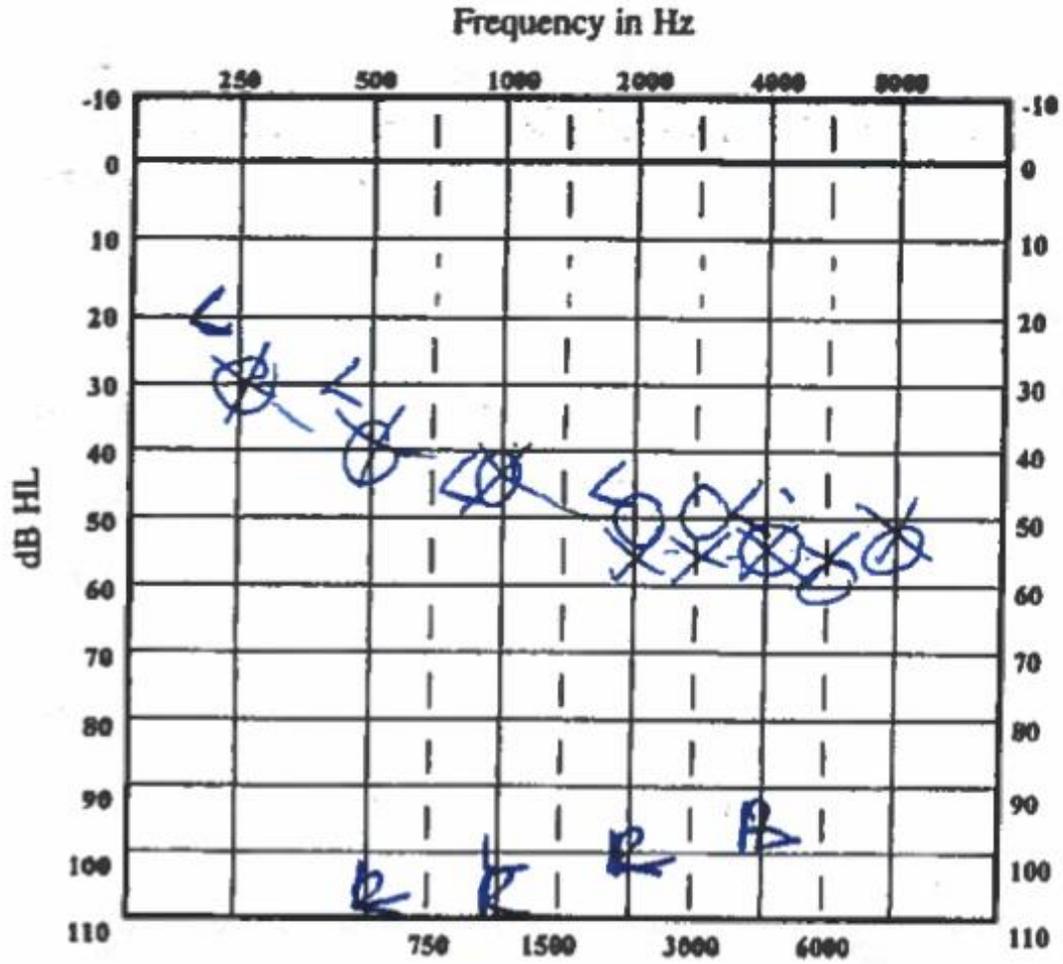


Figure A.4. Participant 171's audiogram. X's and O's, □'s and △'s, >'s and <'s,]'s and '['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

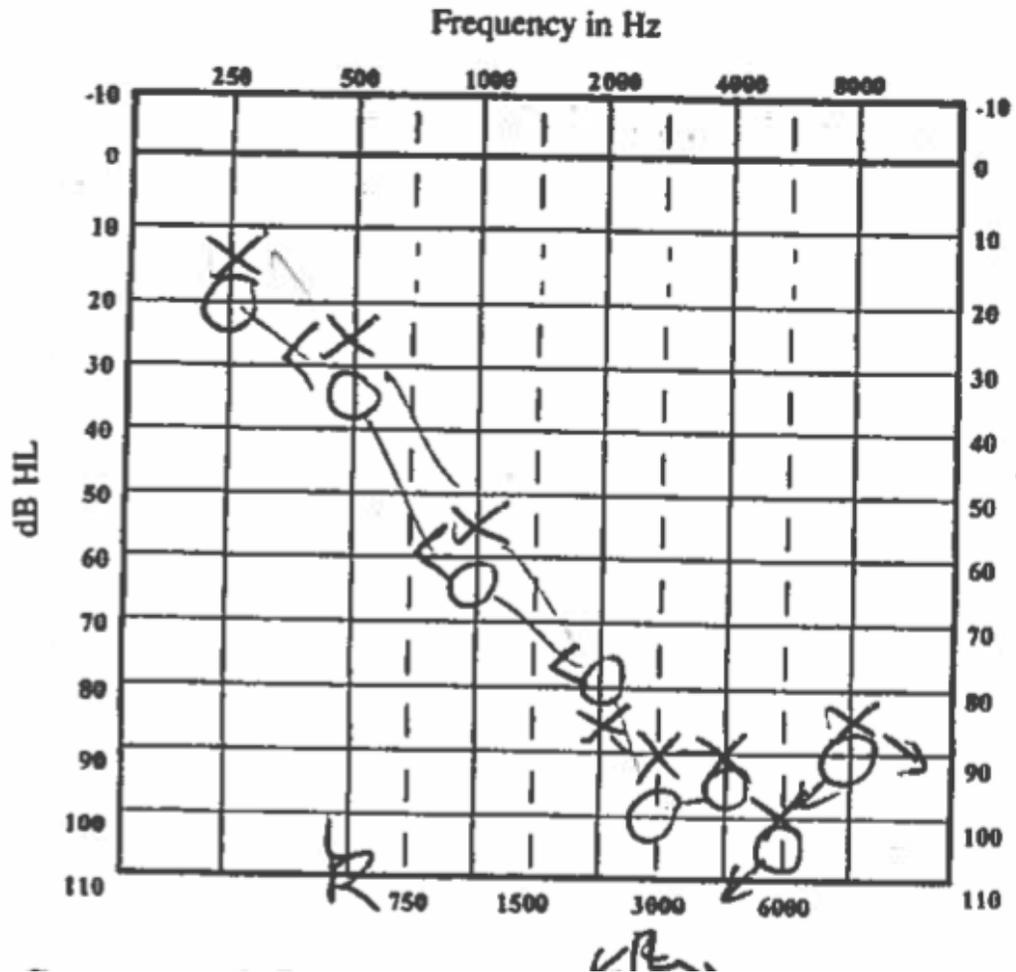


Figure A.5. Participant 223's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

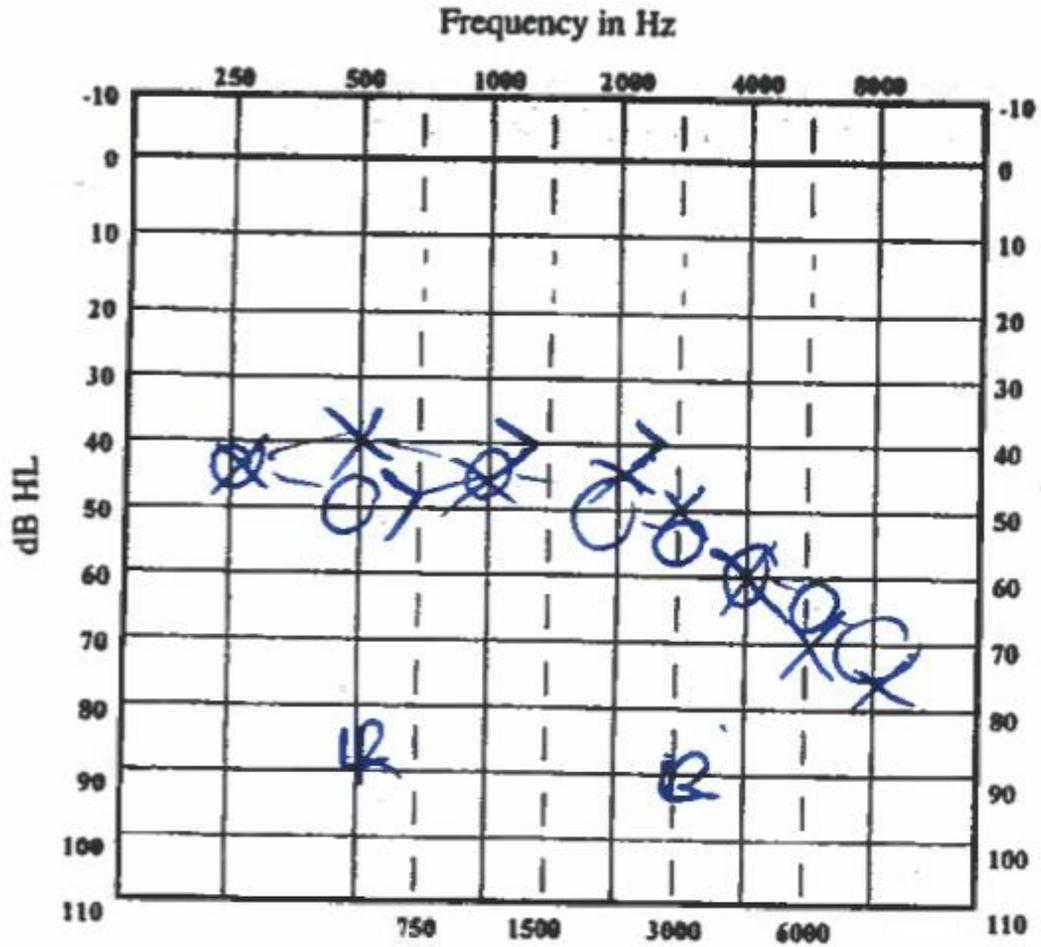


Figure A.6. Participant 314's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

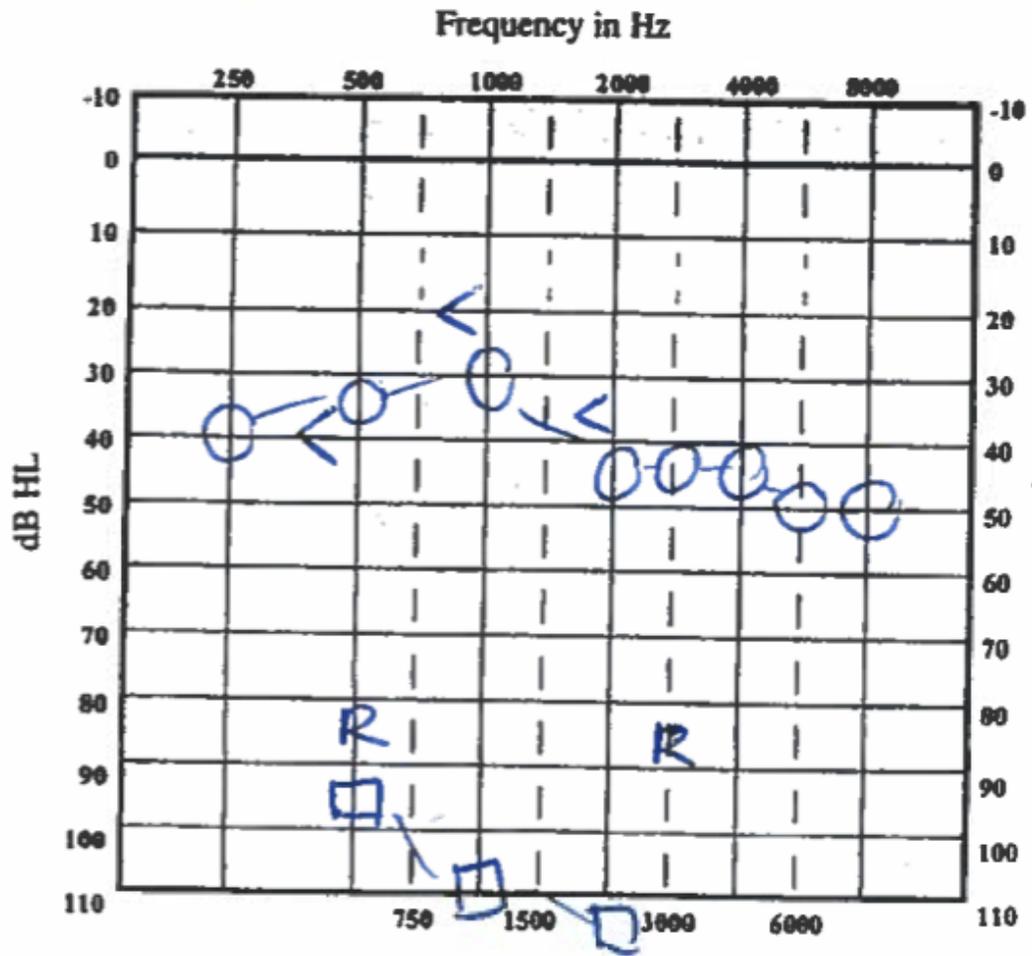


Figure A.7. Participant 449's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and '['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

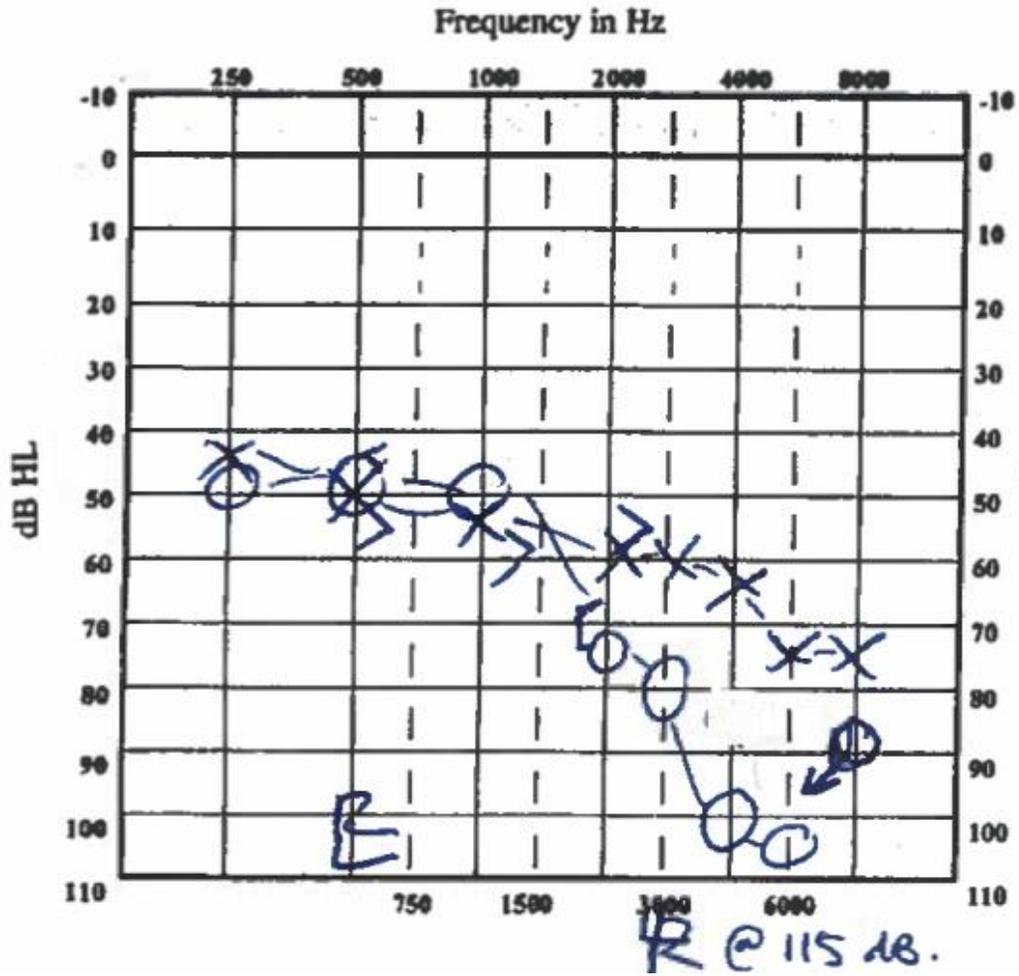


Figure A.8. Participant 505's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

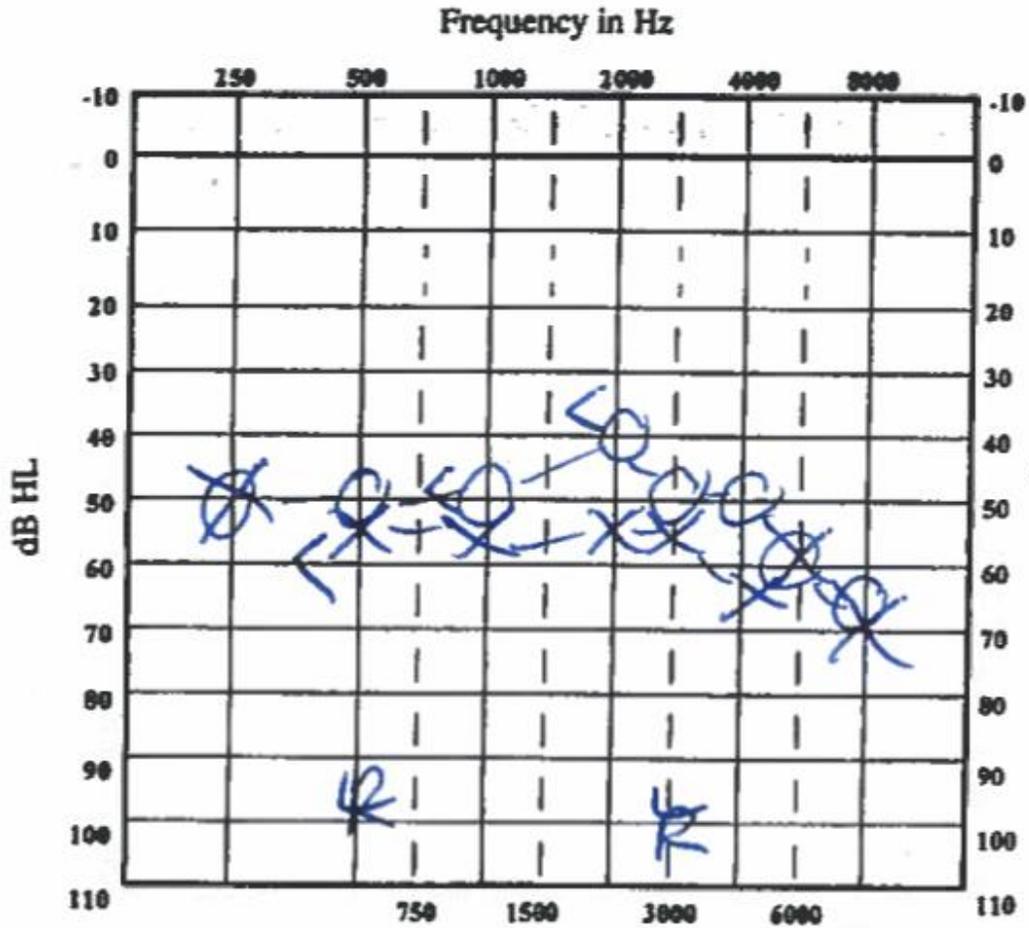


Figure A.9. Participant 687's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

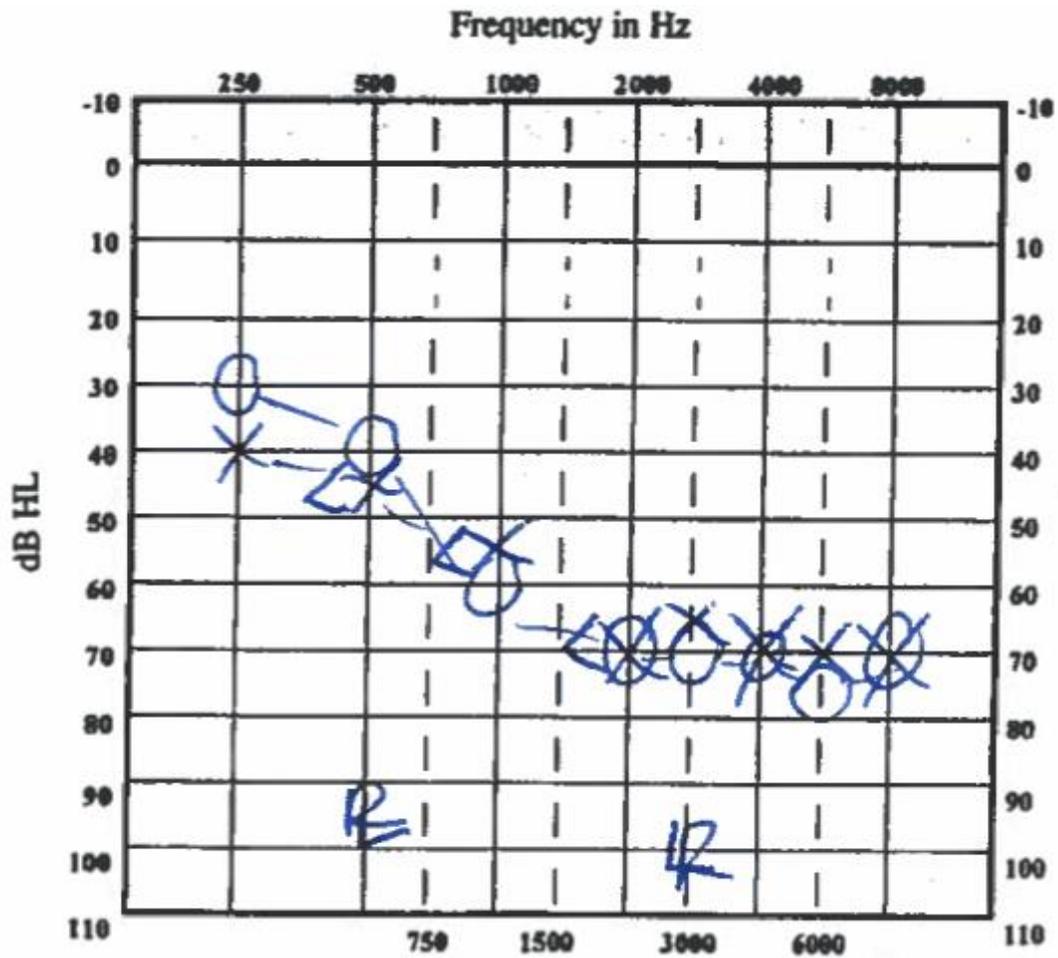


Figure A.10. Participant 733's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

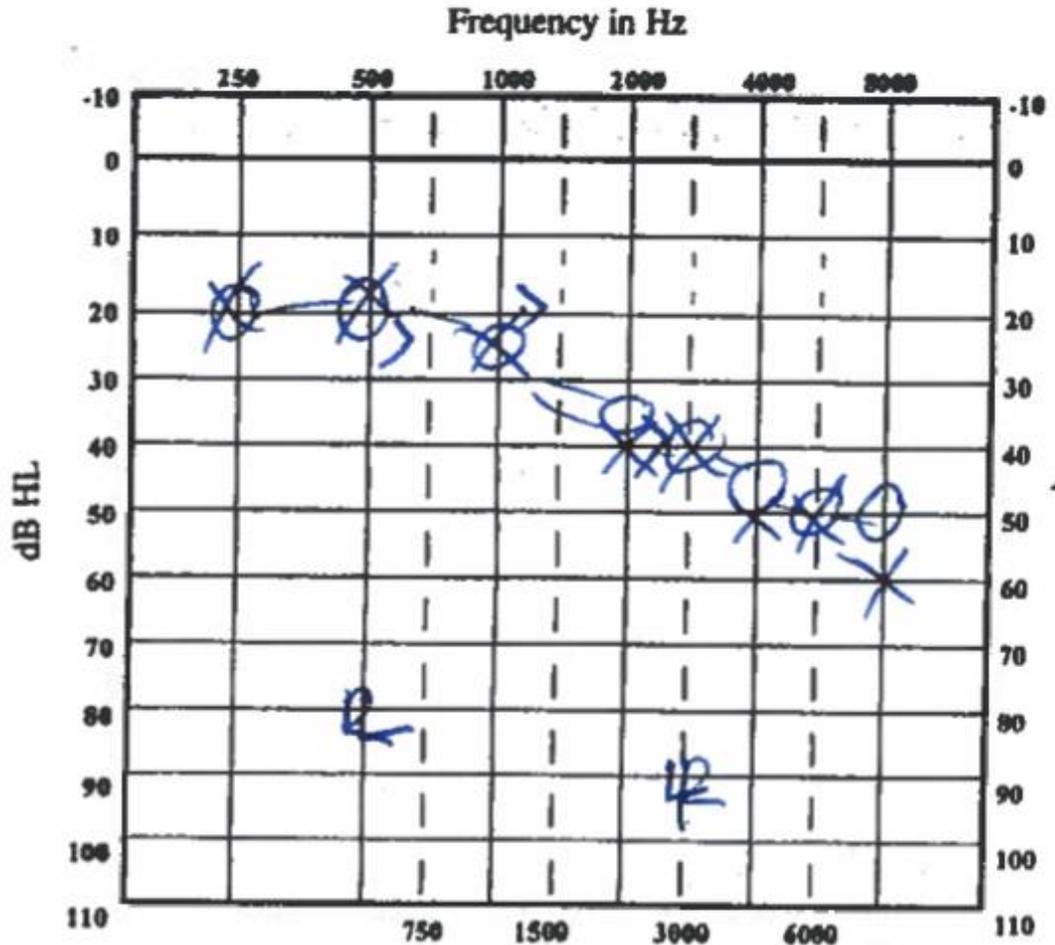


Figure A.11. Participant 881's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

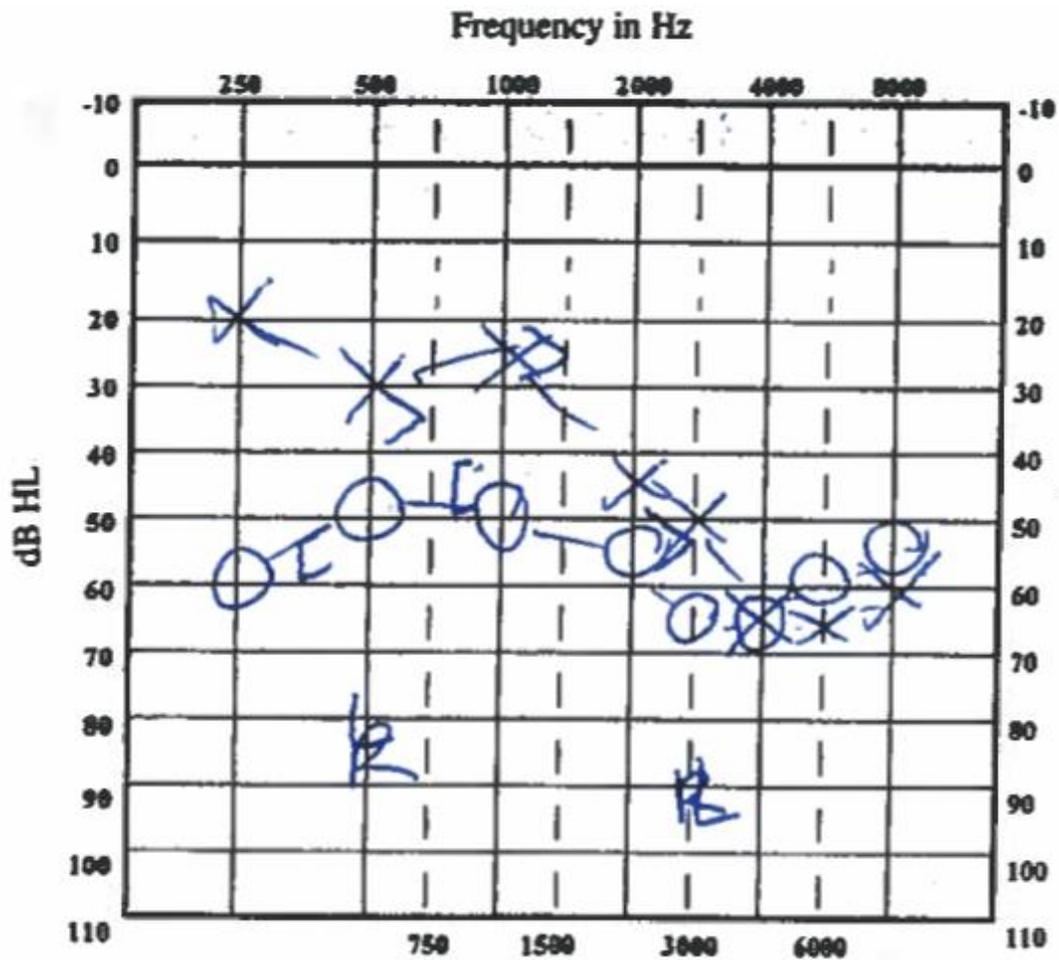


Figure A.12. Participant 906's audiogram. X's and O's, □'s and Δ's, >'s and <'s,]'s and ['s, and L's and R's represent left and right unmasked and masked air conduction thresholds, unmasked and masked bone conduction thresholds, and loudness discomfort levels respectively.

Appendix B Hearing thresholds across participants

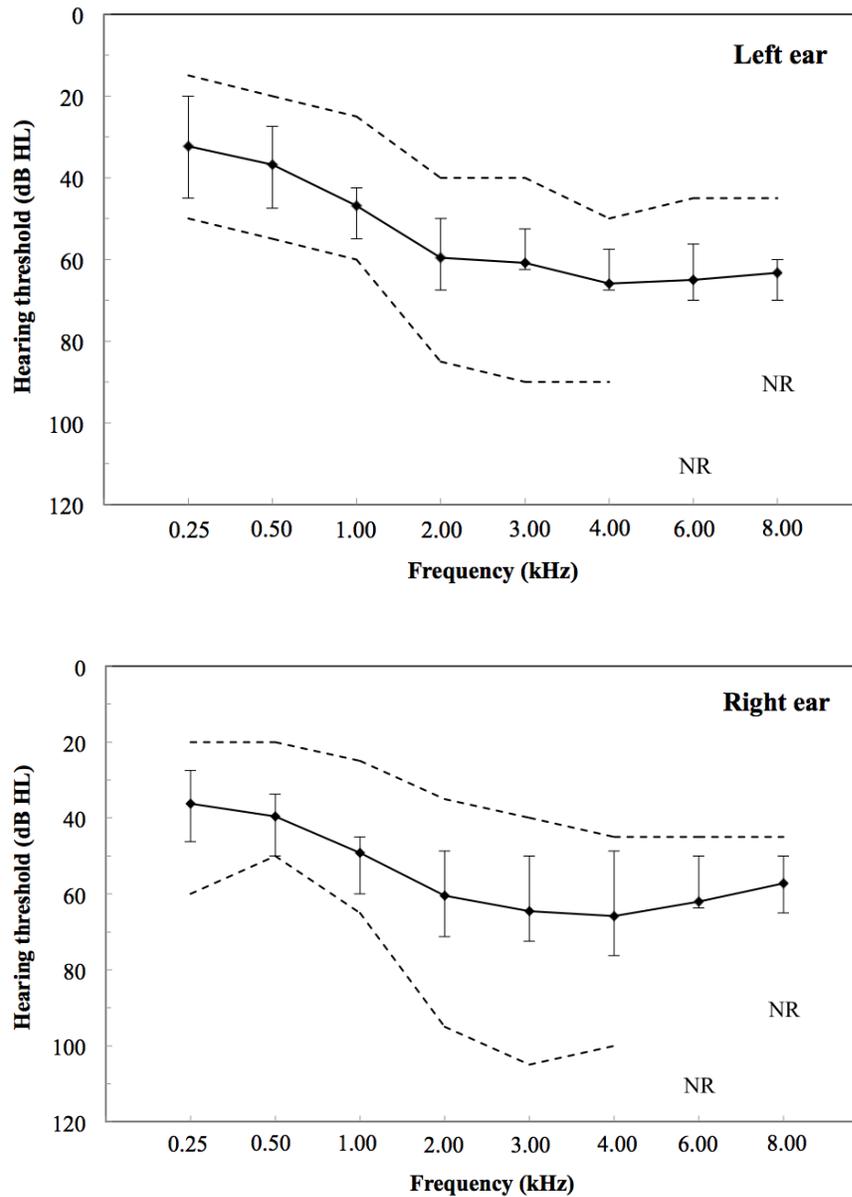


Figure B.1. Mean hearing thresholds for the left and right ear across participants. Data points indicate mean thresholds, dotted lines indicate minimum and maximum thresholds, and error bars indicate the 25th, and 75th percentile pure-tone air conduction thresholds in dB HL (re: American National Standard Institute, 2010). NR = no response at the upper limit of the audiometer.

Appendix C CMA rating instructions

Intelligibility

This is the percentage of spoken words that you can understand. When you make your ratings, think about your ability to understand each word. (Don't make your judgments based on your ability to understand each speech sound, or each sentence.) If you can understand all of the words assign a rating of "All words understood (10)", and if you cannot understand any of the words assign a rating of "No words understood (0)". If you can understand a portion of the words assign a number that corresponds to that portion. Remember, base your rating only on how intelligible the spoken words sound to you; your rating should not be influenced by any other aspects of the speech quality, including loudness.

Noisiness

For this judgment, consider how much noise you hear while listening to the speech. Noise is anything that you hear that is not the speech you are trying to attend. Assign a rating of "No noise (0)" if you do not hear any noise or if the noise is so quiet that it is at the quietest level that you can hear. Assign a rating of "Uncomfortably loud noise (10)" if the noise is so loud that it would be uncomfortable to listen to for more than 5 minutes. Use the numbers between 0 and 10 for noisiness levels between these two end-points.

Listening effort

For this judgment, estimate the amount of effort you need to give to this listening task in order to understand as much of the speech as you possibly can. If you do not have to use any effort in order to understand the speech, assign a rating of "No effort (0)". (You would not need any effort to understand the speech if you were able to complete a difficult task, such as figuring out a math problem, at the same time as you listened to the speech.) If you need to use all your concentration or effort in order to understand the speech, assign a rating of "Maximum effort (10)". For conditions between 0 and 10, assign a rating based on the proportion of your effort required to understand the speech. Remember, base your rating only on how much listening effort is required. Your rating should not be influenced by any other aspects of the speech quality, such as intelligibility.

Loudness

This rating should indicate how loud the speech seems to you. Assign a rating of "Uncomfortably loud (10)" if the speech is so loud that it would be uncomfortable to listen to for more than 5 minutes. Assign a rating of "Very quiet (0)" if the speech is so quiet that it is at the quietest level that you can hear. Use the numbers between 0 and 10 for loudness levels between these two endpoints.

Appendix D Hearing aid fitting distance from NAL-NL1 target gain

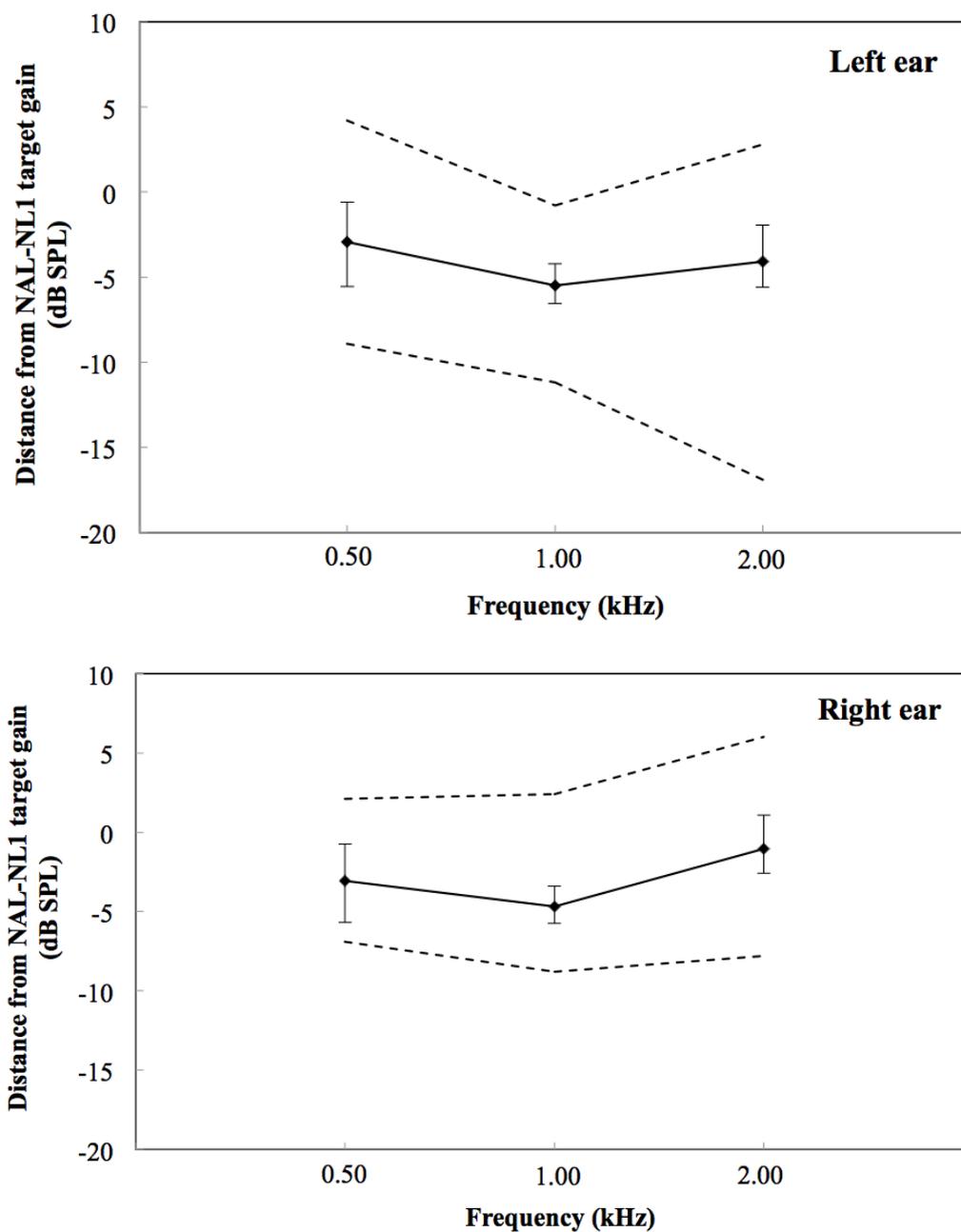


Figure D.1. Mean distance of hearing aid fittings from NAL-NL1 gain targets across participants. Solid lines indicate mean distance from NAL-NL1 targets, dotted lines indicate minimum and maximum distances from targets, and error bars indicate the 25th, and 75th percentile distances from targets in dB SPL.

Appendix E Screenshots of a sample connect-the-dots trial

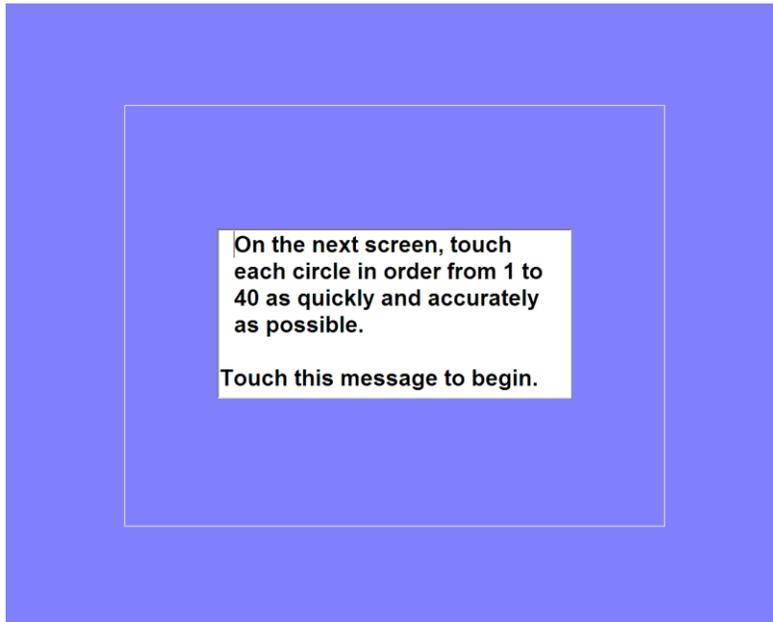


Figure E.1. Connect-the-dots task instructions screen.

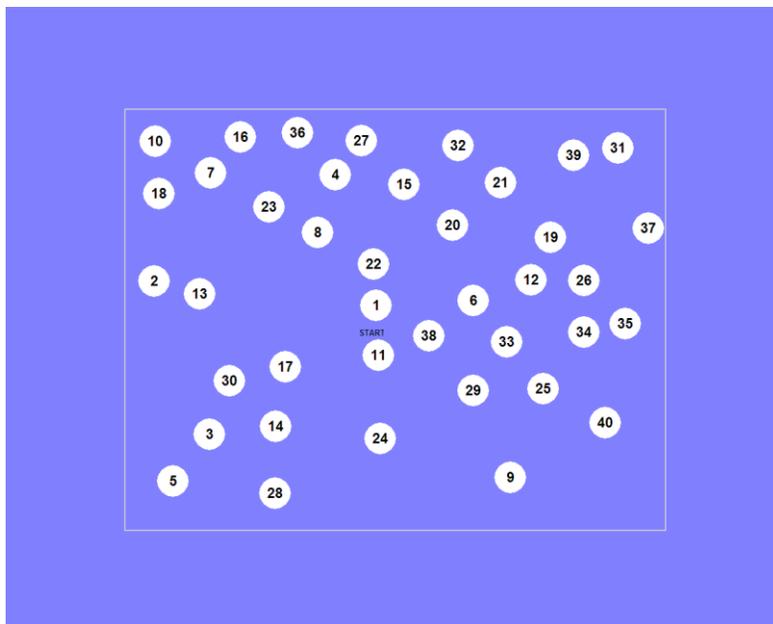


Figure E.2. Connect-the-dots task testing screen.

Appendix F Questionnaire package



QUESTIONNAIRE PACKAGE

Contextual Momentary Assessments by Wearers of Hearing Aids

INSTRUCTIONS: This is a brief questionnaire with four sections (A, B, C and D). Please complete every question in Sections A, B and C that are applicable to you: all of these questionnaires are appropriate for any level of hearing ability. If you currently wear hearing aids, please also complete Section D.



PART A: DEMOGRAPHICS (1 of 1)

Contextual Momentary Assessments by Wearers of Hearing Aids

For the following items, please select one response that is most descriptive of you or fill in the blank as appropriate.

1. Age: _____

2. Gender:

- Female
- Male
- Other: _____
- Prefer not to disclose

3. Which of the following best describes your current living arrangement? Please answer both A and B.

A) Ownership:

- I own the space in which I live
- I rent the space in which I live
- Other (please specify): _____

B) Support:

- I live alone
- I live with only my spouse or common law partner
- I live with family (children, parents, siblings) in the same house
- I live in a retirement community for independent individuals
- I live in an assisted living environment
- Other (please specify): _____

4. Which of the following best describes the highest level of education you have completed?

- Elementary school or less
- Some high school
- High school (secondary school) diploma or equivalent
- Some college/trade school
- College/trade school diploma
- Attended University
- University degree
- Post-graduate degree
- Other (please specify): _____

5. What is your employment status?

- Full Time
- Part Time
- Unemployed
- Retired
- Looking for work
- Other (please specify): _____

6. Please indicate whether your current income meets your needs for daily living, with 1 being less than adequate to meet your needs and 7 being more than adequate to meet your needs.

- Less than adequate More than adequate
- 1 2 3 4 5 6 7

5. What generation Canadian are you?

- New immigrant
- First generation; my parents moved to Canada
- Second generation; my grandparents came to Canada
- More than second generation Canadian



PART B: HEARING STATUS (1 of 2)

Contextual Momentary Assessments by Wearers of Hearing Aids

The purpose of this scale is to identify the problems your hearing loss may be causing you. Circle 'Yes', 'Sometimes', or 'No' for each question. Do not skip any questions. If you use a hearing aid, please answer with the way you hear without a hearing aid in mind.

S-1.	Does a hearing problem cause you to use the phone less often than you would like?	Yes	Sometimes	No
E-2.	Does a hearing problem cause you to feel embarrassed when meeting new people?	Yes	Sometimes	No
S-3.	Does a hearing problem cause you to avoid groups of people?	Yes	Sometimes	No
E-4.	Does a hearing problem make you irritable?	Yes	Sometimes	No
E-5.	Does a hearing problem cause you to feel frustrated when talking to members of your family?	Yes	Sometimes	No
S-6.	Does a hearing problem cause you difficulty when attending a party?	Yes	Sometimes	No
E-7.	Does a hearing problem cause you to feel "stupid" or "dumb"?	Yes	Sometimes	No
S-8.	Do you have difficulty hearing when someone speaks in a whisper?	Yes	Sometimes	No
E-9.	Do you feel handicapped by a hearing problem?	Yes	Sometimes	No
S-10.	Does a hearing problem cause you difficulty when visiting friends, relatives, or neighbors?	Yes	Sometimes	No
S-11.	Does a hearing problem cause you to attend religious services less often than you would like?	Yes	Sometimes	No
E-12.	Does a hearing problem cause you to be nervous?	Yes	Sometimes	No

S-13.	Does a hearing problem cause you to visit friends, relatives, or neighbors less often than you would like?	Yes	Sometimes	No
E-14.	Does a hearing problem cause you to have arguments with family members?	Yes	Sometimes	No
S-15.	Does a hearing problem cause you difficulty when listening to TV or radio?	Yes	Sometimes	No
S-16.	Does a hearing problem cause you to go shopping less often than you would like?	Yes	Sometimes	No
E-17.	Does any problem or difficulty with your hearing upset you at all?	Yes	Sometimes	No
E-18.	Does a hearing problem cause you to want to be by yourself?	Yes	Sometimes	No
S-19.	Does a hearing problem cause you to talk to family members less often than you would like?	Yes	Sometimes	No
E-20.	Do you feel that any difficulty with your hearing limits or hampers your personal or social life?	Yes	Sometimes	No
S-21.	Does a hearing problem cause you difficulty when in a restaurant with relatives or friends?	Yes	Sometimes	No
E-22.	Does a hearing problem cause you to feel depressed?	Yes	Sometimes	No
S-23.	Does a hearing problem cause you to listen to TV or radio less often than you would like?	Yes	Sometimes	No
E-24.	Does a hearing problem cause you to feel uncomfortable when talking to friends?	Yes	Sometimes	No
E-25.	Does a hearing problem cause you to feel left out when you are with a group of people?	Yes	Sometimes	No

Source: Ventry, I.M., and Weinstein, B.E. (1982). The Hearing Handicap Inventory for the Elderly: A new tool. *Ear and Hearing*, 3, 40-46.



PART B: HEARING STATUS (2 of 2)

Contextual Momentary Assessments by Wearers of Hearing Aids

Please circle the answers that come closest to your everyday experience. Notice that each choice includes a percentage. You can use this to help you decide on your answer. For example, if a statement is true about 75% of the time, circle C for that item. If you have not experienced the situation we describe, try to think of a similar situation that you have been in and respond for that situation. If you have no idea, leave that item blank.

*If you do not use hearing aids, only circle answers under "Without My Hearing Aid"

- A Always (99%)
- B Almost Always (87%)
- C Generally (75%)
- D Half-the-time (50%)
- E Occasionally (25%)
- F Seldom (12%)
- G Never (1 %)

	Without My Hearing Aid	With My Hearing Aid
1. When I am in a crowded grocery store, talking with the cashier, I can follow the conversation.	A B C D E F G	A B C D E F G
2. I miss a lot of information when I'm listening to a lecture.	A B C D E F G	A B C D E F G
3. Unexpected sounds, like a smoke detector or alarm bell are uncomfortable.	A B C D E F G	A B C D E F G
4. I have difficulty hearing a conversation when I'm with one of my family at home.	A B C D E F G	A B C D E F G
5. I have trouble understanding dialogue in a movie or at the theater.	A B C D E F G	A B C D E F G

	Without My Hearing Aid	With My Hearing Aid
6. When I am listening to the news on the car radio, and family members are talking, I have trouble hearing the news.	A B C D E F G	A B C D E F G
7. When I am at the dinner table with several people, and am trying to have a conversation with one person, understanding speech is difficult.	A B C D E F G	A B C D E F G
8. Traffic noises are too loud.	A B C D E F G	A B C D E F G
9. When I am talking with someone across a large empty room, I understand the words.	A B C D E F G	A B C D E F G
10. When I am in a small office, interviewing or answering questions, I have difficulty following the conversation.	A B C D E F G	A B C D E F G
11. When I am in a theater watching a movie or play, and the people around me are whispering and rustling paper wrappers, I can still make out the dialogue.	A B C D E F G	A B C D E F G
12. When I am having a quiet conversation with a friend, I have difficulty understanding.	A B C D E F G	A B C D E F G
13. The sounds of running water, such as a toilet or shower, are uncomfortably loud.	A B C D E F G	A B C D E F G
14. When a speaker is addressing a small group, and everyone is listening quietly, I have to strain to understand.	A B C D E F G	A B C D E F G

	Without My Hearing Aid	With My Hearing Aid
15. When I'm in a quiet conversation with my doctor in an examination room, it is hard to follow the conversation.	A B C D E F G	A B C D E F G
16. I can understand conversations even when several people are talking.	A B C D E F G	A B C D E F G
17. The sounds of construction work are uncomfortably loud.	A B C D E F G	A B C D E F G
18. It's hard for me to understand what is being said at lectures or church services.	A B C D E F G	A B C D E F G
19. I can communicate with others when we are in a crowd.	A B C D E F G	A B C D E F G
20. The sound of a fire engine siren close by is so loud that I need to cover my ears.	A B C D E F G	A B C D E F G
21. I can follow the words of a sermon when listening to a religious service.	A B C D E F G	A B C D E F G
22. The sound of screeching tires is uncomfortably loud.	A B C D E F G	A B C D E F G
23. I have to ask people to repeat themselves in one on one conversation in a quiet room.	A B C D E F G	A B C D E F G
24. I have trouble understanding others when an air conditioner or fan is on.	A B C D E F G	A B C D E F G

Source: Cox, R. & Alexander, C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing*, 16(2), 176-186.



PART C: PERSONALITY (1 of 1)

Contextual Momentary Assessments by Wearers of Hearing Aids

How I am in general

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which **you agree or disagree with that statement.**

1	2	3	4	5
Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly

I am someone who...

- | | |
|--|--|
| 1. ___ Is talkative | 23. ___ Tends to be lazy |
| 2. ___ Tends to find fault with others | 24. ___ Is emotionally stable, not easily upset |
| 3. ___ Does a thorough job | 25. ___ Is inventive |
| 4. ___ Is depressed, blue | 26. ___ Has an assertive personality |
| 5. ___ Is original, comes up with new ideas | 27. ___ Can be cold and aloof |
| 6. ___ Is reserved | 28. ___ Perseveres until the task is finished |
| 7. ___ Is helpful and unselfish with others | 29. ___ Can be moody |
| 8. ___ Can be somewhat careless | 30. ___ Values artistic, aesthetic experiences |
| 9. ___ Is relaxed, handles stress well. | 31. ___ Is sometimes shy, inhibited |
| 10. ___ Is curious about many different things | 32. ___ Is considerate and kind to almost everyone |
| 11. ___ Is full of energy | 33. ___ Does things efficiently |

1	2	3	4	5
Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly

- | | | | |
|---------|-------------------------------|---------|---|
| 12. ___ | Starts quarrels with others | 34. ___ | Remains calm in tense situations |
| 13. ___ | Is a reliable worker | 35. ___ | Prefers work that is routine |
| 14. ___ | Can be tense | 36. ___ | Is outgoing, sociable |
| 15. ___ | Is ingenious, a deep thinker | 37. ___ | Is sometimes rude to others |
| 16. ___ | Generates a lot of enthusiasm | 38. ___ | Makes plans and follows through with them |
| 17. ___ | Has a forgiving nature | 39. ___ | Gets nervous easily |
| 18. ___ | Tends to be disorganized | 40. ___ | Likes to reflect, play with ideas |
| 19. ___ | Worries a lot | 41. ___ | Has few artistic interests |
| 20. ___ | Has an active imagination | 42. ___ | Likes to cooperate with others |
| 21. ___ | Tends to be quiet | 43. ___ | Is easily distracted |
| 22. ___ | Is generally trusting | 44. ___ | Is sophisticated in art, music, or literature |



PART D: HEARING AID SATISFACTION (1 of 2)
Contextual Momentary Assessments by Wearers of Hearing Aids

Listed below are questions on your opinions about your hearing aid(s). For each question, please circle the letter that is the best answer for you. The list of words on the right gives the meaning for each letter.

Keep in mind that your answers should show your general opinions about the hearing aids that you are wearing now or have most recently worn.

- A Not At All
- B A Little
- C Somewhat
- D Medium
- E Considerably
- F Greatly
- G Tremendously

1. Compared to using no hearing aid at all, do your hearing aids help you understand the people you speak with most frequently?	A B C D E F G
2. Are you frustrated when your hearing aids pick up sounds that keep you from hearing what you want to hear?	A B C D E F G
3. Are you convinced that obtaining your hearing aids was in your best interests?	A B C D E F G
4. Do you think people notice your hearing loss more when you wear your hearing aids?	A B C D E F G
5. Do your hearing aids reduce the number of times you have to ask people to repeat?	A B C D E F G
6. Do you think your hearing aids are worth the trouble?	A B C D E F G
7. Are you bothered by an inability to get enough loudness from your hearing aids without feedback (whistling)?	A B C D E F G
8. How content are you with the appearance of your hearing aids?	A B C D E F G

9.	Does wearing your hearing aids improve your self-confidence?	A B C D E F G
10.	How natural is the sound from your hearing aids?	A B C D E F G
11.	How helpful are your hearing aids on MOST telephones with NO amplifier or loudspeaker? (If you hear well on the telephone without hearing aids, check here <input type="checkbox"/>)	A B C D E F G
12.	How competent was the person who provided you with your hearing aids?	A B C D E F G
13.	Do you think wearing your hearing aids makes you seem less capable?	A B C D E F G
14.	Does the cost of your hearing aids seem reasonable to you?	A B C D E F G
15.	How pleased are you with the dependability (how often they need repairs) of your hearing aids?	A B C D E F G

Please respond to these additional items.

EXPERIENCE WITH CURRENT HEARING AIDS	LIFETIME HEARING AID EXPERIENCE (includes all old and current hearing aids)	DAILY HEARING AID USE	DEGREE OF HEARING DIFFICULTY
<input type="checkbox"/> Less than 6 weeks	<input type="checkbox"/> Less than 6 weeks	<input type="checkbox"/> None	<input type="checkbox"/> None
<input type="checkbox"/> 6 weeks to 11 months	<input type="checkbox"/> 6 weeks to 11 months	<input type="checkbox"/> Less than 1 hour per day	<input type="checkbox"/> Mild
<input type="checkbox"/> 1 to 10 years	<input type="checkbox"/> 1 to 10 years	<input type="checkbox"/> 1 to 4 hours per day	<input type="checkbox"/> Moderate
<input type="checkbox"/> Over 10 years	<input type="checkbox"/> Over 10 years	<input type="checkbox"/> 4 to 8 hours per day	<input type="checkbox"/> Severe
		<input type="checkbox"/> 8 to 16 hours per day	

Source: Cox, R.M., and Alexander, G.C. (1999). Measuring satisfaction with amplification in daily life: The SADL scale. *Ear and Hearing*, 20, 306-320.



PART D: HEARING AID SATISFACTION (2 of 2)

Contextual Momentary Assessments by Wearers of Hearing Aids

1. Think about how much you used your present hearing aid(s) over the past two weeks. On an average day, how many hours did you use the hearing aid(s)?

none	less than 1 hour a day	1 to 4 hours a day	4 to 8 hours a day	more than 8 hours a day
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Think about the situation where you most wanted to hear better, before you got your present hearing aid(s). Over the past two weeks, how much has the hearing aid helped in those situations?

helped not at all	helped slightly	helped moderately	helped quite a lot	helped very much
<input type="checkbox"/>				

3. Think again about the situation where you most wanted to hear better. When you use your present hearing aid(s), how much difficulty do you STILL have in that situation?

very much difficulty	quite a lot of difficulty	moderate difficulty	slight difficulty	no difficulty
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. Considering everything, do you think your present hearing aid(s) is worth the trouble?

not at all worth it	slightly worth it	moderately worth it	quite a lot worth it	very much worth it
<input type="checkbox"/>				

5. Over the past two weeks, with your present hearing aid(s), how much have your hearing difficulties affected the things you can do?

affected very much	affected quite a lot	affected moderately	affected slightly	affected not at all
<input type="checkbox"/>				

6. Over the past two weeks, with your present hearing aid(s), how much do you think other people were bothered by your hearing difficulties?

bothered very much	bothered quite a lot	bothered moderately	bothered slightly	bothered not at all
<input type="checkbox"/>				

7. Considering everything, how much has your present hearing aid(s) changed your enjoyment of life?

worse	no change	slightly better	quite a lot better	very much better
<input type="checkbox"/>				

8. How much hearing difficulty do you have when you are **not** wearing a hearing aid?

severe	moderately- severe	moderate	mild	none
<input type="checkbox"/>				

Source: Cox, R.M., and Alexander, G.C. (2002). The International Outcome Inventory for Hearing Aids (IOI-HA): psychometric properties of the English version. *International Journal of Audiology*, 41(1), 30-35.

QUESTIONNAIRE PACKAGE COMPLETED

Appendix G Informed consent form



CONSENT FORM

Contextual Momentary Assessments by Wearers of Hearing Aids

Principal Investigator:

Lorienne Jenstad, Ph.D., Aud(C), Registered Audiologist and Hearing Instrument Practitioner
Associate Professor
School of Audiology and Speech Sciences
604-827-3338 (lab); amplab@audiospeech.ubc.ca (email)

Co-Investigators:

Anita DeLongis, Ph.D.
Department of Psychology
adelongis@psych.ubc.ca (email)

Lise Gillen, M.Sc. Candidate
School of Audiology and Speech Sciences
lise.gillen@alumni.ubc.ca (email)

Part of the study is for a Master's thesis.

We are asking you to be in a research study because you have indicated that you think you have hearing loss. The purpose of this consent form is to give you the information you will need to help you decide whether you want to be in the study. Please read the form carefully. You may ask questions about the research, your rights as a volunteer, and anything else about the study that is not clear. When all your questions have been answered, you can decide if you want to be in the study or not. This process is called "informed consent".

SPONSOR

This research is sponsored by a grant from MITACS and is conducted in partnership with Unitron Hearing Ltd.

PURPOSE

Hearing aids are normally fitted in a quiet clinical setting. These settings are not typical of everyday environments. We want to look at how ratings of speech and noise change between different settings. This study is the first of a two-part study to test the participant's ratings in a lab. All participants will be fitted with hearing aids for the study. The lab results will be compared against field results in later studies.

STUDY PROCEDURES

This study will take place at the UBC Amplification Lab. If you choose to be in this study, we would like you to come in for 2 – 3 study sessions.

At the first session, we will ask you general questions about your health. We will also perform a standard hearing test on you. Then, we will ask questions that test your memory and attention. For example, we will ask you to repeat numbers. You do not

have to answer every question. If you do not qualify for this study, this information will be destroyed. This session will last about 1 hour.

If you do qualify for the study, we would like you to return for additional sessions. Each session will last 1 – 2 hours. We will first fit you with hearing aids. You will also be provided with a mobile device. Then, we will ask you to sit comfortably in a small room. Recorded speech and noise will be played through a loudspeaker. We will ask you to listen to the speech or focus on completing a simple task (e.g., a Sudoku puzzle). We want to know how your ratings of the speech and noise change under different conditions. The mobile device will prompt you to answer short questionnaires during and after the testing blocks. We encourage you to take breaks. You can take a break at any time.

Most people take between 2 – 3 two-hour sessions to complete the study. Everyone works at a different pace. It is hard to predict exactly how long it will take you to complete the study.

STUDY RESULTS

Some of the study results will be reported in a graduate thesis. The results may also be published in journal articles and books.

POTENTIAL RISKS, STRESS, OR DISCOMFORT

There are no known physical risks for these study procedures. You may find the hearing aids uncomfortable. The researchers can adjust them for you. Concerns about your privacy are addressed later in this consent form.

POTENTIAL BENEFITS

You will not benefit directly from taking part in this study. The study results may help audiologists in optimizing hearing aids for its use in everyday environments in the future.

CONFIDENTIALITY

Information about you is confidential. We will code study records. The link between the code and your name will be kept at a secured location, separate from the study information. Only lab employees will have access to the link. All researchers will have been trained in privacy and confidentiality. We will keep the link between the study records and your name for five years from the participation date. We will destroy the link after five years. If we publish the study results, we will not use your name.

COMPENSATION

Upon completion of the study, you will be paid \$50. You will also receive reimbursement of parking or transit costs. We will either give you a cheque the day you finish the study or send you a cheque by mail in about 1 week.

OTHER INFORMATION

Being in this study is voluntary. You may decline to enter or withdraw from the study at any time without any consequences.

We may want to re-contact you about taking part in future related studies. Please indicate below whether or not you give your permission to re-contact you. Giving the research team permission to re-contact you does not obligate you in any way.

CONTACT INFORMATION ABOUT THE STUDY

If you have any questions or concerns about what we are asking of you, please contact the study leader or one of the study staff. The names and contact information are listed at the top of the first page of this form.

CONTACT FOR COMPLAINTS

If you have any concerns about your rights as a research subject and/or your experiences while participating in this study, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

Subject's statement

The study described above has been explained to me. I understand that my participation is voluntary. I may decline to enter or withdraw from the study at any time without any consequences. I have had an opportunity to ask questions. I understand that the investigators listed above will answer future questions I may have about the research. If I have concerns about my treatment or rights as a subject, I may call the UBC Office of Research Services at 604-822-8598.

- My signature below indicates that I have received a copy of this consent form for my own records.
- My signature indicates that I consent to participate in this study.

Please check one box below:

"I give my permission for the researchers to re-contact me for future related research."

Yes

No

Participant Signature

Date

Printed Name of the Participant Signing Above

Appendix I MoCA test form

NAME : _____
 Education : _____ Date of birth : _____
 Sex : _____ DATE : _____

VISUOSPATIAL / EXECUTIVE							POINTS
		Copy cube	Draw CLOCK (Ten past eleven) (3 points)				
[]	[]	[]	[]	[]	[]	[]	___/5
NAMING							
						[]	___/3
MEMORY		Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.					No points
		FACE	VELVET	CHURCH	DAISY	RED	
	1st trial						
	2nd trial						
ATTENTION		Read list of digits (1 digit/ sec.). Subject has to repeat them in the forward order [] 2 1 8 5 4 Subject has to repeat them in the backward order [] 7 4 2					___/2
		Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors [] FBACMNAAJKLBAFAKDEAAAJAMOF AAB					___/1
		Serial 7 subtraction starting at 100 [] 93 [] 86 [] 79 [] 72 [] 65 4 or 5 correct subtractions: 3 pts, 2 or 3 correct: 2 pts, 1 correct: 1 pt, 0 correct: 0 pt					___/3
LANGUAGE		Repeat : I only know that John is the one to help today. [] The cat always hid under the couch when dogs were in the room. []					___/2
		Fluency / Name maximum number of words in one minute that begin with the letter F [] _____ (N ≥ 11 words)					___/1
ABSTRACTION		Similarity between e.g. banana - orange = fruit [] train - bicycle [] watch - ruler					___/2
DELAYED RECALL		Has to recall words WITH NO CUE					Points for UNCUEDE recall only
		FACE	VELVET	CHURCH	DAISY	RED	
	Category cue						
Optional	Multiple choice cue						
ORIENTATION		[] Date [] Month [] Year [] Day [] Place [] City					___/6
		© Z.Nasreddine MD Version November 7, 2004					
		Normal ≥ 26 / 30					
		www.mocatest.org					
		TOTAL					___/30
		Add 1 point if ≤ 12 yr edu					

Appendix J MoCA administration instructions

MoCA Version Nov. 2014, Copied from Z. Nasreddine MD at www.mocatest.org

Montreal Cognitive Assessment (MoCA) Administration and Scoring Instructions

The Montreal Cognitive Assessment (MoCA) was designed as a rapid screening instrument for mild cognitive dysfunction. It assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. Time to administer the MoCA is approximately 10 minutes. The total possible score is 30 points; a score of 26 or above is considered normal.

1. Alternating Trail Making:

Administration: The examiner instructs the subject: "Please draw a line, going from a number to a letter in ascending order. Begin here [point to (1)] and draw a line from 1 then to A then to 2 and so on. End here [point to (E)]."

Scoring: Allocate one point if the subject successfully draws the following pattern: 1 □ A- 2- B- 3- C- 4- D- 5- E, without drawing any lines that cross. Any error that is not immediately self-corrected earns a score of 0.

2. Visuoconstructional Skills (Cube):

Administration: The examiner gives the following instructions, pointing to the **cube**: "*Copy this drawing as accurately as you can, in the space below*".

Scoring: One point is allocated for a correctly executed drawing.

- Drawing must be three-dimensional
- All lines are drawn
- No line is added
- Lines are relatively parallel and their length is similar (rectangular prisms are accepted)
- A point is not assigned if any of the above-criteria are not met.

3. Visuoconstructional Skills (Clock):

Administration: Indicate the right third of the space and give the following instructions: "*Draw a clock. Put in all the numbers and set the time to 10 after 11*".

Scoring: One point is allocated for each of the following three criteria:

- Contour (1 pt.): the clock face must be a circle with only minor distortion

- acceptable (e.g., slight imperfection on closing the circle);
- Numbers (1 pt.): all clock numbers must be present with no additional numbers; numbers must be in the correct order and placed in the approximate quadrants on the clock face; Roman numerals are acceptable; numbers can be placed outside the circle contour;
 - Hands (1 pt.): there must be two hands jointly indicating the correct time; the hour hand must be clearly shorter than the minute hand; hands must be centred within the clock face with their junction close to the clock centre.
- A point is not assigned for a given element if any of the above criteria are not met.

4. Naming:

Administration: Beginning on the left, point to each figure and say: *“Tell me the name of this animal”*.

Scoring: One point each is given for the following responses: (1) camel or dromedary, (2) lion, (3) rhinoceros or rhino.

5. Memory:

Administration: The examiner reads a list of 5 words at a rate of one per second, giving the following instructions: *“This is a memory test. I am going to read a list of words that you will have to remember now and later on. Listen carefully. When I am through, tell me as many words as you can remember. It doesn’t matter in what order you say them”*. Mark a check in the allocated space for each word the subject produces on this first trial. When the subject indicates that (s)he has finished (has recalled all words), or can recall no more words, read the list a second time with the following instructions: *“I am going to read the same list for a second time. Try to remember and tell me as many words as you can, including words you said the first time.”* Put a check in the allocated space for each word the subject recalls after the second trial.

At the end of the second trial, inform the subject that (s)he will be asked to recall these words again by saying, *“I will ask you to recall those words again at the end of the test.”*

Scoring: No points are given for Trials One and Two.

6. Attention:

Forward Digit Span: Administration: Give the following instruction: *“I am going to say some numbers and when I am through, repeat them to me exactly as I said them”*. Read the five number sequence at a rate of one digit per second.

Backward Digit Span: Administration: Give the following instruction: *“Now I am going to say some more numbers, but when I am through you must repeat them to me in the backwards order.”* Read the three number sequence at a rate of one digit per second.

Scoring: Allocate one point for each sequence correctly repeated, (N.B.: the correct response for the backwards trial is 2-4-7).

Vigilance: Administration: The examiner reads the list of letters at a rate of one per second, after giving the following instruction: *“I am going to read a sequence of letters. Every time I say the letter A, tap your hand once. If I say a different letter, do not tap your hand”*.

Scoring: Give one point if there is zero to one errors (an error is a tap on a wrong letter or a failure to tap on letter A).

Serial 7s: Administration: The examiner gives the following instruction: *“Now, I will ask you to count by subtracting seven from 100, and then, keep subtracting seven from your answer until I tell you to stop.”* Give this instruction twice if necessary.

Scoring: This item is scored out of 3 points. Give no (0) points for no correct subtractions, 1 point for one correction subtraction, 2 points for two-to-three correct subtractions, and 3 points if the participant successfully makes four or five correct subtractions. Count each correct subtraction of 7 beginning at 100. Each subtraction is evaluated independently; that is, if the participant responds with an incorrect number but continues to correctly subtract 7 from it, give a point for each correct subtraction. For example, a participant may respond “92 – 85 – 78 – 71 – 64” where the “92” is incorrect, but all subsequent numbers are subtracted correctly. This is one error and the item would be given a score of 3.

7. Sentence repetition:

Administration: The examiner gives the following instructions: *“I am going to read you a sentence. Repeat it after me, exactly as I say it [pause]: **I only know that John is the one to help today.**”* Following the response, say: *“Now I am going to read you another sentence. Repeat it after me, exactly as I say it [pause]: **The cat always hid under the couch when dogs were in the room.**”*

Scoring: Allocate 1 point for each sentence correctly repeated. Repetition must be exact. Be alert for errors that are omissions (e.g., omitting "only", "always") and substitutions/additions (e.g., "John is the one who helped today;" substituting "hides" for "hid", altering plurals, etc.).

8. Verbal fluency:

Administration: The examiner gives the following instruction: *“Tell me as many words as you can think of that begin with a certain letter of the alphabet that I will tell you in a moment. You can say any kind of word you want, except for proper nouns (like Bob or Boston), numbers, or words that begin with the same sound but have a different suffix, for example, love, lover, loving. I will tell you to stop after one minute. Are you ready?”*

[Pause] Now, tell me as many words as you can think of that begin with the letter F. [time for 60 sec]. Stop.”

Scoring: Allocate one point if the subject generates 11 words or more in 60 sec. Record the subject’s response in the bottom or side margins.

9. Abstraction:

Administration: The examiner asks the subject to explain what each pair of words has in common, starting with the example: “*Tell me how an orange and a banana are alike*”. If the subject answers in a concrete manner, then say only one additional time: “*Tell me another way in which those items are alike*”. If the subject does not give the appropriate response (fruit), say, “*Yes, and they are also both fruit.*” Do not give any additional instructions or clarification.

After the practice trial, say: “*Now, tell me how a train and a bicycle are alike*”. Following the response, administer the second trial, saying: “*Now tell me how a ruler and a watch are alike*”. Do not give any additional instructions or prompts.

Scoring: Only the last two item pairs are scored. Give 1 point to each item pair correctly answered. The following responses are acceptable:

Train-bicycle = means of transportation, means of travelling, you take trips in both;

Ruler-watch = measuring instruments, used to measure.

The following responses are **not** acceptable: Train-bicycle = they have wheels; Ruler-watch = they have numbers.

10. Delayed recall:

Administration: The examiner gives the following instruction: “*I read some words to you earlier, which I asked you to remember. Tell me as many of those words as you can remember.* Make a check mark for each of the words correctly recalled spontaneously without any cues, in the allocated space.

Scoring: Allocate 1 point for each word recalled freely without any cues.

Optional:

Following the delayed free recall trial, prompt the subject with the semantic category cue provided below for any word not recalled. Make a check mark in the allocated space if the subject remembered the word with the help of a category or multiple-choice cue. Prompt all non-recalled words in this manner. If the subject does not recall the word after the category cue, give him/her a multiple choice trial, using the following example instruction, “*Which of the following words do you think it was, NOSE, FACE, or HAND?*”

FACE:	category cue: part of the body	multiple choice: nose, face, hand
VELVET:	category cue: type of fabric	multiple choice: denim, cotton, velvet
CHURCH:	category cue: type of building	multiple choice: church, school, hospital
DAISY:	category cue: type of flower	multiple choice: rose, daisy, tulip
RED:	category cue: a colour	multiple choice: red, blue, green

Use the following category and/or multiple-choice cues for each word, when appropriate:

Scoring: **No points are allocated for words recalled with a cue.** A cue is used for clinical information purposes only and can give the test interpreter additional information about the type of memory disorder. For memory deficits due to retrieval failures, performance can be improved with a cue. For memory deficits due to encoding failures, performance does not improve with a cue.

11. Orientation:

Administration: The examiner gives the following instructions: *“Tell me the date today”*. If the subject does not give a complete answer, then prompt accordingly by saying: *“Tell me the [year, month, exact date, and day of the week].”* Then say: *“Now, tell me the name of this place, and which city it is in.”*

Scoring: Give one point for each item correctly answered. The subject must tell the exact date and the exact place (name of hospital, clinic, office). No points are allocated if subject makes an error of one day for the day and date.

TOTAL SCORE: Sum all subscores listed on the right-hand side. Add one point for an individual who has 12 years or fewer of formal education, for a possible maximum of 30 points. A final total score of 26 and above is considered normal.

Appendix K Loudness discomfort levels methodology

Loudness Discomfort Levels (LDLs) were performed for 500 and 3000 Hz.

Methodology:

1. Participants were handed a “categories of loudness chart” displaying 7 categories: very soft, soft, comfortable but slightly soft, comfortable, comfortable but slightly loud, loud but o.k., uncomfortably loud.

2. Participants were given the following instructions: “We need to do a test to determine where to set the amplifier on your hearing aid. We want to set it such that sounds do not become uncomfortably loud. If we set it too high, sounds could get too loud, and you may not want to wear the hearing aid. You will hear some sounds through the earphone, and after each one I want you to tell me which of the loudness categories on this sheet (hand loudness chart to patient) best describes the sound to you. So after each sound, tell me if it was “comfortable”, or “comfortable, but slightly loud”, or “loud, but O.K.”, or “uncomfortably loud”, etc. I will be zeroing in on the “uncomfortably loud” category because that is where I want the hearing aid to stop. Think of uncomfortably loud as where you want the hearing aid to stop and not get any louder. We want the hearing aid to keep sounds in the comfortable regions and not let sounds get into the uncomfortable regions. So after each sound, tell me the category that best described it for you.”

3. Starting at 80 dB HL, a pure-tone warble (500 or 3000 Hz) was presented in ascending 5 dB steps until the participant reported “uncomfortable”. The pure-tone was then decreased by 5 steps and then re-ascended in 5 dB steps. The “uncomfortable” threshold was the level where the participant reported “uncomfortable” on 2 out of 3 ascending trials. This value was used as the individual’s LDL.

Appendix L Sentence repetition keywords correct (%) by participant age

Table L.1. Mean sentence repetition performance and standard deviations for each participant listed by age

Age	Sentence repetition scores: % keywords correct								
	50 dB SPL speech level			65 dB SPL speech level			80 dB SPL speech level		
	0 SNR	+5 SNR	+10 SNR	0 SNR	+5 SNR	+10 SNR	0 SNR	+5 SNR	+10 SNR
21	4.1±7.2	12.9±15.2	27.7±21.3	36.0±27.8	77.7±22.5	92.7±10.3	41.0±16.8	71.0±19.6	97.0±6.0
24	5.6±9.7	15.8±16.1	23.8±16.0	28.8±23.0	78.5±13.8	88.9±14.6	30.9±27.3	80.5±17.1	97.9±5.5
30	0.0±0.0	1.6±5.4	6.3±9.6	49.7±22.6	82.1±12.6	94.5±6.5	49.8±15.5	87.2±11.5	97.0±6.0
35	25.9±20.3	68.6±11.4	90.0±10.9	34.8±29.5	86.8±11.7	96.4±5.6	50.0±21.4	88.9±13.3	96.1±10.1
37	1.0±3.4	8.8±8.9	20.8±9.9	39.0±19.8	68.2±19.9	76.6±29.0	50.1±24.1	80.9±20.2	96.4±7.3
42	0.5±1.8	25.0±17.0	47.5±27.1	23.2±18.3	61.4±17.5	77.1±17.1	13.8±15.6	59.8±24.4	71.4±19.6
57	10.3±10.7	64.5±19.4	81.6±13.9	30.3±19.2	82.5±14.6	88.4±12.3	28.1±16.3	65.0±13.8	89.9±15.9
61	1.0±3.6	16.3±11.2	33.7±16.7	18.9±18.8	44.7±24.7	64.1±32.0	14.4±16.5	31.5±28.4	61.7±25.7
66	0.5±1.8	4.1±7.5	7.6±9.9	12.2±14.7	47.8±26.8	73.9±19.5	12.8±19.1	63.7±25.1	95.3±11.0
67	4.1±5.5	16.2±11.3	50.2±12.8	18.9±18.4	67.2±25.9	81.7±11.1	14.3±12.3	61.6±22.0	89.9±13.2
73	0.5±1.7	17.4±24.2	25.4±22.5	22.3±21.6	61.2±17.2	90.2±17.8	6.7±9.4	53.9±33.1	83.9±15.4
76	<u>0.0±0.0</u>	<u>0.0±0.0</u>	<u>0.0±0.0</u>	<u>0.0±0.0</u>	<u>0.0±0.0</u>	<u>0.0±0.0</u>	<u>0.0±0.0</u>	<u>0.0±0.0</u>	<u>2.6±6.2</u>
Overall	4.5±2.1	20.9±6.5	34.6±8.2	26.2±3.8	63.2±6.9	77.0±7.5	26.0±5.2	62.0±7.3	81.6±7.9

Note. Overall means and standard errors across participants are listed at the bottom of the table.

Appendix M CMA rating means across participants

Table M.1. CMA rating means and 95% confidence intervals of the mean across participants for Experiment I

Speech level (dB SPL) SNR (dB)	50			65			80		
	0	+5	+10	0	+5	+10	0	+5	+10
Intelligibility									
During	1.7±1.5	1.6±1.2	3.0±1.9	2.9±1.3	5.3±1.7	6.7±1.5	2.9±1.3	5.9±1.4	7.1±1.6
After	0.9±1.0	2.4±1.5	2.6±1.5	2.6±1.4	4.8±1.5	6.6±1.5	3.0±1.1	5.2±1.5	7.1±1.3
Noisiness									
During	3.6±1.3	2.7±1.2	2.0±1.4	6.3±1.1	5.1±1.3	4.4±1.2	9.0±0.6	8.0±0.8	7.2±0.8
After	3.8±1.4	2.9±1.4	2.2±1.3	6.0±1.5	4.9±1.2	4.5±1.2	8.2±1.1	7.2±1.0	6.5±1.0
Listening effort									
During	9.0±0.8	9.1±0.9	8.6±1.0	9.1±0.7	8.3±0.9	7.7±1.1	9.1±0.9	8.9±0.7	7.7±1.2
After	8.9±1.0	8.8±1.1	8.7±1.1	8.8±1.0	8.3±1.0	7.8±1.1	9.2±0.7	8.6±0.8	7.8±1.1
Loudness									
During	2.3±1.2	1.8±0.8	2.0±1.1	4.8±1.2	5.0±1.3	4.9±1.2	7.9±1.3	7.3±1.1	7.3±0.9
After	2.5±1.3	2.2±1.2	2.0±1.2	4.5±1.6	4.5±1.2	4.4±1.3	7.1±1.3	6.7±1.1	6.5±1.0

Table M.2. CMA rating means and 95% confidence intervals of the mean across participants for Experiment II

Speech level (dB SPL)	50			65			80			
	SNR (dB)	0	+5	+10	0	+5	+10	0	+5	+10
Intelligibility										
SR	1.4±1.1	2.3±1.3	3.3±1.4	2.9±1.3	5.9±1.5	7.0±1.1	2.9±1.2	5.3±1.5	7.2±1.4	
CTD	1.8±1.1	2.5±1.5	3.4±1.5	3.3±1.3	5.3±1.5	6.4±1.3	3.8±1.4	5.5±1.3	7.9±0.9	
Noisiness										
SR	4.2±1.4	3.4±1.5	2.6±1.4	6.1±1.3	5.1±1.2	4.6±1.1	8.6±0.6	7.4±0.8	6.7±0.9	
CTD	4.4±1.4	3.1±1.2	2.7±1.4	5.75±1.3	5.1±1.1	4.6±1.1	8.5±0.8	7.3±1.0	6.6±0.9	
Listening effort										
SR	8.1±1.7	8.4±1.2	8.0±1.3	8.4±1.2	7.7±1.3	7.4±1.3	8.6±1.4	8.6±0.9	7.4±1.3	
CTD	7.5±1.7	7.1±1.6	6.9±1.4	7.4±1.6	6.9±1.3	6.5±1.3	7.4±1.7	6.9±1.5	6.9±1.2	
Loudness										
SR	2.8±1.5	2.6±1.4	2.6±1.2	4.6±1.3	5.1±1.1	4.7±1.1	7.3±1.3	6.9±0.9	6.3±0.9	
CTD	2.4±1.5	2.3±1.0	2.6±1.0	4.4±1.4	4.6±1.1	4.8±1.2	6.9±1.3	6.6±1.1	6.4±0.9	

Note. SR = sentence repetition focus; CTD = connect-the-dots focus.