METHODS FOR ROBUST CHARACTERIZATION OF CONSONANT PERCEPTION IN HEARING-IMPAIRED LISTENERS

BY

WOOJAE HAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Speech and Hearing Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:
Associate Professor Ron D. Chambers, Chair
Associate Professor Jont B. Allen, Director of Research
Associate Professor Cynthia J. Johnson
Associate Professor Mark A. Hasegawa-Johnson
Associate Professor Robert E. Wickesberg

ABSTRACT

Individuals with *sensorineural hearing loss* (SNHL) are prescribed hearing aids and/or a cochlear implant, based on their pure-tone threshold and speech perception scores. Although these assistive listening devices do help these individuals communicate in quiet surroundings, many still have difficulty understanding speech in noisy environments. Especially, listeners with mild-to-moderate SNHL have complained that their hearing aids do not provide enough benefit to facilitate understanding of normal speech. Why is it that the modern hearing aid, even with a high level of technology, does not produce one-hundred percent efficiency? We shall show that the current clinical measurements, which interpret the result as a mean score (e.g., pure-tone average, speech recognition threshold, AI-gram, etc.), do not deliver sufficient information about the characteristics of a SNHL listener's impairment when hearing speech, and thus, result in a poorly fitting hearing aid.

This dissertation addressed three key questions, fundamental to clinical audiology and hearing science: (1) How well do the results of standard clinical tests predict the speech perception ability of SNHL patients? (2) Are the existing methods of hearing aid fitting (e.g., the half-gain rule, NAL-R, etc.) appropriate for modern hearing aid technology? (3) How useful are measured error patterns of speech perception in SNHL patients in addressing these perception errors?

Four sub-studies were conducted for finding answers to the proposed questions:

Study I measured individual consonant errors to quantify how each hearing-impaired (HI) listener perceives speech sounds (e.g., high- vs. low-error consonants), and then compared the individual consonant errors to the results provided by currently used clinical measurements to ascertain the differences. The results of Study I showed that the HI ear had significant errors in receiving only a few consonants. There was a low correlation between the error rates of high-error consonants and either degree and configuration of pure-tone hearing threshold or average consonant scores.

Study II examined how reliably a CV listening test could measure a HI listener's consonant loss using only *zero-error* (ZE) utterances (defined as utterances for which normal hearing (NH) listeners incur zero errors, (Singh and Allen, 2011)) and having a statistically suitable number of presentations in CVs, in order to characterize unique HI consonant loss. We provided graphical as well as statistical analysis to see not only the error rate (%) of a target consonant but also its pattern of specific confusions. As we found in Study I, there was no measurable correlation between pure-tone threshold and the error rate, or no identification of high-error consonants in HI ears. As noise increased, the percentage of error and confusions of target consonants increased. Although some consonants showed significantly higher errors and resulted in more confusion than others, HI ears have a very different consonant confusion pattern than NH ears, which may not be either measured or analyzed by the use of average scores. Comparison between the two (separated) phases of the experiment (Exp. II) showed a good internal consistency for all HI ears.

Study III investigated whether or not NAL-R amplification might offer a positive benefit to speech perception of each HI listener at the consonant level, i.e., differentiates consonants that are distorted with amplification from those that achieve a positive benefit from amplification. The results were then compared to the current clinical measurement to see a relation between consonants which have positive am-

plification benefit and hearing loss. Regardless of NAL-R amplification, HI listeners have their own consonant dependence and the dependence was not predicted by either pure-tone threshold or aided threshold. HI listeners who have symmetrical hearing loss do not have the same positive amplification benefit to the two ears.

Study IV characterized consonant perception errors of each HI listener by identifying missing critical features of misheard consonants as a function of signal-to-noise ratio (SNR), while following the same procedure (i.e., increasing the number of ZE utterance presentations up to 20) as in Study II, yet for the NAL-R amplification condition. As the noise increased, consonant error and confusions were significantly increased, although by applying gains provided by NAL-R amplification correction. The percentage of error and confusions of the target consonants were different across the HI ears, thus could not be averaged. When the results of Study IV were compared with those of Study II, a significant amplification effect is found. Generally, the percentage of error and confusions were decreased in the NAL-R condition as a function of SNRs. However, typical average analysis, using mean score and grouping the HI ears, failed to explain the idiosyncratic characteristics of HI speech perception.

Overall, this series of studies concluded that current average measures and analyses have a serious, even fatal limitation in finding problems of HI speech perception. Therefore, we have explored the use of the nonsense CV test for as a more precise measure. We will show that this can make significant contributions to HI speech perception. We propose that this CV test and its application might be utilized in the clinical setting, to improve the diagnosis of HI speech perception. This research will help HI listeners hear day-to-day conversations more clearly, as well as aid in audiological diagnosis and successful rehabilitation to increase speech perception for HI listeners.

To my parents, Boo-Young Han and Im-Soon Yeum, who mean the world to me.

and

To all my hearing-impaired patients, whom I have loved, do love, and will love.

ACKNOWLEDGMENTS

I thank Professor Jont B. Allen of Electrical and Computer Engineering (ECE), a great mentor and advisor for my doctoral research on hearing-impaired speech perception. He is one of the most passionate researchers and professors I know, and I am very lucky to have had the remarkable opportunity to serve as his graduate research assistant. Whenever I study data of hearing-impaired listeners with Dr. Allen, I have felt that I am a fortunate and blessed person who can help develop better diagnoses and treatment for the hearing-impaired patients. Furthermore, sharing my experience from the trials and challenges in my current work with the hearing-impaired participants has encouraged me seek further accomplishments, and has given me a strong feeling of heading in the right direction. In addition, all the former and current ECE graduate students in the Human Speech Recognition (HSR) group of UIUC (Sandeep, Feipeng, Anjali, Bob, Roger, Andrea, Austin, Cliston, Katie, Sarah, Noori, and my lovely colleague Riya) are wonderful and bright colleagues with whom I enjoyed discussions every Monday and Wednesday. These helped me to deal with problems that I could not solve by myself.

I also thank another advisor, Professor Ron D. Chambers of Speech and Hearing Science (SHS), who guided me in all my doctoral study. Having come from Korea, when I struggled with a different academic system in the United States, he made me see details as well as the big picture - almost like a father. Without him, I could not have survived my four years in Champaign. My friends and colleagues in the SHS department (Hee-Cheong, Fang-Ying, Panying, and Lynn) also were helpful; my joy

was doubled and sadness was halved when I shared my work and concern with them.

I sincerely thank the Doctoral Planning Committee (Dr. Chambers, Dr. Gooler, and Dr. Lansing) and Doctoral Dissertation Committee (Dr. Chambers, Dr. Allen, Dr. Johnson, Dr. Hasegawa-Johnson, and Dr. Wickersberg), both of which supported my study in UIUC.

Special thanks to two unforgettable Korean professors, Dr. Jin-Sook Kim (Master's thesis advisor) and Dr. Jung-Hak Lee (Master's academic advisor and clinical supervisor), who have given me endless and full support since 2001. They are pioneers who opened an Audiology program in Korea and let me understand the importance and value of the field of Audiology and Hearing Science.

I will close with a beautiful story. While preparing this acknowledgement essay, I recalled all the hearing-impaired patients I met in the clinic in Korea. Among them was Eunchae, who was my first client. I clearly remember her story. We met in the clinic when I was a Master's student and she was four years old. When I saw her the first time, I realized that her mother had taught her Korean sign language because the child did not have a hearing aid. I did various hearing tests on Eunchae, fitted hearing aids to her two small ears, and started my first auditory training session under a clinical supervisor. Eunchae made amazing improvement in speech/language discrimination and recognition after only three months of auditory training. Today, Eunchae attends public school with hearing children. She sent me a birthday card from Korea a couple of years ago. Its gist was **"Thanks for fitting my hearing aids. If I don't wear them, I may bother my friends with no hearing as well as no speaking."** Her story still warms my heart and remains my personal inspiration for strongly contributing further in this field.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xiv

LIST OF ABBREVIATIONS xix

CHAPTER 1 GENERAL INTRODUCTION 1
- 1.1 Statement of Problem 1
- 1.2 Literature Review 7
  - 1.2.1 The Theory and Models of Speech Perception 7
  - 1.2.2 Synthetic Speech Cue Research 10
  - 1.2.3 Natural Speech Cue Research 11
    - *Identification of Consonant Cues* 11
    - *Three Dimensional Deep Search (3DDS)* 12
    - *Conflicting Cues* 13
    - *Manipulation of Consonant Cues* 13
    - *Context* 15
- 1.3 Measurements of Speech Perception in Hearing-Impairment 16
  - 1.3.1 Current Clinical Measurements 16
    - *Pure-tone Audiogram (PTA)* 16
    - *Speech Recognition Threshold (SRT)* 17
    - *Word/Sentence Tests* 18
  - 1.3.2 Non-clinical Measurements 20
    - *Articulation Index (AI)* 20
    - *Confusion Matrix (CM)* 20
- 1.4 Purpose of the Study and Hypothesis 21
  - 1.4.1 Study I: Consonant-Loss Profile (CLP) in Hearing-Impaired Listeners 22
  - 1.4.2 Study II: Verification of Consonant Confusion Patterns in Hearing-Impaired Listeners and Test Reliability 22
  - 1.4.3 Study III: Effect of NAL-R Amplification on Consonant-Loss of Hearing-Impaired Listeners 23
  - 1.4.4 Study IV: Verification of Consonant Confusion Patterns with Amplification Condition 24

CHAPTER 2 METHODOLOGY 25
- 2.1 Experiment I 25
  - 2.1.1 Subjects 25
  - 2.1.2 Speech Stimuli 26
  - 2.1.3 Procedure 27
- 2.2 Experiment II 28
  - 2.2.1 Subjects 28
  - 2.2.2 Speech Stimuli 29
  - 2.2.3 Procedure 29
- 2.3 Experiment III 30
  - 2.3.1 Subjects 30
  - 2.3.2 Speech Stimuli 31
  - 2.3.3 NAL-R Amplification Condition 31
  - 2.3.4 Procedure 32
- 2.4 Experiment IV 33
  - 2.4.1 Subjects 33
  - 2.4.2 Speech Stimuli 34
  - 2.4.3 NAL-R Amplification Condition 35
  - 2.4.4 Procedure 35
- 2.5 Bernoulli Trials and Speech Perception 35
  - 2.5.1 Further Considerations 39

CHAPTER 3 RESULTS OF EXP. II 40
- 3.1 Error Pattern Analysis: Subjects 44L, 46L, 36R, and 40L 44
- 3.2 Talker Dependence 52
- 3.3 Internal Consistency 53
- 3.4 Comparison between the PTA and CLP as a Clinical Application 55

CHAPTER 4 RESULTS OF EXP. IV 58
- 4.1 Error Pattern Analysis of NAL-R Amplification 58
- 4.2 Comparison of Exps. II and IV: Flat vs. NAL-R gain 59

CHAPTER 5 RESULTS OF EXPS. I AND III 62
- 5.1 Analysis of Experiment I 62
  - 5.1.1 Comparisons between the PTA and CLP 62
  - 5.1.2 Comparisons between the CRT and CLP 65
- 5.2 Analysis of Experiment III 67
  - 5.2.1 Comparison between the PTA vs. Aided Threshold 67
  - 5.2.2 Consonant-Dependence 67
  - 5.2.3 Listener-Dependence 71
    - Symmetric Hearing Loss 71
  - 5.2.4 Asymmetric Hearing Loss 73

CHAPTER 6 DISCUSSION 76
6.1 Individual differences of HI Consonant Perception 76
6.2 Amplification Effect of Consonant Perception 78
6.3 Relation of PTA and CLP 79
6.4 Relation of CRT and CLP 82
6.5 Limitation of the Studies and Future Directions 83

CHAPTER 7 CONCLUSIONS 86

REFERENCES 88

APPENDIX A: AGE AND PURE-TONE THRESHOLDS OF HI SUBJECTS 95

APPENDIX B: INDIVIDUAL CONSONANT ERRORS OF EXP. II 96

APPENDIX C: INDIVIDUAL CONSONANT ERRORS OF EXP. IV 98

APPENDIX D: IRB DOCUMENTS 100

LIST OF TABLES

2.1 Table summary of the four experimental designs used in the current study. . . . 25

2.2 Example of 6 different utterances per syllable used in Exps. I and III 27

2.3 Number of presentation trials per consonant in Phases I and II of Exps. II and IV, depending on percent error. 30

2.4 Zero-Error utterances which were used in Exps. II and IV. The numbers in parentheses refer to each stimulus' SNR90 (signal-to-noise ratio at which NH listeners perceive on utterance with 90% accuracy). 33

3.1 Percent consonant errors (%) of seven select HI ears (rows) in the **quiet condition** [Exp. II]. **High**(>75%), **medium** (>50% and less than 75%), and *low* (>25% and less than 50%) errors are marked by red, blue, and green, respectively. Empty space indicates no error. For example, as shown by the second row, NH ears had zero error. Note that every HI ear has errors in many individual consonants, but there is high error for only a few of consonants. Note the high /za/ and /ʒa/ errors in HI46R. The two right columns provide typical clinical measures. 3TA (3-tone average, dB HL) is calculated by the average of 0.5, 1, and 2 kHz, and CRT (consonant recognition threshold; dB SNR) is the average consonant threshold at 50% error, similar to the SRT. Although having similar 3TA and CRT, HI01 shows asymmetric consonant perception between left (HI01L) and right (HI01R) ears - /sa/ and /za/ are better perceived in HI01L and /pa/ and /va/ are better in HI01R. 43

3.2 Sub-count matrix at 6 and 0 dB-SNR for HI44L; the frequencies in this table are re-plotted as percentages in Fig. 3.3. Each row is stimulus consonant, while each column is response consonant. Last column is total number of presentations. Cells with only 1 or 2 errors were not displayed because they were considered to be low level random errors. The number in the left top cell indicates the SNR. For example, at 6 dB, /na/ is presented 19 times of which 15 are correct and 4 incorrect (heard as /ma/) responses. 46

3.3 Results of total entropy calculation of 4 SNRs for 17 HI ears in Exp. II: $\mathcal{H} = -\sum_i p(x_i)\log_2 p(x_i)$. $\mathcal{H}$ is a measure of the subject's response uncertainty. When the entropy is zero, there is no subject uncertainty, independent of the scores ($P_{h|s}$). As noise increased, the entropy significantly increases, which means the confusions increased. Bonferroni Post-Hoc test showed there is a significant difference between each of three SNRs (quiet, +12, +6 dB) and 0 dB ($p<0.01$) ($F[3,45]=83.619$, $p<0.01$). Confusions from quiet condition to +6 dB SNR were not increased, but were significantly higher at 0 dB. Group mean of the entropy at quiet, +12, +6, and 0 are 0.242, 0.373, 0.567, and 1.091 bits, respectively. In column six, SNR_1^* indicates the SNR where the entropy is 1-bit, i.e., $\mathcal{H}(\mathrm{SNR}_1^*)=1$. 47

3.4 Sub-count matrix for quiet, +12, +6 and 0 dB for HI46L (see Fig. 3.4). The number in the left top cell indicates the SNR. Each row is a presented consonant (stimuli) and each column is a response. The last column is total number of presentation. Single and double errors are not displayed due to these error. Diagonal entries are correct and off-diagonal is an error. 49

3.5 Sub-count matrix in the quiet, +12, +6 and 0 dB for HI36R, paired with Fig. 3.5. The subject's only errors were for /ba/, /va/, /na/ syllables. Note how the /ba/ errors were confused with /va/ and /da/, and how /va/ was perceived as /pa/. 51

3.6 Sub-count matrix at quiet, +12, +6 and 0 dB for HI40L which is paired with Fig. 3.6. As the noise increases, the number of consonant producing significant high error is increased from 1 (i.e., /fa/) in the quiet condition to 8 at 0 dB. Note how /va/ is represented when /ba, ga, va, ma, na/ are spoken, yet is only recognized 40% of the time. 52

4.1 Results of total entropy calculation of 4 SNRs for 16 HI ears in Exp. IV. Formula of entropy is $\mathcal{H} = -\sum_i p(x_i)\log_2 p(x_i)$. $\mathcal{H}$ is a measure of the subject's response uncertainty. When the entropy is zero, there is no subject uncertainty, independent of the scores ($P_{h|s}$). As noise increased, the entropy was significantly increased ($F[3,45]=100.306$, $p<0.01$). Group mean of entropy at quiet, +12, +6, and 0 was 0.209, 0.345, 0.456, and 0.785 bits, respectively. SNR_1^* indicates 1-bit of entropy for Exps. II and IV. The eighth column is the SNR_1^* difference of two experiment. 59

A.1 Table summary of age and pure-tone thresholds (from .125 to 8 kHz) of HI subjects who were participated in Exps. I to IV 95

B.1 Percent individual consonant error (%) for 17 impaired ears of Exp. II at 12 dB SNR. Each entry represents the error (%) for 14 syllables. Every syllable used in Exp. II is an utterance for which 10 normal hearing listeners have zero error for SNRs ≥ -2 dB, even for 500 trials. Code: **High**(>75%), **medium** (>50% and less than 75%), and *low* (>25% and less than 50%) errors are marked by red, blue, and green, respectively. Empty space indicates zero error. The two right columns display clinical measures; 3TA (3-tone average, dB HL) is calculated by the average of 0.5, 1, and 2 kHz, and CRT (consonant recognition threshold; dB SNR) means the average consonant threshold of 50% error, relative to the SRT calculation. Note how every HI ear makes a high error for a few of consonants. 96
B.2 Percent individual consonant error (%) for 17 impaired ears of Exp. II at 6 dB SNR. Note compared to HI32R and 36L who have same PTA, only HI30R show high error in /sa/, /ba/, and /ʒa/. 97
B.3 Percent individual consonant error (%) for 17 impaired ears of Exp. II at 0 dB SNR. Note as noise increases, HI36L, 32R, and 30R all of same PTA have increased /ba/ error. Yet HI36L has still less error in most consonants except for /pa/ and /ba/. HI36R cannot hear /ba, whereas HI36L misses 50%. 97

C.1 Percent individual consonant error (%) for 16 impaired ears of Exp. IV at quiet. Each entry represents the error (%) for 14 syllables. Every syllable used in Exp. IV is an utterance for which 10 normal hearing listeners have zero error for SNRs ≥ -10 dB. Code: **High**(>75%), **medium** (>50% and less than 75%), and *low* (>25% and less than 50%) errors are marked by red, blue, and green, respectively. Empty space indicates zero error. Note how every HI ear makes a high error for a few of consonants. Order of subject is followed to that of Exp. II. 98
C.2 Percent individual consonant error (%) for 16 impaired ears of Exp. IV at 12 dB SNR. 98
C.3 Percent individual consonant error (%) for 16 impaired ears of Exp. IV at 6 dB SNR. 99
C.4 Percent individual consonant error (%) for 16 impaired ears of Exp. IV at 0 dB SNR. 99

LIST OF FIGURES

1.1 A flow chart of the typical clinical procedure for hearing-impaired listener as a process diagram. Abbreviations used are Tymp = Tympanometry; PTA = Pure-Tone Audiogram; SRT = Speech Recognition Threshold, HINT = Hearing-In-Noise Test; QSIN = Quick Speech-In-Noise test; OAE = Otoacustic Emission; ABR = Auditory Brain Response; NAL-R = the revised National Acoustic Laboratories prescriptive formula; NAL-NL = Nonlinear NAL formula. . 3

2.1 m112 /fa/ token was rendered incomplete by Matlab code designed to automatically cut off the silent part before and after the stimulus. 34

2.2 This figure results from a Monti Carlo (numerical) simulation of a biased coin flip. In this simulation a coin with a bias of $P_{h|s} = 0.95$ was tossed for $N_t = 20$ flips with 10^5 trials. A random variable X was defined as 1 if head and 0 if tails and the mean of the random variable μ and its variance σ_μ was then computed from the trials. A histogram of the outcomes from the 10^5 trials is shown, normalized as a probability. The estimated mean was $\eta = 0.95$, which happened with a probability of $\mu \approx 0.28$, namely 280,000 times. Also show are $\eta - \sigma_\mu$ and $\eta - 3 * \sigma_\mu$. The ratio of the theoretical $\sigma_\mu = \sqrt{P_{h|s}(1 - P_{h|s})/N_t}$ and the actual variance computed by the simulation is ≈ 1 within 0.12%. 38

3.1 Average consonant error for 46 HI ears of Exp. I [(a), solid colored lines] and for 17 HI ears of Exp. II [(b), solid colored lines) as function of signal-to-noise ratio (SNR) in speech-weighted noise: abscissa represents SNR and ordinate is **average percent consonant error (%)** for 16 CVs for Exp. I and 14 CVs for Exp. II. The intersection of the thick horizontal dashed line at the 50% error point and the plotted average error line for each ear, mark the consonant recognition threshold (CRT) in dB. The data for 10 normal hearing (NH) ears are superimposed as solid gray lines for comparison [(a), grey lines]. NH ears have a similar and uniform CRT of -18 to -16 dB (only a 2-dB range), while the CRT of HI ears are spread out between -5 to +28 dB (a 33-dB range). Three out of 46 ears had greater than 50% error in quiet (i.e., no CRT) in panel (a). In panel (b), the CRT for these 17 ears are mostly from the <0 dB CRT region, thus the mean error is much smaller (1% or so) compared to (a) where the mean error is 15%. 41

3.2 Individual consonant loss profiles (CLP) $P_{h|s}$(SNR) for eight consonants of subject HI40L in Exp. II, the consonant scores as function of signal-to-noise ratio (SNR) in speech-weighted noise: the abscissa is the SNR and the ordinate is $P_{h|s}$(SNR) (the probability of responding that consonant h was heard given that consonant s was spoken). The intersection of the thick horizontal dashed line at 50% error point and the plotted average error line for each ear define the consonant recognition threshold (CRT) in dB. HI40L has an average CRT (of the 14 CVs) of 0 dB in Fig. 3.1 (b), while the CRTs of individual consonants range from -5 to 10 dB (-5 is an extrapolative estimate). 42

3.3 Stacked bar plots of HI44L at 4 SNRs: (a) quiet, (b) +12 dB, (c) +6 dB , and (d) 0 dB. In each plot, ordinate indicates percent error of individual consonant and height of each bar means total percent error (P_e, %) which is composed of several confusions in different colors. Abscissa is rank-ordered by total percent error of 14 CVs. The PTA for the subject shown on Fig. 3.9(c) (blue-x). For example, the subject has the highest /ga/ error (50%) at 0 dB; 45% of all /ga/ trials at 0 dB are heard as /va/. Average entropy is listed in the upper left corner of each plot, and each bar has a row entropy and error bar. Subject HI44L has no error in quiet, while at 0 dB SNR /ga/ has 9 /va/ confusions out of 20 presentations ($P_{v|g}(0dB) = 0.4$). For /za/, $P_{t|z}(0dB) = 2/9$ and $P_{z|z}(0dB) = 14/18$. 45

3.4 Stacked bar plots of HI46L of 4 SNRs: (a)-(d). In each plot, ordinate indicates percent error of individual consonant and height of each bar means total percent error (P_e, %) which is consisted of several confusions in different colors. Abscissa is rank-ordered by total percent error of 14 CVs. As noise increases from (a) to (d), total P_e is increased and confusions are higher, consisting of more various colors. 48

3.5 Four stacked bar plots of HI36R of 4 SNRs: (a)-(d). In each plot, y-axis indicates percent error of individual consonant and height of each bar means total percent error (P_e, %) which is consisted of several confusions in different colors. X-axis is rank-ordered by total percent error of 14 CVs. Subject is not affected by noise, showing a few consonant error except for /ba/ syllable. As noise increases, /ba/ had higher percent error (100% at +6 and 0 dB) and confusions is also increased from 1.544 to 2.085 (/ba/ row entropy) 50

3.6 Stacked bar plots of HI40L of 4 SNRs: (a)-(d). Subject's consonant perception is affected from +6 dB. /fa/ perception is always confused to /sa/ regardless of SNR. At 0 dB condition, most consonants make error and row entropy of individual consonant is increased up to about 2.5 (/ga/). 51

3.7 Talker Dependence of Exp. II: panels (a)-(d) for correlation between talker I (female, left plot of each panel) and talker II (male, right plot of each panel) of /ga/, /ba/, /pa/, and /sa/ syllables. X-axis indicates SNR and y-axis is percent correct (%). Numbers above 100% line indicate total number of trials at each SNR. 53

3.8 Internal Consistency. Panels (a)-(d) show the correlation between phase I (abscissa) and phase II (ordinate) of four HI ears. Circle means individual consonant and black, blue, turquoise, and pink colors correspond to quiet, +12, +6, 0 dB SNR, respectively. Panels (e)-(h) show percent correct (%) as a function of SNR for the two phases, for the utterances /ba/, /pa/, /va/, /ka/ in HI32R. Numbers above 100% line indicate total number of trials at each SNR. 54

3.9 The different between two CLPs for two HI subjects from Exp. II are shown. The left and right panels are their PTA and CLP, respectively. On the right panels, curves above the horizontal line (0) indicate a left-ear advantage as a function of SNR, and those below the line show a right-ear advantage as a function of SNR. To reduce the clutter, consonants which have less than 20% ear difference are shown as gray lines. Standard errors are also marked on the significant points. Note how panel (b) shows a large /ba/ advantage (between 30-60%) to the left ear. 56

3.10 The different between two CLPs for two HI subjects from Exp. II are shown. The left and right panels are their PTA and CLP, respectively. On the right panels, curves above the horizontal line (0) indicate a left-ear advantage as a function of SNR, and those below the line show a right-ear advantage as a function of SNR. To reduce the clutter, consonants which have less than 20% ear difference are shown as gray lines. Standard errors are also marked on the significant points. Note how panel (b) and (d) show a strong left ear advantage for many CVs. 57

4.1 Comparison of Exps. II (left panels) and IV (right panels) for utterances /ba/, /va/, /ma/, and /fa/ syllables in HI32. Abscissa indicates SNR and ordinate is percent correct (%). Numbers above 100% line indicate total number of presentation trials at each SNR. 60

5.1 The two left panels show PTA results in the HI subjects and the right panels show their consonant loss profiles in left vs. right ears across the 16 consonants. On the right panels, bar graphs present percent error(%) of each consonant in blue for left ear and red for right ear. The gray bars show left ear vs. right ear advantage: above zero shows a right-ear advantage and below shows a left-ear advantage. Error bars indicate 1 standard error (SE): $SE = \sqrt{\frac{p(1-p)}{N}}$ where p is probability correct, N is the number of presentation trials. Even though these subjects have symmetrical hearing loss (a,c), their consonant perception is asymmetrical and is inhomogeneous across consonants (b,d). PTA cannot predict individual HI ears' consonant-loss. *Due to limitation of creating IPA symbols in MATLAB, the consonants, /θa/, /ʃa/, /ða/, and /ʒa/ are displayed as Ta, Sa, Da, and Za, respectively. 63

5.2 The two left panels show PTA results in the HI subjects and the right panels show their consonant loss profiles across the 16 consonants. On the right panels, bar graphs present percent error(%) of each consonant in blue for the first ear and red for the second ear. The gray bars show first ear vs. second ear advantage: above zero shows a second-ear advantage and below shows a first-ear advantage. Error bars indicate 1 standard error (SE). There is a difference in CLP between two different HI subjects having identical PTA (a). The subject with the asymmetrical pure-tone loss (c) does not have an asymmetrical consonant loss profile (d). 64

5.3 The CRT and CLP of HI ears are compared. The left top panel (a) shows the CRT threshold defined as the SNR at 50% average error, for six pairs of ears showing the same CRT: -3, 0, and 4.5 dB SNR. The right top and two bottom panels show plots of consonant-loss difference between two ears as a function of consonants. Bar graphs present percent error of each consonant as blue for one ear and red for the other ear. The gray bars show left ear vs. right ear advantage: above the zero line one ear has a higher error (disadvantage), and below the line the right ear has the disadvantage. Error bars indicate 1 standard error (SE). Note that one ear is much better than the other in some consonants although they have same CRT. More specifically note the /ba/ syllable of (b) (40% higher error in HI36R), the /ʒa/ syllable of (c) (65% better perception in HI40L), /ʒa/ and /ka/ on (d) (i.e., better in /ʒa/ and worse in /ka/ to HI15L). . 66

5.4 Examples of the comparison between pure-tone audiogram (light dashed grey curve) and aided pure-tone threshold (black solid curve) by applying the NAL-R insertion gain to the hearing aids of 6 HI listeners. Each panel represents a different configuration of hearing loss: Flat hearing loss, low-frequency hearing loss, high-frequency hearing loss, ski-slope high-frequency hearing loss, notched hearing loss (or middle-frequency hearing loss), and reverse-notched hearing loss. 68

5.5 Consonant-dependence in applying no NAL-R condition at most comfortable level (MCL) vs. NAL-R amplification condition across the 16 consonants. The three left panels show PTA results in the HI subjects and the middle and right panels show their consonant loss profiles in left and right ears, respectively. On the middle and right panels, bar graphs present percent error of each consonant in light grey for no-amplification condition and dark grey for with-amplification. Green bars (above zero) mean NAL-R positive benefit and red bars (below zero) show negative benefit. Error bars indicate one standard error (SE). Note some consonants improve when applying NAL-R amplification and some do not, showing a consonant-dependence. 69

5.6 Symmetric bilateral hearing loss and asymmetric benefit of NAL-R amplification. The four left panels show PTA results in the HI subjects and the middle and right panels show their consonant loss profiles in left and right ears, respectively. On the middle and right panels, bar graphs present percent error (%) of each consonant in light grey for no-amplification condition and dark grey for with-amplification. Green bars (above zero) mean NAL-R positive benefit and red bars (below zero) show negative benefit. Error bars indicate one standard error (SE). There is a different positive-benefit in NAL-R amplification in left and right ears in four HI subjects despite a symmetric pure-tone hearing loss, showing that their consonant perception is not homogeneous across consonants. 72

5.7 Consonant perception and NAL-R benefit for the subjects who have asymmetric bilateral hearing loss. The three left panels show PTA results in the HI subjects and the middle and right panels show their consonant loss profiles in left and right ears, respectively. On the middle and right panels, bar graphs present percent error (%) of each consonant in light grey for no-amplification condition and dark grey for with-amplification. Green bars (above zero) mean NAL-R positive benefit and red bars (below zero) show negative benefit. Error bars indicate one standard error (SE). First top panels (a,b,c) show positive benefit in most consonants after applying NAL-R amplification for both left and right ears. Middle panels (d,e,f) show negative benefit in most consonants after applying NAL-R amplification for both ears. The third row panels (g,h,i) show positive benefit in most consonants on her left ear, yet negative in most consonants on her right ear. 74

6.1 Graphical summary of frequency vs. time distribution of the distinctive energy event of English consonants, based on (Li *et al.*, 2010, 2011). 77

LIST OF ABBREVIATIONS

AI	Articulation Index
ASA	Auditory Scene Analysis
ANOVA	Analysis of Variance
CLP	Consonant-Loss Profile
CM	Confusion Matrix
CRT	Consonant Recognition Threshold
cs	centiseconds
CV	consonant-vowel
CVC	consonant-vowel-consonant
DRT	Direct Realist Theory
HA	hearing aid
HI	hearing-impaired
HINT	Hearing In Noise Test
HTL	hearing threshold level
LDC	Linguistic Data Consortium
MCL	most comfortable level
MT	Motor Theory
NAL-R	National Acoustic Laboratories-Revised
NH	normal hearing
NZE	non-zero error

PAL PB-50	Psychoacoustic Laboratory Phonetically Balanced monosyllabic word lists
PTA	pure-tone audiogram
QuickSIN	Quick Speech-In-Noise test
REG	real-ear gain
SII	Speech Intelligibility Index
SNR	signal-to-noise ratio
SNHL	sensorineural hearing loss
SPIN-R	Revised Speech Perception In Noise
SRT	speech recognition threshold
STFT	Short-Time Fourier Transform
3DDS	three-dimensional deep search
3TA	3-tone average
VC	vowel-consonant
WRS	word recognition score
ZE	zero error

CHAPTER 1

GENERAL INTRODUCTION

1.1 Statement of Problem

Unlike normal hearing (NH) listeners who have good ability in separating speech sounds from unwanted surrounding noise and have easy conversation, *hearing-impaired* (HI) listeners with *sensorineural hearing loss* (SNHL) have trouble understanding the basic speech sounds (i.e., consonants and vowels) in a noisy environment, even when they are wearing an assistive listening device. The HI listeners, especially with mild-to-moderate SNHL, complain that their hearing aids do not simulate/approach normal speech perception. According to Kochkin (2000) "Why are my hearing aids in the drawer?", about 30% of hearing aid owners do not wear them. Many of the people that Kochkin surveyed reported that their hearing aids have several serious problems: poor benefit, background noise, and poor fit, and that the hearing aids amplified background noises well, but not human speech (Kochkin, 2000).

Although the topic of *how speech perception for the HI population improves* has been debated for more than a half century in clinical audiology, in hearing science, and in the hearing aid industry, it remains an open and unsolved puzzle. On the side of the clinical research, various diagnostic speech perception tests have been developed using nonsense syllables (Dubno and Dirks, 1982; Dubno *et al.*, 1982; Resnick *et al.*, 1975), words (Plomp, 1986; Ross and Lerman, 1970), and sentence materials (Cox *et al.*, 1988, 1987; Kalikow *et al.*, 1977). In hearing science, there has been fundamental approach while modulating timing and/or frequency of speech sounds (Bacon and

Gleitman, 1992; Moore and Skrodzka, 2002) and changing speech cues and features (Erber, 1975). Yet few to none of these methods have been successful in improving HI speech perception. The hearing aid industry has also developed aids for HI speech perception by signal processing techniques, e.g., wide dynamic range compression circuit (Jenstad *et al.*, 1999) and enhanced localization to reduce unwanted noisy sounds (Carhart, 1958; MacKeith and Coles, 1971; Welker *et al.*, 1997). However, professionals in all three fields have not consolidated their efforts into a single approach and have no united system to data for improving speech intelligibility. Furthermore, despite a body of literature reporting a great improvement of the aided HI speech perception, based on the results of clinical measurements, it is still unclear why two people with a similar hearing loss or the same hearing configuration have significantly different abilities in speech understanding (Tremblay *et al.*, 2006).

Here, therefore, we will address five questions that are fundamental to all three fields: (1) "Do the current clinical measurements diagnose HI speech perception accurately?" (2) "Are current fitting methods (e.g., a half-gain rule, NAL-R, and other prescription formulas) effective?" If yes, then (3) "why do these fitting procedures give unsatisfactory information to the hearing aids wearers?", or (4) "why is it that modern hearing aids are not effective, especially in noise?" If not, (5) "do we need a more accurate and alternative measurement of SNHL listener's loss or impairment?" It seems that these questions underlie an unanswered fascinating problem, that is fundamental to both clinical practice and speech perception research. Hence, we need to scrutinize our current clinical procedures for diagnosis of hearing loss and hearing aid fitting.

Fig. 1.1 illustrates the typical clinical procedure that takes place when individuals visit an audiology clinic. Although speech perception research as related to clinical audiology has developed, the diagnostic speech tests used in a clinic are still very limited, in terms of transferring from research to clinic. For example, based on the

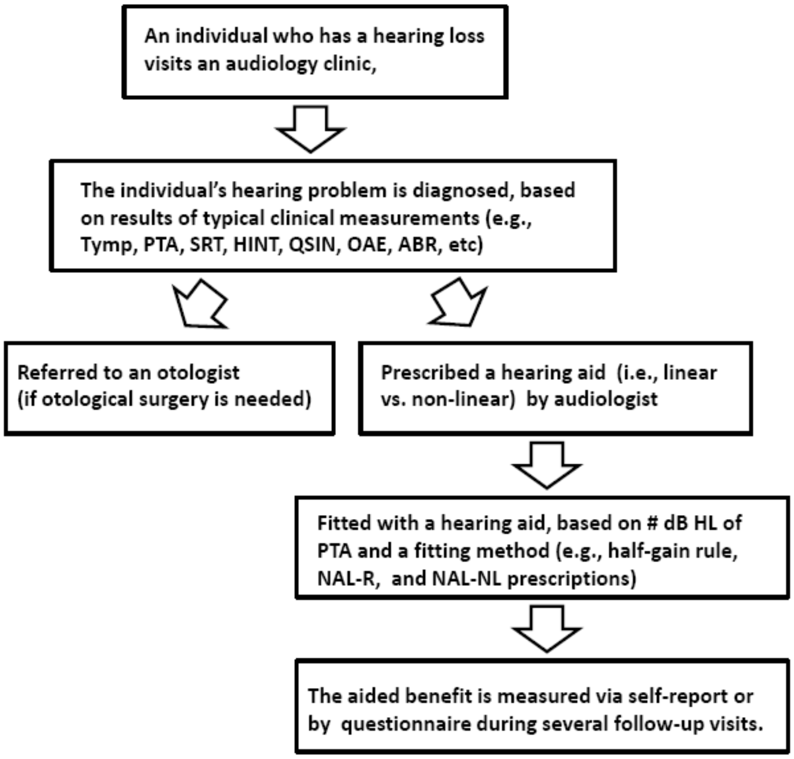


Figure 1.1: A flow chart of the typical clinical procedure for hearing-impaired listener as a process diagram. Abbreviations used are Tymp = Tympanometry; PTA = Pure-Tone Audiogram; SRT = Speech Recognition Threshold, HINT = Hearing-In-Noise Test; QSIN = Quick Speech-In-Noise test; OAE = Otoacustic Emission; ABR = Auditory Brain Response; NAL-R = the revised National Acoustic Laboratories prescriptive formula; NAL-NL = Nonlinear NAL formula.

results of the three most commonly used diagnostic tests, e.g., tympanometry, pure-tone audiometry, and *speech recognition threshold* (SRT), the clinicians typically determine a type, severity, and frequency response of hearing loss. "Type" characterizes the apparent physiological origin of hearing loss as conductive or SNHL. "Severity" is measured in decibels, but may be less precisely categorized as mild, moderate, severe, or profound. "Frequency response" is also measured quantitatively, but may be imprecisely categorized as a flat, low-frequency, or high-frequency hearing loss. In addition, except for two popular speech tests (Hearing-In-Noise Test, or HINT

(Nilsson *et al.*, 1994) and Quick Speech-In-Noise test, or QSIN (Killion *et al.*, 2004)), most measurements using speech materials are not practically accepted in the clinic, due to their being time consuming, complex, or poor in reliability.

Dobie and Sakai addressed common limitations of current clinical tests. They found that the *pure-tone audiogram* (PTA) and *word recognition score* (WRS) are highly correlated, but there is a question as to whether these two predictor variables each explain the variance in self-report about HI listeners' satisfaction with speech perception, or whether the PTA measurement alone is sufficient to predict HI speech perception (Dobie and Sakai, 2001). Dobie and Sakai also discovered a low correlation between current speech tests and self reports of the effect of hearing loss. They suggest that the self-report should be the gold standard. Despite the results of studies like Dobie and Sakai (2001), however clinicians typically still use PTA and WRS for fitting the hearing aid. Fig. 1.1 shows a typical scenario, in which *# dB HL* as a function of testing frequencies, as measured using a PTA, is used for fitting hearing aids to HI patients. The patients then report their hearing aid satisfaction to the clinician, by self-report or a questionnaire in several follow-up visits (Dobie and Sakai, 2001). This dissertation study proposes that the high dissatisfaction with modern hearing aids comes from the averaging scores inherent in PTA and SRT. In other words, existing clinical measurements do not give sufficiently detailed information about the characteristics of the HI listeners' feature loss in speech, to make a useful diagnosis for the hearing aid fitting.

In 2007 and 2008, Phatak and Allen found that the speech perception accuracy of listeners with NH showed significant *consonant dependence*. Although most NH listeners' thresholds are 0 dB HL at all testing frequencies, and thus (we assume) all speech sounds are audible, they perceive the consonants differently (some consonants are more difficult than others). For example, the results for both the experiment having 64 syllables (16 consonants × 4 vowels) for 14 NH listeners in *speech-weighted* *noise* (Phatak and Allen, 2007) and the experiment using 16 syllables (16 consonants × 1 vowel) for 24 NH listeners in *white noise* (Phatak *et al.*, 2008), proved NH consonant-dependence. Although the white noise masked the consonants more uniformly than speech-weighted noise as a function of frequency, those two studies resulted in three NH subgroups of consonant perception: hard, easy, and intermediate groups of sounds. Such findings motivated the present studies.

One year later, Phatak et al. also confirmed that HI listeners have consonant-dependent speech reception accuracy. In their HI experiment, the subjects perceived each consonant with different accuracy, producing either high- or low-error, which indicates that some consonants are more difficult to perceive than others (Phatak *et al.*, 2009). They categorized 26 HI ears into three subgroups according to a level of performance. This dissertation will present findings based on 46 HI ears (of Study I) in a later chapter and will show that our findings support Phatak *et al.* (2009) in several ways: (1) HI listeners have idiosyncratic consonant perception even when they have nearly identical PTA and SRT results; and (2) there even is a significant difference in consonant perception between the left and right ears, even with similar PTAs. However, when the number of subjects was doubled (up to 46 ears), the three-group categorization of sounds seen in Phatak *et al.* (2009) disappeared in our results.

These previous findings suggest the need for a new approach to HI speech perception research, one that is the opposite of the traditional *Articulation Index* (AI) theory proposed by Harvey Fletcher in 1921. Fletcher characterized the information-bearing, frequency dependent regions of speech and modeled nonsense syllable recognition using the *average nonsense phone recognition scores* (Allen, 1996). Fletcher's AI model and theory was highly successful in characterizing the average confusion scores of NH subjects based on a large number of measurements (Allen, 1994). Several variations of the AI model have been used to predict HI speech perception by many researchers

(Brungart, 2001; Dubno *et al.*, 2002, 2003; Pavlovic, 1984; Pavlovic and Studebaker, 1984; Pavlovic *et al.*, 1986), to characterize the *signal-to-noise ratio* (SNR) loss (Killion and Christensen, 1998), and even to fit the hearing aid (Rankovic, 1991). Although the AI theory has done an excellent job in its goal to characterize mean confusion scores, the mean score cannot explain the individual utterance recognition score (Singh and Allen, 2011) or predict each consonant's confusions (Li *et al.*, 2010), because the scores across utterances are idiosyncratic.

This dissertation proposes that knowing the idiosyncratic, consonant-dependent perceptual accuracy of each HI ear should be useful, when diagnosing hearing loss or fitting a hearing aid. We therefore propose when the hearing aid is fitted by the PTA or the mean score, aided HI speech perception will not improve because of the failure to consider the subject-dependent consonant error.

1.2 Literature Review

1.2.1 The Theory and Models of Speech Perception

Over the last 150 years, many theories and models of human speech perception have been proposed and debated. Four theories that have been extensively discussed in the literature are briefly reviewed. In addition, a new concept recently proposed by Singh and Allen (2011), *binary speech masking*, a decision-making on NH speech perception, will be introduced and adapted as the foundation of to our Studies II and IV.

Motor Theory (MT) proposed by Liberman and colleagues in the 1960s theorized that the relation between speech perception and co-articulation results from the coordinated movement of the tongue, lips, and vocal folds (Liberman *et al.*, 1967; Liberman and Mattingly, 1985). The foundation of MT is a one-to-one mapping between individual phoneme and acoustic features (when perceived) and articulation (when produced). For example, when perceiving the /p/ sound, the listener also imagines a speaker's closed lips and a burst release, which necessarily are needed to produce a labial sound. MT claims that the objective of speech perception is an articulatory event rather than acoustic or auditory events, and that to perceive the speech sound requires a special mental module for the recovery of an intended gesture from the acoustic signal of the speech stimuli. However, MT could be an explanation only for the human listener/speaker, not for nonhuman research that uses birds and animals (Kuhl and Miller, 1975, 1978). This line of reasoning has lead to the many ad hoc arguments against MT, which typically debate old ideas and data.

Another theoretical construct related to auditory perception, thorough unrelated to speech perception, is *Auditory Scene Analysis* (ASA). In 1971, Bregman and Campell proposed the ASA theory whereby the human auditory system perceptually organizes sounds into meaningful elements, while needing two stages (i.e., the primitive and the schematic stages) to perceive these elements (Bregman and Campell, 1971). In the first stage, a listener groups energetic events based on proximity in frequency and based on similarity in the change of sound. In the second stage, the listener starts to apply the analysis knowledge learned from the first stage.

Carol Fowler, a Liberman colleague, modified the MT, and proposed the *Direct Realist Theory* (DRT) in 1984 in order to account for the results of many studies related to both birds and humans (Fowler, 1984). The DRT retains the MT premise, that to perceive speech sounds is to perceive the movements of the vocal tract, that structure the acoustic signal, rather than abstract phonemes or events that are causally antecedent to the intended movement or gesture (Fowler, 1984). As with MT, the AI is not considered in DRT.

Marslen-Wilson and Tyler in the late 1980s proposed the *COHORT model*, which was a computational model of human spoken word recognition. According to the COHORT model, a processing of speech perception had three basic functions: Access, selection, and integration (Marslen-Wilson and Tyler, 1980; Marslen-Wilson, 1987). Each function represented a lexical form at the lower level, in order to discriminate the lexical inputs at the next stage, and then combine the syntactic and semantic information at the higher level (called *bottom-up processing*). Among other findings, this model was able to explain speech shadowing, in which the listener correctly repeats what they heard while listening to a sentence. After a revision of early versions of the model (later, resulting in the TRACE Model of spoken word recognition developed by McClelland and Elman in 1986), Marslen-Wilson and Tyler could explain the "bottom-up procedure" of human speech perception when the speech stimulus has a short delay. However, they could not prove the "top-down procedure" of the speech perception with the model when new information having many complicated contexts was presented. Again there is no mention of the AI.

These many persuasive arguments and theories for NH speech perception have continued over a long period. Yet very little is still known about the characteristics and nature of HI speech perception. Research in *how the listeners with hearing loss perceive speech* has not advanced. In their attempts in understanding HI speech perception better, Singh and Allen (2011) explored NH speech perception at a consonant and individual utterance level, and discovered a pinpoint masking phenomenon that they called *binary speech masking* in the NH listeners. Unlike the other studies of human consonant perception, described by the average percentage score across all consonants, Singh and Allen analyzed the error patterns of each individual utterance, for 6 stop consonants and 4 vowels above -10 dB SNR, which they denoted the *low-noise environments*. This terminology reflected the observation that at -2 dB SNR and in quiet, the error was independent of SNR (i.e., the consonant scores saturated at 100% recognition). Most of these utterances (62.8%) had *zero errors* (ZE), and the remaining utterances (37.2%) had *non-zero errors* (NZE) (Singh and Allen, 2011). These latter consisted of three groups: Low (15.8% of the speech sounds), medium (10.7%), and high (10.7%) error groups. The high error group of stop sounds accounted for almost all of the true errors. The low error group resulted from a single random error repeated by a single listener. Thus, 62.8+15.8 = 78.6% had either 'zero' or 'one random error' in about 200 trials. When NH listeners heard masked speech at SNRs below -2 dB, they displayed a *binary decision process*: the error rate for any given utterance sample went from zero to chance (15/16) over a 6 dB SNR range (Singh and Allen, 2011). This same experimental method will be applied to HI subjects in Studies II and IV in this dissertation. Thus, we will only use sounds that NH listeners can identify 100% of the time at SNRs $\geq$ -2dB SNR. Most HI ears experience a high consonant loss (Fig. 3.1), therefore we predict that HI ears will suffer nonzero error rates for the same utterance samples.

1.2.2 Synthetic Speech Cue Research

Starting around 1950, a number of speech scientists from Haskin lab developed a form of speech synthesizer, denoted as the *Pattern Playback* (Slaney, 1995), which they used in several classic studies, which demonstrated that speech was composed of smaller building blocks of narrow band bursts and resonances (Delattre *et al.*, 1955; Liberman *et al.*, 1957). These studies have had a major impact on speech research. Speech synthesis thereafter became a standard method for feature analysis, used in the search of acoustic correlate for stops (Blumstein *et al.*, 1977), fricatives (Heinz and Stevens, 1961; Hughes and Halle, 1956), nasals (Liberman, 1957), as well as distinctive and articulatory features (Blumstein and Stevens, 1979, 1980). An even *more* stylized approach was taken by Remez et al. (1981) to generate highly unintelligible “sine-wave” speech, which was used to study the ability of humans to perceive speech information in signals that only minimally resemble natural speech.

The status quo is now rather confusing, in that many researchers accept that stop consonants are identified by the bursts and transitions (Allen and Li, 2009; Blumstein and Stevens, 1980; Cooper *et al.*, 1952; Heil, 2003; Li *et al.*, 2010), yet they still argue that low-frequency modulations are the key to understanding speech perception (Dau *et al.*, 1997; Drullman *et al.*, 1994; Shannon *et al.*, 1995). All fail to point out that the two views are in conflict.

The argument in favor of the speech-synthesis method is that such features can be precisely controlled. However, the major disadvantage of synthetic speech is that it requires a precise hypothesis about of the cues being sought. Unknown cues can not be made the subject of a hypothesis. Incomplete and inaccurate knowledge about the acoustic cues has led to synthetic speech of low quality; thus it is common that such speech sounds are unnatural and even barely intelligible, which by itself is strong evidence that the critical cues for the perception of target speech sounds are poorly represented. For those cases, an important question is: “How close are the synthetic speech cues to those of natural speech?" (Li and Allen, 2011).

Another key issue is the *natural variability of speech cues* (Hazan and Rosen, 1991) due to the talker, accent, and masking noise, most of which are well beyond the reach of the state-of-the-art speech synthesis technology. To answer questions such as: "Why are /ba/s from some talkers confused with /va/, while others are confused with /ga/?" or "What makes one speech sound more robust in noise than another?", it is necessary to study the acoustic cues of naturally produced speech, not artificially synthesized speech for HI as well as NH listeners.

1.2.3 Natural Speech Cue Research

Although speech perception research is an experimental science, based on psychoacoustic measurement, new insights will only come when carefully controlled experiments are combined with a mathematical analysis of communication (Li and Allen, 2011). However, up to the present, no studies other than Li and Allen (2011) and Hazan and Simpson (1998) have identified invariant speech cues in *natural speech*, which could be manipulated.

Identification of Consonant Cues

To address the large variability of natural speech due to talker effects (e.g., gender, accent, clear articulation) and to explore the perceptual cues of consonant sounds, Li and Allen have developed a systematic psychoacoustic method, called the *three-dimensional deep search* (3DDS), which was shown to work with 16 consonants and 3 vowels in NH listeners (Li and Allen, 2011). Unlike conventional methods using synthetic speech (Cooper *et al.*, 1952), based on a priori hypothese about the speech cues, followed by listener verification, the 3DDS method directly measured the contribution of each sub-component of natural speech by time truncating, high- and low-pass filtering, and masking the speech with noise. The plosive consonants (e.g., /p,t,k,θ,b,d,ð/+/a/) had a well-defined frequency and timing, relative to the onset of the vowel (Li and Allen, 2009; Li *et al.*, 2010). Fricatives (e.g., /s,ʃ,z,ʒ/+/a/) and nasals (e.g., /m,n/+/a/) were determined by the consonant's frication driven resonance center frequency and duration. As the next step, the researchers manipulated these bursts and frication resonances, causing one speech sound to morph to another in a predictable way. These manipulations have now been proven effective in modifying the consonants, not only for nonsense syllables, but also for meaningful words and sentences (Li *et al.*, 2010). They still need to prove the 3DDS method with HI consonant perception.

Three Dimensional Deep Search (3DDS)

According to Li and Allen (2011), the core idea behind 3DDS is to remove a certain time-frequency region of a speech sound and then assess the importance of the removed component from the splattering of confusions. Their 3DDS approach has been found to be a highly quantitative method for identifying cues, which we will rely on for HI studies in the future.

In order to measure the distribution of speech information along the time, frequency, and amplitude dimensions, three independent psycho-acoustical truncation experiments were performed on each speech token: speech sounds were (1) truncated in time, (2) high- and low-pass filtered, and (3) masked with white noise. Each modified sound stimulus was presented to a battery of 20 NH listeners, using randomized trials, across utterances and conditions (Allen and Li, 2009). When an acoustic event was removed by one of these three modifications, the recognition scores of listeners were found to drop abruptly. The experimental results were then presented as confusion patterns, which display the probabilities of possible responses (the target and competing sounds) as a function of the experimental conditions (i.e., truncation time, cutoff frequency, and SNR).

Conflicting Cues

In 2010, Li and Allen's analysis of the Linguistic Data Consortium (LDC) database (Fousek *et al.*, 2004) indicated that due to the physical limitations of the human speech articulator, most stop consonants (e.g., /p,t,k,b,d,g/+/a/) contain combinations of consonant cues that lead to confusions in speech perception under adverse circumstances. It is difficult to naturally produce "ideal" speech sounds, containing only the desired cues. Thus, Li and Allen went on to identify these *conflicting cues*. For example, a talker intends to produce a /ka/ syllable, and listeners report hearing /ka/ 100% of the time. However, the same /ka/ syllable contains both a high-frequency burst above 4 kHz (indicative of a /ta/ production) and a low-frequency burst below 1 kHz (indicative of a /pa/ production). When these two conflicting cues, /ta/ and /pa/, are digitally removed, the speech remains perceptually indistinguishable (Li *et al.*, 2010). In Li and Allen's experiments, the listeners reported a robust /ka/ because the mid-frequency /ka/ burst perceptually overpowered the two interfering cues. Exactly how or why this happens is not yet understood, but it clearly has to do with nonlinear neural processing of the auditory nerve signal. Although such an interesting finding had never been reported in the previous literature, much more needs to be done to quantify how conflicting and primary cues interact in NH listeners (Li and Allen, 2011). Obviously perception by HI listeners must be much more complex.

Manipulation of Consonant Cues

It is widely accepted that human speech perception is a complex multi-level process, where the integration of events is governed by lexical, morphological, syntactic, and semantic context (Allen, 2005; McClelland and Elman, 1986). In order to manipulate the consonants in natural speech, it is convenient to start from nonsense syllables so than the high level contextual constraints on speech perception are maximally controlled (Allen, 2005; Li *et al.*, 2010).

Once the speech cues of each utterance are identified by the 3DDS, they can be easily verified by manipulation (Li *et al.*, 2010; Li and Allen, 2011). As one supporting example, Li and Allen (2010) selected a /ka/ sound from one female talker, using the *Short-Time Fourier Transform* (STFT) method (Allen and Rabiner, 1977). They modified the /ka/ sound by varying the relative levels of three speech cues. When the 1.4∼2 kHz burst [called region 1] was removed, the percept of /ka/ dramatically dropped, and listeners reported either /pa/ or /ta/. When both bursts for /ta/ and /ka/ were removed [regions 2 and 3, respectively], the sound was robustly perceived as /pa/. Boosting the low-frequency burst within 0.5∼0.7 kHz [region 3] strengthened the initial aspiration, making the percept a clearly articulated /pa/. Removing both regions 1 and 3 led to a clear /ta/. Competing cues typically lead to ambiguous stimuli, which they referred to as a form of priming, defined as an *auditory illusion* where prior expectation of the perceived sound affects the sound reported (Li *et al.*, 2010).

Unlike the stop consonants represented by a compact initial burst, the fricatives are characterized by a narrow-band noise cue with varied duration and bandwidth (Li *et al.*, 2010). As an example of such a manipulation, Li and Allen proposed an original sound heard by all listeners as a solid /ʃa/ from one female talker and its perceptual cue which ranged from 13∼38 centiseconds (cs) in time and about 2∼8 kHz in frequency. They demonstrated the manipulation procedure in three steps: (1) High-pass filtering with a cutoff of 4 kHz morphed (or changed) the sound into a robust /sa/, (2) shrinking the duration by 2/3 transformed the sound into a /tʃa/, and (3) combining both manipulations caused most listeners to report /za/. Removing the whole noise patch resulted in /ða/, which can be made robust by amplifying the residual high-frequency burst. Such manipulations need to be tried with HI listeners, who have lost their hearing in specific frequency regions (Allen and Li, 2009; Li and

Allen, 2011).

Context

Of course the difference between words and so-called nonsense syllables is that words have meaning. This semantic constraint (i.e., context) has a major impact on the recognition score (Allen, 2005; Boothroyd and Nittrouer, 1988; Bronkhorst *et al.*, 1993). Despite such powerful context effects, the consonant identity can still be morphed using the technique described above. To demonstrate, Li and Allen chose several words from their speech database and applied the speech-feature modified method (Li *et al.*, 2010; Li and Allen, 2011). As an example, two words /take/ and /peach/ were extracted from a sentence. The /t/ and /k/ were characterized respectively by a high-frequency burst before the vowel and a mid-frequency burst after the vowel. Switching the time location of the two cues morphed the word /take/ into a perceived /kate/. Once the duration between the /p/ burst and the onset of sonorance was shortened, /peach/ robustly morphs to /beach/. Given *a priori* knowledge of specific speech cues, Li and Allen morphed the percept of consonants in natural speech through the manipulation of speech cues in CV syllables, words, and sentences.

The nature of the confusions of individual consonants has yet to be fully analyzed. Both the AI and the *Speech Intelligibility Index* (SII) have been shown to be inaccurate in predicting HI speech perception (Ching *et al.*, 1998). Due to a lack of information about speech cues, no HI speech perception studies have considered individual consonants. Zurek and Delhorne (1987) proposed that simulation of cochlear loss on NH listeners using masking noise showed no consistent difference between the HI listeners and masked NH listeners in terms of average speech intelligibility. However, we have shown the need to characterized loss of HI speech perception based on individual utterances. We would like to show that by analyzing the detail confusion made with each HI ear, we can accurately diagnose the ears cochlear loss. Our preliminary analysis support this hypothesis. We suspect, based on a preliminary analysis, that studying the detailed nature of the confusions will give us a much better understanding of the subject's cochlear loss.

1.3 Measurements of Speech Perception in Hearing-Impairment

1.3.1 Current Clinical Measurements

In this section, we review the pros and cons of three popular clinical measurements commonly used to establish speech perception ability in HI listeners.

Pure-tone Audiogram (PTA)

Pure-tone audiometry is ubiquitously used to measure hearing sensitivity, to determine the degree, type, and configuration of an individual's hearing loss (from 0.125 to 8 kHz in one-octave steps), and to establish either middle-ear or cochlear/auditory nerve damage for both air- and bone-conduction thresholds (Brandy, 2002). This measurement is fast, easy to use, thus widely accepted (Smoorenburg, 1992).

However, audiometry does *not* directly evaluate the ability of the HI listener to perceive speech sounds. In fact, it is widely accepted that the PTA correlates poorly with HI speech perception (Carhart, 1946; Smoorenburg, 1992). Many studies have reported that for listeners with moderate-to-severe SNHL, there is *no correlation* between hearing threshold and speech perception, while others report a partial positive correlation for listeners with normal to mild SNHL (Festen and Plomp, 1986; Fletcher, 1950; Smoorenburg, 1992). We shall show that while an elevated threshold does predict that there will be some speech loss, it gives no diagnostic information as to the nature of that speech loss.

Many studies have attempted to develop predictions of a listener's ability to understand speech on the basis of his pure-tone sensitivity. For example, Fletcher (1950) and later Smoorenburg (1992) developed a formula for predicting the HI listener's ability to perceive speech from the three-frequency average of hearing thresholds at the most important frequencies (i.e., 3-tone average (3TA) of 0.5, 1, and 2 kHz). They found that there was a very large across-subject variance, that depends on audiometric configuration. In particular, the 3TA had much lower (better) thresholds than speech scores for a non-flat audiogram (e.g., high-frequency ski-slope hearing loss) (Carhart, 1946; Fletcher, 1950). The fact that there is such loose relation between the PTA and speech perception has serious clinical ramifications.

Speech Recognition Threshold (SRT)

The SRT was introduced by Plomp (1986), who defined it as the signal-to-noise ratio (SNR) at which the listener achieves 50% accuracy for recognizing syllables, words, or sentences (Plomp, 1986). The SRT has been widely accepted, due to its convenience and speed, and has become a preeminent "speech test." While distinct from pure-tone audiometry, it clinically correlates well with PTA in quiet (Brandy, 2002; Dobie and Sakai, 2001).

The SRT has three serious limitations. *First*, this measure evaluates a listener's speech threshold, not the ability to recognize speech. Simply said, it is a wide-band threshold test using speech, instead of narrow-band tones, quantified via a VU meter in 5-dB steps (Brandy, 2002). Like the PTA, the SRT has equally limited ability to predict the listener's speech recognition ability. The problem of HI speech perception is not the deficit in detection, but rather poor recognition (Turner *et al.*, 1992). In this dissertation, we shall prove that detection is a necessary, but not a sufficient condition, for consonant recognition.

Second, the SRT uses 20 homogeneous *spondee words* (with doubly-stressed meaningful syllables; e.g., air-plane, birth-day, cow-boy) having high context, because tests based on spondee words are easier and faster to administer than those based on sentences (Brandy, 2002; Carhart, 1946). It is a problem that when the spondee words are used, patients say what they *guess*, not what they actually perceive. As a result, *high context* creates a bias, that raises the score. Thus due to context, the spondee test is *not* a sensitive measure of the speech perception, as it depends on the language skill and performance ability of the patient. It seems safe to say that little or no information on individual phone scores can be inferred from the SRT, at least not in a reliable way.

Third, the SRT considers only average speech scores instead of focusing on individual consonant scores. Being an average measure, it ignores valuable information about what a listener hears, that is, detailed consonant articulation scores that contain essential, even critical information about acoustic cues of the speech stimuli that the HI ear can or cannot hear. Averaged scores remove not only the wide variance of speech perception, but also the key characteristics of hearing loss. By not recording the errors of maximum entropy consonants, as Fletcher and Galt (1950); French and Steinberg (1947); Miller and Nicely (1955) did, we are losing out on our main opportunity to understand the cochlear loss in the HI ear.

Word/Sentence Tests

Apart from the PTA and SRT measurements, various word/sentence tests have been used to diagnose the degree of impairment and to evaluate the benefits of hearing aids (HAs). These tests have become increasingly popular over the years, in part because standardized versions have become available, such as the Psychoacoustic Laboratory Phonetically Balanced monosyllabic word lists (PAL PB-50) (Egan, 1948), the Hearing In Noise Test (HINT) (Nilsson *et al.*, 1994), the Revised Speech Perception In

Noise (SPIN-R) (Bilger *et al.*, 1984), and the Quick Speech-In-Noise test (QuickSIN) (Killion *et al.*, 2004).

Although the tests all differ slightly in composition, research shows that a common advantage of these tests is to simulate everyday listening conditions which are realistic for measuring the speech perception ability of HI listeners (Kalikow *et al.*, 1977; Killion *et al.*, 2004). However, these tests fail to fully reflect HI speech perception in terms of the acoustic and speech cues, because a contextual bias is inherent in these word/sentence tests (Miller *et al.*, 1951; Phatak *et al.*, 2009). Nor is this even desirable. Boothroyd clearly demonstrated that hearing-impaired listeners decode CVCs based on both direct sensory evidence and indirect *contextual* evidence as they decode the speech sound (Boothroyd, 1994). Thus we must separate our measures of consonant perception from the contextual effect. Versfeld et al. (2000) also insisted that redundancy in speech makes hearing-impaired listeners' perceptual scores improve more than one would predict from their hearing loss. As with the SRT, familiar words or topics make it even easier to understand a conversation, whereas new or unfamiliar ones make it more difficult (Connine *et al.*, 1990; Miller *et al.*, 1951). Of course, the contextual linguistic skills are essential and natural in communication, but they are not appropriate in a speech hearing test. Since these features allow the HI listeners to guess the target words (Boothroyd and Nittrouer, 1988; Bronkhorst *et al.*, 1993), the test scores *do not address listeners' core and unique individual consonant errors* (Pichora-Fuller *et al.*, 1995). This observation is further supported by the Phatak *et al.* (2009) study, which found a poor correlation between the results of consonant-loss and the QuickSIN test in 26 hearing-impaired listeners.

In summary, none of these three current clinical tests provides detailed feedback that can predict or identify an individual HI listener's consonant perception loss, as demonstrated by the *Consonant-Loss Profile* (CLP), over a set of 16 English consonants, the likelihood of misperceiving individual consonant (Phatak *et al.*, 2009).

We view this as a serious weakness of these popular clinical measures. The CLP is an alternative that we believe overcomes all these weaknesses.

1.3.2 Non-clinical Measurements

As reviewed by Allen (2005), the *Articulation Index* (AI) and the *Confusion Matrix* (CM) are important measurements of speech perception.

Articulation Index (AI)

In 1921, Harvey Fletcher created the AI, which is used in the prediction of the average phone error (Fletcher, 1950; French and Steinberg, 1947). Although Fletcher revised the calculation for clinical application, his revised method has not been extensively used in practice because the AI provides no diagnostic or frequency-dependent information (Allen, 2005). In addition, the complexity of the AI led to its disuse in clinic settings although it can be useful in choosing the gain of a hearing aid (Souza *et al.*, 2000). Recently, Singh and Allen re-visited the AI, computing an empirical average score,and showed that the average phone score as a function of SNR is very poorly correlated to *individual* consonant errors (Singh and Allen, 2011).

Confusion Matrix (CM)

In 1955, Miller and Nicely developed the consonant CM, which is a useful quantitative and visual representation, displaying the stimulus versus the response score in a table (Miller and Nicely, 1955). The CM allows for greater understanding of an individual's CLP (i.e., over a set of the 16 English consonants, the likelihood of misperceiving each consonant), because it gives detailed information about consonant confusions - that is, (1) which sounds an HI listener can or cannot hear (i.e., diagonal entries) and (2) which sounds are confused with other sounds (off-diagonal entries) (Han *et al.*, 2011a,b). Nevertheless, Miller and Nicely (1955)'s CM method has clinical shortcomings because it is complex, time consuming, and difficult to interpret. In 1976, Bilger and Wang studied an average consonant CM measure in 22 SNHL patients, using a combination of 16 consonants and 3 vowels averaged across SNRs (Bilger and Wang, 1976). They reported that HI listeners made specific consonant errors, with error patterns that depend on the degree and configuration of the hearing loss, as well as the level of noise. While measuring CV and VC confusions, they only reported mean scores (% correct) of 4 CV subsets. Their findings strongly suggest the need for further research into the detailed characteristics of consonant perception error, which are idiosyncratic across HI listener (Han *et al.*, 2011a,b; Phatak *et al.*, 2009).

1.4 Purpose of the Study and Hypothesis

The main goal of this dissertation is to gain precise insight into HI consonant perception and to provide a bridge between speech perception research for clinical practice and hearing aid fitting. We understand that there are many unresolved problems that HI listeners have when trying to recognize speech in noisy surroundings, even when wearing the hearing aid. Therefore, we would confirm an advanced speech perception diagnostic measurement for the HI patient, thereby developing greatly improved compensation for the HI listener's speech/consonant loss. Our approach is to generate the consonant confusion matrix and use a detailed graphical analysis (i.e., stacked bar plots showing both diagonals and off-diagonal entries) - at the level of individual consonant, signal-to-noise ratio (SNR), and individual HI ear. Our tasks will be addressed sequentially in Studies I, II, III, and IV.

1.4.1 Study I: Consonant-Loss Profile (CLP) in Hearing-Impaired Listeners

In Study I, we measured individual consonant error to quantify how each SNHL listener confuses the consonants. Next, we compared these consonant percent errors (%) to the current and commonly used clinical measurements (e.g., PTA and SRT) to determine the clinical power of a confusion.

Two key hypotheses were posed: (1) When a HI listener misses resolving a solitary acoustic cue (e.g., voice onset time, presence or duration of a burst, etc.), the result is a high error rate (Pe, %) for only a few consonants. This measurement is defined in the *Consonant-Loss Profile* (CLP), as quantified by a small but significant subset of consonant errors, unique to each ear (i.e., defined by the diagonal entries of the consonant confusion matrix); and (2) Neither the PTA nor SRT measurements can quantify such a unique profile, because all average measures do not parse out perceptional differences at the consonant level. Consequently, hearing aids fitted on the basis of the PTA or SRT necessarily provide less benefit than those fitted on the basis of a small number of high-error consonants, identified by the CLP. Only with such detailed idiosyncratic knowledge, based on speech feature loss unique to that ear, can we hope to proceed with the most beneficial fitting of modern hearing aids.

1.4.2 Study II: Verification of Consonant Confusion Patterns in Hearing-Impaired Listeners and Test Reliability

There was a specific purpose of Study II: To repeat Exp. I on a smaller set of listeners, with a much greater number of trials, in order to increase the number of utterances per syllable up to 20 for the purpose of improved statistical power. Increased statistical power should allow us to verify the test-retest reliability of the CV syllable scores, while calculating internal consistency (e.g., correlation between sessions, or between phases I and II). Study II also characterized speech recognition errors of every SNHL listener by identifying missing critical features of misheard consonants as a function of SNR, focusing on not only the error rate (%) of the target consonant, but also confusions with neighboring consonants. Hypotheses are the same as in Study I, with the additional hypothesis that (3) the results show good consistency.

1.4.3 Study III: Effect of NAL-R Amplification on Consonant-Loss of Hearing-Impaired Listeners

The dissertation also investigated whether or not NAL-R amplification could positively benefit the speech perception of each SNHL listener at the consonant level. It is possible that there is no net benefit of NAL-R amplification because some consonants are distorted by amplification, while others achieve a positive amplification benefit. In particular, Study III sought to find an answer to the question: Why do HI listeners have trouble listening to speech after being fitted with a hearing aid? We could expect that (1) NAL-R amplification does not offer a full positive benefit to all 16 English consonants; some consonants improve and some do not, because of idiosyncratic consonant-dependence in many HI ears. Based on consonant confusion, this dissertation will address *why* HI listeners are not fully satisfied with their amplification. We further observed that (2) the benefits of the NAL-R amplification are also idiosyncratic for each HI listener. However, we show a low correlation between NAL-R benefit and pure-tone threshold and configuration (or hearing loss pattern). Our results suggest that we need an alternative fitting method in order to take advantage of the large individual differences across listeners, thus to enhance the speech perception of those HI listeners who do not receive a fully positive amplification benefit from NAL-R.

1.4.4 Study IV: Verification of Consonant Confusion Patterns with Amplification Condition

There are two purposes of Study IV: (1) To find the unique errors of HI consonants in the *aided condition* by concentrating on only ZE utterances (in NH listeners), while increasing the number of utterances per syllable up to 20; and (2) to compare the results to Study II (of flat gain) to ascertain how much HI error pattern would be changed after applying a frequency specific amplification.

First, we anticipated that although overall consonant error could be reduced when applying the NAL-R amplification formula, some consonants would exhibit higher error rates due to inappropriate amplification (based on PTA). In fact, we find that some consonant error rates decrease, while others increase, following NAL-R amplification. Such changes are large individual differences across the HI ear, which we shall show in Studies I and II. Second, we show that HI listeners sometimes find it difficult to select their response among several competing and confusable consonants, given amplification. We suggest it may be this increased uncertainty that makes them uncomfortable when listening to the speech with a hearing aid. In other words, to determine whether the HI listeners receive a positive or negative benefit depends on the confusion of consonants, even though the number of and degree of error for the lost consonants is very small. We extend this reasoning by suggesting that, to reduce the number of confusable consonants, an entropy measure should be adopted as criterion for setting the amplification, in place of the average score.

CHAPTER 2

METHODOLOGY

This dissertation consists of 4 studies (i.e., experiments as described in table 2.1).

2.1 Experiment I

2.1.1 Subjects

From July 2009 to October 2009, twenty-seven HI subjects (17 females and 10 males) were recruited from the Urbana-Champaign community. All subjects were native speakers of American English and all were paid for their participation. They ranged in age from 21 to 88 years (mean = 54.96 years, SD = 20.28, see the Appendix A, Table A.1). Subjects were chosen based on normal middle-ear status (type A tympanogram) and mild-to-moderate SNHL at 3TA (3-tone average in hearing threshold at 0.5, 1, and 2 kHz). Informed consent was obtained and approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign.

Table 2.1: Table summary of the four experimental designs used in the current study.

	3 kinds of utterances (N=12) 6 SNRs, 16 CVs	ZE utterances (N=20) 4 SNRs, 14 CVs
No NAL-R condition	Exp. I	Exp. II
NAL-R condition	Exp. III	Exp. IV

The etiologies of individual subjects varied, in terms of the degree and configuration of hearing loss. Of the 27 subjects, 21 had symmetrical bilateral, 4 had asymmetrical bilateral, and 2 had unilateral hearing loss. A total of 48 HI ears were selected for Exp. I. Of these, 10 ears had flat audiograms, with 3 mild, 4 mild-to-moderate, and 3 moderate SNHL. Another 16 ears showed high-frequency SNHL varying in the degree of impairment, with 8 mild, 6 moderate, and 2 moderate-to-severe in hearing loss. A mild-to-moderate high frequency SNHL was present in 11 ears, with a ski-slope loss at either 1 or 2 kHz. The following atypical configurations were also included: 2 ears with low-frequency hearing loss, 2 with cookie-bite (middle-frequency) hearing loss, 3 with reversed cookie-bite (low- and high-frequencies) hearing loss, and 4 with mild hearing loss accompanied by a notch at 4 kHz (see the Appendix A, Table A.1 to see subjects' pure-tone threshold).

2.1.2 Speech Stimuli

Isolated English *consonant-vowel* (CV) syllables were chosen from the Linguistic Data Consortium (LDC) 2205S22 database (Fousek *et al.*, 2004), spoken by eighteen native speakers of American-English. The CV syllables consisted of sixteen consonants (six stops /p, b, t, d, k, g/, eight fricatives /f, v, s, ʃ, z, ʒ, ð, θ/, and two nasals /m, n/) followed by the /a/ vowel (Miller and Nicely, 1955). All stimuli used were digitally recorded at a sampling rate of 16 kHz. They were presented monaurally in quiet and at five different SNRs (+12, +6, 0, -6, -12 dB) in speech-weighted noise. The presentation level of the syllables was set to the subject's *most comfortable level* (MCL) initially, and then adjusted so that the CVs were equally loud independent of SNR. A specific overall attenuator setting (i.e., 0, +10, +20 dB) was maintained for each listener throughout the experiment, while minor variations in intensity (+3 to -3 dB) were made via numerical scaling of a sound card.

Exp. I was intentionally designed to include two low, two medium, and two high error utterances (a total of six different utterances per syllable, provided in Table 2.2), in order to create a more realistic listening situation. In retrospect, this turned out to be a poor strategy, since 66% of the utterances were not robust (i.e., NH ears also had some errors when listening to these utterances). Once we evaluated the seriousness of these confusions, we designed Exps. II and IV to include only "zero-error (ZE)" tokens, on the basis of averaged data from 10 NH listeners (Singh and Allen, 2011).

Table 2.2: Example of 6 different utterances per syllable used in Exps. I and III

pa	f103	f106	f109	m104	m114	m118
ta	f105	f106	f108	m104	m112	m115
ka	f103	f105	f119	m111	m114	m118
fa	f101	f103	f105	m111	m112	m117
θa	f108	f109	f113	m102	m112	m115
sa	f108	f109	f113	m111	m112	m117
ʃa	f103	f106	f109	m111	m115	m118
ba	f101	f105	f119	m107	m111	m118
da	f103	f119	m104	m111	m115	m118
ga	f108	f109	f119	m104	m111	m112
va	f103	f105	f108	m104	m111	m120
ða	f103	f108	f119	m102	m112	m120
za	f105	f108	f109	m104	m118	m120
ʒa	f103	f108	m107	m114	m117	m118
ma	f101	f103	f105	m102	m115	m118
na	f101	f109	f113	m102	m112	m120

2.1.3 Procedure

The test procedures for the CV measurements were very similar to those used in a previous study by Phatak *et al.* (2009). All subjects had one practice session consisting of 10 syllables in quiet to familiarize each subject with the experiment. Subjects were asked to identify each presented consonant of the CV syllable, by selecting 1 of 16 software buttons on a computer screen, each labeled with an individual consonant sound. A 'noise only' button was available for the subjects to specify if they heard *only noise*. A pronunciation for each consonant was provided using an example word below its button to avoid possible confusions from any orthographic similarity between consonants. The subjects were allowed to hear each utterance a maximum of 3 times before making their decision. Once a response was entered, the next syllable was automatically presented after a short pause. The subjects were tested in one session, but they were asked to take several breaks. The experiment took a total of 1 to 1.5 hours per ear.

Each syllable presentation was randomized with respect to consonants and speakers, but not across SNRs. The test proceeded from the *easiest* to *the most difficult* noise conditions - quiet first, followed by +12 to -12 dB SNR. This was done in order to gradually increase the difficulty from the onset, so that subjects were not pushed beyond their limits in terms of performance level. In our pilot studies, we found that when the noise levels were not randomized, the performance at a given SNR improved, which is an additional benefit to this procedure.

A maximum of 1152 trials were presented (16 consonants × 6 utterances × 2 presentations × 6 different noise conditions) to every subject. When the score was less than or equal to 3/16 (18.75%, or three times chance) for any given consonant, the consonant was not presented at subsequent (lower) SNRs.

2.2 Experiment II

2.2.1 Subjects

Seventeen HI ears of Exp. I were tested for Exp. II from April 2010 to May 2010. Each participant again passed a middle-ear examination and was confirmed to have the same hearing level (HL), as measured in Exp. I, which means their audibility had not changed.

2.2.2 Speech Stimuli

The CV syllables consisted of 14 consonants (6 stops, 6 fricatives, and 2 nasals) followed by the /a/ vowel. Two fricatives, /θ/ and /ð/, were not used in the experiment, as they have high error, even for NH ears (Li *et al.*, 2010; Phatak and Allen, 2007). To reduce the time of administration, only 2 talkers (1 male and 1 female) were selected per consonant. The consonants were chosen from those for which there was less than 10% error in data of NH listeners (see the Table 2.4). In total, there were $14 \times 2 = 28$ different utterances. All 28 utterances had zero-error for SNR $\geq$ -2 dB (SNR90 $\geq$ -2) across the 14 NH listeners in the Phatak and Allen (2007) study.

2.2.3 Procedure

All of the subjects had one practice session, with 14 syllables in quiet, before they began the experiment. These 14 syllables were always different from the practice syllables, to limit learning effects. Syllable presentation was randomized over consonants, speakers, and even SNRs. Three SNRs - 12, 6, and 0 dB - and quiet conditions were tested. The experiment consisted of two sessions. In the first session (Phase I), each of the 28 utterances was presented 4 times at each SNR. This resulted in 28 utterances × 4 SNRs × 4 presentations = 448 trials. For each utterance at each SNR, the correct score percentage was calculated. The possible scores were 0% (0/4), 25% (1/4), 50% (2/4), 75% (3/4) and 100% (4/4). In the second session (Phase II), the number of trials depended on the subject's performance in the first session. Across the two sessions each utterance was presented between 5 and 10 times, depending on the error rate in the first session, and therefore each consonant was presented between 10 and 20 times at each SNR (see Table 2.3).

The rationale behind this experimental design was to increase the sample size as

Table 2.3: Number of presentation trials per consonant in Phases I and II of Exps. II and IV, depending on percent error.

	Exp. II			Exp. IV		
# of error (P_e)	Phase I	Phase II	Total	Phase I	Phase II	Total
0 (0%)	8	2	10	8	4	12
1 (25%)	8	4	12	8	4	12
2 (50%)	8	10	18	8	10	18
3 (75%)	8	12	20	8	12	20
4 (100%)	8	12	20	8	12	20

a function of the score and to obtain more data when there are more errors being made. The total number of trials per consonant (sum of sessions I and II) was not same for all subjects. About 800-1000 trials were presented to each subject and the experiment took a total of 30-40 mins per ear.

2.3 Experiment III

2.3.1 Subjects

Twenty HI subjects recruited from the Urbana-Champaign community participated. All subjects were native speakers of American-English and all were paid. Informed consent was obtained from all subjects, and all procedures of the study were approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign. Subjects had normal middle-ear status (type A tympanogram) and sensorineural hearing-loss (SNHL). The etiologies of subjects' hearing loss varied. The results of the hearing screening tests varied in terms of the degree and configuration of individual hearing loss. Of the 20 subjects, 9 had symmetrical and 11 had asymmetrical bilateral hearing loss. They ranged in age from 21 to 84 years (mean = 55.45 years, SD=20.42).

2.3.2 Speech Stimuli

Speech stimuli used in Exp. III were exactly same as in Exp. I (Table 2.2): sixteen consonants followed by the /a/ vowel (Miller and Nicely, 1955).

2.3.3 NAL-R Amplification Condition

To compare the consonant error (P_e, %) and confusions between the flat gain at the most comfortable level (MCL) and NAL-R amplification (also at MCL, but gain was frequency dependent based on pure-tone threshold) conditions, all subjects were tested in the two conditions, called 'no NAL-R condition' and 'NAL-R amplification condition.' When simulating the NAL-R condition, its formula was calculated in two steps for each subject, by obtaining the required *real-ear gain* (REG) as a function of frequency (Dillon, 2001).

Step 1:

Calculate $X(dB) = 0.15 \times (HTL_{500} + HTL_{1000} + HTL_{2000})/3$, where HTL_f is the hearing threshold level (HTL) of the ear at frequency f.

Step 2:

Calculate the prescribed REG at each frequency:

$REG_{250}(dB) = X + 0.31 \times HTL_{250} - 17$

$REG_{500}(dB) = X + 0.31 \times HTL_{500} - 8$

$REG_{1000}(dB) = X + 0.31 \times HTL_{1000} - 3$

$REG_{1500}(dB) = X + 0.31 \times HTL_{1500} + 1$

$REG_{2000}(dB) = X + 0.31 \times HTL_{2000} + 1$

$REG_{3000}(dB) = X + 0.31 \times HTL_{3000} - 1$

$REG_{4000}(dB) = X + 0.31 \times HTL_{4000} - 2$

$REG_{6000}(dB) = X + 0.31 \times HTL_{6000} - 2$

where REG_f is the real-ear gain at frequency f.

2.3.4 Procedure

The test procedures for the CV measurements were very similar to those used in Study I. All subjects had one practice session consisting of 10 syllables in quiet to familiarize each subject with the test. Subjects were asked to identify the consonant in the presented CV syllable by selecting one of 16 software buttons on a computer screen, each labeled with an individual consonant sound. A ‘noise only’ button was allowed for the subjects to choose if they heard only noise without any speech. A pronunciation for each consonant was provided below its button to avoid possible confusions from any orthographic similarity between consonants (e.g., ʃ of shoes). The subjects were allowed to hear each utterance a maximum of 3 times before making their decision. Once a response was entered, the next syllable was automatically presented after a short pause.

Each syllable presentation was randomized with respect to consonants and speakers, but not with respect to SNR. The test proceeded from the easiest to the most difficult noise conditions - quiet first, followed by +12 to -12 dB SNR. This was done in order to gradually increase the difficulty from the onset, so that subjects were not pushed beyond their limits in terms of performance level.

Each subject heard a maximum of 1152 trials (16 consonants × 6 utterances × 2 presentations × 6 different noise conditions). When the score was less than or equal to 3/16 (18.75%, or three times chance) for each consonant, that consonant was not presented at subsequent (lower) SNRs. The experiment took a total of 1 to 1.5 hours per ear.

2.4 Experiment IV

2.4.1 Subjects

To characterize unique HI consonant loss when amplified speech sounds are present and to investigate the amplification effect of ZE utterances having a statistically suitable number of presentations per CV, the subjects (total 16 ears) who were involved in Exp. II were contacted by email and phone in May 2011. All of the Exp. II subjects also participated in Exp. IV, except for one subject who had moved away. All subjects had the same pure-tone hearing threshold as in the previous year within 5-dB in the testing frequencies (from 1.25-8 kHz), and all subjects had no history of middle ear pathology.

Table 2.4: Zero-Error utterances which were used in Exps. II and IV. The numbers in parentheses refer to each stimulus' SNR90 (signal-to-noise ratio at which NH listeners perceive on utterance with 90% accuracy).

	Exp. II		Exp. IV	
pa	f103 (-20)	m118 (-16)	f103 (-20)	m118 (-16)
ta	f108 (-16)	m112 (-20)	f108 (-16)	m112 (-20)
ka	f103 (-10)	m111 (-16)	f103 (-10)	m111 (-16)
fa	f109 (-16)	m112 (-10)	f109 (-16)	m107 (-10)
sa	f103 (-16)	m120 (-10)	f103 (-16)	m120 (-10)
ʃa	f103 (-16)	m118 (-16)	f103 (-16)	m118 (-16)
ba	f101 (-10)	m112 (-2)	f101 (-10)	m112 (-2)
da	f105 (-16)	m118 (-10)	f105 (-16)	m118 (-10)
ga	f109 (-10)	m111 (-16)	f109 (-10)	m111 (-16)
va	f101 (-10)	m118 (-2)	f101 (-10)	m111 (-10)
za	f106 (-20)	m118 (-16)	f106 (-20)	m118 (-16)
ʒa	f105 (-16)	m107 (-10)	f105 (-16)	m111 (-20)
ma	f103 (-16)	m118 (-16)	f103 (-16)	m118 (-16)
na	f101 (-10)	m118 (-2)	f101 (-10)	m112 (-16)

2.4.2 Speech Stimuli

The CV syllables consisted of 14 consonants (6 stops, 6 fricatives, and 2 nasals), followed by the /a/ vowel. Two fricatives, /θ/ and /ð/, were not used in Exp. IV, as they have a high error, even for normal hearing ears which is the same reasoning as for Exp. II (Phatak and Allen, 2007; Li *et al.*, 2010). Most utterances in Exp. IV were the same as those in Exp. II, but some utterances were changed to ones having much lower SNR90 (SNR90 is the signal-to-noise ratio at which NH listeners perceive an utterance with 90% accuracy). The selected utterances and their SNR90 are provided in Table 2.4.

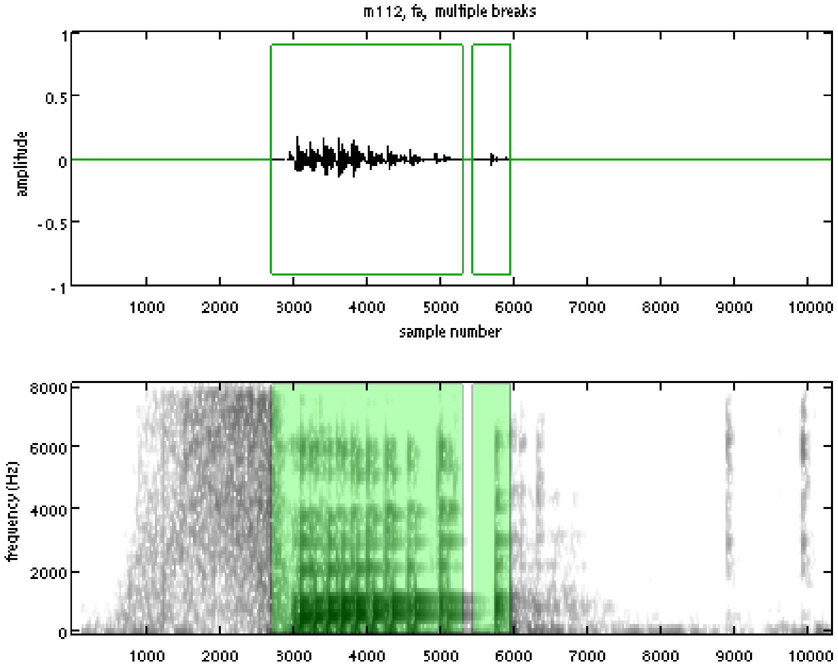


Figure 2.1: m112 /fa/ token was rendered incomplete by Matlab code designed to automatically cut off the silent part before and after the stimulus.

However, we realized that one token, m112 fa, was broken when presented through the software of Exp. II (a script in Matlab). During the filtering of silent parts existing before and after the speech stimulus, the frication energy of the /fa/ was removed (Fig. 2.1). Consequently, we had to remove the token from all statistical analysis, and the number of /fa/ utterances was not the same as the number of other utterances. We fixed this problem for Exp. IV, resulting in a stimulus set with no broken tokens.

2.4.3 NAL-R Amplification Condition

The same amplification procedure as in Exp. III was followed.

2.4.4 Procedure

All test procedures were the same as in Exp. II. After the subjects took the practice session having 28 tokens, syllable presentation was randomized over 14 consonants, 2 speakers, and 4 SNRs. Like Exp. II, the experiment consisted of two sessions. In the first session (Phase I), each of the 28 utterances was presented 4 times at each SNR. This resulted in 28 utterances × 4 SNRs × 4 presentations = 448 trials. For each token at each SNR, the correct score percentage was calculated. The possible scores were 0% (0/4), 25% (2/4), 50% (2/4), 75% (3/4) and 100% (4/4). In the second session (Phase II), the number of trials depended on the subject's performance in the first session. Across the two sessions each token was presented between 5 and 10 times, depending on the error rate in the first session, therefore each consonant was presented between 10 and 20 times at each SNR (see Table 2.3). The total number of trials (sum of sessions I and II) was not same for every subject. About 800-1000 tokens were presented and it took a total of 30-40 mins per ear.

2.5 Bernoulli Trials and Speech Perception

In this section we deal with the difficult problem of determining the number of trials required to quantify speech perception, when building CV confusion matrices (or a count matrix). The problem may be simply stated: *What number of Bernoulli trials* N_t *of a particular consonant-vowel sound is required in order to determine the* *probability* $\eta = P_{h|s}$ *with a specified confidence (e.g.,* 3σ*), that consonant /h/ was* heard *when consonant /s/ is spoken?*

To address this problem one must make a minimum of two assumptions. The first is that the subject is consistent. In fact since the subjects are human, and fall asleep, become bored, exhibit learning effects, or even play games during tedious psychological experiments in the booth, etc, one can never be sure that this is not violated. However there are well know methods to keeping the subject attentive, such as frequent breaks, and by monitoring the subject during the experiment. This may be a fragile assumption, but it is a necessary one.

The second assumption is that we may model the outcomes using Bernoulli trials with binomial outcomes. In fact the experiment by its very nature is multi-nomial. For example, in the experiments here, and those of Miller and Nicely (1955), where the CM is 16x16, the response space is a 16 dimensional vector space. While it would be nice to deal with such 16 dimensional model of the data, it is not possible, given the restrictions on time and the practical limitations of the size of N_t achievable in a real-world experiment. Thus we limit ourselves to the Binomial[1] probability weights "n choose k"

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}$$

applied to outcome probabilities $P_{h|s}^k (1 - P_{h|s})^{n-k}$.

Given the above basic assumptions we may apply well know results to compute estimates of confidence intervals for N_t as a function of $P_{h|s}$. We state these well known results in a series of three related statements.

1. The best estimate of the true probability $P_{h|s}$ given N_t Bernoulli trials is the mean

$$\mu = \frac{1}{N_t} \sum_{n=1}^{N_t} X_n,$$

---

[1] `http://en.wikipedia.org/wiki/Binomial_distribution`

where X_n is the random variable of binary outcomes, of the n^{th} trial, with $X_n = 1$ when $h = s$ (a *hit*) and 0 otherwise (a *miss*).

2. The standard deviation of the the above estimator of the mean μ is

$$\sigma_\mu = \sqrt{\frac{P_{h|s}(1 - P_{h|s})}{N_t}}.$$

3. According to the *Vysochanskij–Petunin inequality*,[2] the 95% confidence interval of this estimator is given by $3\sigma_\mu$.

These three results are well documented in the statistical literature, and well known. The most available source for these results may be found on *Wikipedia*, which in turn has excellent references into the literature.

Next we consider the application of these formulas to our basic question: What value of N_t is required to provide a 95% confidence for a given probability. Of course the problem here is that we do not know $P_{h|s}$. Furthermore this probability is a function of the the speech to noise ratio *SNR*. The standard approach is a bit of a bootstrap method. First estimate the score from N_t trials, and then justify this number of trials based on the value of $3\sigma_\mu$. To gain some confidence in such a procedure we only need to take some examples, and consider the nature of what we are trying to do.

As an example, let us assume that a normal hearing person responds $N_t = 20$ times in a row with the correct answer /ga/ when played /ga/. Given 20 such trials (correct in this case) trials, what bound may we place on the probability of $P_{h|s}$?

The smallest possible value of μ that is within the confidence interval is simply

$$\mu_B \equiv \mu_N - 3\sigma_\mu = \mu_N - 3\sqrt{\frac{\mu_N(1 - \mu_N)}{N_t}}.$$

---

[2] `http://en.wikipedia.org/wiki/Vysochanskii-Petunin_inequality`

Given N_t and μ_N we may use the above formula to calculate μ_B, the smallest value of the mean that is within the confidence interval.

For our example of $\mu = 0.9$ and $N = 20$, the above evaluates to $\mu_B \approx 0.7$. If we assume $\mu = 0.95$ and $N = 10$ then $\mu_B = .74$.

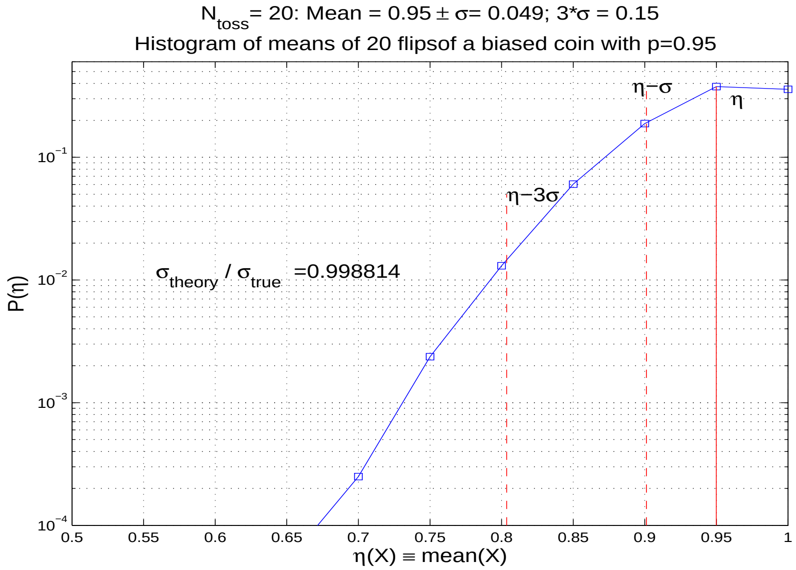


Figure 2.2: This figure results from a Monti Carlo (numerical) simulation of a biased coin flip. In this simulation a coin with a bias of $P_{h|s} = 0.95$ was tossed for $N_t = 20$ flips with 10^5 trials. A random variable X was defined as 1 if head and 0 if tails and the mean of the random variable μ and its variance σ_μ was then computed from the trials. A histogram of the outcomes from the 10^5 trials is shown, normalized as a probability. The estimated mean was $\eta = 0.95$, which happened with a probability of $\mu \approx 0.28$, namely 280,000 times. Also show are $\eta - \sigma_\mu$ and $\eta - 3 * \sigma_\mu$. The ratio of the theoretical $\sigma_\mu = \sqrt{P_{h|s}(1 - P_{h|s})/N_t}$ and the actual variance computed by the simulation is ≈ 1 within 0.12%.

A somewhat more satisfying measure is to ask, given a bias of $P_{h|s}$, how many flips N_t of the coin would it take to assure that with equal probability we would see N_t heads or $N_t - 1$ heads and one tail?

To solve this problem we set the first (N_t hits) and second binomial coefficient ($N_t - 1$ hits and 1 miss) equal and solve for $P_{h|s}$. Doing this gives

$$\binom{N_t}{N_t} P_{h|s}^{N_t} = \binom{N_t}{N_t - 1} P_{h|s}^{N_t} (1 - P_{h|s})$$

which results in a value of

$$P_{h|s} = \frac{N_t}{1 + N_t}.$$

This number is obviously related to the quantization inherent in estimating a probability given N_t flips. For our example this estimate gives $P_{h|s} = 20/21 = 0.9524$.

2.5.1 Further Considerations

It is our observation that the confusions for individual utterances form small groups. The scores are either above 90% (no error), or they form a group of 2 or three confusions. These small groups typically have error rates given by one over the size of the groups. For example, when the confusions are 1 of two sounds, then the scores are near 50%, and when there are 3 confusions, the scores are near 33%. Following the presentation of the results of Experiment II we shall discuss the importance of this observation in terms of estimating N_t. When these conditions are valid, the problem becomes one of separating cases having $P_{h|s}$ of $> 90\%$, $\approx 50\%$ and $\approx 33\%$.

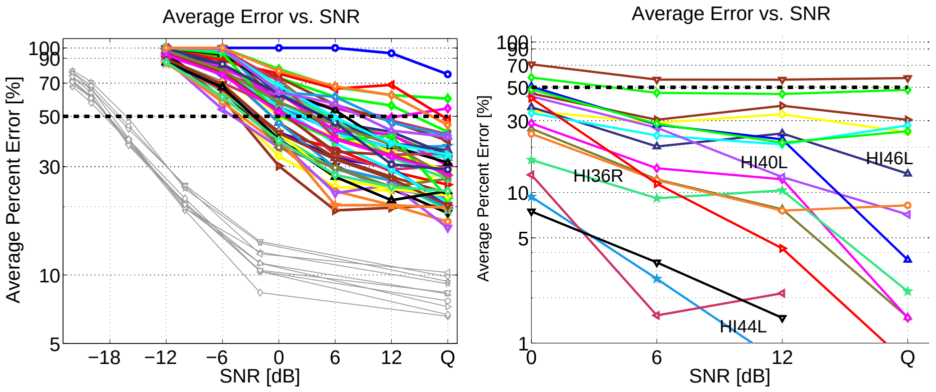
CHAPTER 3

RESULTS OF EXP. II

To assist the reader to understand our results of the 4 experiments, rather than following a chronological order (e.g.,I-IV), we organize the presentation into three chapters, 3, 4, and 5 where we present the results of Exp. II, Exp. IV, and pilot Exps. I and III, respectively. In the chapters 3 and 4, we show to analyze individual consonant and confusions as a function of SNRs for individual HI ears. In the chapter 5, we analyze the data using a large number of subjects and a small number of stimulus presentation. The small N for these two experiments require that we average the 5 SNRs, conforming to the clinical application of CV test.

According to Phatak *et al.* (2009), HI ears exhibit large individual differences in their average consonant loss, even given a similar PTA. In Fig. 3.1, we show the average consonant loss as a function of SNR, as measured in 46 and 17 ears (solid colored lines, (a) and (b), respectively) and 10 NH ears (solid gray lines). These data were collected using a procedure very similar to that of Miller and Nicely (1955), where full-rank confusion matrices were collected, from -12 to +12 dB SNR and in quiet (no noise).

When we retested eight of the 46 ears of Exp. I, we discussed that the number of stimulus trials was too small (N, the number of presentation per consonant was between 2 and 8 in Exp. I when presented to zero-error utterances). Once we appreciated the need for much a large N, we designed Exp. II to have 10 times the number of stimuli per utterance at each SNR (see the analysis in Ch. 2, section 2.5). As outlined in the Exp. II Methods section, to reduce possible errors, only utterances



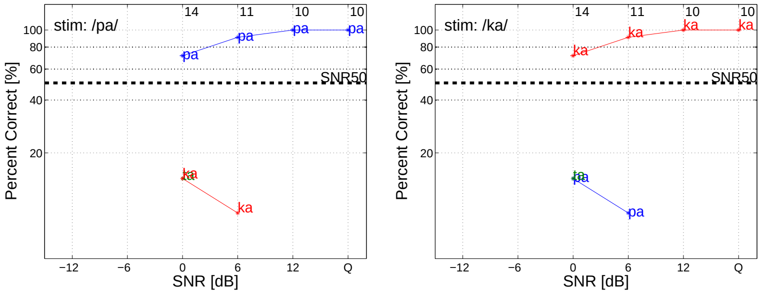
(a) 46 HI ears of Exp. I

(b) 17 HI ears of Exp. II

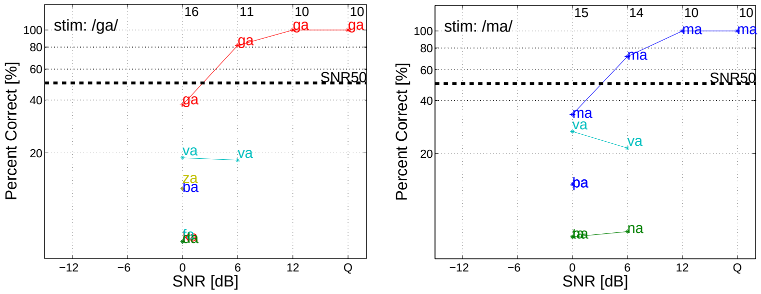
Figure 3.1: Average consonant error for 46 HI ears of Exp. I [(a), solid colored lines] and for 17 HI ears of Exp. II [(b), solid colored lines) as function of signal-to-noise ratio (SNR) in speech-weighted noise: abscissa represents SNR and ordinate is **average percent consonant error (%)** for 16 CVs for Exp. I and 14 CVs for Exp. II. The intersection of the thick horizontal dashed line at the 50% error point and the plotted average error line for each ear, mark the consonant recognition threshold (CRT) in dB. The data for 10 normal hearing (NH) ears are superimposed as solid gray lines for comparison [(a), grey lines]. NH ears have a similar and uniform CRT of -18 to -16 dB (only a 2-dB range), while the CRT of HI ears are spread out between -5 to +28 dB (a 33-dB range). Three out of 46 ears had greater than 50% error in quiet (i.e., no CRT) in panel (a). In panel (b), the CRT for these 17 ears are mostly from the <0 dB CRT region, thus the mean error is much smaller (1% or so) compared to (a) where the mean error is 15%.

having no error for NH ears above -2 dB SNR were presented. Theses CVs were randomly presented at 4 SNRs (-12 and -6 dB SNRs were dropped given the high error seen for most HI ears in Exp. I), and the responses of 14 CVs were collected, excluding the two consonants /θa/ and /ða/, since these always had error rate as high as 40% for NH ears.

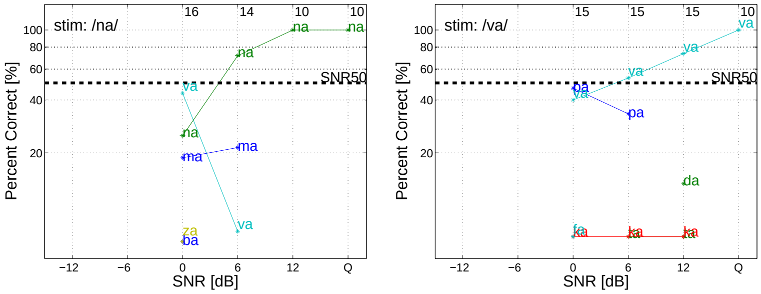
To verify that the consonant loss profiles (CLP) of HI ears which we define as $P_{h|s}(\mathrm{SNR})$, the probability of hearing consonant h in response to spoken consonants s, were stable given the larger N, we retested a subset of our original HI subjects. When all 46 ears were asked to participate in Exp. II, 17 HI ears having the *consonant recognition thresholds* (CRT) below 0 dB SNR self-selected. As shown in Fig. 3.1 (a) and (b), there is a wide disparity in the CRTs, defined as the SNR required by a particular HI ear for a 50% average recognition score. For Exp. I, these CRTs range from -5 to >+28 dB SNR. The SRT, which uses spondee words (i.e., context infor-



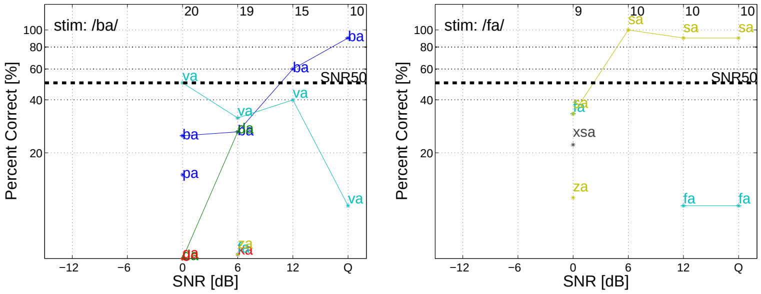
(a) $P_{h|s}$ /pa/ (b) $P_{h|s}$ /ka/



(c) $P_{h|s}$ /ga/ (d) $P_{h|s}$ /ma/



(e) $P_{h|s}$ /na/ (f) $P_{h|s}$ /va/



(g) $P_{h|s}$ /ba/ (h) $P_{h|s}$ /fa/

Figure 3.2: Individual consonant loss profiles (CLP) $P_{h|s}$(SNR) for eight consonants of subject HI40L in Exp. II, the consonant scores as function of signal-to-noise ratio (SNR) in speech-weighted noise: the abscissa is the SNR and the ordinate is $P_{h|s}$(SNR) (the probability of responding that consonant h was heard given that consonant s was spoken). The intersection of the thick horizontal dashed line at 50% error point and the plotted average error line for each ear define the consonant recognition threshold (CRT) in dB. HI40L has an average CRT (of the 14 CVs) of 0 dB in Fig. 3.1 (b), while the CRTs of individual consonants range from -5 to 10 dB (-5 is an extrapolative estimate).

mation) (Brandy, 2002), is significantly different from the CRT which uses nonsense CVs. The slope of the average error as a function of SNR [see Fig. 3.1, (a)] varies widely when compared to the 10 NH ears (gray lines), which only have a 2-dB CRT range (-18 to -16 dB) and a uniform slope. From Fig. 3.1 (a), we see that three of the HI ears never reached 50% error rate in quiet, thus the CRT is undefined.

Our results agree with the earlier findings of Phatak *et al.* (2009) with some minor differences. Both studies found that average consonant scores for the HI ears are poorly correlated with the PTA. For example, the HI ears with the lowest and highest slopes for average consonant error in Fig. 3.1 did not have the best and worst PTAs, respectively. However, there are some differences between the two results. In Phatak *et al.* (2009), 26 HI ears could be simply divided into 3 subgroups: high, medium, and low performance. When the number of HI ears was increased to 46 in Exp. I, the responses showed a continuum in performance (Fig. 3.1 (a), solid colored lines). There are four labeled HI ears in [Fig. 3.1, (b), i.e., HI36R, HI40L, HI44L, and HI46L]. These four subjects represent a sampling of performance for 17 ears of Exp. II. We will discuss these four subjects in the following section.

Table 3.1: Percent consonant errors (%) of seven select HI ears (rows) in the **quiet condition** [Exp. II]. **High** (>75%), **medium** (>50% and less than 75%), and *low* (>25% and less than 50%) errors are marked by red, blue, and green, respectively. Empty space indicates no error. For example, as shown by the second row, NH ears had zero error. Note that every HI ear has errors in many individual consonants, but there is high error for only a few of consonants. Note the high /za/ and /ʒa/ errors in HI46R. The two right columns provide typical clinical measures. 3TA (3-tone average, dB HL) is calculated by the average of 0.5, 1, and 2 kHz, and CRT (consonant recognition threshold; dB SNR) is the average consonant threshold at 50% error, similar to the SRT. Although having similar 3TA and CRT, HI01 shows asymmetric consonant perception between left (HI01L) and right (HI01R) ears - /sa/ and /za/ are better perceived in HI01L and /pa/ and /va/ are better in HI01R.

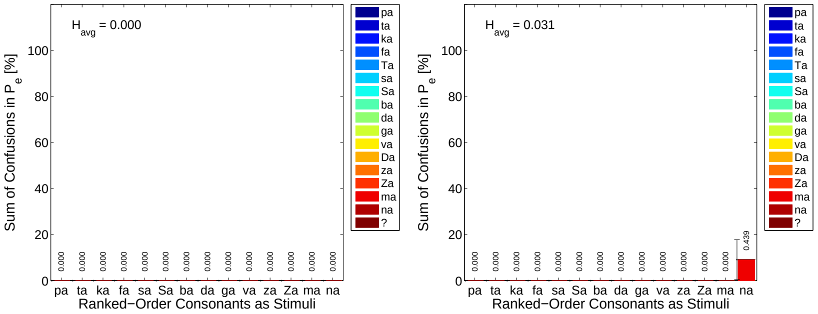
Ear	/pa/	/ta/	/ka/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/	3TA	CRT
NH														0	0
HI46R						10		*40*	9	**89**	**95**	9		16.7	1
HI40R						9			*27*				10	23.3	.5
HI30R				**60**	9	*27*			**56**	**72**	**69**			26.7	4.5
HI36R						*37*								28.3	-3
HI01L	**60**		**90**	*40*	20	9	9	**90**	**75**	20	**70**	**50**		45	14
HI01R	10		**100**	**100**	*36*		**67**	**100**	10	**67**	**95**	10	*35*	46.7	14.5
HI14R	9	*27*	9	*27*		9	*39*	9	**60**	10	14	**56**	*27*	73.3	12

A closer examination of Fig. 3.1 [either (a) or (b)] reveals a major weakness of the CRT in that it fails to quantify the individual differences. Therefore, we desire an improved metric of the individual differences of consonant scores. A typical example of $P_{h|s}$(SNR) for subject HI40L, whose average score of 0 dB [Fig. 3.1(b)], is presented in Fig. 3.2, which shows a large range of individual CRTs from -5 (<0 where is the lowest SNR in Exp. II) to 10 dB.

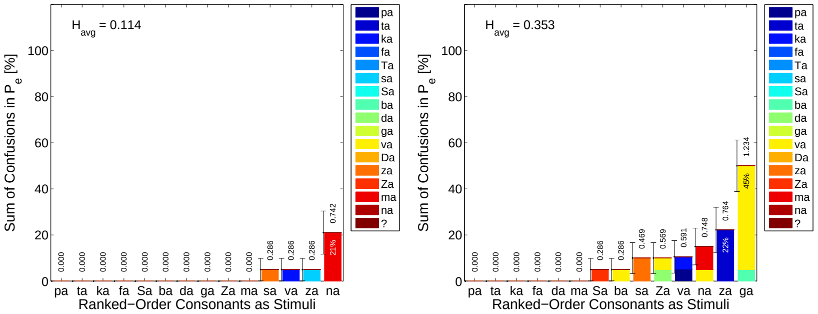
In Table 3.1, we provide detailed consonant error (%) from a sample of 7 out of the 17 impaired ears from Exp. II. HI46R has the lowest 3TA among 7 ears, yet this subject shows very high error for /za/ and /ʒa/. Such errors do not occur for HI40R, which has the second best 3TA of the 7 ears. HI36R and HI30R have similar 3TAs, but show very different consonant errors. HI36R has only /ba/ error, whereas HI30R shows substantial error for /sa/, /va/, /za/, and /ʒa/. Ears HI01L (left) and 01R (right) from subject HI01 show a symmetrical hearing loss. Both ears have poor perception on /ka/ and /ga/ (i.e., 90∼100% error), yet they have asymmetrical CLPs, with high /pa/ and /va/ errors only in the left ear (HI01L) and high /sa/, /da/, and /za/ errors in the right (HI01R). Another interesting case, HI14R does not reveal any particularly high error rate for consonants (>75%), despite having the worst 3TA. Again, each ear discussed here has a different consonant profile, poorly represented by the average scores, 3TA and CRT. The consonant error for all 17 HI ears (Exp. II) are presented in the Appendix B (+12, +6, and 0 dB in Tables B.1, B.2, and B.3, respectively). A reasonable summary of these seven ears is to emphasize the huge individual differences.

3.1 Error Pattern Analysis: Subjects 44L, 46L, 36R, and 40L

To show in greater detail the percentage of error for each consonant and its error pattern (or confusion), we generated color stacked bar plots in Figs. 3.3, 3.4, 3.5,



(a) Quiet condition (b) +12 dB SNR condition



(c) +6 dB SNR condition (d) 0 dB SNR condition

Figure 3.3: Stacked bar plots of HI44L at 4 SNRs: (a) quiet, (b) +12 dB, (c) +6 dB , and (d) 0 dB. In each plot, ordinate indicates percent error of individual consonant and height of each bar means total percent error (P_e, %) which is composed of several confusions in different colors. Abscissa is rank-ordered by total percent error of 14 CVs. The PTA for the subject shown on Fig. 3.9(c) (blue-x). For example, the subject has the highest /ga/ error (50%) at 0 dB; 45% of all /ga/ trials at 0 dB are heard as /va/. Average entropy is listed in the upper left corner of each plot, and each bar has a row entropy and error bar. Subject HI44L has no error in quiet, while at 0 dB SNR /ga/ has 9 /va/ confusions out of 20 presentations ($P_{v|g}(0dB) = 0.4$). For /za/, $P_{t|z}(0dB) = 2/9$ and $P_{z|z}(0dB) = 14/18$.

and 3.6 for the four highlighted ears in Fig. 3.1 (b).

In each of Figs. 3.3, 3.4, 3.5, and 3.6, there are four sub-plots, corresponding to 4 SNRs conditions: Quiet, +12, +6, and 0 dB. The abscissa of each plot is sorted by rank-order: easiest (or lowest error) consonant to highest error consonant. The ordinate shows total percent error (%) of the individual consonant. The error bar is calculated using $SE = \sqrt{\frac{p(1-p)}{N}}$ where p is probability correct, N is total number of trials, and SE is then provided on each bar. The total height of each bar indicates the fractional error for that consonant. The colors on the bar show consonants with which the target consonant has been confused. For example, a bar which has many different colors means that the target consonant has been confused with many others (i.e., many confusions), while a bar having a solid color shows that the target consonant has only been mistaken for one other. The number in the left upper corner of each plot is the total entropy for all consonant confusions (i.e., total confusions, given SNR). This entropy can be between 0-4 bits. The largest entropy shown is 1.895 bits for HI01R at 0 dB. Each bar has a number showing the consonant entropy (i.e., row confusions). The formula for the entropy is $H = -\sum_i p(x_i) \log_2 p(x_i)$. Higher entropy indicates greater uncertainty. The entropy of the 17 HI ears and 4 SNRs is summarized in Table 3.3.[1]

Table 3.2: Sub-count matrix at 6 and 0 dB-SNR for HI44L; the frequencies in this table are re-plotted as percentages in Fig. 3.3. Each row is stimulus consonant, while each column is response consonant. Last column is total number of presentations. Cells with only 1 or 2 errors were not displayed because they were considered to be low level random errors. The number in the left top cell indicates the SNR. For example, at 6 dB, /na/ is presented 19 times of which 15 are correct and 4 incorrect (heard as /ma/) responses.

6	na	ma	Σ
na	15	4	19

0	ga	za	va	ta	Σ
ga	10		9		20
za		14		4	18

The error gradually changed with SNR. A one-way Analysis of Variance (ANOVA) showed that the percent of errors is significantly increased as noise increases (F[3,45]=56.428, p<0.01), therefore the quiet, +12, +6, and 0 dB SNR conditions each differ. Although the percent of error was very different, and the order of error across consonants varied by the subject, the HI ears had significant difference in consonant perception (F[13,195]=5.451, p<0.01). In general, /da/ had the lowest error and /ta, ʃa/ were next lowest error, while /ʒa, ba, va, za, fa/ had the highest errors.

The entropy was also statistically analyzed with ANOVA to find its significant predictors. As noise increased, the entropy significantly increased (F[3,45]=83.619, p<0.01). The Bonferroni Post-Hoc test resulted in a significant difference between

---

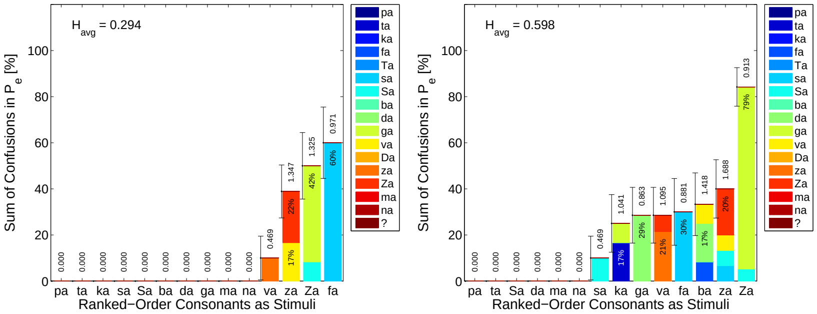
[1] SNR90^*_1 is a convenient measure for 1-bit of confusion in terms of the average SNR as indicated by table 3.3. This is convenient way of sorting the HI ears in terms of their average confusion.

each of the three SNRs (quiet, +12, +6 dB) and 0 dB ($p<0.01$). Entropies under quiet condition vs. +6 dB SNR was not significant, but entropy was significantly higher at 0 dB. /ba/ and /va/ syllables had the highest entropy at 1.044 and 1.062 bits, respectively, and /da/ was the consonant with the lowest entropy (0.174 bits) (F[13,195]=10.755, $p<0.01$)..

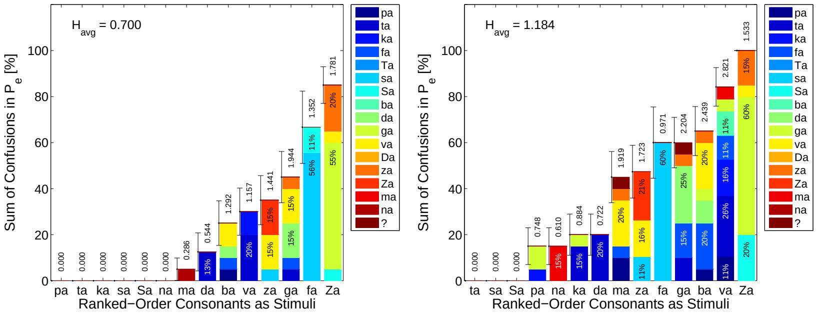
Table 3.3: Results of total entropy calculation of 4 SNRs for 17 HI ears in Exp. II: $\mathcal{H} = -\sum_i p(x_i)\log_2 p(x_i)$. $\mathcal{H}$ is a measure of the subject's response uncertainty. When the entropy is zero, there is no subject uncertainty, independent of the scores ($P_{h|s}$). As noise increased, the entropy significantly increases, which means the confusions increased. Bonferroni Post-Hoc test showed there is a significant difference between each of three SNRs (quiet, +12, +6 dB) and 0 dB ($p<0.01$) (F[3,45]=83.619, $p<0.01$). Confusions from quiet condition to +6 dB SNR were not increased, but were significantly higher at 0 dB. Group mean of the entropy at quiet, +12, +6, and 0 are 0.242, 0.373, 0.567, and 1.091 bits, respectively. In column six, SNR_1^* indicates the SNR where the entropy is 1-bit, i.e., $\mathcal{H}(\mathrm{SNR}_1^*)=1$.

	Quiet	12 dB	6 dB	0 dB	SNR_1^*
HI32L	0.067	0.194	0.403	0.931	<0
HI36L	0	0.062	0.144	0.317	<0
HI36R	0.06	0.146	0.162	0.545	<0
HI40R	0.209	0.213	0.49	0.835	<0
HI44L	0	0.031	0.114	0.353	<0
HI44R	0	0.074	0.065	0.57	<0
HI14R	0.863	0.656	0.877	1.093	0
HI32R	0.065	0.342	0.5	1.053	0
HI46R	0.348	0.332	0.636	1.109	1
HI40L	0.067	0.191	0.598	1.284	2
HI46L	0.294	0.598	0.7	1.184	2
HI30L	0.362	0.578	0.734	1.167	2
HI34L	0.031	0.164	0.477	1.506	3
HI30R	0.577	0.647	0.763	1.418	4
HI34R	0.137	0.623	0.926	1.656	5
HI01L	0.802	0.768	0.979	1.633	6
HI01R	0.852	1.012	1.375	1.895	12

We again claim that average error rates, as used in most clinical tests, cannot explain that the many highly significant individual differences seen in these consonant confusions. To better characterize these individual differences, we generated sub-confusion matrices that included only the consonants having confusions (Tables 3.2, 3.4, 3.5, and 3.6 which are to be paired with Figures 3.3, 3.4, 3.5,and 3.6). Like the classic confusion matrix of Miller and Nicely, a row represents the presented stimulus, a column is the response, and the total number of presentations is marked in the last column. Errors that occurred only once or twice were considered inconsequential and removed from the count matrix.



(a) Quiet condition (b) +12 dB SNR condition



(c) +6 dB SNR condition (d) 0 dB SNR condition

Figure 3.4: Stacked bar plots of HI46L of 4 SNRs: (a)-(d). In each plot, ordinate indicates percent error of individual consonant and height of each bar means total percent error (P_e, %) which is consisted of several confusions in different colors. Abscissa is rank-ordered by total percent error of 14 CVs. As noise increases from (a) to (d), total P_e is increased and confusions are higher, consisting of more various colors.

- **HI Case I** The left ear of HI44 (Figure 3.3 and Table 3.2) was the subject with fewest errors (see Fig. 3.1 (b)). She did not make any consonant errors in quiet, and a small error of /na/ (1 of 12, 8%) at 12 dB and (4 of 19, 21%) at 6 dB SNR. She showed a significant 50% /ga/ error (10 out of 20 trials) and reported this /ga/ as /va/ (45%). /ga/ has burst spectral energy at about 1.4-2 kHz, near that of /va/

having frication energy in the 0.6-1.3 kHz range. This 50% /ga/ error resulted in the discovering of talker effect in a more detailed analysis.

Table 3.4: Sub-count matrix for quiet, +12, +6 and 0 dB for HI46L (see Fig. 3.4). The number in the left top cell indicates the SNR. Each row is a presented consonant (stimuli) and each column is a response. The last column is total number of presentation. Single and double errors are not displayed due to these error. Diagonal entries are correct and off-diagonal is an error.

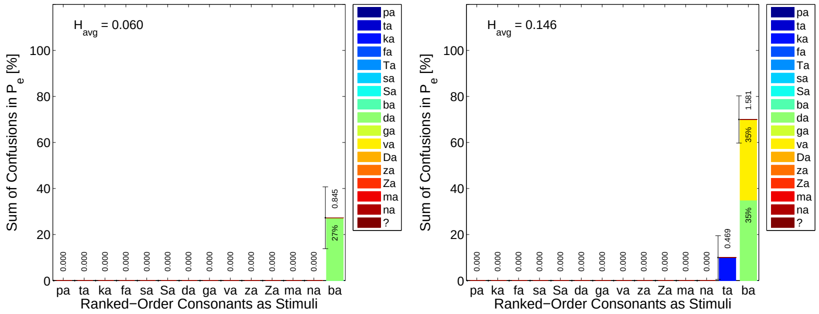
Q	fa	za	ʒa	sa	ga	va	Σ
fa	4			6			10
za		11	4			3	18
ʒa			6		5		12

12	fa	ga	va	za	ʒa	da	sa	Σ
fa	7						3	10
ga		10				4		14
va			10	3				14
za				9	3			15
ʒa		15			3			19

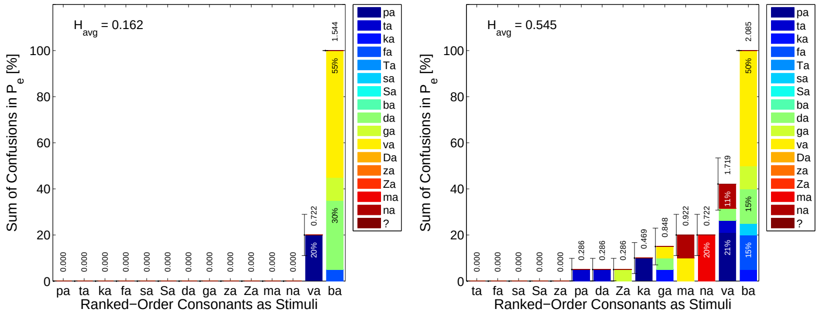
6	fa	ga	va	za	ʒa	da	ta	sa	Σ
fa	3							5	9
ga		11	3			3			20
va			14				4		20
za			3	13	3				20
ʒa		11		4	3				20

0	ka	fa	ba	da	ga	va	za	ʒa	ma	na	ta	sa	ʃa	Σ
ka	16										3			20
fa		4										6		10
ba		4	7			4								20
da				16							4			20
ga		3		5	8									20
va	3					3					5			19
za						3	10	4						19
ʒa					12		3						4	20
ma						4			11					20
na									3	17				20

- **HI Case II** In Figure 3.4, HI46L made errors given /fa, ʒa, za/, even in the quiet condition. At +12, +6, and 0 dB SNR, /ʒa/ was the highest error consonant, and /za, ba, fa, ga/ were also commonly confused and reported. Importantly, the subject has some consistent error patterns as a function of SNRs (Table 3.4). That is, /fa/ (energy between 1.3 and 2.8 kHz) was confused with /sa/ (3.8-8.0 kHz), /ga/ (1.4-2 kHz) with /da/ (>4 kHz), /za/ (3.5-8.0 kHz) with /ʒa/ (2.0-3.2 kHz), and /ʒa/ (2.0-3.2 kHz) with /ga/ (1.4-2 kHz) in most SNRs. The frequencies in parentheses are the support energy region for each these consonants (Li *et al.*, 2010, 2011), as summarized by Fig. 6.1. It seems that this subject did not have a voicing or burst



(a) Quiet condition (b) +12 dB SNR condition



(c) +6 dB SNR condition (d) 0 dB SNR condition

Figure 3.5: Four stacked bar plots of HI36R of 4 SNRs: (a)-(d). In each plot, y-axis indicates percent error of individual consonant and height of each bar means total percent error (P_e, %) which is consisted of several confusions in different colors. X-axis is rank-ordered by total percent error of 14 CVs. Subject is not affected by noise, showing a few consonant error except for /ba/ syllable. As noise increases, /ba/ had higher percent error (100% at +6 and 0 dB) and confusions is also increased from 1.544 to 2.085 (/ba/ row entropy)

perception problem, but rather had trouble distinguishing among consonants with support in the frequency region above 1.4 kHz. Having a high frequency hearing loss beyond 3 kHz, his consonant perception results are poorly correlated with his PTA.

- **HI Case III** In Figure 3.5, the right ear of HI36 made a significant /ba/ error, and the error dramatically increases with noise. Table 3.5 shows that /ba/ was reported as /va/ and /da/, while the /va/ was reported as /pa/. It seems that HI36 has trouble using timing cues to distinguish voicing and continuity (see the discussion in Fig. 6.1).

- **HI Case IV** In Fig. 3.6 and Table 3.6, HI40L showed a high entropy for spoken sounds /fa, va, sa, pa/ at 0 dB, so we cannot simply say that HI40L had either a burst, modulation, or frequency resolution processing problem. Specifically he could

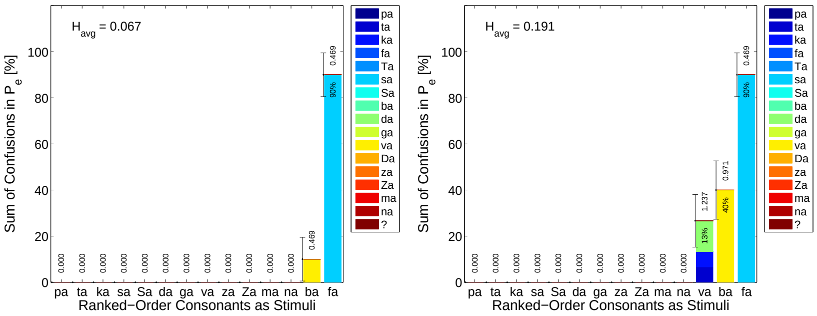
Table 3.5: Sub-count matrix in the quiet, +12, +6 and 0 dB for HI36R, paired with Fig. 3.5. The subject's only errors were for /ba/, /va/, /na/ syllables. Note how the /ba/ errors were confused with /va/ and /da/, and how /va/ was perceived as /pa/.

Q	ba	da	Σ
ba	8	3	11

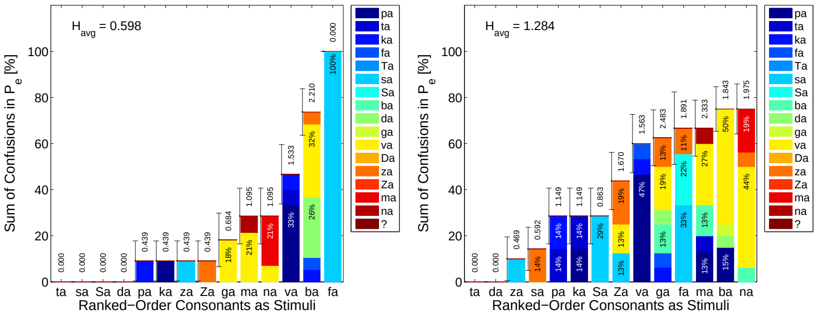
12	ba	da	va	Σ
ba	6	7	7	20

6	ba	va	pa	da	Σ
ba		11		6	20
va		16	4		20

0	ba	va	na	fa	pa	da	ma	Σ
ba		10		3		3		20
va		11			4			19
na			16				4	20



(a) Quiet condition (b) +12 dB SNR condition



(c) +6 dB SNR condition (d) 0 dB SNR condition

Figure 3.6: Stacked bar plots of HI40L of 4 SNRs: (a)-(d). Subject's consonant perception is affected from +6 dB. /fa/ perception is always confused to /sa/ regardless of SNR. At 0 dB condition, most consonants make error and row entropy of individual consonant is increased up to about 2.5 (/ga/).

not distinguish the burst from frication in the /ba/ and /va/, while /fa/ was confused with /sa/ as a function of SNR.

Table 3.6: Sub-count matrix at quiet, +12, +6 and 0 dB for HI40L which is paired with Fig. 3.6. As the noise increases, the number of consonant producing significant high error is increased from 1 (i.e., /fa/) in the quiet condition to 8 at 0 dB. Note how /va/ is represented when /ba, ga, va, ma, na/ are spoken, yet is only recognized 40% of the time.

Q	fa	sa	Σ
fa	1	9	10

12	fa	ba	sa	va	Σ
fa	1		9		10
ba		9		6	15

6	fa	ba	va	ma	na	da	pa	sa	Σ
fa								10	10
ba		5	6			5			19
va			8				5		15
ma			3	10					14
na				3	10				14

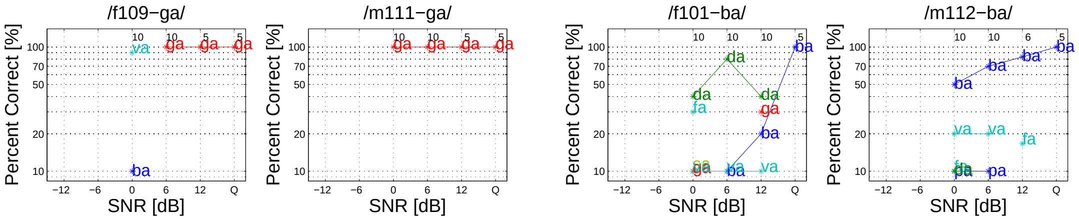
0	fa	ʃa	ba	ga	va	ʒa	ma	na	pa	sa	za	Σ
fa	3									3		9
ʃa		10								4		14
ba			5		10				3			20
ga				6	3							16
va					6				7			15
ʒa						9					3	16
ma					4		5					15
na					7		3	4				16

Again, to use an average score or do a typical statistical analysis would require us to treat HI listeners as one homogenous group, which means we could miss all these detailed individual differences.

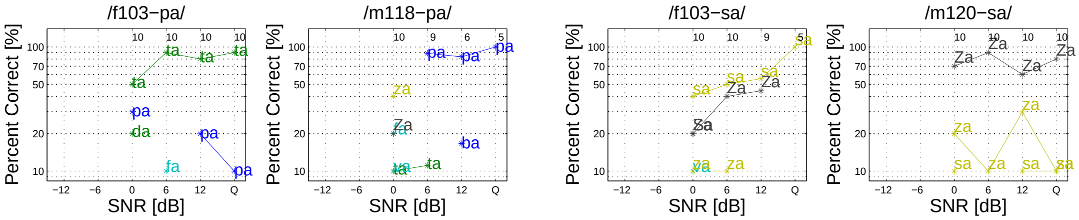
3.2 Talker Dependence

Consonant perception accuracy was tested for utterance (talker) dependence. This analysis is shown here in Fig. 3.7 for only four listeners. We observed in Table 3.2 that listener HI44L showed exactly 50% error when the /ga/ syllable was presented at 0 dB SNR, and almost all confusions were reported as /va/. Panel (a) of Fig. 3.7 verifies that she never correctly perceived the /ga/ syllable spoken by a female talker at 0 dB SNR, whereas the /ga/ of a male talker was 100% correct. Panel (b) also shows that

HI34R had difficulty perceiving the female /ba/. Subject HI01L (panel (c)) reported that a female speaker generally is not easy to understand in her everyday conversation, and her introspection is confirmed by the accuracy of her /pa/ perception. Subject HI31L of panel (d) could not correctly perceive the male /sa/ syllable at any SNR.



(a) Comparison of Talkers:HI44L/ga/ (b) Comparison of Talkers:HI34R/ba/



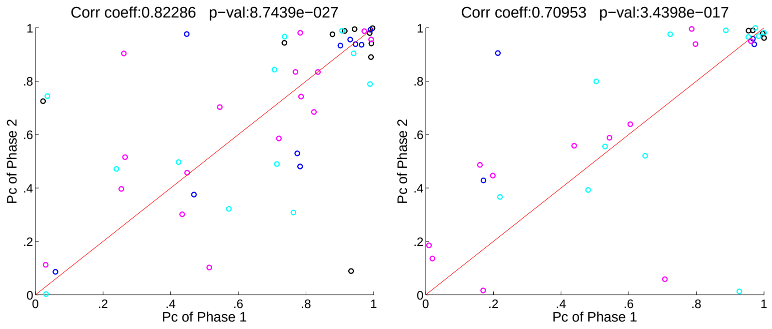
(c) Comparison of Talkers:HI01L/pa/ (d) Comparison of Talkers:HI31L/sa/

Figure 3.7: Talker Dependence of Exp. II: panels (a)-(d) for correlation between talker I (female, left plot of each panel) and talker II (male, right plot of each panel) of /ga/, /ba/, /pa/, and /sa/ syllables. X-axis indicates SNR and y-axis is percent correct (%). Numbers above 100% line indicate total number of trials at each SNR.

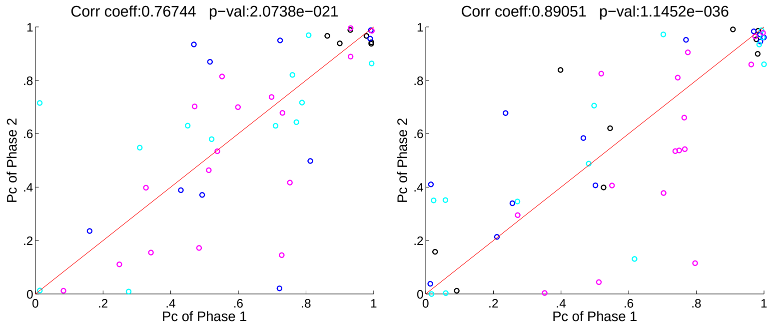
Since this analysis was not designed into the experiment, it was difficult to isolate any consistent pattern of differences, thus more careful analysis and an additional studies are needed to explore the complex issue of talker dependence.

3.3 Internal Consistency

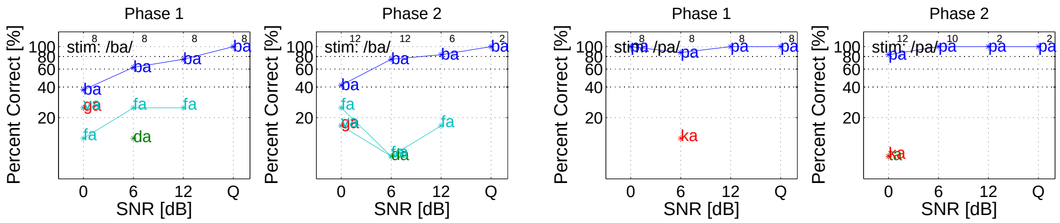
To test for the reliability of Exp. II, we computed an internal consistency, typically a measure based on the correlations between different items on the same test (or the same sub-scale on a larger test). Specifically, we computed the correlation coefficient and its p-value between Phase I (the first session; abscissa) and Phase II (the second session; ordinate) (see the 4 examples of (a)-(d) in Fig. 3.8). Every subject showed a



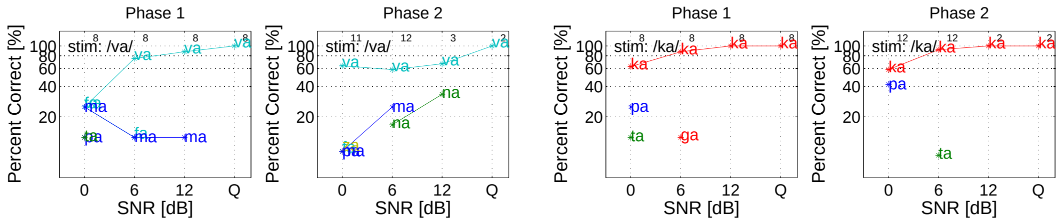
(a) HI34L (b) HI36L



(c) HI36R (d) HI44L



(e) Comparison of Phases: /ba/ (f) Comparison of Phases: /pa/



(g) Comparison of Phases: /va/ (h) Comparison of Phases: /ka/

Figure 3.8: Internal Consistency. Panels (a)-(d) show the correlation between phase I (abscissa) and phase II (ordinate) of four HI ears. Circle means individual consonant and black, blue, turquoise, and pink colors correspond to quiet, +12, +6, 0 dB SNR, respectively. Panels (e)-(h) show percent correct (%) as a function of SNR for the two phases, for the utterances /ba/, /pa/, /va/, /ka/ in HI32R. Numbers above 100% line indicate total number of trials at each SNR.

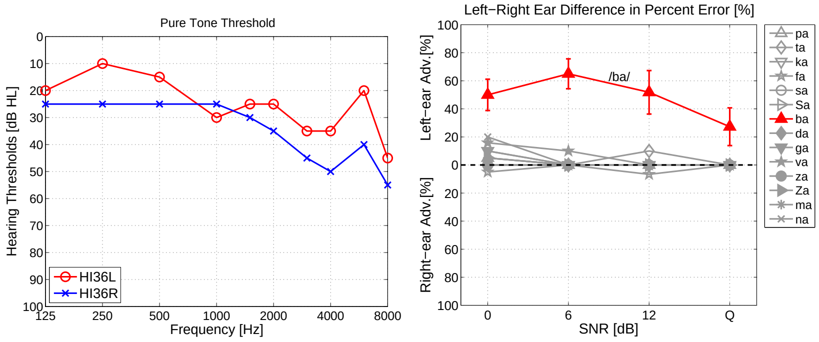
high correlation between the two phases (range of correlation coefficient: 0.425-0.89, mean: 0.67, std: 0.125).

Furthermore, Fig. 3.8 (e-h) shows the level of consonant-by-consonant correlation. At each phase, the percentage of correct responses rises smoothly with increasing SNR. However, when the number of presentations changed, the percentage also slightly changed (less than 20%) for some consonants (panels (e) and (g)). Separating into two phases shows that the specific confusion patterns as well as percent correct (%) are internally consistent. For example, /ba/ in panel (e) was confused with /fa/ in both Phases I and II, while /va/ (panel (g)) was confused with the nasals, /ma/ and /na/, in both phases. Necessarily the N must be split for this comparison, making the curves less certain.

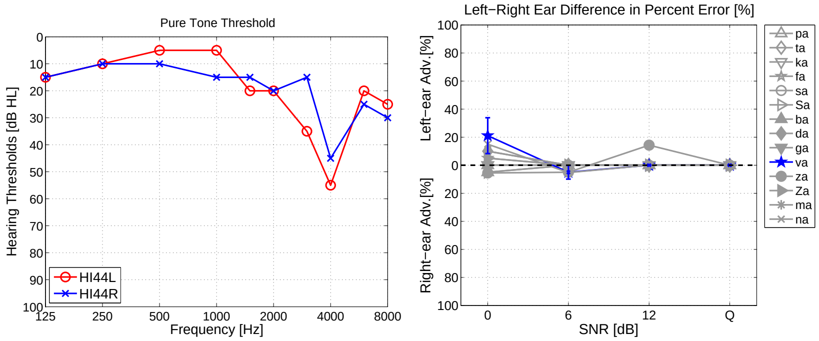
3.4 Comparison between the PTA and CLP as a Clinical Application

Figs. 3.9 and 3.10 show the PTA for four HI subjects and their CLP from Exp. II as a function of SNR. These data again confirm that the impaired ears are very different, and their consonant loss profile is poorly correlated with their average scores for both the PTA and CRT.

HI36 in Fig. 3.9 (a,b) had 10-20 dB better thresholds in the left ear (blue-x), as shown in panel (a), and has a large left-ear advantage for /ba/ as shown in (b). The advantage peaks at 6 dB SNR with a 60% left-ear advantage, and is 30% even in the quiet condition (Q on the abscissa). The subject heard most consonants similarly in both ears (less than 20% difference in left versus right ear) with no difference in /pa/, whose burst spectrum has energy in the same frequency range of .3∼2 kHz with /ba/. In fact, the results for HI36 in Exp. I show little difference in consonant loss between left and right ears (Figure 3.10 [b]). However, when SNRs were separated, a left-ear



(a) Audiogram: HI36L/36R (b) Consonant-loss of HI36L/36R



(c) Audiogram: HI44L/44R (d) Consonant-loss of HI44L/44R

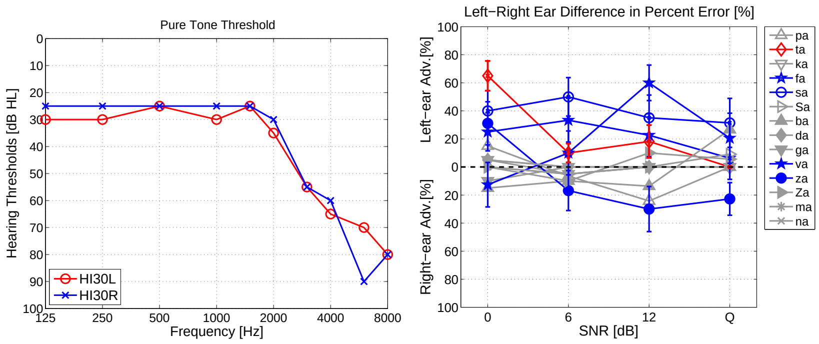
Figure 3.9: The different between two CLPs for two HI subjects from Exp. II are shown. The left and right panels are their PTA and CLP, respectively. On the right panels, curves above the horizontal line (0) indicate a left-ear advantage as a function of SNR, and those below the line show a right-ear advantage as a function of SNR. To reduce the clutter, consonants which have less than 20% ear difference are shown as gray lines. Standard errors are also marked on the significant points. Note how panel (b) shows a large /ba/ advantage (between 30-60%) to the left ear.

advantage in the /ba/ syllable was indicated. This illustrates the importance of the SNR when measuring the consonant-loss in HI listeners.

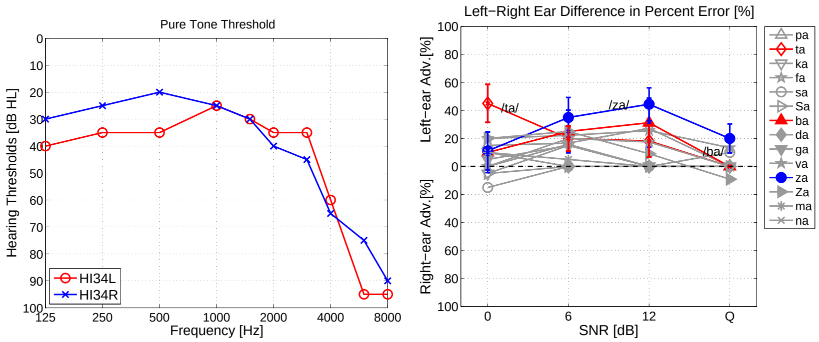
In Fig. 3.9,(c,d), HI44 has almost no difference in consonant perception between her ears as a function of SNR, and the PTA is very similar.

Fig. 3.10 (a,b), HI30 has a 20 dB HL difference at 6 kHz (worse in the left ear), yet has a distinct and significant left-ear advantage for syllables /va/, /sa/, and /fa/, and up to a 30% right ear advantage for /za/.

In Fig. 3.10 (c,d), subject HI34 has slightly 10 dB better thresholds in the right ear (red-o), with a steep loss between 3-5 kHz, reaching 90 dB HL. However, we see in (d) that her left ear has the clear advantage for all consonants, although the



(a) Audiogram: HI30L/30R (b) Consonant-loss of HI30L/30R



(c) Audiogram: HI34L/34R (d) Consonant-loss of HI34L/34R

Figure 3.10: The different between two CLPs for two HI subjects from Exp. II are shown. The left and right panels are their PTA and CLP, respectively. On the right panels, curves above the horizontal line (0) indicate a left-ear advantage as a function of SNR, and those below the line show a right-ear advantage as a function of SNR. To reduce the clutter, consonants which have less than 20% ear difference are shown as gray lines. Standard errors are also marked on the significant points. Note how panel (b) and (d) show a strong left ear advantage for many CVs.

advantage for most consonants was <20%, with the important exceptions of /ta/, /ba/, and /za/, for which there was an advantage of up to 45%. When asked if she had any ear preference, she reported always using her left ear for the telephone (she is right-handed).

Interestingly, subject HI36 (a,b) of Fig. 3.9 has a better pure-tone threshold for the left ear in the low (.125-.5 kHz) and high (6-8 kHz) frequencies, resulting in a /ba/ advantage in the same ear while subject HI34 (c,d) of Fig. 3.10 showed a discrepancy between the audiograms (better pure-tone threshold in the right ear) and her consonant-dependence (better perception in the left ear). Such findings strongly support the view that the PTA and the CLP are in serious disagreement.

CHAPTER 4

RESULTS OF EXP. IV

4.1 Error Pattern Analysis of NAL-R Amplification

Like Exp. II, one-way ANOVA resulted in the total percent of error (%) of the consonants significantly increasing as a function of SNR (F[3,45]= 73.680, $p<0.01$). Mean of quiet, +12, +6, and 0 dB was 8.659 (SE=2.405), 13.002 (SE=2.837), 16.942 (SE=3.082), and 26.739 (SE=3.306), respectively. In addition, there was a statistically significant difference in the percent of error among 14 CVs (F[13,195]= 8.245, $p<0.01$). /ta, da, na, ʃa/ were the low error consonants, whereas /fa/ had the highest error and /ʒa/, /ba/, /va/, and /ga/ followed as the next highest errors. Compared to Exp. II, /za/ was excluded from the list of the high error consonants, and /na/ lay in the low error consonant group.

In the ANOVA results for entropy, the total entropy significantly increased as noise increased, which means the number of confusions increased (Table 4.1(F[3,45]= 100.306, $p<0.01$). The result of the Bonferroni correction showed that there was significant difference between 0 dB and quiet, +12, and +6 dB. That is, the HI subjects were affected by high noise, such as 0 dB, and had high entropy even though they had a NAL-R correction. Based on a Repeated Measure ANOVA to see the entropy variance of individual consonants, the /fa/ syllable had the highest entropy at 1.122 bits, and /ta/ and /da/ were consonants with the lowest entropy, 0.067 and 0.077 bits, respectively, (F[13,195]=10.755, $p<0.01$).

In addition, the consonant error profile for all 16 HI ears are presented in the

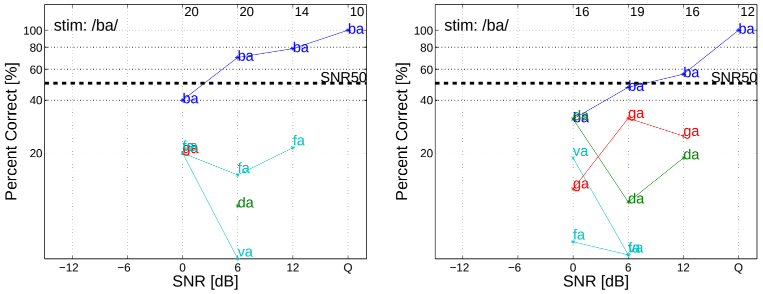
Table 4.1: Results of total entropy calculation of 4 SNRs for 16 HI ears in Exp. IV. Formula of entropy is $\mathcal{H} = -\sum_i p(x_i) \log_2 p(x_i)$. $\mathcal{H}$ is a measure of the subject's response uncertainty. When the entropy is zero, there is no subject uncertainty, independent of the scores ($P_{h|s}$). As noise increased, the entropy was significantly increased (F[3,45]=100.306, p<0.01). Group mean of entropy at quiet, +12, +6, and 0 was 0.209, 0.345, 0.456, and 0.785 bits, respectively. SNR$_1^*$ indicates 1-bit of entropy for Exps. II and IV. The eighth column is the SNR$_1^*$ difference of two experiment.

	Quiet	12 dB	6 dB	0 dB	SNR$_1^*$ for Exp. IV	SNR$_1^*$ for Exp. II	Diff. of Exps. II and IV
HI01L	0.545	0.816	0.979	1.232	6	6	0
HI01R	0.678	0.911	0.975	1.205	6	12	6
HI30L	0.2	0.304	0.521	0.79	<0	2	>2
HI30R	0.379	0.55	0.619	1.064	0	4	4
HI32L	0.166	0.356	0.541	0.932	0	<0	<0
HI32R	0.214	0.459	0.541	1.048	0	0	0
HI34L	0.225	0.353	0.536	0.906	0	3	3
HI34R	0.223	0.427	0.611	1.248	2	5	3
HI36L	0.03	0.11	0.128	0.264	<0	<0	0
HI36R	0.03	0.129	0.2	0.433	<0	<0	0
HI40L	0.059	0.175	0.3	0.735	<0	2	>2
HI40R	0	0.189	0.322	0.548	<0	<0	0
HI44L	0.059	0.059	0.089	0.362	<0	<0	0
HI44R	0.03	0.03	0.141	0.292	<0	<0	0
HI46L	0.254	0.322	0.41	0.782	<0	2	>2
HI46R	0.257	0.337	0.38	0.715	<0	1	>1

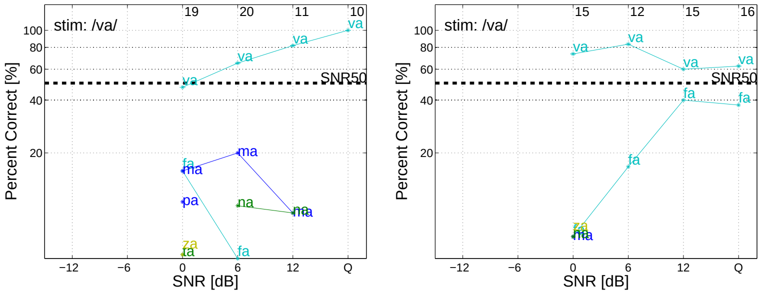
Appendix C (quiet, +12, +6, and 0 dB in Tables C.1, C.2, C.3, and C.4, respectively).

4.2 Comparison of Exps. II and IV: Flat vs. NAL-R gain

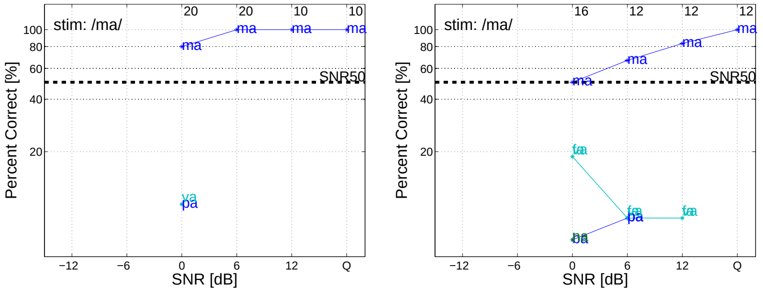
First, we confirmed any change in total percent of errors per individual consonant (i.e., height of each bar), while doing the typical statistical analysis. According to the results of the Repeated Measures ANOVA (2 experiments × 4 SNRs × 14 consonants), there was a significant difference between the two experiments, Exps II and IV (F[1,15]= 6.491, p=0.023). The consonant percent error for Exp. IV (mean=16.336, SE=2.821) was smaller than the one for Exp. II (mean=20.097, SE=3.706). There was a significant difference in SNRs (F[3,45]= 8.0213, p<0.00). The mean of quiet, +12, +6, 0 dB SNR was 9.972 (SE=2.917), 13.990 (SE=3.054), 18.633 (SE=3.384), and 30.271 (SE=3.843), respectively. Error rate differed significantly as a function of consonant (F[13,195]= 8.001, p<0.00) and there was no significant interation



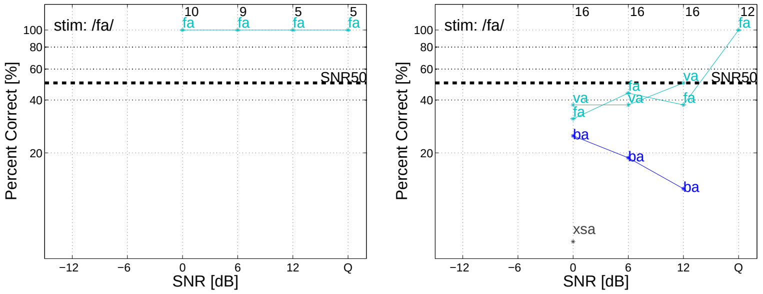
(a) Exp. II: /ba/ (b) Exp. IV: /ba/



(c) Exp. II: /va/ (d) Exp. IV: /va/



(e) Exp. II: /ma/ (f) Exp. IV: /ma/



(g) Exp. II: /fa/ (h) Exp. IV: /fa/

Figure 4.1: Comparison of Exps. II (left panels) and IV (right panels) for utterances /ba/, /va/, /ma/, and /fa/ syllables in HI32. Abscissa indicates SNR and ordinate is percent correct (%). Numbers above 100% line indicate total number of presentation trials at each SNR.

between consonant and experiment. Percent errors for /na/ greatly reduced from 16.419 (Exp. II) to 4.393 (Exp. IV), and /ʃa/ was not much changed between Exp. II (mean=6.413) and Exp. IV (mean=4.447). However, /fa/ and /sa/ were up to 10 % worse when the NAL-R formula was applied.

Second, we measured the dependence of entropy on experiment, SNR, and consonant. Based on the results of a Repeated Measures ANOVA, there was a significant difference between Exps. II and IV (F[1,15]= 13.414, p=0.002). Entropy was 0.568 (SE=0.083) for Exp. II and 0.449 (SE=0.06) for Exp. IV, meaning that after applying a NAL-R correction, consonant confusion was slightly reduced. Also there was a significant difference among the four 4 SNRs (F[3,45]= 106.944, p<0.00). Quiet, +12, +6, 0 were 0.226 (SE=0.058), 0.359 (SE=0.066), 0.511 (SE=0.076), and 0.938 bits (SE=0.099), respectively. The consonants had a significant entropy difference (F[13,195]= 15.057, p<0.00). /va/ and /na/ were less likely to be confused in Exp. IV than in Exp. II, whereas /fa/ was more confused in the NAL-R amplification condition (0.636 for Exp. II and 1.122 for Exp. IV). The /ka/ and /sa/ syllables were similar for the two experimental conditions.

Although NAL-R provides significant benefit on average, Exp. IV has uncovered many specific cases in which NAL-R fails, and in which adjustments in signal strength based on the CLP would provide much greater benefit to HI patients.

Fig. 4.1 shows a comparison between Exps. II and IV at the utterance level. Compared to the left four panels that had no NAL-R amplification correlation, the paired right panels (i.e., NAL-R amplification condition) show that, for these four consonants, the percent accuracy was worse, and that NAL-R created confusions which were not present in the non NAL-R condition. NAL-R improved error rates on average, but if we do not look at the error rates for individual consonants, the negative impact of NAL-R on some consonants will never be discovered, and we will never identify the unsolved problem of HI speech perception.

CHAPTER 5

RESULTS OF EXPS. I AND III

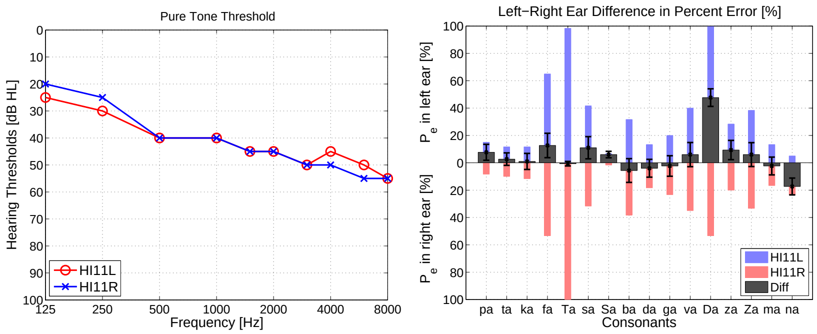
5.1 Analysis of Experiment I

5.1.1 Comparisons between the PTA and CLP

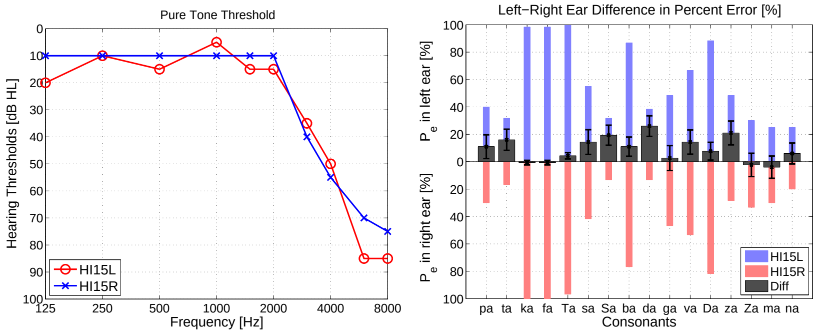
Both Fig. 5.1 and Fig. 5.2 show two PTAs (left panels) along with their CLP (right panels). In Fig. 5.1, two HI subjects show a symmetrical hearing-loss in the left and right ears: (a) high-frequency and (c) high-frequency ski-slope hearing loss. In the Fig. 5.2, panel (a) shows two different HI listeners with nearly identical PTAs, while the HI subject of panel (c) has an asymmetrical PTA.

Each of the right panels shows percent error for each consonant in both left and right ears as blue and red bars from the baseline, respectively. The difference in the percent error of consonant identification between the left and right ears across 16 consonants is presented as block wide bar graphs. The gray bar located above the horizontal axis indicates a right-ear advantage, while the bar below the horizontal axis indicates a left-ear advantage for that consonant.

Since the number of presentations at each SNR was small in Exp. I (N=12), we averaged the error over five noise conditions (quiet, 12, 6, 0, -6 dB) for each consonant, raising the number of utterances from 12 to 60 for each consonant. We did not include -12 dB SNR in this average, since at this level most HI subjects had 100% error in all 16 consonants. We calculated standard error of the mean: $SE = \sqrt{\frac{p(1-p)}{N}}$ where p is probability correct, N is the total number of trials. Three of the four comparisons



(a) Gradual slope: HI11L/11R (b) Consonant-loss of HI11L/11R

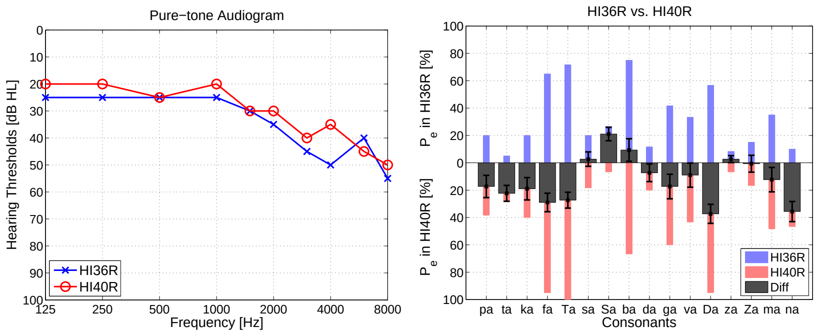


(c) Ski slope: HI15L/15R (d) Consonant-loss of HI15L/15R

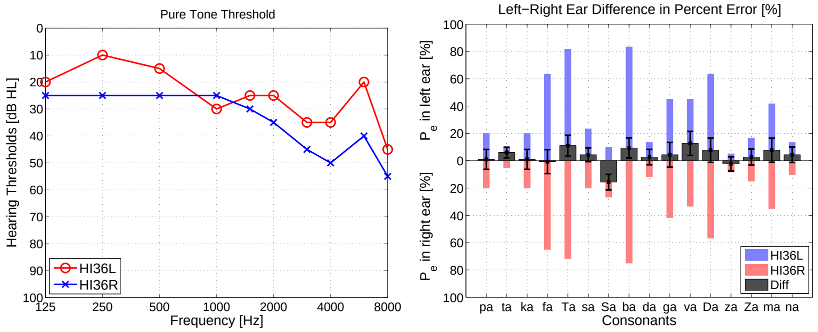
Figure 5.1: The two left panels show PTA results in the HI subjects and the right panels show their consonant loss profiles in left vs. right ears across the 16 consonants. On the right panels, bar graphs present percent error(%) of each consonant in blue for left ear and red for right ear. The gray bars show left ear vs. right ear advantage: above zero shows a right-ear advantage and below shows a left-ear advantage. Error bars indicate 1 standard error (SE): $SE = \sqrt{\frac{p(1-p)}{N}}$ where p is probability correct, N is the number of presentation trials. Even though these subjects have symmetrical hearing loss (a,c), their consonant perception is asymmetrical and is inhomogeneous across consonants (b,d). PTA cannot predict individual HI ears' consonant-loss. *Due to limitation of creating IPA symbols in MATLAB, the consonants, /θa/, /ʃa/, /ða/, and /ʒa/ are displayed as Ta, Sa, Da, and Za, respectively.

showed significantly different consonant-loss profiles between ears. Though we found less than 20% difference for most consonant scores, we consistently found a large difference for a few consonants. Of 20 listeners (40 ears) with a symmetric pure-tone hearing-loss (functionally identical), 17 (85%) had an asymmetrical consonant-loss. Most importantly, three listeners of Fig. 5.1 and 5.2 showed a different consonant-loss between ears except for HI36 (panels [c,d] in Fig. 5.2). These cases are discussed in the following paragraphs.

- **Gradual Sloping High Frequency Hearing Loss** Subject HI11 in Fig. 5.1



(a) Identical PTA: HI36R/40R (b) Consonant-loss of HI36R/40R



(c) Different PTA: HI36L/36R (d) Consonant-loss of HI36L/36R

Figure 5.2: The two left panels show PTA results in the HI subjects and the right panels show their consonant loss profiles across the 16 consonants. On the right panels, bar graphs present percent error(%) of each consonant in blue for the first ear and red for the second ear. The gray bars show first ear vs. second ear advantage: above zero shows a second-ear advantage and below shows a first-ear advantage. Error bars indicate 1 standard error (SE). There is a difference in CLP between two different HI subjects having identical PTA (a). The subject with the asymmetrical pure-tone loss (c) does not have an asymmetrical consonant loss profile (d).

(a,b) had high error rate in /fa/,/θa/, and /ða/ for both ears. The /θa/ syllable had 100% error in both ears. She could not perceive /ða/ with her left ear, but correctly perceive it at 50% in her right ear. The 4 consonants /ta/, /ka/, /ga/, and /ma/ resulted in low error rate and also elicited no significant difference between ears. HI11 has a left-ear advantage of about 18% for /na/, a 46% right-ear advantage in /ða/ and a small 10∼15% right-ear advantage for the /fa/, /sa/, and /za/ syllables.

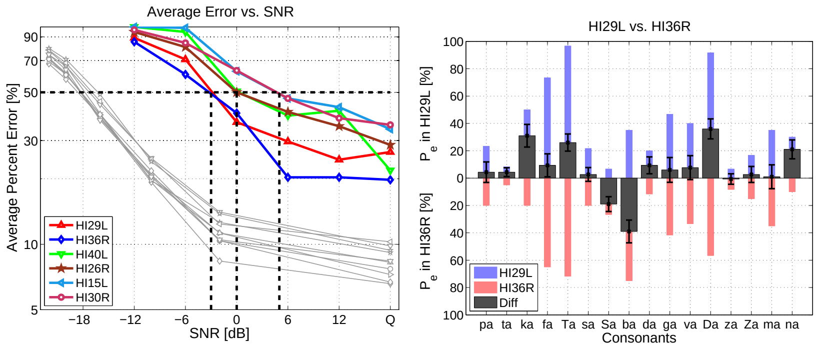
- **Ski-slope High Frequency Hearing Loss** Subject HI15 Fig. 5.1 (c,d) showed 100% error rate for /ka/, /fa/, and /θa/ syllables and about 80% error rate for /ba/ and /ða/ in both ears. Compared to subject HI11, this subject has higher error rates in many consonants although she has a better pure-tone threshold below 4 kHz. In spite of her symmetrical PTA, the subject HI15 showed a right-ear advantage for 12 out of 14 consonants (about 2∼25%). Even though the PTA threshold was 10∼15 dB HL higher (worse) in the right ear than the left ear at 6-8 kHz, her HL could not explain better performance in the right ear even for syllables containing low frequency consonants, /pa/ and /ba/.

- **Identical Audiogram and Different Consonant-loss** Two subjects with identical pure-tone threshold, HI36R and HI40R Fig. 5.2(a,b), show dissimilar error rates and patterns in their consonant perception. HI36R has a lower consonant error rate overall (excluding /ʃa/), compared to HI40R who has almost 100% error rate for /fa/, /θa/, and /ða/ syllables. The largest difference in consonant error rate between the two subjects was for the /ða/ and /na/ syllables, about 38%. Again, their obviously different CLPs are not predicted by their nearly identical audiograms.

- **Dissimilar Audiogram and Same Consonant-loss** Subject HI36 Fig. 5.2(c,d) has an asymmetrical pure-tone hearing loss and about 20 dB HL better audibility in the left ear. However, his consonant-loss profile is not consistent with this difference. Overall, he poorly perceives the /fa/, /θa/, /ba/, and /ða/ syllables, with less than a 20% difference between the two ears. The better audiogram in the left ear does not lead to a left-ear advantage in consonant perception; instead, there is a small right-ear advantage for a number of consonants.

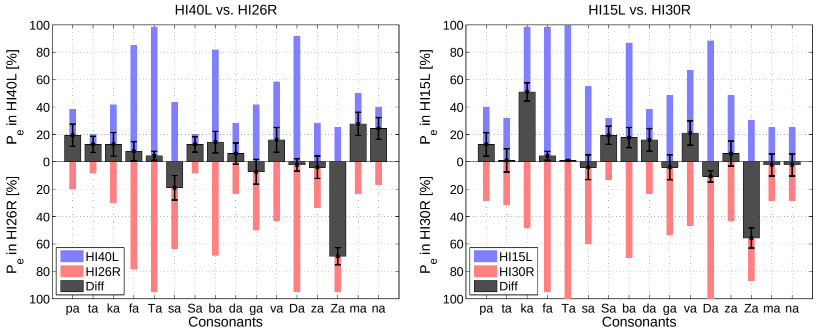
5.1.2 Comparisons between the CRT and CLP

Figure 5.3 shows that consonant-loss and the CRT can be poorly correlated. In Fig. 5.3(a), six HI ears are paired in terms of their CRTs (-3, 0, and 4.5 dB SNR), shown by black dashed lines. Their consonant-loss is shown in sub-figures (b), (c), and (d). Note that the paired ears do not have the same CLP, even though they have the same average consonant scores. In Fig. 5.3(b), although both ears have a CRT of -3 dB SNR, HI29L heard /ba/ 40% better than HI36R. The difference in



(a) 16 consonant average score; CRT

(b) Consonant-loss @ CRT= -3dB



(c) Consonant-loss @ CRT= 0dB

(d) Consonant-loss @ CRT= 4.5dB

Figure 5.3: The CRT and CLP of HI ears are compared. The left top panel (a) shows the CRT threshold defined as the SNR at 50% average error, for six pairs of ears showing the same CRT: -3, 0, and 4.5 dB SNR. The right top and two bottom panels show plots of consonant-loss difference between two ears as a function of consonants. Bar graphs present percent error of each consonant as blue for one ear and red for the other ear. The gray bars show left ear vs. right ear advantage: above the zero line one ear has a higher error (disadvantage), and below the line the right ear has the disadvantage. Error bars indicate 1 standard error (SE). Note that one ear is much better than the other in some consonants although they have same CRT. More specifically note the /ba/ syllable of (b) (40% higher error in HI36R), the /ʒa/ syllable of (c) (65% better perception in HI40L), /ʒa/ and /ka/ on (d) (i.e., better in /ʒa/ and worse in /ka/ to HI15L).

/ba/ perception was up to 60% at 0, 6, and 12 dB SNR (not shown). The ear also performed 20% better for /ʃa/. However, the same ear (HI29L) showed 20∼38% poorer performance for /ka/, /θa/, /ða/, and /na/, when compared to HI36R. In Fig. 5.3(c), HI26R was better than HI40L in most of the CVs. Interestingly, however, HI26R could not correctly perceive /ʒa/ at all, while HI40L could (a 70% difference). Of the two HI ears having a 4.5 dB CRT (Fig. 5.3[d]), HI15L was much better with /ʒa/, while the other ear was better with /ka/.

While the CRTs in this example are consistent with the extent of consonant-loss, they cannot explain the detailed distribution of the CLP. The audiogram configurations were mild flat, mild-to-moderate gradual high frequency, and mild-to-moderate ski-slope high-frequency hearing loss in (b), (c), and (d), respectively. While there was no difference in the average scores and PTAs for the paired ears, their consonant losses profiles differ dramatically as shown by CLP measures. In summary, the ears' consonant perception abilities seem to differ in a manner uncorrelated with their PTA and SRT.

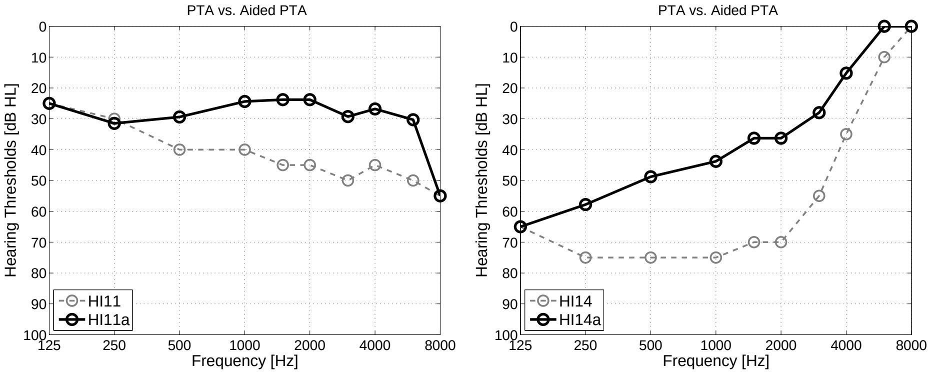
5.2 Analysis of Experiment III

5.2.1 Comparison between the PTA vs. Aided Threshold

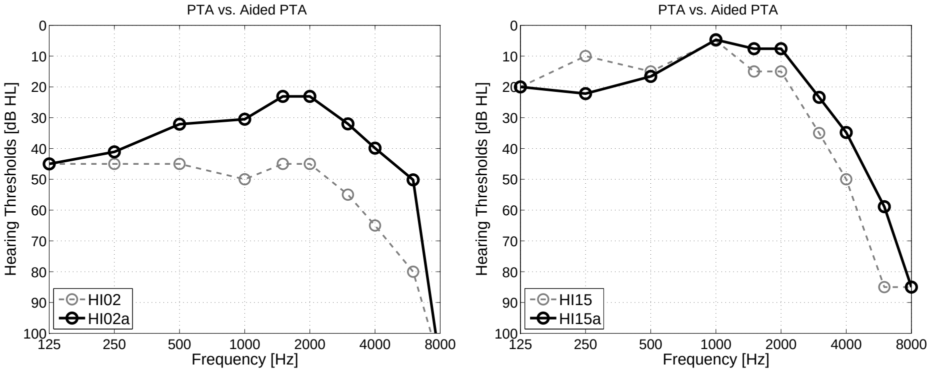
Fig. 5.4 demonstrates how much pure-tone audibility is shifted after applying the NAL-R prescriptive method. Each panel has two audibility curves: a light dashed grey curve for *pure-tone audiogram* (PTA) and a black solid curve for *aided PTA*. Because of no real-ear gain (REG) for .125 and 8 kHz in the NAL-R formula, there was a greater audibility change in the middle frequencies including .5, 1, and 2 kHz. However, there was also an individual difference between *PTA* and *aided PTA* depending on the subject's PTA and calculated REG. Compared to the other subjects, the subject in panel (d) Fig. 5.4 of did not get a change of the aided PTA except for 25 dB at 6 kHz.

5.2.2 Consonant-Dependence

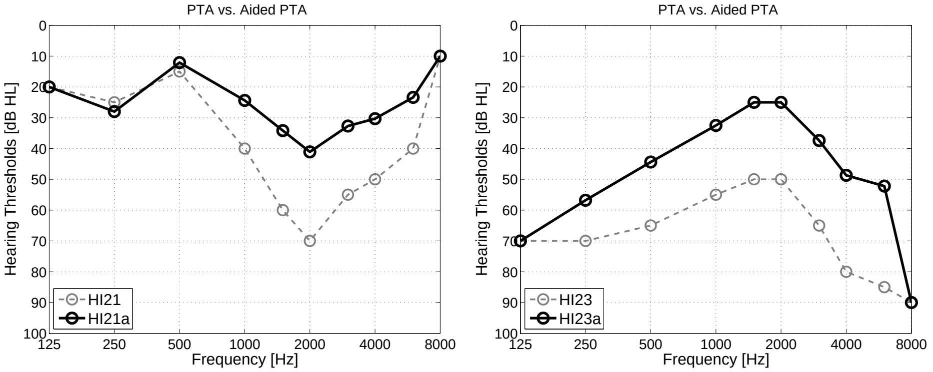
Fig. 5.5 shows three PTAs (left panels) along with their consonant loss profile (middle and right panels). Each of the middle and right panels shows percentage error for each consonant in left and right ears, respectively, as light grey bars from the baseline for no NAL-R condition (using flat gain with MCL) and dark grey bars for



(a) Flat Hearing Loss (b) Low-frequency Hearing Loss

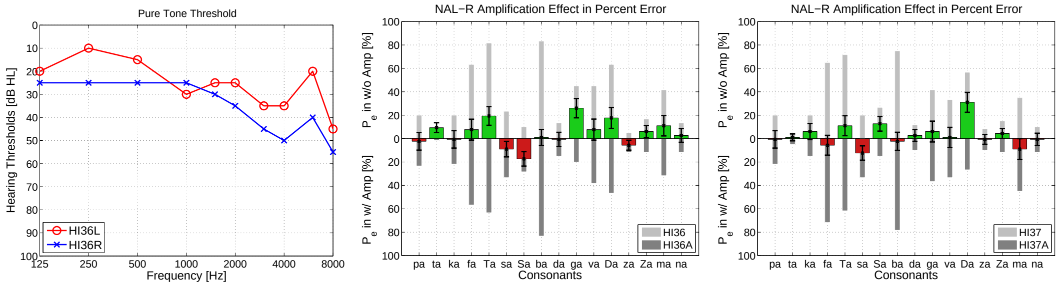


(c) High-frequency Hearing Loss (d) Ski-slope HF Hearing Loss

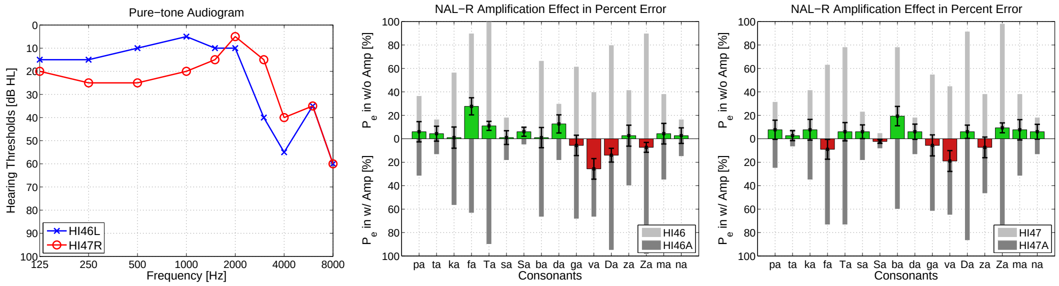


(e) Notched Hearing Loss (f) Reverse-Notched Hearing Loss

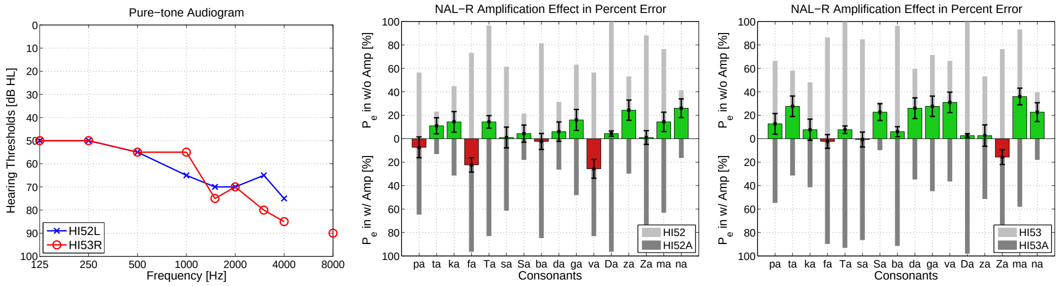
Figure 5.4: Examples of the comparison between pure-tone audiogram (light dashed grey curve) and aided pure-tone threshold (black solid curve) by applying the NAL-R insertion gain to the hearing aids of 6 HI listeners. Each panel represents a different configuration of hearing loss: Flat hearing loss, low-frequency hearing loss, high-frequency hearing loss, ski-slope high-frequency hearing loss, notched hearing loss (or middle-frequency hearing loss), and reverse-notched hearing loss.



(a) PTA: HI36 (b) HI36L (c) HI36R



(d) PTA: HI46 (e) HI46L (f) HI46R



(g) PTA: HI52 (h) HI52L (i) HI52R

Figure 5.5: Consonant-dependence in applying no NAL-R condition at most comfortable level (MCL) vs. NAL-R amplification condition across the 16 consonants. The three left panels show PTA results in the HI subjects and the middle and right panels show their consonant loss profiles in left and right ears, respectively. On the middle and right panels, bar graphs present percent error of each consonant in light grey for no-amplification condition and dark grey for with-amplification. Green bars (above zero) mean NAL-R positive benefit and red bars (below zero) show negative benefit. Error bars indicate one standard error (SE). Note some consonants improve when applying NAL-R amplification and some do not, showing a consonant-dependence.

NAL-R amplification condition. The difference in the percentage error of consonant identification between the no NAL-R and NAL-R amplification conditions across 16 consonants is presented as block wide green and red bar graphs. The green bar located above the horizontal axis indicates a NAL-R positive benefit; the red bar below the horizontal axis indicates a NAL-R negative benefit for that consonant. Since the number of presentations at each SNR was not statistically sufficient in the low SNRs, we averaged the error rates over five SNRs (tested at quiet, 12, 6, 0, and -6 dB) for each consonant, raising the number of presentation trials from 12 to 60. We did not include -12 dB SNR in this average, since at this level most HI subjects had 100% error in all 16 consonants.

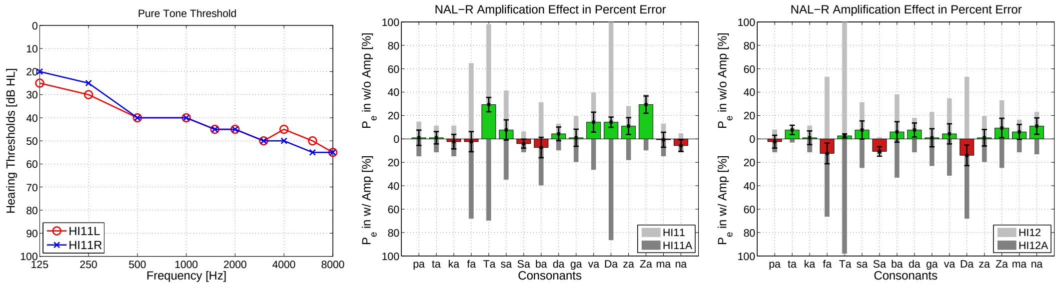
Most importantly, three listeners showed different NAL-R amplification positive/negative benefits at different consonants: some consonants improved up to 38% (positive), yet some were worse 20% or more (negative). In the top panels (a,b,c) of Fig. 5.5, subject HI36 showed positive benefits of 10% or more in/ta/ and /ma/ and 20% or more in /θa/, /ga/, and /ða/ at the left ear (HI36L), and for 10-15% in /θa/ and /ʃa/ and 30% in /ða/ at the right ear (HI36R), whereas there was negative benefit (about 20%) for /ʃa/ and /sa/ sounds for left and right ears, respectively. The /ʃa/ sound resulted in 16% positive benefit for the right ear; in contrast it showed 18% negative benefit for the left ear. In the middle panels (d,e,f), subject HI46 showed the positive benefit for /fa/(28%), /θa/ (12%), and /da/(14%) in the left ear (HI46L) and for /ba/(20%), and /ʒa/(11%) in the right ear (HI46R), whereas /va/(25%) and /ða/ (18%) sounds in left ear and /va/(20%) sound in the right ear were worse in the NAL-R condition than in the no-amplification condition. In the bottom panels (g,h,i), subject HI52 had highly positive benefit in most consonants, with a maximum benefit of 38% for /ma/ (52R). That is, his results showed positive benefit for /ta/, /ka/, /θa/, /ga/, /za/, /ma/, and /na/ in the left ear and /pa/, /ta/, /ʃa/, /da/, /ga/, /va/, /ma/, and /na/ in the right ear, although he also had negative benefit for /fa/ and /va/ in the left ear and /ʒa/ in the right ear. Note that all 20 subjects (40 ears) had different positive/negative benefits of NAL-R amplification for different consonants, even though the amplification condition was fitted to each ear under the same procedure.

5.2.3 Listener-Dependence

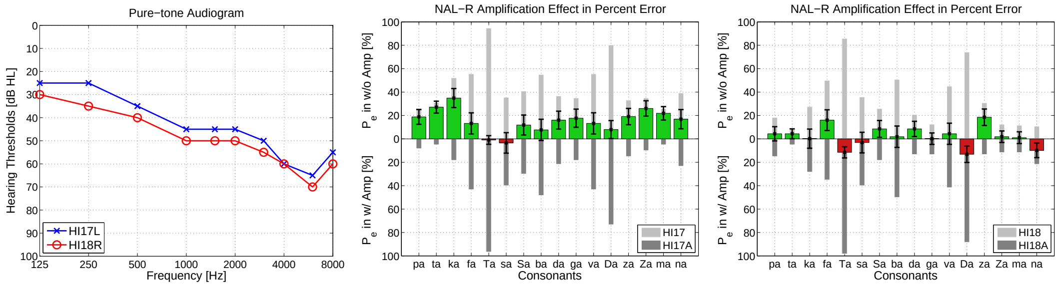
Symmetric Hearing Loss

Fig. 5.6 explains that the subjects who have symmetric bilateral hearing loss (criterion is less than a 10-dB difference of pure-tone threshold between left and right ears at all testing frequencies) do not receive the same benefit of NAL-R amplification for consonants in left vs. right ear. In the first row of panels (a,b,c), the subject HI11 has symmetric mild-to-moderate gradual high frequency hearing loss. She reported an 18-30% positive benefit with NAL-R amplification for /θa/, /va/, /ða/, /za/, and /ʒa/ in her left ear (HI11L) and 10% or more positive benefit for /ta/, /sa/, /da/, /ʒa/, and /na/ in the right ear (HI11R). Although having no negative NAL-R amplification benefit of any consonant on her left ear, three sounds, /fa/, /ʃa/, and /ða/, were worse up to 17% in her right ear after applying the NAL-R amplification. Interestingly, /ða/ sound gave 18% positive benefit to her left ear, but an 18% negative benefit to her right ear.

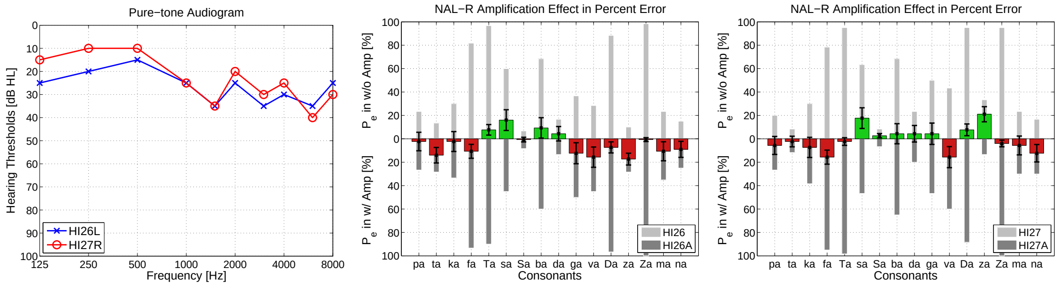
In the second row of panels (d,e,f) of Fig. 5.6, subject HI17 showed positive benefit in most consonants in her left ear (HI17L), whereas her right ear (HI17R) results in about 15% negative amplification benefit for /θa/, /ða/, and /na/; all three of these improved in the left ear, especially /na/(18%-positive). Her left ear seems to be an ideal candidate for a hearing aid. Although her left and right ears showed a very similar degree (41-46 dB HL) and configuration (gradual high frequency sloping hearing loss) in the PTA result, the application of NAL-R amplification to her right



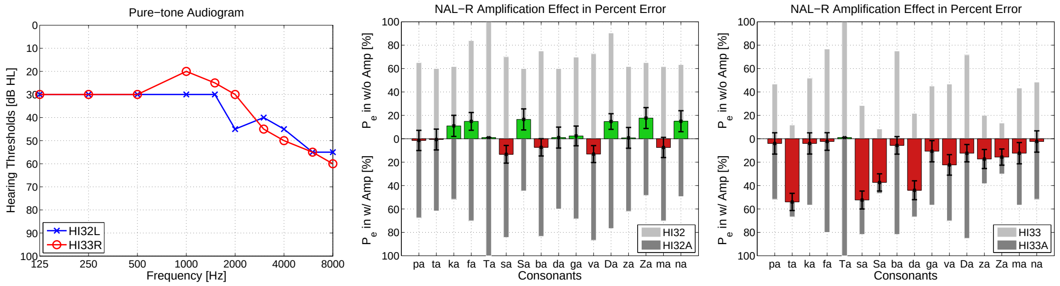
(a) PTA: HI11 (b) HI11L (c) HI11R



(d) PTA: Hi17 (e) HI17L (f) HI17R



(g) PTA: HI26 (h) HI26L (i) HI26R



(j) PTA: HI32 (k) HI32L (l) HI32R

Figure 5.6: Symmetric bilateral hearing loss and asymmetric benefit of NAL-R amplification. The four left panels show PTA results in the HI subjects and the middle and right panels show their consonant loss profiles in left and right ears, respectively. On the middle and right panels, bar graphs present percent error (%) of each consonant in light grey for no-amplification condition and dark grey for with-amplification. Green bars (above zero) mean NAL-R positive benefit and red bars (below zero) show negative benefit. Error bars indicate one standard error (SE). There is a different positive-benefit in NAL-R amplification in left and right ears in four HI subjects despite a symmetric pure-tone hearing loss, showing that their consonant perception is not homogeneous across consonants.

ear did not result in uniformly enhanced speech perception having the amplified sounds.

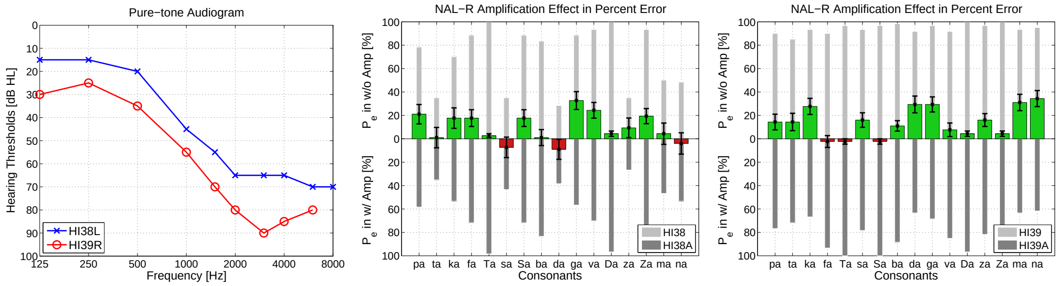
Subject HI26 in the third row panels (g,h,i) showed a 10-17% positive amplification benefit for /sa/ and /ba/ sounds and 20% positive benefit for /sa/ and /za/ sounds in left (HI26L) and right (HI26R) ears, respectively. Although /za/ showed a 20% positive benefit in the right ear, her left ear responded to it with an 18% negative benefit. Including /za/ sounds, subject MB also has negative benefit for /ta/, /fa/, /ga/, /va/, and /ma/ in her left ear, whereas the right ear had negative benefit for /fa/, /va/, and /na/. Compared to the positive benefit in only two consonants per ear after using the amplification condition, her consonant perception was worse overall.

In the last row of panels (j,k,l) of Fig. 5.6, subject HI32 had a positive benefit for /ka/, /fa/, /ʃa/, /ða/, /ʒa/, and /na/, and a negative benefit for /sa/ and /va/ in her left ear (HI32L). Remarkably, her right ear (HI32R) did not have positive benefit for any consonant. Further, the /ʃa/ and /ʒa/ sounds, which showed a positive amplification benefit in her left ear, showed 38% and 26% negative benefit in her right ear, respectively. In addition, she had more than 40% negative benefit for /ta/, /sa/, and /da/ sounds. Despite these findings, her right ear, which she felt had much more difficulty in consonant perception and made high errors in the CV measurement, was not much different from the left ear in terms of PTA results.

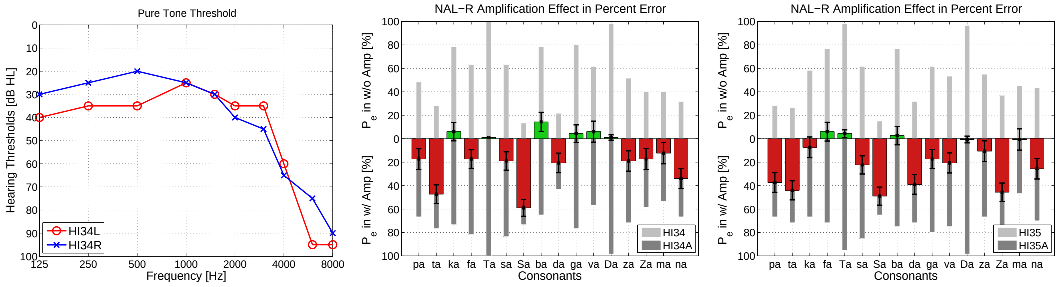
5.2.4 Asymmetric Hearing Loss

Fig. 5.7 shows that subjects who have asymmetric bilateral hearing loss (criteria are at least a 15-dB or greater difference at two or more frequencies) also exhibit consonant perception results that are not predicted by the PTA. The three subjects display obviously different results of consonant loss profiles and positive/negative amplification benefits in left vs. right ears.

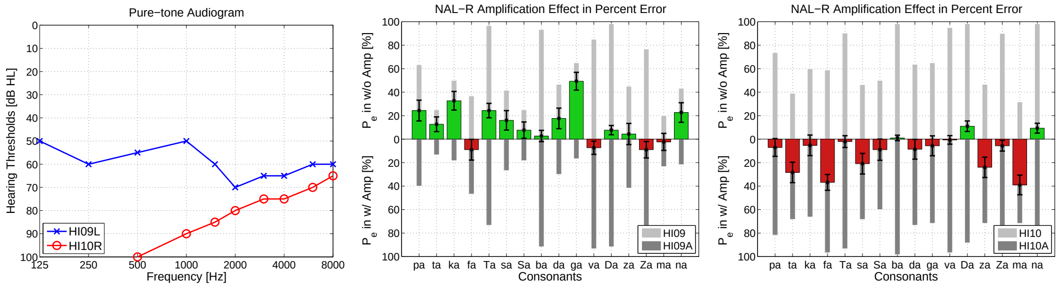
Subject HI38, in the first row of panels (a,b,c) of Fig. 5.7, received an NAL-



(a) PTA: HI38 (b) HI38L (c) HI38R



(d) PTA: HI34 (e) HI34L (f) HI34R



(g) PTA: HI09 (h) HI09L (i) HI09R

Figure 5.7: Consonant perception and NAL-R benefit for the subjects who have asymmetric bilateral hearing loss. The three left panels show PTA results in the HI subjects and the middle and right panels show their consonant loss profiles in left and right ears, respectively. On the middle and right panels, bar graphs present percent error (%) of each consonant in light grey for no-amplification condition and dark grey for with-amplification. Green bars (above zero) mean NAL-R positive benefit and red bars (below zero) show negative benefit. Error bars indicate one standard error (SE). First top panels (a,b,c) show positive benefit in most consonants after applying NAL-R amplification for both left and right ears. Middle panels (d,e,f) show negative benefit in most consonants after applying NAL-R amplification for both ears. The third row panels (g,h,i) show positive benefit in most consonants on her left ear, yet negative in most consonants on her right ear.

R benefit of about 20% in most consonants for the left (except for /sa/ and /da/ sounds) and right ears. Even though the right ear (HI38R) is 10-25 dB HL higher than the left ear (HI38L) in the PTA result, her right ear had more benefit, especially in /ma/ and /na/ sounds. In contrast, subject HI34, who has a similar configuration of ski-slope high frequency hearing loss, had different results from subject HI38. In the second row panels (d,e,f), HI34 had negative benefit in all consonants after applying NAL-R amplification. Except for a positive benefit of the /ba/ sound in the left ear, she heard distorted consonants, resulting in up to 60% worse perceptual accuracy. This result could not be predicted with only the PTA result and NAL-R fitting based on the PTA result, indeed, this result predicts her dissatisfaction with the hearing aid.

As an interesting case, HI09 in the third row of panels (g,h,i) had a positive benefit for /pa/, /ta/, /ka/, /θa/, /sa/, /da/, /ga/(50%), and /na/ in her left ear (HI09L), but negative benefit for /ta/, /fa/(38%), /sa/, /za/, and /ma/(40%) in her right ear (HI09R). Her worse ear (according to the PTA result) did not perceive the consonants clearly with-amplification, contrary to the experience of subject HI38.

CHAPTER 6

DISCUSSION

6.1 Individual differences of HI Consonant Perception

CV syllable scores of HI listeners reveal their CLP.

All SNHL listeners have a loss of both sensitivity and speech clarity (Killion, 1997; Plomp, 1986). The loss of sensitivity is represented by the PTA and can be easily evaluated. However, as Plomp's distortion function and later Killion's SNR-loss express, clarity is not completely described by either PTA or SRT measurements. Our results show poorer consonant perception for most HI listeners in quiet as well as lower SNR thresholds than for NH listeners, with respect to the average scores. This defines an SNR-loss for HI listeners and is consistent with Killion's 1997 results. As shown in Fig. 3.1, all 63 (= 46 + 17) HI ears have significantly higher average error consonant scores versus SNR than NH listeners.

In the current study, we look beyond the average SNR-loss of the HI listener, by investigating individual consonants versus SNR. The consonant recognition does not vary among NH listeners (Phatak and Allen, 2007; Phatak *et al.*, 2008), whereas each HI listener has his own profile; that is, consonant loss is not homogeneous across all consonants. HI consonant confusion is diverse; we have shown that SNHL listeners have SNR-loss with a consonant-dependence, necessitating our new term, *consonant-loss profile* (CLP). The CLP cannot be predicted from existing clinical measurements. To fully diagnose the HI ear, we believe that a CV syllable confusion test must become essential in the clinic and in the research laboratory. Since acoustic features are now known for these sounds (Li *et al.*, 2010, 2011), knowledge of the CLP will allow us to precisely pinpoint cochlear lesions.

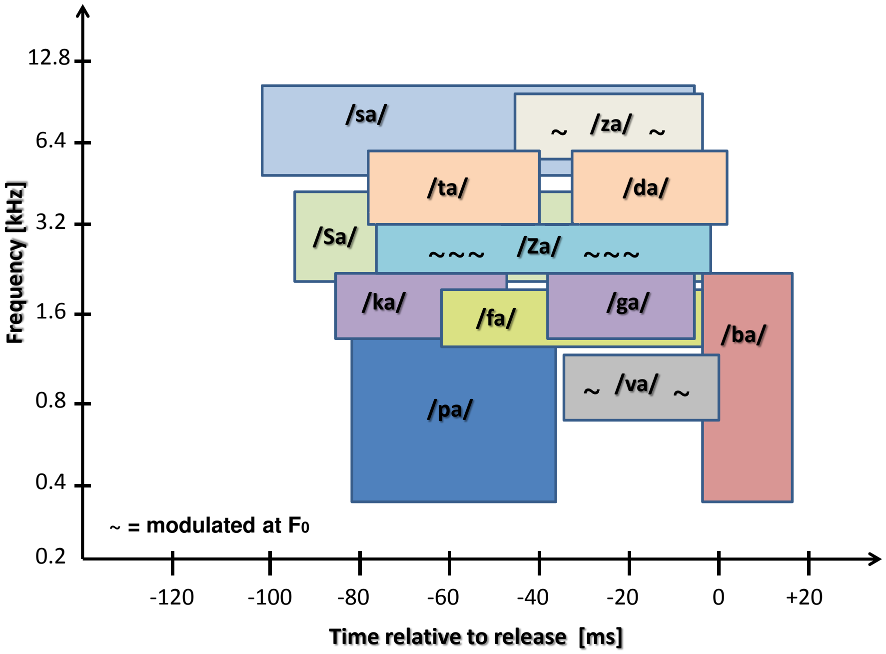


Figure 6.1: Graphical summary of frequency vs. time distribution of the distinctive energy event of English consonants, based on (Li *et al.*, 2010, 2011).

Earlier studies have not fully supported these results. A major problem with many current HI studies is how to analyze the data, specifically, how to quantify the effect of hearing loss and masking noise on the perception of speech cues, assuming that a speech sound is intelligible if the dominant cue is audible. Due to the lack of accurate information about speech cues, most studies can only look at the perceptual score of speech on average, or draw some general conclusions about the correlation between the configuration of hearing loss and the confusion patterns in speech perception without touching the detail.

We propose that, using Fig. 6.1 based on Li et al.'s recent two papers, it may be possible to use the CLP as a replacement for PTA in the fitting of HAs. Specifically, we may hypothesize that there is a specific relation between the inability of SNHL listeners to correctly understand specific frequency and timing cues and their individual consonant confusion matrices (Table 3.2, 3.4, 3.5, and 3.6). By using the CLP together with Fig. 6.1, therefore, it may be possible to get a better understanding of HI speech perception, for individual patients and in general.

The CLP measurement shows a high degree of internal consistency when the number of utterances is increased up to 20, proving the consistency in individual HI consonant perception and the data reliability.

Interestingly, we found that some HI ears show a statistically significant talker-dependence in the consonant perception when data is divided by 2 different talkers (e.g., female vs. male in our experiment) in Exp. II. Some HI subjects did not hear utterances (up to 100%) produced by a female talker, whereas some subjects did not understand the utterance of a male talker (Fig. 3.7). We speculate that different talker genders (or even obvious different voices of the same gender), having different voicing cues and energy in different frequency regions will possibly have different effects on HI consonant/speech perception, and may therefore give detailed diagnosis of cochlear dead regions. For example, the subject of panel (a) of Fig. 3.7 has a problem in perception of /ga/ produced by only the female, but not by the male talker.

6.2 Amplification Effect of Consonant Perception

Current hearing aid fitting formulas including NAL-R do not fully improve the HI consonant perception because their calculations depend on PTA.

When we applied the additional audibility calculated by the NAL-R amplification formula to HI ears (Exps. III and IV), pure-tone audibility was enhanced (Fig. 5.4). In addition, there was a statistically significant difference of HI consonant perception scores between no NAL-R and NAL-R amplification conditions. Overall consonant percent errors were decreased with NAL-R correction, compared to no NAL-R (flat gain) conditions. However, most HI subjects did report that to understand the consonants was not much different between the two conditions, and they sometimes complained that it was more difficult to understand the consonants with NAL-R correction. In other words, when we look at the aided audibility and average consonant errors (or scores) after fitting a hearing aid, the HI speech perception seems better than before wearing the hearing aid. However, we claim that the average score is an insufficient description of the effect of NAL-R.

As we already confirmed in the results of Exp. III, there is a significant difference among consonants: some consonants obtain great benefit from NAL-R and others do not. Also, subjects who have similar pure-tone audibility do not receive the same benefit from the amplification. Therefore, we conclude that although current amplification fitting methods can offer positive benefit on average to the speech perception of HI listeners, they cannot offer equally positive benefits to every consonant and every HI listener. We propose that a more consistent benefit could be obtained by using the CLP measurement for detecting problems of HI speech perception and for a better strategy in fitting hearing aids.

6.3 Relation of PTA and CLP

The fact that there is a significant difference in consonant scores between left and right ears in listeners with symmetrical hearing loss implies that PTA has limitations.

We have confirmed the findings of Phatak *et al.* (2009), that HI individuals have highly CLPs variable across HI listeners that are not well predicted by their PTA scores. In other words, HI individuals with symmetric hearing loss can have asymmetric consonant perception, as shown in Figure 3.9 (a-d), whereas individuals who have asymmetric PTAs can show little differences in CLP between two ears (Fig. 3.10 [a,b]). Earlier studies have supported these results. Killion (1997) states that pure-tone threshold is limited in its utility for predicting speech perception because the loss of audibility and loss of speech clarity (i.e., SNR-loss) are functionally separated. In other words, there is a major difference between hearing speech (i.e., audibility of speech) and understanding it (i.e., intelligibility of speech). Theoretically speaking, patients with *outer hair cell* (OHC) and/or *inner hair cell* (IHC) loss could show the same hearing threshold, yet have different symptoms. This is because damage to the OHCs reduces the active vibration of the basilar membrane at the frequency of the incoming signal, resulting in an elevated detection threshold. Damage to the IHCs reduces the efficiency of transduction (Kujawa and Liberman, 2009). Given identical detection thresholds, it might be that OHCs and IHCs impact speech perception differently. For example, some individuals have a much greater loss of intelligibility in noise than might be expected from their audiogram (Killion, 1997). In order to avoid the limitations of pure-tone audiometry, Killion suggests that the graphic Count-the-Dot Audiogram Method be used to estimate the AI (Mueller and Killion, 1990). This method provides an easy and practical way to clinically measure the degree of the HI patient's loss of speech clarity by computing the number of dots on the audiogram (Killion, 1997). Yet, the method can not give an estimate of the inhomogeneous of audiogram shape on speech perception. The Count-the-Dot Audiogram, like the AI, does not provide information regarding an asymmetry in speech perception between two ears.

Our CV syllable test may explain HI individuals' ear preference when using the telephone. Ten subjects with symmetrical hearing loss were asked about their phone ear preference by e-mail survey. Eight of them reported having an ear preference while using a cell phone, which correlates with their CLP results. Subjects HI15

(Fig. 5.1 [c]) and HI34 (Fig. 3.10 [d]), use their right and left ears, respectively, on the phone, and these are the ears that show an advantage in consonant perception (Fig. 5.1 [d] and Fig. 3.10 [d]), even though both subjects have a symmetric hearing loss as measured by PTA. The CLP may be useful in deciding which ear to fit in cases of monaural hearing aids (HAs), for which purpose it might replace hearing loss threshold as the main variable considered when fitting hearing aids (e.g., NAL-R). The CV test may also predict problems in listeners who have normal hearing but complain that speech is unclear under specific noisy circumstances. Our findings are similar to those of Danhauer and Lawarre (1979), who found no relationship between PTA and CLP.

Dubno and Schaefer (1992) found a correlation between frequency selectivity and consonant recognition using 66 CV and 63 VC syllables for both HI and masked NH listeners (Dubno and Schaefer, 1992). Their results showed that frequency selectivity is poorer for HI listeners than for masked NH listeners. However, there is no difference in consonant recognition between two groups having equal speech-spectrum audibility. A major study completed by Zurek and Delhorne (1987) also revealed that the average consonant reception performance is arguably not significantly different from that of masked NH listeners. They conclude that audibility is the primary variable in speech scores (Zurek and Delhorne, 1987). Note that their argument is based entirely on *average* consonant scores. Here we argue that perception of individual consonants is not dependent on PTA, even when thresholds between the two groups are matched. Thus, our conclusion is the opposite to that of Dubno and Schaefer (1992) and Zurek and Delhorne (1987). We argue for the use of the CLP rather than average scores. The large differences between ears imply a significant cochlear-specific deficiency. Such a difference could be due to specific cochlear lesions, such as a cochlear dead region.

In conclusion, PTA correlates poorly with consonant accuracy, and has never yet been shown to be sufficient in predicting the frequency regions in which consonant perception is damaged. Our speech test precisely identifies the consonant errors, which, when compared to our knowledge of the key frequencies of each speech feature, should allow one to precisely pinpoint dysfunctional frequency regions in that ear.

6.4 Relation of CRT and CLP

Each HI ear has a unique CLP that is not correlated with the CRT/SRT measures.

Although Plomp and his colleagues (1986) proposed the SRT test to connect PTA with speech perception, the SRT is not actually a perception test in clinical practice, rather it is a speech audibility test. Turner et al. (1992) found that consonant detection of HI listeners in a suprathreshold-level masking noise was not different from that of NH listeners. In addition, they explained that HI listeners' poor speech perception might be due to their inability to efficiently utilize audible speech cues (Turner *et al.*, 1992). CRT is well correlated with average consonant recognition, but average recognition is insufficient to describe details of the distribution and is supporting Turner's study, while we have not obtained data from spondee SRT measurements (as is typically used in the clinics). Fig. 5.3 (b-d) shows that consonants may not have the same errors in two ears having the same average CRT scores. We have also demonstrated that the HI ears are difficult to classify, in contrast to 3 performance groupings from the previous study (Phatak *et al.*, 2009). A table summarizing the consonant errors from Exp. II is shown in the Appendix B. It is apparent that the individual consonant errors are not well predicted by the CRT and 3TA. Since HI ears show errors in only a few sounds, average scores or word/sentence scores obscure these unique and relevant errors.

6.5 Limitation of the Studies and Future Directions

We have successfully developed full-rank consonant-confusion matrices as a function of SNR to provide a new clinical diagnostic test for quantifying speech perception in HI listeners. Results of HI speech perception tests indicate that different configurations of hearing loss, such as flat, sloping and cochlear dead region, have a distinct impact on consonant identification. It is generally true that a HI listener cannot hear a sound because the dominant cue that defines the sound is distorted or inaudible due to the hearing loss or masking noise. Under certain circumstances, the HI listener may learn to use a set of minor cues that are ignored by the average normal hearing listeners because of the existence of the dominant cue. This is one of the reasons for which we need to measure HI speech perception using ZE utterances (utterances perceived with ZE by NH listeners), in order to avoid confounding NH and HI problems (or mistakes) and to find unique HI problems.

All the results of the series of studies (I-IV) should make significant contributions to our understanding of HI speech perception. These findings should be applicable to the clinical settings to improve hearing aid fittings and design in the future. This CV test is too time-consuming for clinical use in its current format, but by reducing the number of syllables presented and carefully selecting exceptional tokens, it should be possible to develop a convenient, fast, and statistically viable speech prescription test for clinical HA fitting. It should also be a motivation for further studies in speech perception research related to the clinical practice. Methodology on how to best classify the inhomogeneous HI listeners' error patterns on a consonant-by-consonant basis is difficult. Another concern remains regarding how to effectively amplify consonants having high error rate, without distorting the perceptual cues for an HI listener's intact sound sensitivity (e.g., utterances for which NH subjects have no error in environments where the noise is as low as -2 dB SNR). We will continue to develop a categorical model of HI speech intelligibility, establishing a new "no distortion" amplification formula that is based on individual prescriptive speech scores.The research will help HI listeners hear day-to-day conversations more clearly in both quiet and in noise, and aid in audiological diagnosis and successful rehabilitation to increase speech perception for the HI population.

Future research and several ongoing studies related to the consonant confusion measures will seek to address several possible future goals. The first is to find the relationship between consonant error and cochlear dead regions, analyzing the confusions for clues on specific feature loss. It may be possible to use a test based on the consonant confusion matrices to detect cochlear dead regions as an alternative to existing psychoacoustic measurements (e.g., psychophysical tuning curve and TEN, by Moore *et al.* (2004)), which are not functional for clinical use. We will also study the reverse mapping from confusions to distorted features, given consonant-loss in the CLP.

A second goal is to examine the benefit of amplified speech through our individual consonant-loss measure, our gold standard. Linear and non-linear multi-band amplification, corresponding to a dead region, may not be beneficial and may even impair speech intelligibility (Moore and Alcantara, 2001). Our ongoing studies will explore the problem of speech perception in noisy situations.

Finally, we continue to work on establishing a delicate amplification formula that is based on individual speech scores, applying differential amplification (i.e., manipulating both frequency loudness and feature detection). The goal is to use features in the HI ears to provide no-distortion amplification. Our approach differs considerably from the current clinical amplification formulae because it is very efficient in manipulating relevant speech features; hence, it might benefit both experienced hearing-aid patients and new wearers in terms of auditory plasticity. The study could thus contribute significantly to helping HI listeners hear conversations more clearly and could further aid in audiological diagnosis and successful rehabilitation in the future.

CHAPTER 7

CONCLUSIONS

This series of studies constitutes a step toward better measures for finding detailed characteristics of HI speech perception, using English nonsense CVs. The key conclusions of the studies are as follows:

1. Regardless of similar losses of audibility and consonant average scores, HI ears differ significantly in their consonant loss profiles (CLPs), showing a large variance. This information is unavailable in the clinical PTA and SRT measurements. In other words, PTA or SRT are a necessary, but not a sufficient measurement for hearing-impaired (HI) speech perception. The CV syllable test has much greater utility than existing clinical measurements because it gives detailed information about the characteristics of the HI listeners' feature loss in speech, thus better characterizing hearing loss.

2. Percent error (%) and confusions (i.e., entropy) are significantly increased as noise increases in either flat gain (no NAL-R amplification) or NAL-R condition. Compared to quiet, +12, and +6 dB SNR, the 0 dB condition highly affects consonant percent error (%) of HI listeners. However, confusions are very sensitive to noise, and increase even when the quiet condition is changed to +12 dB.

3. Although average scores of consonants and typical statistical analysis (i.e., grouping HI subjects) in research might give a general summary of HI speech perception, it fails to explain that there is a huge individual HI listener difference when analyzing the accuracies of individual utterances and generating sub-count matrices

per individual HI ear. In addition, required SNR levels are consonant specific to HI ears.

4. Good internal consistency is confirmed for all subjects. We have demonstrated good reliability of the current CV measurement.

5. Preliminary results suggest that some HI ears show talker dependence of consonant perception. This preliminary result should be verified in a future study.

6. When Studies II and IV only use ZE utterances (SNR90 $<$ -2 or -10, respectively) and statistically increase the number of presentations up to 20, we can compute HI consonant error as a function of SNR. SNR-dependent differences in CLP might be developed as a good speech perception tool, and possibly adapted for clinical testing.

REFERENCES

Allen, J. (**1994**), "How do humans process and recognize speech?" Speech and Audio Processing, IEEE Transactions **2**, 567–577.

Allen, J. (**1996**), "Harvey Fletchers role in the creation of communication acoustics," J. Acoust. Soc. Am. **99**(4), 1825–1839.

Allen, J. (**2005**), *Articulation and intelligibility* (Morgan & Claypool, Lexington, KY).

Allen, J. and Rabiner, L. (**1977**), "A unified approach to short-time Fourier analysis and synthesis," Proceedings of the IEEE **65**, 1558–1564.

Allen, J. B. and Li, F. (**2009**), "Speech perception and cochlear signal processing," IEEE Signal Processing Magazine **26**, 73–77.

Bacon, S. and Gleitman, R. (**1992**), "Modulation detection in subjects with relatively flat hearing losses," J. Speech Hear. Res. **35**, 642–653.

Bilger, R. and Wang, M. D. (**1976**), "Consonant confusions in patients with sensorineural hearing loss," J. Speech Hear. Res. **19**, 718–748.

Bilger, R. C., Nuetzel, J. M., and Rabinowitz, W. M. (**1984**), "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Blumstein, S. E. and Stevens, K. N. (**1979**), "Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**, 1001–1017.

Blumstein, S. E. and Stevens, K. N. (**1980**), "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**, 648–666.

Blumstein, S. E., Stevens, K. N., and Nigro, G. N. (**1977**), "Property detectors for bursts and transitions in speech perceptions," J. Acoust. Soc. Am. **61**, 1301–1313.

Boothroyd, A. (**1994**), "Speech perception by hearing-impaired listeners," J. Acoust. Soc. Am. **95**(5), 2998.

Boothroyd, A. and Nittrouer, S. (**1988**), "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**(1), 101–114.

Brandy, W. T. (**2002**), *Handbook of clinical audiology* (Lippincott Williams and Wilkins, Baltimore), chap. Speech Audiometry, 5th ed., pp. 96–110.

Bregman, A. S. and Campell, J. (**1971**), “Primary auditory stream segregation and perception of order in rapid sequences of tones,” J. Exp. Psych. **89**, 244249.

Bronkhorst, A. W., Bosman, A. J., and Smoorenburg, G. F. (**1993**), “A model for context effects in speech recognition,” J. Acoust. Soc. Am. **93**(1), 499–509.

Brungart, D. S. (**2001**), “Informational and energetic masking effects in the perception of two simultaneous talkers,” J. Acoust. Soc. Am. **109**, 1101–1109.

Carhart, R. (**1946**), “Speech reception in relation to pattern of pure tone loss,” J. Speech Disorders. **11**, 97–108.

Carhart, R. (**1958**), “The usefulness of the binaural hearing aid,” J. Speech Hear. Disorder **23**, 42–51.

Ching, T. Y., Dillon, H., and Byrne, D. (**1998**), “Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency,” J. Acoust. Soc. Am. **103**, 1128–1140.

Connine, C. M., Mullennix, J., Shernoff, E., and Yelen, J. (**1990**), “Word familiarity and frequency in visual and auditory word recognition,” J. Exp. Psychol. **16**(6), 1084–1096.

Cooper, F., Delattre, P., Liberman, A., Borst, J., and Gerstman, L. (**1952**), “Some experiments on the perception of synthetic speech sounds,” J. Acoust. Soc. Am. **24**, 597–606.

Cox, R., Alexander, G., Gilmore, C., and Pusakulich, K. (**1988**), “Use of the Connected Speech Test (CST) with Hearing-Impaired Listeners,” Ear Hear. **9**, 198.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**), “Development of the Connected Speech Test (CST),” Ear Hear. **8**(5), 119–126.

Danhauer, J. L. and Lawarre, R. M. (**1979**), “Dissimilarity ratings of English consonants by normally-hearing and hearing-impaired indivisuals,” J. Speech Hear. Res. **22**(2), 236–246.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**), “Modeling auditory processing of amplitude modulation II: Spectral and temporal integration,” J. Acoust. Soc. Am. **102**, 2906–2919.

Delattre, P., Liberman, A., and Cooper, F. (**1955**), “Acoustic Loci and Translational cues for consonants,” J. Acoust. Soc. Am. **27**, 769–773.

Dillon, H. (**2001**), *Hearing Aids* (Boomerang Press, Australia).

Dobie, R. A. and Sakai, C. S. (**2001**), *Noise induced hearing loss basic mechanisms, prevention, and control* (NRN publications, London, UK), chap. Estimation of hearing loss severity from the audigram, pp. 351–363.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**), “Effect of temporal envelope smearing on speech reception,” J. Acoust. Soc. Am. **95**, 1053–1064.

Dubno, J. R. and Dirks, D. D. (**1982**), “Evaluation of Hearing-Impaired Listeners Using a Nonsense-Syllable Test I. Test Reliability,” J. Speech Hear. Res. **25**, 135–141.

Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**), “Evaluation of hearing-impaired listeners using a nonsense-syllable test. II. syllable recognition and consonant confusion patterns,” J. Speech Hear. Res. **25**(1), 141–148.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**), “Benefit of modulated maskers for speech recognition by younger and older adults with normal-hearing,” J. Acoust. Soc. Am. **111**, 2897–2907.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**), “Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal-hearing,” J. Acoust. Soc. Am. **113**, 2084–2094.

Dubno, J. R. and Schaefer, A. B. (**1992**), “Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners,” J. Acoust. Soc. Am. **91**(4), 2110–2121.

Egan, J. (**1948**), “Articulation testing methods,” Laryngoscope **58**, 955–991.

Erber, N. (**1975**), “Auditory-visual perception of speech,” J. Speech Hear. Disorder **40**, 481–492.

Festen, J. and Plomp, R. (**1986**), “Speech-reception threshold in noise with one and two hearing aids,” J. Acoust. Soc. Am. **79**(2), 465–471.

Fletcher, H. (**1950**), “A method of calculating hearing loss for speech from an audiogram,” J. Acoust. Soc. Am. **22**, 1–5.

Fletcher, H. and Galt, R. H. (**1950**), “The perception of speech and its relation to telephony,” J. Acoust. Soc. Am. **22**, 89–151.

Fousek, P., Svojanovsky, P., Grezl, F., and Hermansky, H. (**2004**), “New nonsense syllables database - Analyses and preliminary ASR experiments,” in *The International Conference on Spoken Language Processing (ICSLP)*.

Fowler, C. A. (**1984**), “Segmentation of coarticulated speech in perception,” Perception and Psychophysics **36**, 359–368.

French, N. R. and Steinberg, J. C. (**1947**), “Factors governing the intelligibility of speech sounds,” J. Acoust. Soc. Am. **19**(1), 90–119.

Han, W., Singh, R., and Allen, J. (**2011**a), "Consonant-Loss in Hearing-Impaired Listeners," in *American Auditory Society Scientific and Technology Meeting* (American Auditory Society).

Han, W., Singh, R., and Allen, J. (**2011**b), "The Limitations of using Average Scores in Speech Perception Studies," in *American Auditory Society Scientific and Technology Meeting* (American Auditory Society).

Hazan, V. and Rosen, S. (**1991**), "Individual variability in the perception of cues to place contrasts in initial stops," Perception and Psychophysics , 187–200.

Heil, P. (**2003**), "Coding of temporal onset envelope in the auditory system," Speech Communication **41**, 123–134.

Heinz, J. and Stevens, K. (**1961**), "On the perception of voiceless fricative consonants," J. Acoust. Soc. Am. **33**, 589–596.

Hughes, G. W. and Halle, M. (**1956**), "Spectral properties of fricative consonants," J. Acoust. Soc. Am. **28**, 303–310.

Jenstad, L., Seewald, R., Cornelisse, L., and Shantz, J. (**1999**), "Comparison of linear gain and wide dynamic range compression hearing aid circuits: Aided speech perception measures," Ear Hear. **20**, 117–126.

Kalikow, D. N., Stevens, K. N., and Elliot, L. L. (**1977**), "Development of a test of speech intelligibility in noise using sentene materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Killion, M. and Christensen, L. (**1998**), "The case of the missing dots: Al and SNR loss," Hear. J **51**, 3247.

Killion, M. C. (**1997**), "SNR Loss: I can hear what people say, but I can't understand them," Hear. Reveiw **4**(12), 8–14.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**), "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**(4), 2395–2405.

Kochkin, S. (**Feburary 2000**), "MarkeTrak V: "Why my hearing aids are in the drawer": The consumers' perspective," Hear. J. **53**(2), 34,36,39–41.

Kuhl, P. and Miller, J. (**1975**), "Speech perception by the chinchila: Voiced-voiceless distinction in alveolar plosive consonants," Science **190**, 6972.

Kuhl, P. and Miller, J. (**1978**), "Speech perception by the chinchila: Identi?cation functions for synthetic VOT stimuli," J. Acoust. Soc. Am. **63**, 905–917.

Kujawa, S. and Liberman, M. (**2009**), "Adding insult to injury: cochlear nerve degeneration after temporary noise-induced hearing loss," J Neurosci. **29**(45), 14077–14085.

Li, F. and Allen, J. B. (**2009**), "Multiband product rule and consonant identification," J. Acoust. Soc. Am. **126**, 347–353.

Li, F. and Allen, J. B. (**2011**), "Manipulation of Consonants in Natural Speech," IEEE Trans. Audio, Speech and Language processing , 496–504.

Li, F., Menon, A., and Allen, J. (**2010**), "A psychoacoustic method to find the perceptual cues of stop consonants in natural speech," J. Acoust. Soc. Am. **127**(4), 2599–2610.

Li, F., Menon, A., and Allen, J. (**2011**), "A psychoacoustic method for studying the necessary and sufficient perceptual cues of fricative consonants in noise," Submitted to JASA at June 2011.

Liberman, A. (**1957**), "Some results of research on speech perception," J. Acoust. Soc. Am. **29**, 117–123.

Liberman, A., Cooper, F., Shankweiler, D., and Studdert-Kennedy, M. (**1967**), "Perception of the speech code," Psychol. Review **74**, 431–461.

Liberman, A., Harris, K., Hoffman, H., and Griffith, B. (**1957**), "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psych. **54**, 358–368.

Liberman, A. and Mattingly, I. (**1985**), "The motor theory of speech perception revised," Cognition **21**, 1–36.

MacKeith, N. W. and Coles, R. R. A. (**1971**), "Binaural advantages in hearing of speech," J. Laryng. Otology **85**, 213–232.

Marslen-Wilson, W. and Tyler, L. (**1980**), "The temporal structure of spoken language understanding," Cognition , 1–71.

Marslen-Wilson, W. D. (**1987**), "Functional parallelism in spoken word-recognition," Cognition **25**, 71–102.

McClelland, J. and Elman, J. (**1986**), "The TRACE model of speech perception* 1," Cognitive Psyc. **18**, 1–86.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**), "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**(5), 329–335.

Miller, G. A. and Nicely, P. (**1955**), "An Analysis of Perceptual Confusions among some English Consonants," J. Acoust. Soc. Am. **27**(2), 338–352.

Moore, B. C. and Alcantara, J. I. (**2001**), “The use of psychophysical tuning curves to explore dead regions in the cochlea,” Ear Hear. **22**(4), 268–278.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**2004**), “New version of the TEN test with calibrations in dB HL,” Ear Hear. **25**(5), 478–487.

Moore, B. C. J. and Skrodzka, E. (**2002**), “Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation,” J. Acoust. Soc. Am. **111**, 327–335.

Mueller, H. G. and Killion, M. C. (**1990**), “An easy methold for calculating the articulation index,” Hear. J. **43**(9), 14–17.

Nilsson, M., Soli, S., and Sullivan, J. (**1994**), “Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise,” J. Acoust. Soc. Am. **95**, 1085–1099.

Pavlovic, C. V. (**1984**), “Use of the articulation index for assesing residual auditory function in listeners with sensorineural hearing impairment,” J. Acoust. Soc. Am. **75**, 1253–1258.

Pavlovic, C. V. and Studebaker, G. A. (**1984**), “An evaluation of some assumptions underlying the articulation index,” J. Acoust. Soc. Am. **75**, 1606–1612.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**), “An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,” J. Acoust. Soc. Am. **80**, 50–57.

Phatak, S. and Allen, J. (**2007**), “Consonant and vowel confusions in speech-weighted noise,” J. Acoust. Soc. Am. **121**(4), 2312–2326.

Phatak, S., Lovitt, A., and Allen, J. (**2008**), “Consonant confusions in white noise,” J. Acoust. Soc. Am. **124**(2), 1220–1233.

Phatak, S., Yoon, Y., Gooler, D., and Allen, J. (**2009**), “Consonant recognition loss in hearing impaired listeners,” J. Acoust. Soc. Am. **126**(5), 2683–2694.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**), “How young and old adults listen to and remember speech in noise,” J. Acoust. Soc. Am. **97**(1), 593–608.

Plomp, R. (**1986**), “A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired,” J. Speech Hear. Res. **29**, 146–154.

Rankovic, C. (**1991**), “An application of the articulation index to hearing aid fitting,” J. Acoust. Soc. Am. **34**, 391–402.

Resnick, S. B., Dubno, J. R., Hoffnung, S., and Levitt, H. (**1975**), “Phoneme errors on a nonsense syllable test,” J. Acoust. Soc. Am. **58**(S1), S114–S114.

Ross, M. and Lerman, J. (**1970**), “A picture identification test for hearing-impaired children,” J. Speech Hear. Res. **13**, 44–53.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**), “Speech recognition with primarily temporal cues,” Science **270**, 303–304.

Singh, R. and Allen, J. (**2011**), “Sources of stop consonant errors in low-noise environments,” Under review in J. Acoust. Soc. Am.

Slaney, M. (**1995**), “Pattern playback from 1950 to 1995,” in *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on* (IEEE), vol. 4, pp. 3519–3524.

Smoorenburg, G. (**1992**), “Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram,” J. Acoust. Soc. Am. **91**, 421–437.

Souza, P. E., Yueh, B., Sarubbi, M., and Loovis, C. (**2000**), “Fitting hearing aids with the Articulation Index : Impact on hearing aid effectiveness,” J. Rehabil. Res. Dev. **37**(4), 473–481.

Tremblay, K., Billings, C., Friesen, L., and Souza, P. (**2006**), “Neural representation of amplified speech sounds,” Ear Hear. **27**, 93–103.

Turner, C. W., Fabry, D. A., Barrett, S., and Horwitz, A. R. (**1992**), “Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners,” J. Speech Hear. Res. **35**, 942–949.

Welker, D., Greenberg, J., Desloge, J., and Zurek, P. (**1997**), “Microphone-array hearing aids with binaural output. II. A two-microphone adaptive system,” Speech and Audio Processing, IEEE Transactions on **5**, 543 – 551.

Zurek, P. and Delhorne, L. (**1987**), “Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment,” J. Acoust. Soc. Am. **82**(5), 1548–1599.

APPENDIX A: AGE AND PURE-TONE THRESHOLDS OF HI SUBJECTS

Table A.1: Table summary of age and pure-tone thresholds (from .125 to 8 kHz) of HI subjects who were participated in Exps. I to IV

Sub.	Age	Frequency									
		.125 Hz	.25 Hz	.5 Hz	1 Hz	1.5 Hz	2 Hz	3 Hz	4 Hz	6 Hz	8 Hz
HI01L	82	40	40	45	45	45	45	45	45	65	75
HI01R	82	45	45	45	50	45	45	55	65	80	110
HI05L	52	45	45	40	45	40	30	25	50	55	65
HI05R	52	25	25	35	40	35	30	25	20	40	60
HI09L	21	50	60	55	50	60	70	65	65	60	60
HI09R	21	110	105	100	90	85	80	75	75	70	65
HI11L	44	25	30	40	40	45	45	50	45	50	55
HI11R	44	20	25	40	40	45	45	50	50	55	55
HI13R	25	10	15	25	20	10	5	40	60	70	75
HI14R	25	65	75	75	75	70	70	55	35	10	0
HI15L	63	20	10	15	5	15	15	35	50	85	85
HI15R	63	10	10	10	10	10	10	40	55	70	75
HI17L	27	25	25	35	45	45	45	50	60	65	55
HI17R	27	30	35	40	50	50	50	55	60	70	60
HI19L	34	35	35	45	45	40	35	40	45	45	55
HI19R	34	35	40	50	40	35	30	35	35	50	50
HI21L	50	20	25	15	40	60	70	55	50	40	10
HI21R	50	15	20	15	35	55	65	55	50	35	20
HI23L	53	70	70	65	55	50	50	65	80	85	90
HI24L	61	50	50	65	30	10	25	35	55	55	50
HI24R	61	45	45	55	30	15	15	10	30	50	65
HI26L	57	25	20	15	25	35	25	35	30	35	25
HI26R	57	15	10	10	25	35	20	30	25	40	30
HI28R	67	20	15	10	5	15	30	55	65	70	75
HI29L	60	30	25	20	5	15	20	35	35	45	30
HI30L	66	30	30	25	30	25	35	55	65	70	80
HI30R	66	25	25	25	25	25	30	55	60	90	80
HI32L	74	30	30	30	30	30	45	40	45	55	55
HI32R	74	30	30	30	20	25	30	45	50	55	60
HI34L	84	40	35	35	25	30	35	35	60	95	95
HI34R	84	30	25	20	25	30	40	45	65	75	90
HI36L	72	20	10	15	30	30	35	35	35	20	45
HI36R	72	25	25	25	25	30	35	45	50	40	55
HI38L	88	15	15	20	45	55	65	65	65	70	70
HI38R	88	30	25	35	55	70	80	90	85	80	120
HI40L	79	20	15	25	20	30	20	35	50	45	65
HI40R	79	20	10	25	15	30	30	40	35	45	50
HI42L	24	60	60	55	65	65	70	60	70	75	75
HI42R	24	60	50	50	50	45	55	45	50	60	70
HI44L	65	15	10	5	5	20	20	35	55	20	25
HI44R	65	15	10	10	15	15	20	15	45	25	30
HI46L	67	15	15	10	5	10	10	40	55	35	60
HI46R	67	20	25	25	20	15	5	15	40	35	60
HI48L	26	50	55	50	60	55	55	50	50	80	70
HI48R	26	40	35	40	60	60	60	60	90	80	75
HI50L	59	15	15	10	5	0	-5	10	5	45	40
HI50R	59	10	0	5	5	5	-5	-5	25	40	45
HI52L	59	50	50	55	65	70	70	65	75	120	120
HI52R	59	50	50	55	55	75	70	80	85	120	90

APPENDIX B: INDIVIDUAL CONSONANT ERRORS OF EXP. II

Table B.1: Percent individual consonant error (%) for 17 impaired ears of Exp. II at 12 dB SNR. Each entry represents the error (%) for 14 syllables. Every syllable used in Exp. II is an utterance for which 10 normal hearing listeners have zero error for SNRs ≥ -2 dB, even for 500 trials. Code: **High** (>75%), **medium** (>50% and less than 75%), and *low* (>25% and less than 50%) errors are marked by red, blue, and green, respectively. Empty space indicates zero error. The two right columns display clinical measures; 3TA (3-tone average, dB HL) is calculated by the average of 0.5, 1, and 2 kHz, and CRT (consonant recognition threshold; dB SNR) means the average consonant threshold of 50% error, relative to the SRT calculation. Note how every HI ear makes a high error for a few of consonants.

Sub.	/pa/	/ta/	/ka/	/fa/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/	3TA	CRT
HI46L			*25*	**60**	10		*33*		*29*	*29*	*40*	**84**			8.3	4
HI44L				**67**										9	10	-2
HI44R				**67**							14			9	15	-2
HI46R			20	**69**			18		9		*47*	**95**			16.7	1
HI40L				**95**			*40*			*27*					21.7	0
HI40R	*47*			**69**			14							10	23.3	.5
HI36L				**67**			18								26.7	-2.5
HI32R				**69**	18		21			18	*27*			**50**	26.7	1.5
HI30R	9	18		**70**	**68**		**60**			**60**	**50**	**95**			26.7	4.5
HI36R		10		**60**			**70**								28.3	-3
HI34R	*27*	18		**79**			**56**			*39*	*44*	9			28.3	5.5
HI30L	*33*			*47*	*33*		**74**				**80**	**85**			30	3.5
HI34L				**53**			*25*			21					31.7	7
HI32L				**50**	10		**50**		15	20	20			*47*	35	-
HI01L	**56**		**74**	**75**	*25*	14	9		**90**	**55**	10	**75**	**65**		45	14
HI01R	10	*33*	**100**	**68**	**84**	*27*			**100**	**55**	**73**	**90**	18	*40*	46.7	14.5
HI14R				**69**	9		*40*	18	*33*	*47*	*39*	9	*33*	9	73.3	12

Table B.2: Percent individual consonant error (%) for 17 impaired ears of Exp. II at 6 dB SNR. Note compared to HI32R and 36L who have same PTA, only HI30R show high error in /sa/, /ba/, and /ʒa/.

Sub.	/pa/	/ta/	/ka/	/fa/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/	3TA	CRT
HI46L				**72**			*25*	13	*45*	*30*	*35*	**85**	5		8.3	4
HI44L				**50**	5					5	5			21	10	-2
HI44R				**50**	5									16	15	-2
HI46R			*30*	**70**	10		*30*		*30*	*35*	**50**	**100**	5	5	16.7	1
HI40L	9		9	**100**			**74**		18	*47*	9	9	*29*	*29*	21.7	0
HI40R	*28*			**67**			20		**50**				*30*	21	23.3	.5
HI36L				**50**			*35*			10					26.7	-2.5
HI32R	6	5	10	**50**	20		*30*		10	*35*	*25*			**53**	26.7	1.5
HI30R	16	10		**72**	**75**		**75**			**50**	**63**	**90**		5	26.7	4.5
HI36R				**50**			**100**			20					28.3	-3
HI34R	*28*	20	15	**72**	10		**60**		*30*	**65**	**55**	*35*	10	16	28.3	5.5
HI30L	22		5	*39*	*25*	5	**85**			*40*	**80**	**100**	5	5	30	3.5
HI34L	11			**50**	10		*35*		15	*45*	20	10	5		31.7	7
HI32L				**50**	10		**50**		15	20	20			*47*	35	-
HI01L	**58**		**65**	**80**	**60**	20	*35*		**95**	**70**	*31*	**68**	**80**		45	14
HI01R	**58**	15	**95**	**70**	**90**	**69**	5	*38*	**100**	**60**	*42*	**89**	*30*	**53**	46.7	14.5
HI14R	9			**73**	14		21	*47*	**50**	**50**	*33*	*33*		**63**	73.3	12

Table B.3: Percent individual consonant error (%) for 17 impaired ears of Exp. II at 0 dB SNR. Note as noise increases, HI36L, 32R, and 30R all of same PTA have increased /ba/ error. Yet HI36L has still less error in most consonants except for /pa/ and /ba/. HI36R cannot hear /ba, whereas HI36L misses 50%.

Sub.	/pa/	/ta/	/ka/	/fa/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/	3TA	CRT
HI46L	15		20	**70**			**65**	20	**60**	**84**	*47*	**100**	*45*	15	8.3	4
HI44L				**50**	10	5	5		**50**	11	22	10		15	10	-2
HI44R				**55**	10			10	**50**	*32*	17	15	10	*30*	15	-2
HI46R	20		*40*	**75**	5		**70**	5	**80**	*37*	*28*	**100**	*25*	20	16.7	1
HI40L	*29*		*29*	**84**	14	*29*	**75**		**63**	**60**	10	*44*	**67**	**75**	21.7	0
HI40R	*30*	5	10	**55**	5		**55**	5	**70**	10	5	*25*	**50**	**55**	23.3	.5
HI36L				**50**			**50**	5	5	*26*			15		26.7	-2.5
HI32R	10	15	*40*	*40*	15	10	**60**	5	**50**	**53**	*28*	15	*25*	**65**	26.7	1.5
HI30R	*35*	**65**	*25*	**85**	**75**		**85**	20	*25*	*40*	**70**	**85**	*25*	*30*	26.7	4.5
HI36R	5		10	*45*			**100**	5	15	*42*		5	20	20	28.3	-3
HI34R	**65**	**60**	*45*	**90**	*30*		**75**	*25*	**70**	**74**	**83**	*35*	*40*	*45*	28.3	5.5
HI30L	20		20	**60**	*35*		**100**	15	*35*	**53**	*39*	**85**	*25*	*25*	30	3.5
HI34L	**50**	15	**50**	**70**	*45*	5	**65**	15	**65**	**74**	**72**	15	*30*	*45*	31.7	7
HI32L	5		*30*	**50**	15		**70**	15	*40*	**68**	**50**	15	20	*35*	35	-
HI01L	**85**	5	**65**	**85**	**85**	**61**	**50**	10	**95**	**80**	*42*	**70**	**75**	*25*	45	14
HI01R	**65**	**55**	**100**	**75**	**95**	**100**	**50**	15	**100**	**68**	**67**	**100**	**55**	**65**	46.7	14.5
HI14R	18		9	**67**			**80**	**63**	**68**	**90**	*27*	**67**	*25*	**80**	73.3	12

APPENDIX C: INDIVIDUAL CONSONANT ERRORS OF EXP. IV

Table C.1: Percent individual consonant error (%) for 16 impaired ears of Exp. IV at quiet. Each entry represents the error (%) for 14 syllables. Every syllable used in Exp. IV is an utterance for which 10 normal hearing listeners have zero error for SNRs ≥ -10 dB. Code: **High** (>75%), **medium** (>50% and less than 75%), and *low* (>25% and less than 50%) errors are marked by red, blue, and green, respectively. Empty space indicates zero error. Note how every HI ear makes a high error for a few of consonants. Order of subject is followed to that of Exp. II.

Sub.	/pa/	/ta/	/ka/	/fa/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/
HI46L		8							8	8	**55**	**85**		
HI44L		8						8						
HI44R											8			
HI46R			17						8	8	*40*	**79**		
HI40L				8						8				
HI40R														
HI36L				8										
HI32R				8	17					*37*	*40*			
HI30R				*27*	**62**					*27*	*25*	**50**		
HI36R											8			
HI34R			**50**	17	17		8							
Hi30L				13	**56**					8	8	8		
HI34L			**69**	17	**62**					17				
HI32L					17		20			**62**				
HI01L	8		8	**56**	8	8	13		**90**	**85**	8	20		8
HI01R	**50**		*31*	**65**	**53**		8		**100**	**95**	*27*	**68**		8

Table C.2: Percent individual consonant error (%) for 16 impaired ears of Exp. IV at 12 dB SNR.

Sub.	/pa/	/ta/	/ka/	/fa/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/
HI46L			8				*27*		8	8	*44*	**95**		
HI44L					8							8		
HI44R	8													
HI46R			8	20						20	*33*	**100**	8	
HI40L				**50**	8		*25*							
HI40R				13			*37*			8				
HI36L				20			*25*							
HI32R				**62**	*25*	8	*44*			*40*	*44*			
HI30R	8			**53**	**68**		17				*44*	**75**		
HI36R				17			*33*							
HI34R			8	*31*	*44*		**68**			8	17	8		
Hi30L				8	**56**		47					*27*		
HI34L			**62**	*47*	**58**		8		8	20				
HI32L				**62**	8		**58**				*40*		17	
HI01L	**62**		8	**62**	17	8	**50**		**90**	**100**	8	**69**	**53**	8
HI01R	**62**	13		**65**	*33*	17	*25*	8	**95**	**95**	*25*	**69**	8	17

Table C.3: Percent individual consonant error (%) for 16 impaired ears of Exp. IV at 6 dB SNR.

Sub.	/pa/	/ta/	/ka/	/fa/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/
HI46L				*47*			*47*		8	8	*40*	**100**		
HI44L				8							8	8		
HI44R				*40*		8								
HI46R			13	*31*			*25*	8	8	8	8	**100**	8	
HI40L				**69**	8		*33*				8		20	
HI40R	8		8	**50**			*40*			8			*31*	
HI36L				*40*	8		8							
HI32R				**62**	**58**		**53**			17	**56**	8	*27*	8
HI30R	19			**75**	**63**		**53**			17	*28*	**75**		
HI36R				*25*			**70**			8				
HI34R	8		8	*25*	*37*		**53**		8	*37*	*44*	17		
Hi30L				*47*	*47*		**62**			*25*	8	*33*	8	8
HI34L	8		**56**	**50**	*25*		*47*		17	17		8		
HI32L				**56**	*47*		*44*			17	**56**		*33*	8
HI01L	**62**	17	8	**84**	8	*40*	*37*		**95**	**100**	*25*	**90**	*44*	
HI01R	**75**	17	*44*	**85**	*33*	*37*		8	**100**	**100**	17	**68**	8	*25*

Table C.4: Percent individual consonant error (%) for 16 impaired ears of Exp. IV at 0 dB SNR.

Sub.	/pa/	/ta/	/ka/	/fa/	/sa/	/ʃa/	/ba/	/da/	/ga/	/va/	/za/	/ʒa/	/ma/	/na/
HI46L			*27*	*40*	8	8	*42*		*42*	17	*27*	**95**	*27*	
HI44L				*40*	8				*40*	17		8		
HI44R				**62**			8		*27*	17				
HI46R			*33*	**56**			*25*	8	**58**	17	*25*	**100**	8	20
HI40L			8	**62**	*25*		**68**	8	**50**		8	**50**	**62**	*33*
HI40R	17			**62**			*25*		**56**	*33*			*25*	
HI36L				**56**	8		**75**			8				
HI32R			*44*	*25*	*33*	13	**69**		20	*27*	**56**	**62**	**50**	**58**
HI30R	*37*		*40*	**70**	**62**		**61**		*44*	*40*	*42*	**85**	8	8
HI36R			8	*47*			**80**		*40*	17		13	8	
HI34R	17	8	**56**	**79**	**55**		**62**	17	**60**	**62**	**74**	*40*	8	17
Hi30L			17	**53**	**69**	8	**69**	8	*27*	*31*	*27*	**50**	8	8
HI34L	*31*		**74**	**63**	**62**		*25*	8	**74**	*47*		*40*	17	13
HI32L	8		20	**69**	*25*	8	**62**	8	*27*	*27*	*44*	*31*	**50**	
HI01L	**90**		*40*	**90**	8	**62**	*44*	8	**100**	**95**	*27*	**95**	**50**	
HI01R	**69**	20	**69**	**70**	**62**	**56**	8		**100**	**95**	17	**85**	8	*31*

APPENDIX D: IRB DOCUMENTS

The material in Appendix D may be found in a supplemental file named, *IRB documents*.