

© 2017 Clifton Luther Cole

ON THE EFFECTS OF MASKING OF PERCEPTUAL CUES IN
HEARING-IMPAIRED EARS

BY

CLISTON LUTHER COLE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Jont B. Allen, Chair
Professor Mark Hasegawa-Johnson
Professor Douglas L. Jones
Professor Stephen E. Levinson
Associate Professor Robert E. Wickesberg

ABSTRACT

One of the goals of the Human Speech Recognition (HSR) group is to understand the strategy of the hearing-impaired (HI) ear in detecting consonants. It has been uniformly assumed that audibility is the main factor in speech perception, for both normal-hearing (NH) and HI listeners (Zurek and Delhorne, 1987). Based on an entropy measure, Trevino and Allen (2013) have shown that *at most comfortable level* (MCL) audibility is *not* the main issue for the HI ear. This observation is counter-intuitive. In this research group, we hope to find answers to the following questions: What is the strategy of each HI ear in detecting consonants? How can we determine the subject's strategy? From the 3DDS findings of perceptual cues (Li and Allen 2011; Li et al. 2012), results from two perceptual masking experiments (Li and Allen 2011; Kapoor and Allen 2012), and analysis of work by Han (2011) and Trevino and Allen (2013), we generalize the errors made by an HI ear with up to *four strategies*. *S1*: The frequency of the consonant's primary cue is varied by changing the vowels, which slightly moves the cue frequency. *S2*: The conflicting cues are varied. Different tokens of the same consonant have different confusions, due to conflicting cues. *S3*: The masking of the primary cue is varied. The primary cue for many tokens of the same consonant-vowel is highly correlated with the NH SNR_{90} . *S4*: The number of conflicting cues is varied, as measured by the error entropy. The entropy of a token tells us something about the number of conflicting cues and/or about the ambiguity of the primary cue. In this research, we focus on one strategy, the masking of the primary cue on HI ears, and hope it will lead us in a positive direction of generalization. An extension of three consonant identification experiments is proposed, derived from Miller and Nicely (1955), Li and Allen (2011), and Kapoor and Allen (2012). Both Li and Kapoor showed that masking of primary cue and/or removing the conflicting cues can improve speech perception for NH ears. To determine the strategy of the HI ear in detecting consonants, we study consonant group error patterns. If we can establish error generalizability in the HI ears, we will gain insight into that ear's

decoding strategy.

To my parents, for their love and support.

ACKNOWLEDGMENTS

The author would like to thank Professor Jont Allen for his valuable time and expertise during his continuous advice, and for providing encouragement, support, and sincere co-operation during the research process. The author also wishes to express heartfelt thanks, for their advice and contributions, to members of the author's doctoral committee: Professor Mark Hasegawa-Johnson, Professor Douglas L. Jones, Professor Stephen E. Levinson, and Professor Robert E. Wickesberg. He would like to express his gratitude and thanks to fellow students and members of the Human Speech Recognition group and faculty of Electrical and Computer Engineering for their support during this research. The author would like to express his sincere thanks to those who made it all possible for him: his parents, siblings, nieces, nephews, family, friends, and more importantly, his Lord and Savior, Jesus Christ.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION	1
1.1	What Are Perceptual Cues in Speech Perception?	2
1.2	Current Normal Hearing and Hearing-Impaired Speech Perception	4
1.3	Finding Perceptual Cues	7
1.4	Thesis Outline	8
CHAPTER 2	RELEVANT BACKGROUND IN SPEECH PERCEPTION	10
2.1	Speech Perception of the Communication System	10
2.2	Normal Hearing Speech Perception	12
2.2.1	Acoustic Cues	12
2.2.2	Articulation-Index	13
2.2.3	Three-Dimensional Deep Search	14
2.2.4	3DDS Findings of Plosive and Fricative Cues	15
2.2.5	Normal Hearing Error Patterns for Plosives	16
2.3	Hearing Impaired Speech Perception	17
2.4	Relevant Clinical Speech Perception Measurements and HAs Gain Prescriptions	19
2.4.1	Pure Tone Audiometry	19
2.4.2	Speech Tests Use in Clinics	20
2.4.3	Prescriptive Procedures for Hearing Aids	21
2.5	Modification of Perceptual Cues	22
CHAPTER 3	METHODS	29
3.1	Methods for the Case Study	30
3.1.1	Subject	30
3.1.2	Speech Materials	31
3.1.3	Experimental Procedures	31
3.2	General Methods for Exp. I and Exp. II	33
3.2.1	Subjects	33
3.2.2	Speech Materials	33
3.2.3	Experimental Procedures	35
3.3	Experiment I: Methods for Varying the Masking of the Primary Cue	37
3.4	Experiment II: Methods for the Removal of Conflicting Cues	37

CHAPTER 4	RESULTS	40
4.1	Preliminary Results of a Case Study	40
4.1.1	Case Study Count Matrix – Results	41
4.1.2	Average Probability of Error and Token Errors Analysis for Subject CC	44
4.1.3	Entropy for Subject CC	46
4.1.4	Sorted Error for Subject CC	47
4.1.5	Confusion Patterns for Subject CC	49
4.2	Experiment I: Primary Cue is Varied	51
4.2.1	Sorted Error	51
4.2.2	Analysis of Exp. I	54
4.2.3	Confusion Patterns	60
4.3	Experiment II: The Removal of the Conflicting Cues	63
4.3.1	Analysis of Exp. II	64
4.3.2	Effect of Conflicting Cues for the Average NH Ears	71
4.3.3	Effect of Conflicting Cues for LEG Ears	73
4.3.4	Effect of Conflicting Cues for HEG Ears	74
4.3.5	Effect of Conflicting Cues for f103 /ka/ and m115 /ta/ on the Average NH Ears	76
4.3.6	Effect of Conflicting Cues for f103 /ka/ and m115 /ta/ on the LEG Ears	80
4.3.7	Effect of Conflicting Cues for f103 /ka/ and m115 /ta/ on the HEG Ears	80
4.4	Decoding Strategy for m115 /ta/	91
CHAPTER 5	DISCUSSION	93
5.1	Discussion of Case Study CC	93
5.2	Discussion of Experiment I	95
5.2.1	Sorted Error	96
5.2.2	Confusion Patterns	97
5.3	Discussion of Experiment II	97
5.3.1	Sorted Error	98
5.3.2	Confusion Patterns	98
5.4	Discussion of the Case Study	100
CHAPTER 6	CONCLUSIONS	103
BIBLIOGRAPHY		105
APPENDIX A	EXPERIMENT I RESULTS	111
A.1	Modifications	111
A.2	Sorted Errors	113
A.3	Average Token Error Analysis	116
A.4	Confusion Patterns	121

APPENDIX B	EXPERIMENT II RESULTS	145
B.1	Modifications	145
B.2	Sorted Errors	147
B.3	Average Token Error Analysis	150
B.4	Confusion Patterns	155

CHAPTER 1

INTRODUCTION

Out of 311.5 million Americans living today, there are approximately 12%, or 38 million people, who have significant hearing loss. For every 1000 babies born in the United States, approximately 2-3 have hearing loss. One out of three people over the age of 65 are living with hearing loss in the United States. At least 40% of people aged of 12-35 are exposed to unsafe levels of sound from personal audio devices, clubs, and bars. Having hearing loss as a child has a huge impact on communication skills, including speech production, speech perception, basic education skills, social skills, etc. Specific ways in which hearing loss affects children, according to my personal childhood experiences and friends with hearing loss, are 1) inability to speak clearly, 2) repeating and pronouncing a word when someone could not understand what was said, 3) misunderstanding speech in both quiet and noisy environments, 4) poor reading skills: 50% of classroom discussion may not be understood, and 5) poor interaction skills with peers with normal hearing (NH).

One of the most important factors affecting communication is noise. Even small amounts of noise can cause speech to be heard incorrectly. The characteristics of these noises are either known or unknown; however, they all distort, disrupt, or disguise the quality of speech signals. Therefore, background noise and noisy environments are likely to affect many people, but mostly people with severe hearing loss.

The hearing impaired (HI) usually complain about the performance of hearing aids because the devices do not help them understand speech in both quiet and noisy environments. Hearing aids help most HI-listeners to *decode* the sounds of the noisy speech, but not so much to *understand* them. We shall show that the reason why the HI cannot understand speech in noise is that they cannot hear critical speech cues, due to both hearing loss and the masking effect of the noise. Now that we know the speech cues used by NH ears, in this study we investigate the degree to which that HI ears use these same cues (Li and Allen 2011, Li et al.

2010, Li et al. 2012, and Kapoor and Allen 2012).

In the past, numerous types of analog and digital hearing aids used a variety of signal processing algorithms to improve speech perception (Dillon 2012). Today commercial hearing aids use multiband compression, noise reduction techniques, feedback cancellation, directional processing, adaptive signal processing, and environment classification. However, from my personal experience, and that of close personal friends who wear hearing aids, the signal processing algorithms that I have mentioned above have not significantly improved speech perception. One of the common problems of the current signal processing techniques for hearing aids is the methodology of the amplification strategy. For example, some amplification strategies for hearing aids use an amplification formula that amplifies high, mid, and low frequencies (i.e. National Acoustic Laboratory - Revised (NALR)). A hearing impairment with only high-frequency hearing loss (HL) necessitates amplification only at the high frequencies and not the mid or low. Therefore, for hearing aids to better enhance speech perception, it is essential that signal processing engineers understand the necessary and sufficient perceptual cues that an HI ear uses for correct recognition. We shall show that the failure of past strategies is due to inadequate metrics of performance due to a poor understanding of speech cues.

1.1 What Are Perceptual Cues in Speech Perception?

It is essential to understand the basic concepts of perceptual cues used in speech perception for both the NH and HI ears. Perceptual cues are time-frequency energy patterns of spoken utterances. Perceptual cues can be visualized with analytical tools such as the spectrogram or the Articulation Index (AI) gram (Allen and Li 2009). Details of the AI, or AI-gram, will be explained in Chapter 2. An example of an AI-gram of a /ka/ token spoken by f103 is shown in Fig. 1.1.¹ The ordinate and abscissa represent the frequency [kHz] and time [cs] (1 cs = 10 ms), respectively. All visual time-frequency energy patterns are potential cues. All acoustic features in the AI gram are potential perceptual cues. Only a few are actually used by the auditory system in decoding speech sounds. The audibility of a token is dependent on the acoustic cues. NH-listeners can hear and identify a

¹<http://jontalle.web.engr.illinois.edu/Public/InterspeechDemosAug25.13/ka2ta2pa.m4v>

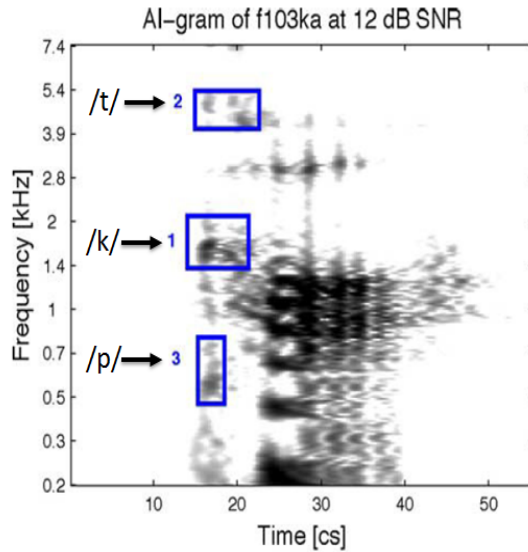


Figure 1.1: AI-gram of a /ka/ token spoken by f103. Adapted from (Li and Allen 2011). The white space is the noise floor (i.e. the background signal-to-noise ratio (SNR)). The three blue outlines of the time-frequency energy patterns are special types of acoustic cues. These cues carry the important information in speech perception.

consonant based on these cues. The white space is the noise floor (i.e. the background signal-to-noise ratio (SNR)). The three blue outlines of the time-frequency energy patterns are special types of acoustic cues. These cues carry the important information in speech perception.

The time-frequency energy patterns associated with the label ‘1’ constitute the primary cue for the /ka/ token. The primary cue of a spoken token is the critical information that the NH listener uses to identify the utterances. Any acoustic cues outside the primary cue are secondary cues. Labels ‘2’ or ‘3’, /t/ and /p/ acoustic cues, respectively, are a unique type of secondary cues for /k/, called conflicting cues. Conflicting cues are opposing acoustic cues to the primary cue. When a talker generates and transmits a /ka/ token, the listener can be confused with either conflicting cue, /t/ or /p/, especially in the presence of noise, typically in the same confusion group of the primary (Allen and Li 2009). Noise can mask or weaken the acoustic cues, and as a result the intended token can end up ambiguous. If either the /ta/ or /pa/ token was spoken, the label ‘2’ and ‘3’, respectively, would be the primary cue. Since we now know from Li and Allen (2011) the location of several tokens’ primary cues that the average NH listeners use, this research demonstrates that HI ears use the primary cue to make their choice, when they

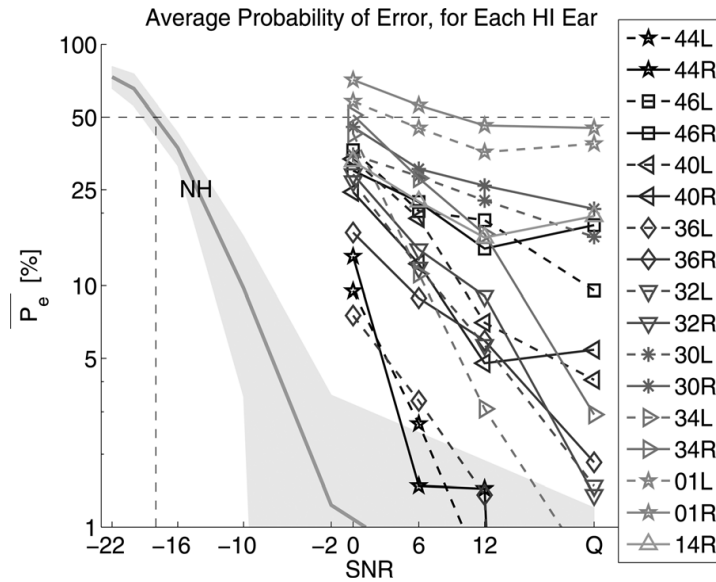


Figure 1.2: The average consonant error for 16 NH and 17 HI listeners (Adapted from Trevino and Allen 2013). The ordinate and abscissa represent the log error [%] and SNR [dB]. The error shaded region with the 'NH' label represents the normal hearing that is one standard deviation relative to the mean of 16 NH listeners. The 17 curves with symbols are the 17 HI ears. The solid lines represent the right ears, while the dashed lines represent the left ears.

are correct. Also, we show that some HI ears depend on a secondary cue to make their selection. This typically results in specific patterns of errors. We view this as a fundamental and important insight into HI speech processing, in quiet and in noise.

1.2 Current Normal Hearing and Hearing-Impaired Speech Perception

One of the goals of the Human Speech Recognition (HSR) group is to understand the strategy of the hearing-impaired (HI) ear in detecting consonants in the presence of noise. Several past studies had attempted to understand and explore why it is difficult for HI listeners to understand speech in noise. What do we presently know about speech perception in an HI ear? It is generally assumed that audibility is the main factor in speech perception, for both NH and HI listeners (Zurek and Delhorne 1987). However, based on an entropy measure, Trevino and Allen (2013) have shown that *at the most comfortable level (MCL)*, audibility is

not the main issue for the HI ear. They argue that if the speech were inaudible, the entropy would be at chance level, while the data indicate a much higher entropy. This research has fundamentally changed the metrics on audibility. As stated on page 1, hearing is not the same as understanding. Furthermore, the errors depend strongly on the token (Trevino and Allen 2013). One /ba/ may have 100% error at a given SNR while another has zero error at the same SNR. They also observed, using data of Han (2011), that a little less than 12 of all target tokens have errors in low noise environments. There is an important result. An analysis of psychoacoustic data gathered from NH and HI listeners is shown in Fig. 1.2, for the average consonant error for 14 naturally spoken American English consonants. The ordinate and abscissa represent the log error [%] and SNR [dB], respectively. This figure shows a systematic difference between the 16 NH listeners and 17 HI ears. Label 'NH' shows the error shaded region for the normal hearing that is one standard deviation relative to the mean of 16 NH listeners. The 17 curves with symbols are the 17 HI ears. The solid and dashed lines represent the right and left ears, respectively. Results show NH listeners have <1% error for SNRs greater than 0 [dB] (i.e. $SNR_{90} > 0$ [dB], meaning that NH listeners have <10% error for SNRs greater than 0 [dB]) in speech-shaped thermal noise (Phatak and Allen 2007; Régnier and Allen 2008). Only four HI ears (44L/R, 36L, 34L) fall within the error shaded region when $SNR > 0$ [dB]. Using the f $SNR_{90} > 0$ [dB] metric, Trevino and Allen (2013) concluded the differences in noise robustness for HI ears are correlated to the noise robustness of consonants for NH listeners. Given what we know about speech perception in an HI ear, the entropy measurement tells us something important about the audibility and ambiguity of the token. For a given consonant, the entropy will be high when the errors spread across other non-target consonants. Given the role of conflicting cues for masked NH listeners (Kapoor and Allen 2012), it is likely that conflicting cues play a substantial role in HI confusions, where the entropy is less than one bit.

Therefore, the main question of the research is: How can we generalize the strategy of each HI ear in detecting consonants? Based on results of two perceptual masking experiments on the findings of perceptual cues (Li and Allen 2011; Kapoor and Allen 2012), we may generalize the error patterns made by an HI ear with up to *four strategies*. *S1*: The frequency of the consonant's primary cue may be varied by changing the vowels, thereby slightly moving the cue frequency. *S2*: The conflicting cues may be varied. Different tokens of the same consonant have different confusions, depending on the relative strengths of conflicting cues. *S3*:

The masking of the primary cue may be varied. The primary cue for many tokens of the same consonant-vowel is highly correlated with the NH SNR₉₀. *S4*: The number of conflicting cues may be varied, as measured by the token error entropy. At least for NH ears, the entropy of a token tells us something about the number of conflicting cues and/or the ambiguity of the primary cue.

In this research, after some preliminary work, we have focused on strategy *S3*, the masking of the primary cue on HI ears, with the hope it will lead us to an understanding of the confusions in HI ears. If we can find patterns in the HI ears, at the token level, this should allow us to generalize the nature of HI errors. An extension of three consonant identification experiments is proposed, derived from Miller and Nicely (1955), Li and Allen (2011), and Kapoor and Allen (2012). Both Li and Kapoor showed that the masking of a primary cue, and/or removing the conflicting cues, can change and even improve speech perception in NH ears. Therefore, in this research, we hope to find answers to the following key unanswered questions:

1. What is the impact of varying the masking of the plosive's primary cue that can be error prone in an HI ear?
2. Do the conflicting cues play any role in increasing the entropy of confusions in HI ears?

This study establishes *error generalizability* by examining errors made by an HI ear due to the masking of the primary and conflicting cues. Error generalizability is our term used to describe natural error patterns in an HI ear's errors. For example, are the errors due to 1) masked primary cues, or 2) the influence of the conflicting cues, or 3) simple audibility? The answers to the research questions will lead us to insights into the HI ear's decoding strategy. If we can determine the strategy of an HI ear identifying a consonant, then may we be able to develop a better amplification strategy to improve speech perception. For example, the present goal of most hearing aid fitting methods is to improve audibility by equalizing loudness across critical bands. An alternative strategy might be to reduce entropy due to the conflicting cues, or to slightly boost the cues where the ear has a critical loss and attenuate the speech where the ear has less loss when the audibility is not the issue. This strategy could lead to better hearing aids by amplifying the vital acoustic cues in speech perception.

1.3 Finding Perceptual Cues

Since the early 1940s, the speech community has conducted studies to investigate perceptual cues for speech recognition (Fletcher and Galt 1950). The conclusion drawn from these studies is that the burst released and transitional cues play an important role for recognizing plosives (Cole and Scott 1974). One may fairly question the utility of the crude spectrograms of this time. Also, synthetic speech has a large disparity in the perceptual cues compared to human speech.

With these questions in mind, the Human Speech Recognition (HSR) group from the University of Illinois at Urbana-Champaign (UIUC) conducted a large number of studies to find the necessary and sufficient cues for speech recognition on NH ears. Unlike the other studies, their method used natural speech and, equally important, the concept of Harvey Fletcher's Articulation Index (Régnier and Allen 2008). Lobdell and Allen (2011) developed the AI-gram model that visualizes the audibility of noise-masked speech, based on the AI. Later, Li and Allen (2011) developed a robust tool for finding perceptual cues called 3DDS, which uses three experiments: time truncation, low and high pass filters, and wideband masking. The identification of the perceptual cues for plosives and fricatives that utilized the AI-gram and 3DDS can be found in two publications (Li and Allen 2011; Li and et. al 2012). These results determined regions in time and frequency for perceptual cues, for correct identification for the plosives and fricatives. There is significant evidence that for NH ears, certain pattern recognition strategies are operating. First are the narrow band bursts (*/p, t, k, b, d, g/*). Second is timing (voiced vs. unvoiced detection, */p/ vs /b/* and */k/ vs. /g/*). Third is low frequency edge detection (*/f, ð, s, /j/*). As with plosives, duration plays a role in */j/ vs. /θ/* and */s/ vs. /ch/*, also in */z/ vs /j/* due to noise modulation.

Given these basic findings, Li and Allen (2011), and Kapoor and Allen (2012) conducted key studies to examine the effects of the burst released, conflicting cues and modification of the primary cue in NH ears. Kapoor and Allen used results from both 3DDS and the AI-gram to locate and then modify the time and frequency regions of the burst release for plosives. To examine the effects of the burst released on NH ears, the time and frequency region of each stimulus was modified in three ways: The burst released was amplified (+6 [dB]), attenuated (-6 [dB]), and removed. Their results clearly show increased identification for tokens when applying amplification and decreased identification when applying attenuation. Li and Allen conjectured that the removal of conflicting cues, and

amplification of necessary and sufficient perceptual cues, can modify speech intelligibility in the presence of masking noise, for both normal and HI ears. Kapoor concluded that for NH ears, the conflicting cue conjecture was true. The goal of the present research is to show that the conjecture is true for HI ears as well.

What is the strategy of the HI ear in detecting consonants? This is actually an open question. All we know is the result of one study on one subject. As discussed in the previous section, we have been exploring four strategies that could lead to the generalization of HI ear error patterns. One of those strategies, varying the masking of the primary cue, has already been conducted on NH ears, and needs to be performed on HI ears. In this research, we will extend both studies, Li and Allen (2011), and Kapoor and Allen (2012), for HI ears. I will examine how well HI listeners perform when the primary cue is amplified, attenuated, and removed, as compared to unmodified. If the HI ears use the same cues as NH ears, then the removal of conflicting cues and amplifying the primary cue should enhance speech intelligibility, as it did for the NH listeners.

Research Goals:

The objectives of the research reported in this thesis are:

1. To determine whether the results of varying the masking of the plosive's primary cue on an HI ear are consistent with the results on NH ears.
2. To examine the masking effects of a plosive's primary cue in HI ears.
3. To determine if the conflicting cues play a significant role when errors are made by HI ears.

1.4 Thesis Outline

This research is organized as follows. In Chapter 2, we review the human auditory system and provide relevant background on speech perception in three areas: NH ears, HI ears, and speech clinics. We describe the AI-gram and its utility in finding perceptual cues with respect to the 3DDS, and summarize key papers related to HI speech perception. Next, we briefly discuss speech measurements and tests used in speech clinics for prescribing hearing aids. Subsequently, we present previous HI experiments and show several methods for analyzing HI

psychoacoustics data from a case study that includes the average error, consonant error, sorted error, response distribution, confusion patterns, and the entropy. Finally, we provide key details of Li and Kapoor, the 3DDS findings and speech modification experiment on the NH ears, respectively.

In Chapter 3, we describe the methodology for the two experiments of our research. We give details about the participants, speech materials, and modification of the stimuli for both experiments.

In Chapter 4, based on the analysis of the case study from the preliminary analysis section, we determine which analysis we will perform for our study to answer the research questions and meet the goals of this research. Next, we examine the results of varying the masking of the primary cue and the removal of conflicting cues for both NH and HI ears. We review a unique and important case study and demonstrate a possible decoding strategy for recognizing m115 /ta/ by one subject having a strange pattern of errors.

In Chapter 5 we begin with a discussion of the preliminary results of a case study. Then we discuss the results of Experiment I (primary cue varied) and Experiment II (conflicting cue removed). Finally, we discuss another case study that demonstrates a possible decoding strategy.

Finally in Chapter 6, we review the research questions, and draw conclusions from our research.

CHAPTER 2

RELEVANT BACKGROUND IN SPEECH PERCEPTION

2.1 Speech Perception of the Communication System

Every speech communication situation involves at least two people, a talker and a listener. An example of a typical speech communication system is shown in Fig. 2.1. To have a conversation with another person, there are two stages to transmit the message to the person receiving the message. The first stage is labeled as the information source. For the case of speech, the information source is the talker's brain. The second stage is the transmitter. In our case acoustically, the message is transmitted using the speaker's vocal apparatus. These stages are the mechanism of speech production.

The beginning of speech perception, or understanding the encoded transmitted speech signal from the speaker's mouth, starts at the ear canal. The receiver represents the human ear of the listener as depicted in Fig. 2.2. As the low impedance induced sound waves (i.e. $\rho c = 403$ [Rayles]) traveling at a speed of 343 m/s impinge the pinna, the impedance starts to increase in the ear canal. The sound waves in the ear canal travel along the auditory canal and vibrate a thin membrane, called the ear drum, that separates the outer ear from the middle ear. About 30% of the energy of the induced sound waves is reflected at the ear drum. The three bones (called the ossicles) in the middle ear match and convert the relatively low impedance of the vibration to high impedance of the fluid wave at the oval window in the inner ear. At the oval window the acoustic wave enters the most complicated and interesting organ in the human auditory system, the cochlea, or inner ear.

In 1863, Herman von Helmholtz compared the cochlea to a harp, where the strings of the harp are like the highly tuned resonators on the basilar membrane (BM), as shown in Fig. 2.3. The harp can only produce sounds when somebody places their finger on the string. Similarly, if a sound is applied at the ear drum,

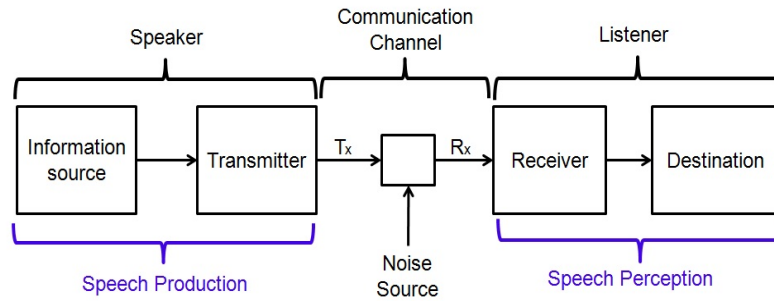


Figure 2.1: Basic speech communication system (adapted from Shannon 1948). Speech production consists of two stages: the information source and the transmitter represent the talker’s brain and mouth, respectively. Speech perception consists of two stages, the receiver and destination represent the listener’s ear and auditory cortex in the brain.

the place of the fluid wave will vibrate on the BM. The location on the BM will vibrate according to frequency input at the stapes; the high frequencies will resonate closer to the stapes than the low frequencies.

The cochlea consists of three major chambers: the scala media, scala vestibuli, and scala tympani. Scala media contains fluid that is very high in potassium called endolymph. Both scala tympani and scala vestibuli contain fluid called perilymph. The inner hair cells (IHC) and the outer hair cells (OHC) sit between the BM and the tectorial membrane (TM), as illustrated in Fig. 2.4. There are approximately 3000 IHC and 9000 OHC that sit on the approximately 35 mm of the BM. The role of the cochlea is to convert acoustic waves at the stapes into tuned neural signals. When the vibration from the sounds excites the BM, the TM shears against the reticular lamina, which causes the stereocilia to move back and forth. This stereocilia movement opens and closes ion channels at its tips, allowing positive ions to depolarize the hair cells. Neurotransmitter is thus released to the afferent auditory nerve fibers, or VIII nerves, which carry the signal to the cochlear nucleus.

The cochlear nucleus is divided into two parts: dorsal (DCN) and ventral (VCN). There are three projections from the cochlear nucleus: the superior olivary complex, lateral lemniscus, and inferior colliculus. Each projection has cells that perform useful functions for hearing speech and music. The signals from the inferior colliculus travel to the medial geniculate nucleus and then terminate in the auditory cortex, where the signals are decoded. It is here where we suspect

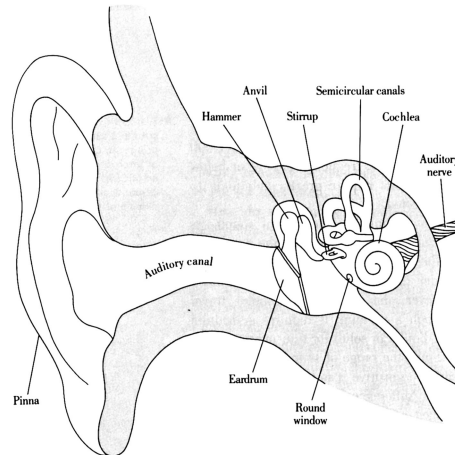


Figure 2.2: A anatomy of the human ear (adapted from Dewey 2007).

speech features are decoded. The “Destination” in Fig. 2.1 represents the auditory cortex.

2.2 Normal Hearing Speech Perception

2.2.1 Acoustic Cues

The fundamental problem of speech perception is the relationship between the acoustic cues and perceptual units (Fletcher and Galt 1950; Allen 1996, Allen and Li 2009). The first search for acoustic cues goes back to the 1940s at Bell Laboratories when Potter et al. (1966) initialized their visible speech project. At Haskins laboratory during the early 1950s, Cooper et al. (1952) and colleagues developed a speech synthesizer called pattern playback to study the acoustic cues of consonants. Based on their observations from the spectrograms, the acoustic cues of plosive consonants can be found in their initial burst and the delay to the following vowel. Starting from the mid-1950s, speech synthesis was the standard tool used for investigating a variety of acoustic features such as the plosive (Blumstein et al. 1977), fricative (Hughes and Halle 1956; Heinz and Stevens 1961), nasal (Liberman 1957) and various articulatory features (Stevens and Blumstein 1978; Blumstein and Stevens 1980). Due to the large variability of acoustic cues in naturally produced speech, speech synthesis was the standard method for investigating the acoustics cues. However, the quality of this synthetic speech was very low, and

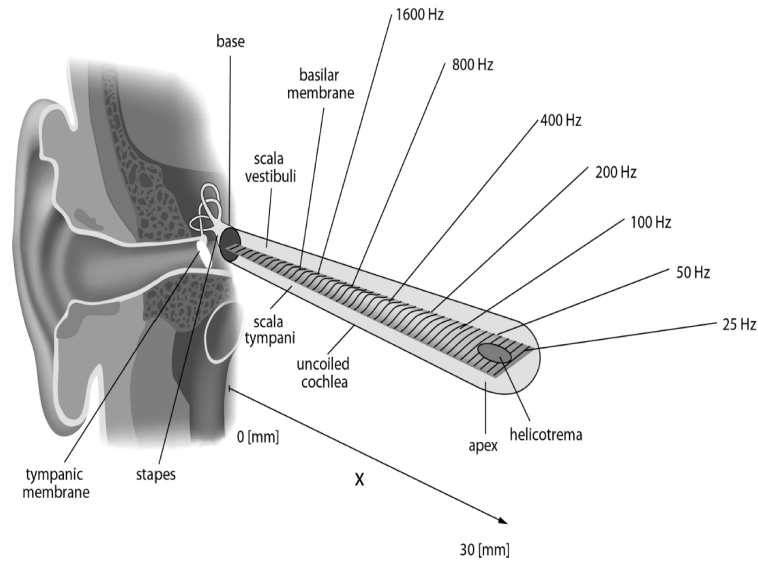


Figure 2.3: Uncoiled cochlea (adapted from Augustine et al. 2001).

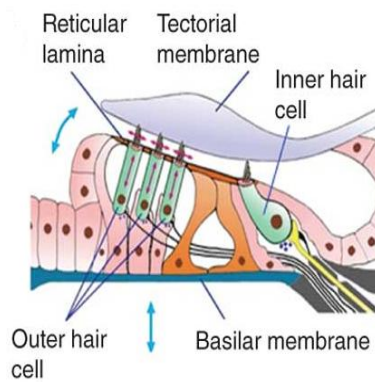


Figure 2.4: Organ of Corti.

was almost unintelligible. To deal with this problem, only two-way comparisons were made, namely a 1 bit task. Consonant and vowel recognition is close to a 4.3 bit task. In the following section, we look into AI-gram and 3DDS methods, as first discussed by Régnier and Allen (2008), and Allen and Li (2009), for finding consonant cues of natural English spoken consonant-vowel pairs.

2.2.2 Articulation-Index

Dating back to the early 1920s, Fletcher and his colleagues initiated and conducted several noise masking experiments on low and high pass filtered consonant-vowel (CV) pairs, to investigate the effects and contributions of different fre-

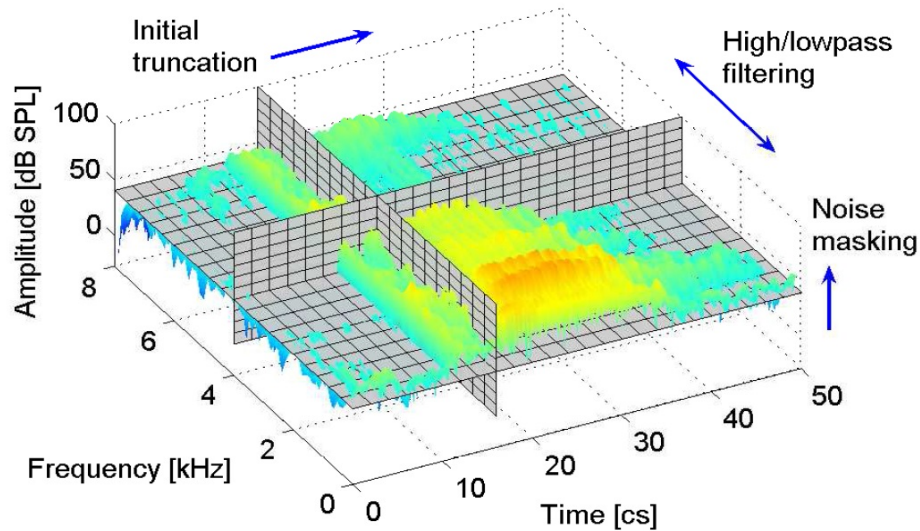


Figure 2.5: Schematic diagram of 3DDS (adapted from Li 2010).

quency bands in speech perception. These classic experiments are the basis for the American National Standards Institute (ANSI) standard for the articulation index (AI). From Fletcher’s AI model, a computational model and visualization tool called the AI-gram, that simulates the human auditory peripheral processing of the speech signal and the impact of noise on audibility, was developed by Lobdell and Allen (Lobdell et al. 2011). The AI-gram may be viewed as an image of a time-frequency representation of a speech stimulus with added noise. This computational model was used to develop additional tools for cue identification experiments, as discussed next.

2.2.3 Three-Dimensional Deep Search

On the BM, there are time-varying energy patterns that represent the corresponding speech sounds. Any sound that is audible by the central auditory system contributes to speech perception. The AI-gram is an important tool for analyzing these audible components by providing an image of speech sounds as represented on the BM. However, the AI-gram cannot be used alone to identify those audible speech components without additional experiments, because only a subset of features are perceptually relevant.

Previously at the University of Illinois at Urbana-Champaign (UIUC), Li and Allen had conducted extensive speech tests on NH participants to identify plosive,

fricative, and nasal consonants cues; they called this method the 3-Dimensional Deep Search (3DDS). It consisted of three different independent speech perception experiments, to identify the perceptually significant audible components for speech perception, as shown in Fig. 2.5 (Li et al. 2010). The purpose of these speech tests was to identify the acoustic cues that listeners use for consonant identification. These acoustic cues are time-frequency energy patterns that carry the consonant code. The image of the audible component may be visualized by the AI-gram. The 3DDS experiments modified and isolated the sound in time, frequency, and intensity (i.e. SNR). The 3-D approach quantifies the strengths of the acoustic cues.

2.2.4 3DDS Findings of Plosive and Fricative Cues

The 3DDS technique has been utilized to investigate the perceptual cues of plosive consonants by Li et al. (2010), fricatives by Li et al. (2012) and nasal consonants. The important cues that are necessary for speech perception of plosives are shown in Fig. 2.6-left. The voice onset time (VOT) defines an essential feature when recognizing and separating the voiced plosives (/ba, da, ga/) from the unvoiced plosives (/pa, ta, ka/). The relationship between the fundamental cue region and the robustness to noise has been shown to play a key role in speech perception (Régner and Allen 2008; Li et al. 2010). Régner and Allen demonstrated that the strength of the burst feature region (voice onset) determines the consonant thresholds to noise. Both Li and Kapoor then showed that most natural speech sounds contain conflicting cue regions that lead to confusions (Li and Allen 2011; Kapoor and Allen 2012). Note that synthetic speech does not contain conflicting cues. Li and Allen, and Kapoor and Allen conducted experiments that improved speech perception in NH ears by manipulation of the conflicting cues and primary burst feature region.

Li et al. (2012) then followed up with the 3DDS technique on six fricative American English consonants (/f/, /s/, /ʃ/, /v/, /z/, /ʒ/), followed by the /a/ (Miller and Nicely 1955). The 3DDS findings for these cues are highlighted in Fig. 2.6-right. The sustained frications for alveolar consonants /sa, za/ and palato-alveolar consonants /ʃa, ʒa/ have their cue region lower than 2 [kHz] and between 1.3 and 3.6 [kHz], respectively. The cue region is between 0.6 and 1.7 [kHz] for the voiceless and voiced labiodentals /fa, va/. The frication noise

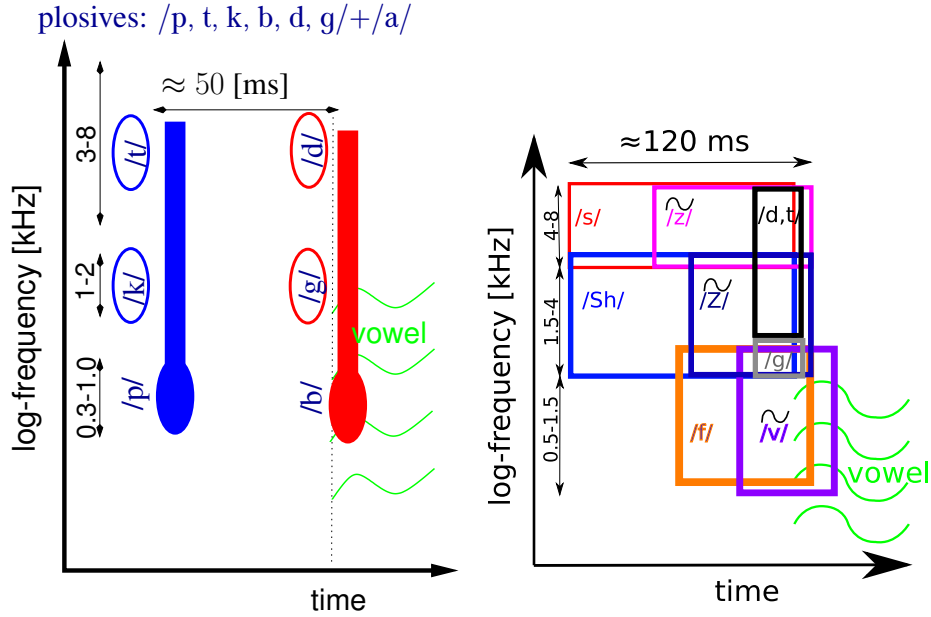


Figure 2.6: Schematic summary plots for the results found by the 3DDS method applied to plosives (left panel) and fricatives (right panel). On the left is the time and frequency location of primary cue for the voiced plosives (/ba, da, ga/) and the unvoiced plosives (/pa, ta, ka/), while on the right is the primary cue for the fricatives (/f/, /s/, /ʃ/, /v/, /z/, /ʒ/).

for the voiced sibilants is modulated by the pitch fundamental. Li et al. (2012) concluded, by applying a high pass filter and removing the entire low frequency spectral region of voiced fricative, that the high-frequency modulation is sufficient for speech perception. Therefore, high-frequency modulation is an important cue for the voice fricative.

2.2.5 Normal Hearing Error Patterns for Plosives

In the presence of noise, how do the NH ears perform when identifying plosives? Using results of Phatak and Allen (2007), Singh and Allen (2012) showed that $\approx 80\%$ of the plosives have zero-error (ZE) for $\text{SNR} \geq -2$ [dB]. Furthermore, the error patterns for the NH ears behave as a binary process or step function as a function of SNR. The errors are essentially very low before their rate increases dramatically, at a specific critical low SNR threshold. Above SNR threshold, the NH ears perform very well at locating and listening to critical acoustic cues of each sound.

2.3 Hearing Impaired Speech Perception

In speech communication, 58% of the words spoken are consonants (Mines et al. 1978). However, in the speech research community, the subject of which speech sound, consonant or vowel, plays a more important role for speech perception is still being debated (Hood and Poole 1977; Burkle et al. 2004). For example, there are several papers that favor vowels as more important than consonants (Kewley-Port et al. 2007; Cole et al. 1996), while others favor consonants (Miller 1951, Li and Allen 2011). Since our study focuses on experiments that relate to HI consonant perception, this section examines other relevant works that relate to HI consonant perception.

Several classical papers explored the impact of HI consonant recognition utilizing naturally spoken phonemes (Lawrence and Byers 1969; Bilger and Wang 1976; Owens 1978; Wang et al. 1978; Dubno and Dirks 1982; Boothroyd 1984; Fabry and Van Tasell 1986; Dreschler 1986; Gordan-Salant 1987; Zurek and Delhorne 1987). It is well known that for speech perception in HI ears, the performance of recognizing consonants is highly correlated with the SNR. The average consonant correct score is often utilized in past studies. In our studies, we use log error [%] as a more useful error, in keeping with the latest AI studies. However, using speech as a measure is not yet well established (Wilson et al. 2007; Killion and Gudmundsen 2005).

Owens (1978) examined consonant errors and consonant confusions for HI listeners using low context stimuli. He developed the California Confusion Test (CCT), utilizing CVC sounds in quiet condition, to investigate consonant confusions. His speech test was a multiple choice, with four selections to choose from. One of the conclusions was that the consonant confusion groups for the HI listeners were similar to NH listeners, except for two or three other consonants. Also, a second conclusion of this study regards the talkers of the same study. HI listeners tended to perform better when the talker was more closely related to the HI listener, such as relative, friend, or co-worker.

Dubno et al. (1984) examined the reliability of their nonsense speech test on HI subjects using a statistical analysis. Thirty-eight HI listeners with mild-to-moderate sensorineural hearing loss were divided into three groups corresponding to their audiogram configurations. Pure tone audibility was assumed to be the most important variable for speech, a questionable assumption. The stimuli were consonant-vowel (CV) and vowel-consonant (VC). For each trial, 91 differ-

ent nonsense syllables were presented at 90 [dB] SPL, with cafeteria background noise at +20 [dB] SNR. Eight trials were used to examine the reliability and variability of the test. For each group of HI listeners, the responses were averaged. To investigate the reliability of their test, they measured the HI listener's ability to recognize nonsense syllables in the presence of background noise. The off-diagonal elements of the confusion matrices were also analyzed to further investigate the variability of the test. Results of the eight trials indicated that there were no systematic differences in scores. They concluded that the nonsense speech tests were highly reliable for evaluating consonant confusion patterns of the HI listeners.

Gordan-Salant (1987) conducted an experiment utilizing nonsense syllables, to investigate the correlation between consonant recognition or confusion patterns with hearing loss (HL) in 30 elderly listeners ranging from 65 to 75 years. Based on the audiometric configuration for each of the HI listeners, they were divided into either of the three HL groups: flat, gradually, and sharply-sloping (Dubno et al. 1984). A total of 57 stimuli were used in the experiment: Nineteen consonants paired with each of three vowels. The stimuli were presented at two levels, 75 and 90 DB SPL, with a +6 [dB] signal-to-babble (S/B) ratio. Due to the many methodological similarities between Gordon-Salant and Dubno studies, results may be compared. Several results in the Gordan-Salant study contradict Dubno. For example, one of the findings of Dubno et al. (1982) that relates to consonant recognition is that listeners with gradually sloping audiometric configurations outperformed listeners with sharply-sloping audiometric configuration for every manner and place category. This result may indicate that younger HI listeners with the same audiometric configurations as elderly HI listeners do not have similar characteristics. Younger HI listeners with moderate HL (40 -70 [dB]) may perform better than elderly HI listeners with flat HL (0-20 [dB]). A second possibility was that S/B ratio for each study had a 14 [dB] difference. One point of agreement was that the confusions do not differ among the three groups.

Zurek and Delhorne (1987) explored why it is difficult for HI listeners to understand speech in noise. Zurek studied the role of audibility by matching each HI ear with a masked normal. They then measured $P_c(SNR)$ for each pair. They had grouped the listeners by severity of HL. They found that HI listeners were arguably better at the task than NH. Overall, their finding suggested that the reduced audibility of speech cues due to HL and external noise was the key factor in HI listeners' difficulty understanding speech in noise. Therefore, they concluded that the impact of supra-threshold deficits is insignificant compared to the impact

of the loss of audibility. Thus they questioned the results of Plomp and his colleagues that HI ears had an additional deficit beyond audibility, which they called “distortion,” or “SNR loss.” Thus Zurek and Delhorne concluded that audibility explained the HI loss. One serious flaw in this approach is the use of averaging across all the sounds. By failing to report confusion, they may have come to the wrong conclusion, as we shall explain in the present study.

2.4 Relevant Clinical Speech Perception Measurements and HAs Gain Prescriptions

When an individual experiences some difficulty understanding speech or feeling discomfort within the ears, he/she may have some type of HL. Usually when someone experiences symptoms that relate to hearing, they visit the audiology clinic where the hearing problem is diagnosed based on the results of clinical auditory threshold measurements and evaluations. Several types of measurements may be conducted to determine if the patient has HL. In this research we are specifically concerned with measurements that use speech (Han 2011). The common tests that use speech are Speech Recognition Thresholds (SRT), HINT, QuickSIN, BKB-SIN, and WIN (Wilson et al. 2007). A flow chart of the typical basic audiological clinical procedures for patients that are experiencing HL is shown in Fig. 2.7.

2.4.1 Pure Tone Audiometry

In the audiology clinics, the first test that a patient would be given is pure-tone audiometry (PTA), which measures audible thresholds of particular frequencies. The results of the PTA are displayed on a graph called an audiogram. Examples of an audiogram of a right ear and left ear is shown in Fig. 2.8(a). The audiogram shows the hearing level vs. frequency. The PTA can be labeled as NH or HI. A patient is considered NH if he/she has HL between 20 [dB-HL] and -10 [dB-HL]. Depending on the amount of significant HL, patients may be classified as mildly, moderately, severely, or profoundly HL. The audiogram can be used to determine which sounds are audible by comparing with the “speech banana,” which shows the hypothetical range of speech sounds in conversational speech. An example

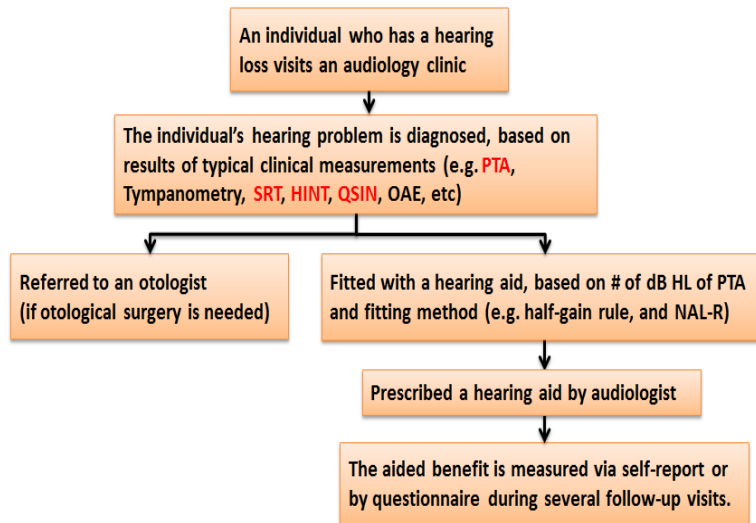


Figure 2.7: Chart of typical clinical procedures (adapted from Han 2011).

of the audiogram and audibility are shown in Fig. 2.8(b). The PTA data superimposed on the speech banana is helpful in providing information about modern amplification strategy for hearing aids, FM system, and directional microphone systems (Han 2011). Unfortunately, this method has been largely unsuccessful in properly fitting modern signal processing hearing aids. The reasons for this are unclear.

2.4.2 Speech Tests Use in Clinics

After a patient takes the PTA, he/she may also take a speech test. The speech tests are designed to see how well the patients can understand speech in babble noise. Speech tests presently used in audiology clinics are one or more of HINT, QuickSIN, BKB-SIN, and WIN. A comparison of speech tests is shown in Table 2.1. There are seven basic parameters that distinguish between the four speech tests. Those parameters are sentences, talker, target words, step size, presentation level varied, noise level varied, and adaptive. One method that is used to determine the patient performance is the Spearman-Karber method (Wilson et al. 2007).

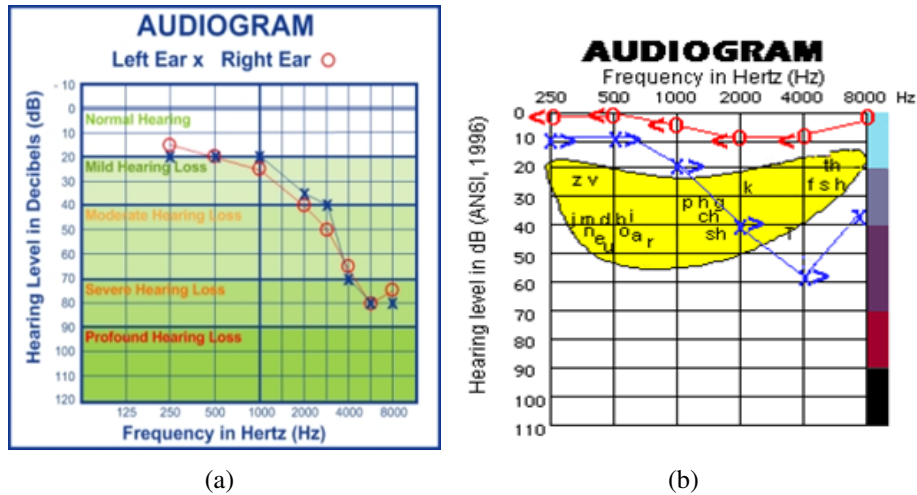


Figure 2.8: Pure-tone audiometry data: (a) audiogram (b) example of speech audibility.

2.4.3 Prescriptive Procedures for Hearing Aids

The purpose of a hearing aid is to maximize speech intelligibility. In the past 80 years, several publications of methods for selecting the gain and frequency response of a hearing aid have been proposed (McCandless and Lybarger 1983; Cox (1983, 1985, 1988); Byrne and Dillon 1986; Allen et al. 1990). While methodology continues to evolve, it is never properly evaluated. Part of the problem is the standard metrics that are used, such as the average score $P_c(SNR)$. Due to the time required for the test, rarely is the confusion matrix test used (Miller and Nicely 1955). Several “prescriptive formulas” have been used for improving speech perception. These formulas include a “combination of theoretical approaches” and measurements of the hearing threshold levels (PTA, aka audiometric configuration), most comfortable levels (MCL) based on the subject’s loudness preference, and loudness discomfort levels (LDL). However, there is no accepted optimal formula or measurement that the speech community agrees on (Byrne and Dillon 1986). All the formulas are heuristic and typically have some theoretical target, such as optimal audibility, or uniform loudness across frequency (Allen et al. 1990). A reliable metric, such as a confusion matrix, is never used because it is too time-consuming. Many of the formulas were derived from earlier formulas, such as the half-gain rule. The most popular prescribed gains for modern hearing aids that are worn by hearing aid users today are NAL-NL2 (National Acoustic Laboratory-Nonlinear 2, Keidser et al. 2011), NAL-NL1 (National Acoustic

Table 2.1: Four speech-recognition tests.

Speech Tests	Sentences	Talker	Target Words	Step Size	Presentation Level Varied	Noise Level Varied	Adaptive
HINT	10	Male	None	Adaptive	Yes	Yes	Yes
QuickSIN	6	Woman	5	fixed	No	Yes	No
BKB-SIN	10	Male	3	fixed	No	Yes	No
WIN	6	Woman	5	fixed	Yes	No	No

Laboratory-Nonlinear 1, Byrne et al. 2001), MSU (Memphis State University, Cox 1988), and POGO (Prescriptive of Gain and Maximum Output, McCandless and Lybarger 1983). Another approach is to measure the aided insertion gain of the hearing aid using “real ear” methods. This requires placing a microphone in the ear canal and measuring the canal-to-free-field spectral gain as a function of frequency. A serious drawback of this method is the existence of a standing wave in the ear canal.

2.5 Modification of Perceptual Cues

Before Li, Kapoor and Allen investigated the impact of plosive burst features on NH ears, several papers examined burst and transitional features of both natural and synthetic speech. Some of these papers provided limited information on the impact of burst and transitional features on speech perception because their studies did not include masking noise (Cole and Scott 1974; Dorman et al. 1977; Blumstein et al. 1977). However, these studies still came to some important conclusions. Without providing any experimental results, Cole and Scott (1974) explicitly stated that the characteristics of the burst and transitional features are essential for speech perception. Dorman et al. (1977) challenged Cole and Scott’s notion of the importance of the burst and transitional features by investigating natural speech of three plosives, followed by nine different vowels. They trun-

cated each CVC stimulus into three parts: burst, devoiced transition, and VC. Due to the difficulty of the experiment tasks and the limited analysis tools, Dorman et al. (1977) concluded that the transitional information was important for plosives, fricatives, and nasals. In the same year, Blumstein et al. (1977) also investigated the burst and transitional features using the same set of consonants as Dorman et al. and discovered an important concept that had a huge impact in the speech research community, which they called “conflicting cues”. However, their experiments involved synthetic speech. Thus, some might reasonably question the validity of this conclusion.

Several other studies explored the effects of burst and transitional features on speech perception, but unlike the studies above, masking noise was incorporated (Gordan-Salant 1986; Harzan and Simpson 1998). Gordan-Salant conducted a study of 19 natural spoken consonants, paired with each of the three vowels, in CV format, modified to improve speech recognition. Her experiment was conducted on young and elderly NH subjects. There were four sets of stimuli used for evaluating the efficiency of the acoustic modifications. Three sets of stimuli were modified differently: 1) consonant duration increased by 100%, 2) consonant-vowel ratio increased by +10 [dB], and 3) both consonant duration and consonant-vowel ratio increased. It was demonstrated that amplification of the consonant region can increase recognition score. Also, the set of stimuli that contained consonant-vowel outperformed other modification schemes.

Harzan and Simpson studied the effects of modifying both burst and transitional features using techniques similar to those of Gordan-Salant. Their stimulus corpus consisted of plosive, fricative, and nasal in VCV context. Also, four different modification techniques were used for speech enhancement. These modifications included amplifying the burst and transition regions, and applying a bandpass filter on the relevant burst while the formant transition increased by +6 [dB]. For the plosive, each burst was increased by +12 [dB] and then presented at 0 and -5 [dB]. Their results may be the first to show that given prior information of the perceptual cue location, as determined by NH listeners, speech enhancement can increase intelligibility under various noise conditions.

The Li, Kapoor, and Allen studies were influenced by the results of Harzan and Simpson (1998), Ohde et al. (1995), and Ohde and Stevens (1983). However, there were two key differences between the Li and Allen, and Kapoor and Allen studies: 1) short-time Fourier transform signal processing methods were used, that allow for precise control of the gain over time and frequency (Allen 1977), and 2)

each stimulus's format transition was unmodified. Their focus was primary and conflicting cues, within the burst duration. According to several studies conducted by the Human Speech Recognition group at UIUC, the plosive bursts are the primary acoustic feature for correct identification (Allen and Li 2009; Li and Allen 2011; Li et al. 2010; Régnier and Allen 2008), in contrast to previous studies of plosives (Kewley-Port et al. 1983; Stevens and Blumstein 1978).

We now direct our attention to the Li and Allen studies, which include the 3DDS findings and the modifications of specific time-frequency regions at the burst frequency. These specific time-frequency regions are highly correlated with /p, t, k/ and /b, d, g/ identification. However, some of these acoustic features in the burst time regions have positive or negative effect in speech perception. The burst feature is characterized by a wideband click for /p, b/, but narrow band for /g, d, t, k/ and may have cues that can be confused or misheard as another sound when the primary cue is masked.

The burst feature of a target stimulus usually has two types of acoustic cue: the primary cue and conflicting cues. The primary cue is a critical time-frequency region for correct identification, whereas the conflicting cues are opposing acoustic cues often having frequency regions above and/or below the primary cue. The AI-grams for the 3DDS findings of the primary and conflicting cues of 16 consonants are illustrated in Fig. 2.9. The red rectangular shape represents the primary cue of each target consonant while the blue ellipse represents the conflicting cues. When the target stimulus is a /ka/ sound, the time-frequency region of the primary cue is located around 1.4-2 kHz, while it contains two conflicting cues, one at 5 kHz (often heard as /ta/) and the other at 0.4-0.7 kHz (often hear as /pa/). Unlike Kapoor and Allen (2012), Li and Allen (2011) conducted a case study to examine the impact of the primary and conflicting cues of plosive consonants on three NH listeners (Li and Allen 2011) and one HI listener (not published, Li 2010) utilizing speech modification as shown in Fig. 2.10.

The Kapoor study involved the role of burst features. He was able to identify and modify the burst region (i.e. primary cue) for 4 plosive consonants using the AI gram. Later, he published an experiment that involved 21 NH listeners, identifying plosives that had their burst feature regions: unmodified, amplified, attenuated, and removed. Some results from this experiment are shown in Fig. 2.11. Each panel represents the correct identification as a function of SNR, for a particular target token. The 6 [dB] amplification of the burst region for each token improved consonant identification significantly, while removing the burst released

resulted in large errors, as expected. Kapoor concluded that the burst features for the plosives are vital perceptual cues for correct recognition for the NH ears. He also showed significant interaction between amplified conflicting cues and attenuated primary cues. In summary, Kapoor and Allen have provided a tool that could prove useful in investigating perceptual cues using HI ears.

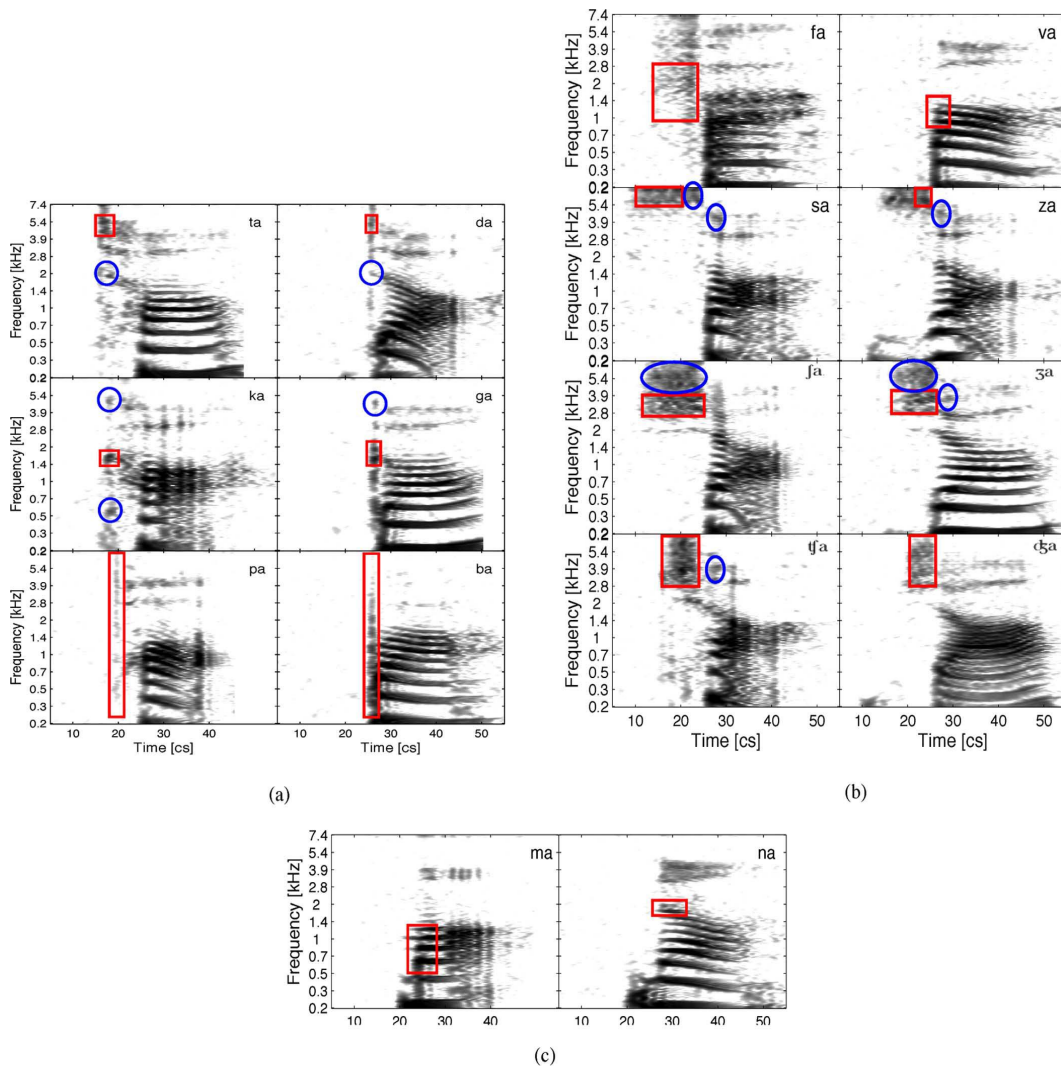


Figure 2.9: 3DDS findings: (a) plosive consonants, (b) fricative consonants, (c) nasal consonants (adapted from Li and Allen, 2011). The red rectangular symbols are the primary cue, while the blue ellipse symbols are the conflicting cue.

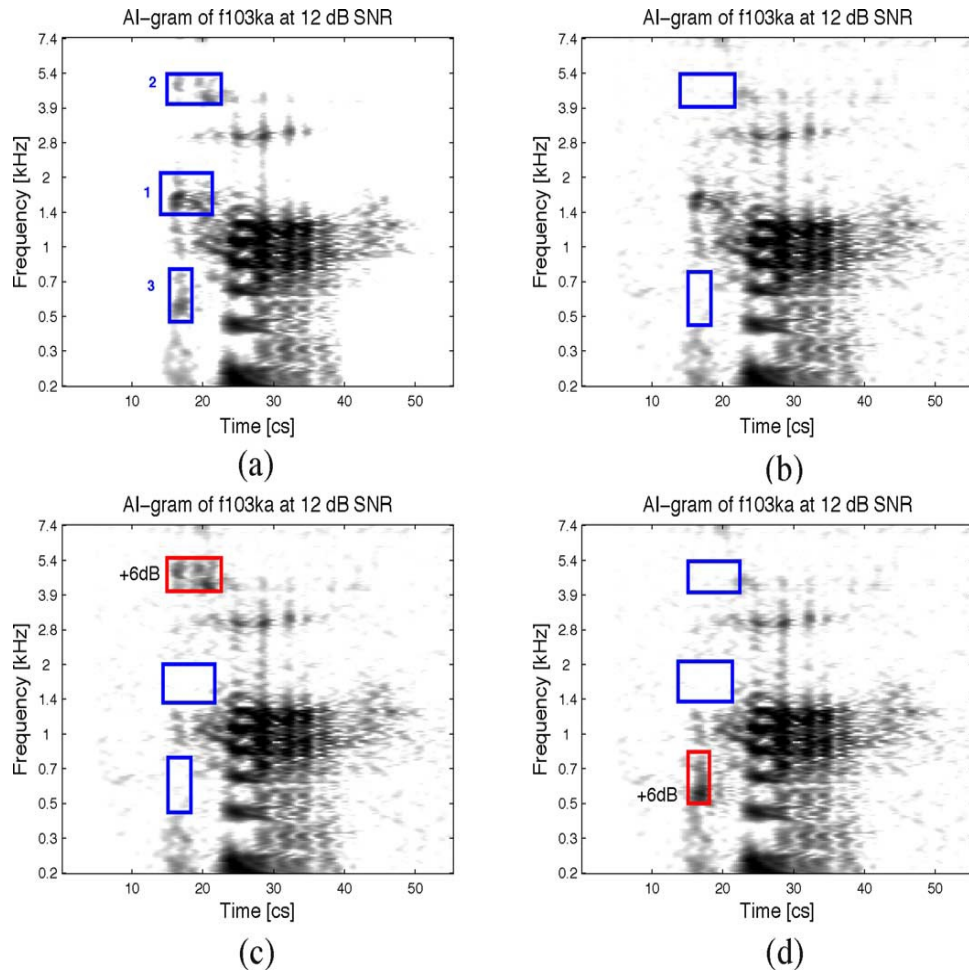


Figure 2.10: Three-way manipulation of unvoiced stop consonant /ka/: (a) unmodified, (b) both conflicting cues (features 2 and 3) removed, (c) primary and conflicting cues (features 1 and 3) are removed while the other conflicting cue (features 2) is amplified by + 6 [dB], (d) primary and conflicting cues (features 1 and 2) are removed while the other conflicting cue (feature 3) is amplified by + 6 [dB] (adapted from Li and Allen 2011).

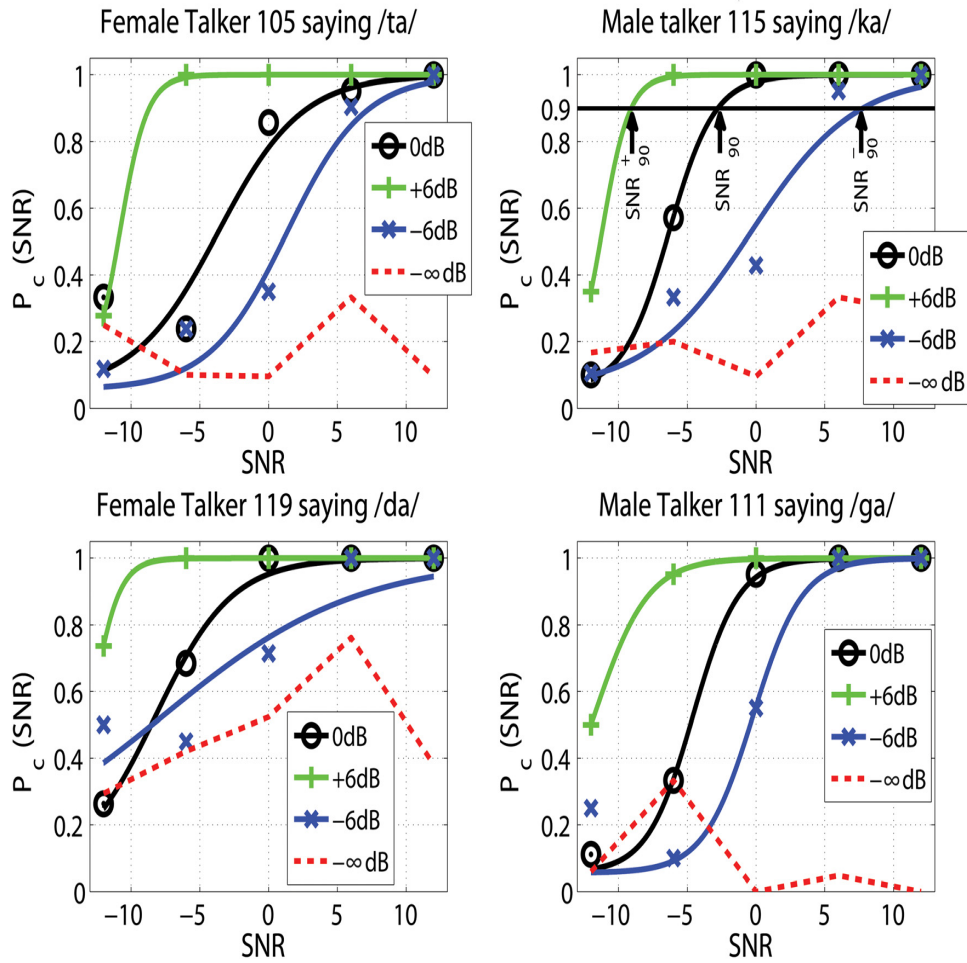


Figure 2.11: $P_c(SNR)$ for the unmodified and three modified versions of three burst released CV sounds (adapted from Kapoor and Allen 2012). The removal of the burst feature shows the average NH ears perform poorly, while the amplification of + 6 [dB] shows improvement.

CHAPTER 3

METHODS

In the last two chapters, we defined perceptual cues and their relationship to Human Speech Recognition (HSR). We then explained the differences between the primary and conflicting cues and their effects in HSR. Li and Allen (2011) describe the locations of both primary and conflicting cues for 16 naturally American English consonants along with the Articulation Index (AI) gram and the 3-Dimensional Deep Search (3DDS) method. Two perceptual modification experiments (Li and Allen 2011, Kapoor and Allen 2012) showed the effects of the primary and/or conflicting cues in the NH ears. The results from Kapoor and Allen (2012) showed that the majority of normal hearing (NH) listeners depend on the primary cue for correct identification of plosives up to a signal-to-noise level (Singh and Allen 2012). The noise level reaches to a token dependent critical signal-to-noise ratio (SNR) threshold once the primary cue is masked.

In this chapter, we explain the methodology of a case study and two perceptual modification experiments on the HI ears: by varying the masking of the primary cue, and by removing the conflicting cues. The case study is two psychoacoustic experiments from the Han (2011) study that evaluate two basic prescribed gains for hearing aids: flat gain (FG) and Natural Acoustic Laboratory - Revised (NALR).

For the first experiment, we are repeating the Kapoor and Allen (2012) experiment on hearing impaired (HI) ears. The experiments will lead us to understand 1) the plosive's primary cue for HI ears, and 2) the significance of the conflicting cues' role in modifying the entropy of confusions in HI ears. An understanding of the roles 1) and 2), could aid us in explaining the error patterns made by an HI ear. These error patterns are the key in decoding the strategies of an HI ear in identifying consonants. The basic methodologies of the two perceptual modification experiments are given next. We describe modification of the masking of the primary cue for Exp. I, and the removal of the conflicting cues for Exp. II. The idea behind the two experiments is to show that 1) the HI ears depend on the

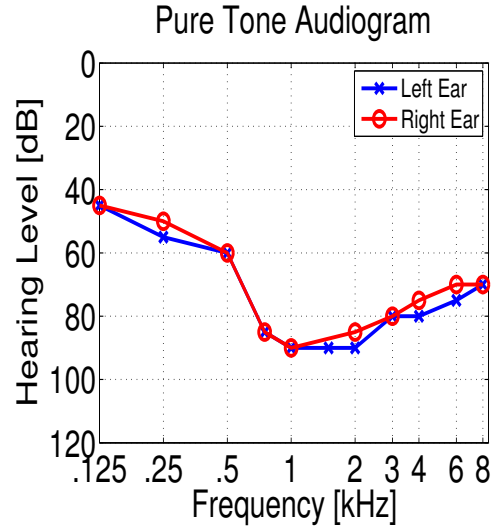


Figure 3.1: Subject CC’s PTA: Subject has a 90 [dB] hearing loss at 1 [kHz] and 85 [dB] HL at 2 and 3 [kHz]. These frequencies are critically important for speech perception.

primary cue, and 2) when the primary cues are masked, the secondary cues can play a central role. The main point we wish to show is that HI ears are using the same cues as NH ears. While other factors can play a role, as they do for NH ears, the main factors are the primary and secondary cues used by NH ears.

3.1 Methods for the Case Study

The basic methodology consists of two HI perceptual experiments, and techniques for analyzing the data are given next. Additional details of the methodology of HI psycho-acoustic experiments can be found in Han (2011).

3.1.1 Subject

The case study is based on the experiments from Han (2011) on 8 subjects with mild to moderate HL. For the case study, we added a 9th subject, the author, who has profound hearing loss, up to 90 [dB] HL, at the most important speech frequencies (i.e. 1000, 1500, and 2000 [Hz]), as shown in Fig. 3.1.

3.1.2 Speech Materials

All speech stimuli used in this study are CVs and selected from the Linguistic Data Consortium (LDC) 2205S22 database (Fousek et al. 2004). The speech stimuli are presented by both male and female voices of native English speakers and digitally recorded at a sampling rate of 16 kHz. This study focuses on fourteen common consonant phonemes (/p/,/t/,/k/,/f/,/s/,/ʃ/,/b/,/d/,/g/,/v/,/z/,/ʒ/,/m/,/n/), followed by the /a/ (Miller and Nicely 1955). Each consonant was spoken by one male and one female which gives a total of 28 test tokens (14 consonants x 2 talkers = 28 tokens). Table 3.1 provides a list of tokens used in this study. Only tokens having an $SNR_{90} > 0$ [dB] are investigated, meaning that NH listeners have <10% error for SNRs greater than 0 [dB], in SWN. Linear interpolation was done between measurements taken at -22, -20, -16, -10, and -2 [dB] to compute all SNR_{90} values. In order to ensure that the speech stimuli were audible, all stimuli were presented at MCL.

3.1.3 Experimental Procedures

The experimental test procedures of the two perceptual experiments, HI Exp. II (flat gain) and HI Exp. IV (NALR), are similar to those used in a previous study by Han (2011). The differences between the two experiments are the two gain conditions which will be explained later in this section.

The study was conducted in the speech lab at UIUC. In order to prevent any distraction or external noise interference from the speech lab, the subject is seated in an enclosed single-walled soundproof booth. A MATLAB GUI was designed to facilitate the experiment. Since each ear had to be tested, a total of four tests were done (right and left ears using flat gain, right and left ears using NALR).

Before the experiment began, the MCL for each ear was set in order to maximize consonant identification. The MCL used for Subject CC's left and right ears was 130 [dB] and 120 [dB], respectively. Subsequently, a practice session was mandatory to familiarize the subject with the MATLAB GUI, and its experimental task. Each of the fourteen consonants was repeated twice giving a total of 28 tokens for the practice session. After each response, the subject was presented with a feedback of the correct consonant. The tokens used for the practice session are different from those used for the experiment, to prevent memorization of any tokens that were used in the experiment.

Table 3.1: List of tokens used in experiments II and IV.

CV	Male Talker	SNR_{90}	Female Talker	SNR_{90}
ba	m112	-2	f101	-10
da	m118	-7	f105	-13
fa	m112*	-5*	f109	-12
ga	m111	-12	f109	-3
ka	m111	-13	f103	-11
ma	m118	-14	f103	-11
na	m118	-4	f101	-7
pa	m118	-14	f103	-17
sa	m120	-10	f103	-13
ja	m118	-16	f103	-15
ta	m112	-17	f108	-14
va	m118	-3	f101	-10
ʒa	m107	-7	f105	-17
za	m118	-17	f106	-18

The data collection of the experiment consisted of two phases. In the first phase, each consonant was presented in quiet and at varying SNR levels. The SNR levels used will be +0 [dB], +6 [dB], +12 [dB], and quiet, when in the presence of speech weighed noise. A total of 448 different presentations were used, 28 tokens x 4 SNR conditions x 4 presentations. There were 8 presentations for each consonant at a particular SNR. Each trial can be presented two more times to assist in decision-making. After all presentations, the percent error was calculated for each token at each SNR, ranging from 0/4= 0%, 1/4=25%, 2/4=50%, 3/4=75%, and 4/4=100%. These results determined the number of presentations per consonant in Phases II of Experiments II or IV. At each SNR, the number of presentations for each consonant in Phase II was presented between 2 and 12 times (see Table 3.2). The theory behind the design of the second phase was to satisfy the statistical power test by increasing the number of presentations. The range of maximum total number of presentations per consonant for each HI ear is 40-80. Singh and Allen (2012) verified and showed that number of presentations was sufficient to determine correct perception within a 95% confidence interval using by the Vysochanskij-Petunin inequality (Vysochanskij and Petunin 1980).

Table 3.2: Number of tokens presented in each phase in Experiments II and IV.

# of error ($\mathcal{P}_\mathcal{E}$)	Exp. II or Exp. IV		
	Phase I	Phase II	Σ
0(0%)	4	2	6
1(25%)	4	2	6
2(50%)	4	5	9
3(75%)	4	6	10
4(100%)	4	6	10

3.2 General Methods for Exp. I and Exp. II

The subjects, speech materials, and experimental procedures are the same for the two experiments.

3.2.1 Subjects

A total of 5 NH and 10 HI subjects were recruited from the Urbana-Champaign community by advertisement and were paid to participate in the study. All were born in the United States, their primary language was English, and they ranged in age between 26 and 65. All NH subjects had normal hearing (self-reported). All HI subjects agreed to take an audiological evaluation to obtain their audiometric profile as shown in Fig. 3.2 and Table 3.3. Each subject read, understood and signed the consent form. This study was approved by the UIUC IRB.

3.2.2 Speech Materials

Four plosive consonants /t,k,d,g/ paired with vowel /a/ in CV context were used as target speech stimuli. Table 3.4 shows the list of stimuli and specific time-frequency regions that were manually selected for modification based on Li and Allen, and Kapoor and Allen, using the AI-gram as a guide. Due to the small number of target consonants, an additional six non-plosive consonants, paired with vowel /a/ in CV context, were added to the list of stimuli, as seed sounds. The list of seed sounds is shown in Table 3.5. All sounds used in this study were selected from the Linguistic Data Consortium, University of Pennsylvania, Fousek et al. (2004). It was confirmed by Phatak et al. (2008) that all stimuli had 0% recognition error at and above 12 [dB] SNR.

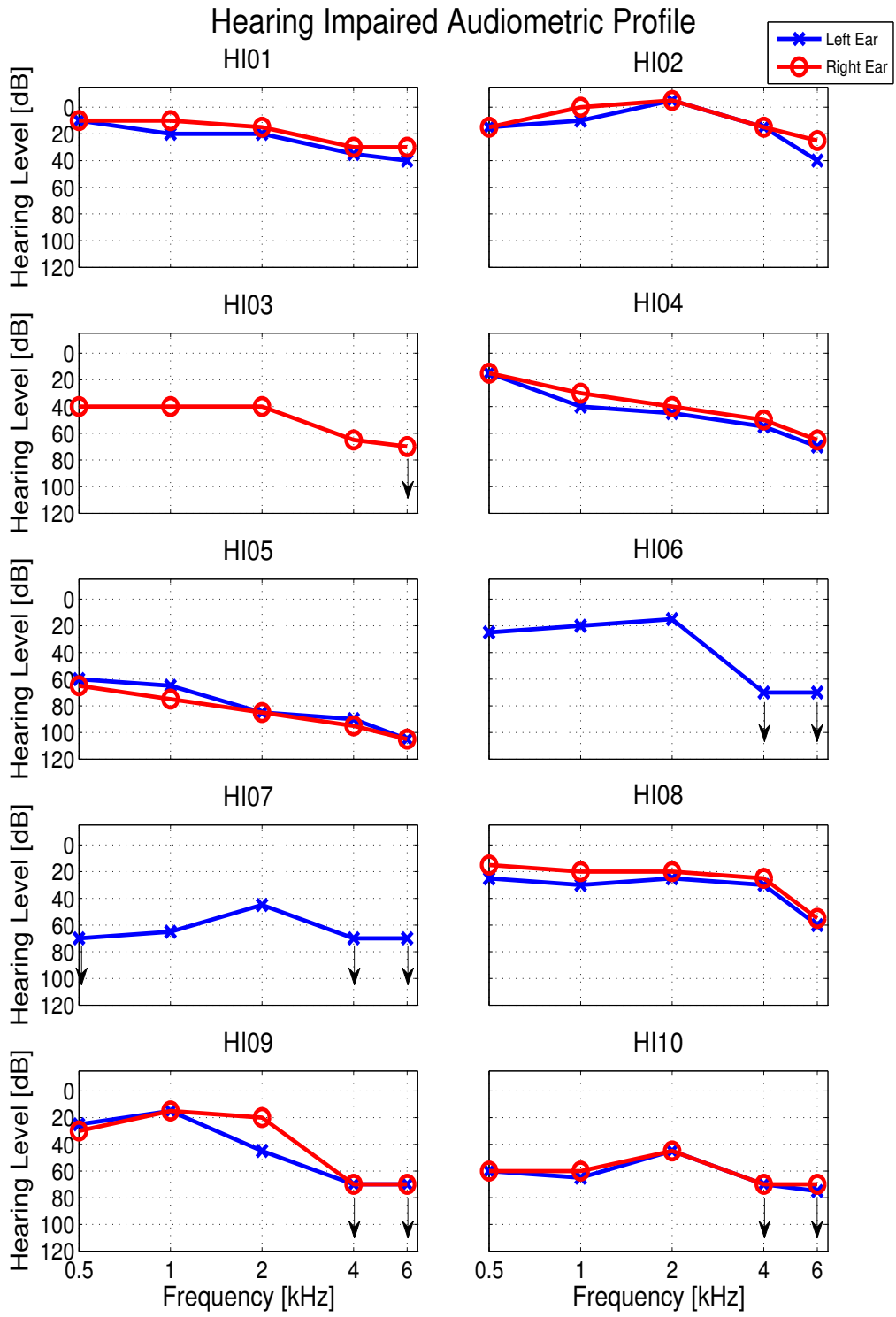


Figure 3.2: PTA for 10 HI subjects: The black arrows indicate ears that have HL greater than 70 [dB]. Hearing loss greater than 70 [dB] is considered severe. HI05, HI06, HI07, HI09, and HI10 have at least one ear that has severe HL at 4 and 6 [kHz]. HI05 has profound HL (i.e. ≥ 90 [dB]) for both ears at 4 and 6 [kHz].

Table 3.3: Hearing impaired audiometric profile.

Subject Ear	500 Hz	1000 Hz	2000 Hz	4000 Hz	6000 Hz
HI01-R	10 dB	10 dB	15 dB	30 dB	30 dB
HI01-L	10 dB	20 dB	20 dB	35 dB	40 dB
HI02-R	15 dB	0 dB	-5 dB	15 dB	25 dB
HI02-L	15 dB	10 dB	-5 dB	15 dB	40 dB
HI03-R	40 dB	40 dB	40 dB	65 dB	>70 dB
HI04-R	15 dB	30 dB	40 dB	50 dB	65 dB
HI04-L	15 dB	40 dB	45 dB	55 dB	70 dB
HI05-R	65 dB	75 dB	85 dB	95 dB	105 dB
HI05-L	60 dB	65 dB	85 dB	90 dB	105 dB
HI06-L	25 dB	20 dB	15 dB	>70 dB	>70 dB
HI07-L	>70 dB	65 dB	45 dB	>70 dB	>70 dB
HI08-R	15 dB	20 dB	20 dB	25 dB	55 dB
HI08-L	25 dB	30 dB	25 dB	30 dB	60 dB
HI09-R	30 dB	15 dB	20 dB	>70 dB	>70 dB
HI09-L	25 dB	15 dB	45 dB	>70 dB	>70 dB
HI10-R	60 dB	60 dB	45 dB	>70 dB	>70 dB
HI10-L	60 dB	65 dB	45 dB	70 dB	>70 dB

3.2.3 Experimental Procedures

The subject was seated in the sound booth and instructed on the use of the MATLAB GUI interface. Each subject provided personal information such as first and last names, birth date, and the type of accent (i.e. Southern, Midwestern, Eastern, Western). Their personal information was saved via the MATLAB GUI interface. The subjects began by listening to trials, to set the level to their most comfortable level (MCL). All stimuli were presented via an Etymotic ER3 insert earphone.

Prior to the experiment session, a short practice session was provided to allow subjects to become familiar with the MATLAB GUI interface that displayed 20 options comprised of 18 unmodified CV syllables and “Only Noise” and “Other”. The practice sounds are listed in Table 3.6. The second column (Talker 1) is the primary list of tokens that was presented. However, the subject had an option to repeat any CV up to three times. If the subject wished to repeat the token, the first repeat presented the original token (Talker 1) again and the second and third repeats presented the other two talkers, respectively (third column, Extra Talkers). Feedback was provided to ensure each subject was familiar with the correct sounds. If a token was not identified correctly, the token was placed at the end of the practice list and this procedure was done a maximum of three times for

Table 3.4: Basic parameters for each modified sound used in this study.

Token	Δt [cs]	F_{lo} [kHz]	F_{hi} [kHz]
m115ta	8.5	1.5	7.4
f119ta	7.3	1.7	7.4
m111ka	5.0	0.6	2.7
f103ka	5.3	1.1	2.1
f105da	7.5	1.6	7.4
f119da	7.8	2.1	7.4
m111ga	7.5	0.3	2.4
f103ga	4.5	0.5	2.4

Table 3.5: Unmodified seed sounds used in this study.

CV	Talkers
sa	m118, f103
fa	m120, f113
ma	m118, f108
za	m112, f109
ʃa	m115, f119
va	m118, f106

Table 3.6: Extra talkers used during the practice session.

CV	Talkers 1	Extra Talkers
ma	m118	f119, m120
va	m120	f105, m111
sa	f109	f106, f119
θa	m114	m118, f103
za	f105	m107, f101
ra	f105	Not repeated
ka	f113	m111, f119
ða	f101	m118, f103
pa	f105	f109, m102
ʒa	m107	m111, m104
fa	f101	f105, m104
ba	m115	m114, f103
ta	f119	m118, f119
klæ	m118	Not repeated
na	m111	m112, f105
ha	m114	f103, m120
da	f106	m102, f108
ga	m111	m115, f109
ʃa	f103	f101, m114
la	f108	m104, m112

each token. All stimuli in the practice session were presented at 18 [dB] SNR.

The experiment consisted of one session. The total presentation for the NH subjects was 156 tokens (4 plosive x 2 talkers x 3 SNRs x 5 modification types + 6 non-plosive x 2 talkers x 3 SNRs) and a maximum of 468 tokens for the HI subjects (4 plosive x 2 talkers x 3 SNRs x 5 modification types x 3 repeats + 6 non-plosive x 2 talkers x 3 SNRs x 3 repeats). The number of target tokens (also the minimum of presentation assuming no repeats) was 120 for the NH subjects, and 360 for the HI subjects. To prevent learning effects such as learning the sounds, there was no feedback during the experiment. However, each presentation could be repeated up to 3 times. The subject was encouraged to take a few minutes break to avoid fatigue.

3.3 Experiment I: Methods for Varying the Masking of the Primary Cue

Masking of the primary cue at a particular time-frequency feature region was based on the 3DDS method (Li and Allen 2011). Verification of the efficiency of each modification was conducted by several pilot experiments using NH listeners. An example of four types of modifications utilizing AI-gram of female talker f103 saying /ka/ is shown in Fig. 3.3. The top left end panel (a) is the unmodified version ($\times 1$ or 0 [dB]). The rest of the panels are modified (b) primary cue removed ($\times 0$ or $-\infty$), (c) primary cue attenuated ($\times -1/2$ or -6 [dB]), and (d-e) primary cue amplified ($\times 2$ or +6 [dB]; $\times 4$ or +12 [dB]). Three different wideband SNR conditions (0, 9, and 18 [dB]) were used. All SNR conditions used white ($\frac{Watts}{Hz} = \text{Constant}$) noise. In order to ensure that the noise is the same across all modifications, except in the burst region, the SNR is based on the unmodified sound.

3.4 Experiment II: Methods for the Removal of Conflicting Cues

Examples of AI-grams for the removal of the conflicting cues for f103 /ka/ are shown in Fig. 3.4. The unmodified version as shown in (a) is compared to the modified versions as shown in (b)-(f)(see caption). The red outlines that are above

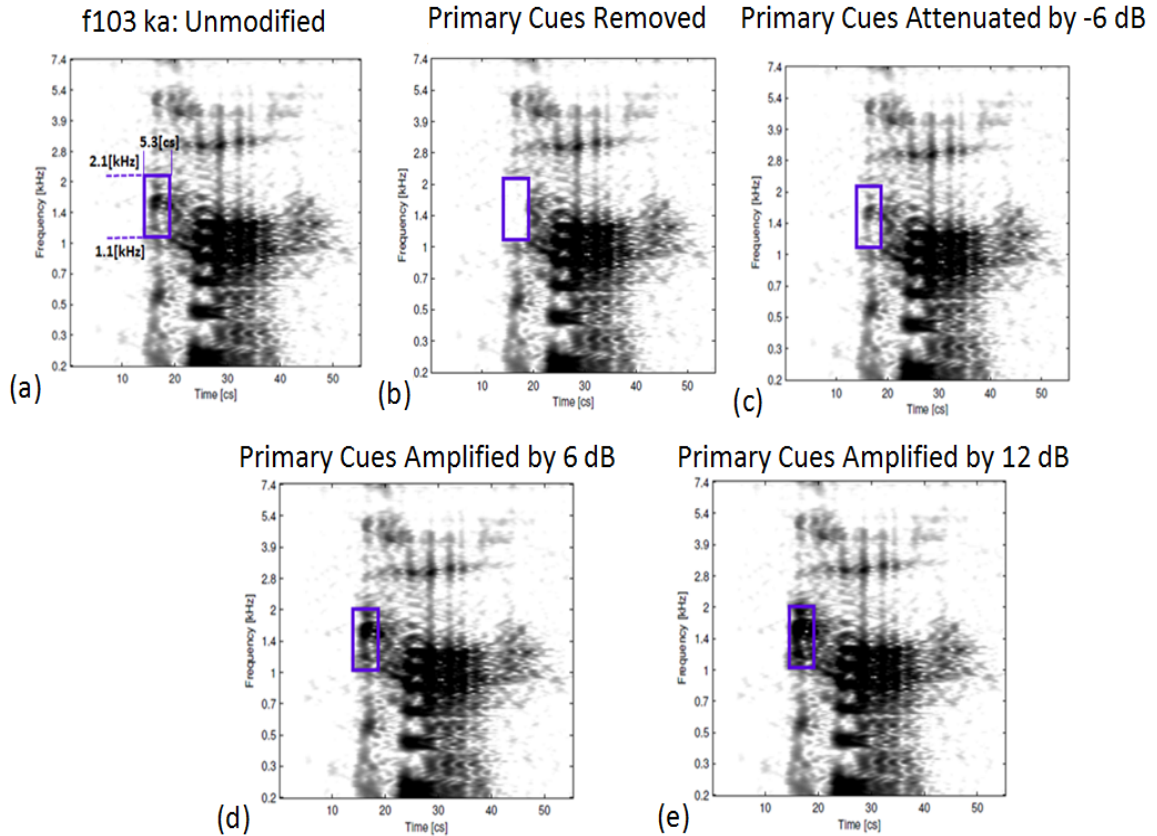


Figure 3.3: Varying the masking of the primary cue for female talker f103: (a) Unmodified version. (b) Primary cue removed. (c) Primary cue attenuated. (d) Primary cue amplified by 6 [dB]. (e) Primary cue amplified by 12 [dB].

and below the primary cue (i.e. blue outline) are the parameters in which the conflicting cues are removed. Figure 3.4(b) shows the primary cue unmodified and the conflicting cues removed. The blue outline in Fig. 3.4(c)-(f) is modified similar to Fig. 3.3 (b)-(e): the removal of the primary cue (Fig. 3.4(c)), the primary cue attenuated -6 [dB] (Fig. 3.4(d)), the primary cue amplified by 6 [dB] (Fig. 3.4(e)), and the primary cue amplified the primary cue by 12 [dB] (Fig. 3.4(f)). The efficiency of the modification of the stimuli for the removal of the conflicting cues experiment was verified by several pilot tests. The SNR conditions for Experiment II are the same as Exp. I (varying the masking of the primary cue).

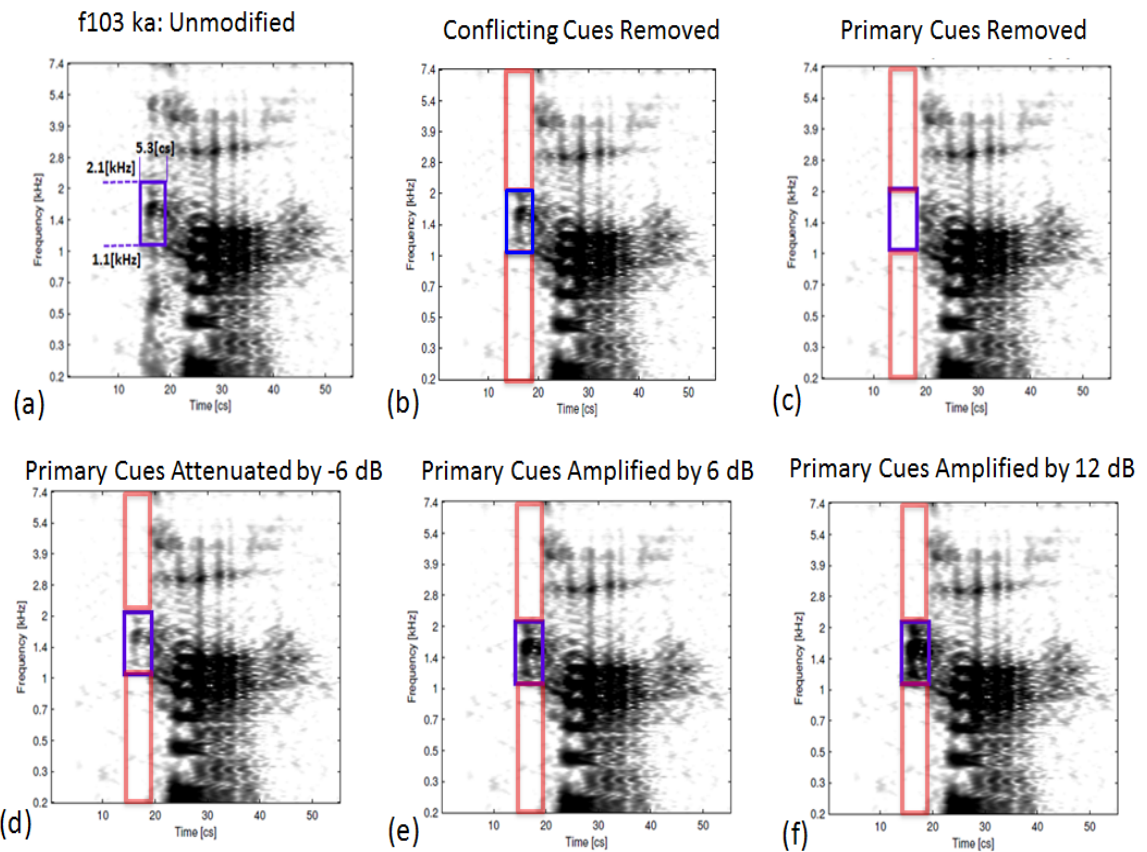


Figure 3.4: The removal of the conflicting cues for female talker f103: (a) Unmodified version. (b) Primary cue unmodified or amplified by 1 (0 [dB] gain). (c) Primary cue removed or amplified by 0 ($-\infty$). (d) Primary cue attenuated or amplified by $-1/2$ (-6 [dB] gain). (e) Primary cue amplified by 2 (6 [dB] gain). (f) Primary cue amplified by 4 (12 [dB] gain).

CHAPTER 4

RESULTS

In the previous two chapters we presented details of our studies that relate human speech perception in normal hearing (NH) and hearing impaired (HI) ears, for human speech sounds which have been modified by manipulating critical speech cues. We discussed the Li and Kapoor studies that inspired this research, provided the details of our study's objectives, and then discussed the methods of a case study on Subject CC, as well as those for the two main experiments, Exp. I and Exp. II, aimed at addressing our basic research questions.

In this chapter, we present results of the case study of Subject CC (Section 4.1), which details 5 techniques for analyzing the psycho-acoustic data. In Section 4.2 we analyze the results of Exp. I, where we masked the plosive's primary cues for eight tokens, allowing us to investigate the impact of the primary cue for 5 NH ears and 10 HI ears. In Section 4.3 we review the results of Exp. II, where we removed both the primary and conflicting cues, to see if there is any improvement relative to Exp. I when the conflicting cues are removed. Finally, in Section 4.4 (p. 91), we discuss a case study on decoding strategies, and discuss a possible method for mitigating high frequency loss.

4.1 Preliminary Results of a Case Study

The prescribed gain for hearing aids (HAs) is normally based on pure-tone audiometry (PTA), and occasional speech-in-noise tests (WINS), that use meaningful words for evaluating HAs (Wilson et al., 2007). While WINS is able to do what PTA does (i.e., identify HI ears), it has not been successful in quantifying degrees of speech loss. We shall show that, unlike WINS, the sorted error method can titrate out the degrees of speech loss.

In a previous study, the Human Speech Recognition (HSR) group at the University of Illinois at Urbana-Champaign (UIUC), conducted two psycho-acoustic

experiments to evaluate two basic prescribed gains for hearing aids, flat gain (FG) and National Acoustic Laboratory - Revised (NALR), on 8 subjects (16 HI ears), with mild to moderate HL (Han, 2011). These ears were tested using 14 natural spoken English consonants, with a total of 24 tokens. Details of this methodology of HI psycho-acoustic experiments may be found in Han (2011), Trevino and Allen (2013), and Scheidiger et al. (2017).

We begin our analysis using the Han 2011 protocol, as a case study for a 9th subject, the author. Since there are 8 subjects (16 HI ears) in the Han (2011) study, this provides an interesting comparative case study. The objective of this analysis is to study subject confusions when applying FG and NALR insertion gains. We seek a better understanding of the subject’s confusions, to determine the next step for reducing them via improved amplification strategies. Since this 9th subject is the author, a more general analysis was applied to the collected data. To determine which provide the best description of the errors, we evaluate these results using 5 techniques:

1. First, we show the average probability, $\overline{P}_e(\text{SNR})$, averaged across tokens, and explain its fundamental limitations.
2. Next, we look at the average probability at the token level $P_e(CV_i, \text{SNR})$ for token CV_i , as a function of SNR.
3. Third, we look at the entropy, a measure of consistency, as a function of the error $\mathcal{H}(\mathcal{P}_e)$.
4. Fourth, we examine the *sorted-error* technique, that describes the errors of each token, sorted from the minimum to maximum error.
5. Finally, we analyze the confusion patterns (CPs), $P_{h|s}(\text{SNR})$, using a graphical presentation of a row of the confusion matrix as a function of SNR (Allen, 2005).

By this comparison we conclude that sorted-errors (4) and CPs (5) are the best analysis methods.

4.1.1 Case Study Count Matrix – Results

CM tables: First we examine the raw data from the case study of subject CC (# 9) for the protocol of Han (2011). After completing the two experiments (FG

Table 4.1: Total count matrix in 18 [dB] SNR for the left ear: talkers were females for FG.

18 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	6														6
ta		6													6
ka		4	5												9
da								5	1						6
ga							1	1	4						6

Table 4.5: Total count matrix in 18 [dB] SNR for the right ear: talkers were females for FG.

18 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	4	1	1												6
ta		6													6
ka		2	7												9
da								6							6
ga									6						6

Table 4.2: Total count matrix in 12 [dB] SNR for the left ear: talkers were females for FG.

12 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	3	2	1												6
ta		6													6
ka		9	1												10
da								6							6
ga							1	3	5						9

Table 4.6: Total count matrix in 12 [dB] SNR for the right ear: talkers were females for FG.

12 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	0	1	9												10
ta		6													6
ka		2	4												6
da								6							6
ga									6						6

Table 4.3: Total count matrix in 6 [dB] SNR for the left ear: talkers were females for FG.

6 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	2	5	3												10
ta		6													6
ka		5	4												9
da								6							6
ga							1	4	1						6

Table 4.7: Total count matrix in 6 [dB] SNR for the right ear: talkers were females for FG.

6 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	1	4	5												10
ta		6													6
ka		4	6												10
da								6							6
ga							1	5							6

Table 4.4: Total count matrix at 0 [dB] SNR for the left ear: female talker for FG.

0 dB	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	6	2	1												9
ta		6													6
ka	6	2	2												10
da								6							6
ga							3	1	1	5					10

Table 4.8: Total count matrix at 0 [dB] SNR for the right ear: female talker for FG.

0 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	2	4	4												10
ta		6													6
ka		5	5												10
da								6							6
ga							2	1	2	4					9

Table 4.9: Total count matrix in 18 [dB] SNR for the left ear: talkers were females for NALR.

18 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	6														6
ta		6													6
ka			6												6
da							5	1							6
ga									6						6

Table 4.13: Total count matrix in 18 [dB] SNR for the right ear: talkers were females for NALR.

18 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	4	4	2												10
ta		6													6
ka			6												6
da							6								6
ga								6							6

Table 4.10: Total count matrix in 12 [dB] SNR for the left ear: talkers were females for NALR.

12 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	4	4	1												9
ta		6													6
ka			6												6
da							5	1							6
ga									6						6

Table 4.14: Total count matrix in 12 [dB] SNR for the right ear: talkers were females for NALR.

12 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	5		4												9
ta		6													6
ka			4												6
da							6								6
ga								6							6

Table 4.11: Total count matrix in 6 [dB] SNR for the left ear: talkers were females for NALR.

6 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	2	7													9
ta		6													6
ka			5												6
da							6								6
ga								1	6						6

Table 4.15: Total count matrix in 6 [dB] SNR for the right ear: talkers were females for NALR.

6 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	3	3	4												10
ta		6													6
ka			5												6
da							5				1				6
ga								1	5						6

Table 4.12: Total count matrix at 0 [dB] SNR for the left ear: female talker for NALR.

0 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	6	2	1												9
ta		6													6
ka			6												9
da							6								6
ga								2	1	4					9

Table 4.16: Total count matrix at 0 [dB] SNR for the right ear: female talker for NALR.

0 [dB]	pa	ta	ka	fa	sa	fa	ba	da	ga	va	za	ʒa	ma	na	Σ
pa	4		2												6
ta		6													6
ka			3												6
da							6								6
ga								2	3	3		1			10

vs. NALR), we generated the count matrix (CM) and the *confusion matrix* (the row normalized CM) for each condition. There are a total of 32 CMs, corresponding to 2 gain conditions (FG, NALR), 2 talkers (male, female), 2 ears (right, left) and 4 SNRs (0, 6, 12, and 18 [dB]). Tables 4.1-4.16 show the CM results for the female talker (male talker not shown), for the left (4.1-4.4, 4.9-4.12) and right (4.5-4.8, 4.13-4.16) ears, for the FG (4.1-4.8) and NALR (4.9-4.16) insertion gains, at each of the 4 SNRs. The rows of each CM correspond to the stimuli (spoken CVs: /pa/, /ta/, /ka/...). The columns are the responses (reported heard). The highlighted numbers in blue (on the CM diagonal) indicate correct responses. Blank entries indicate zero responses.

Average and Token errors: Further analysis of the CM has been provided using *MATLAB*TM to compute the average probability of error $\bar{P}_e(\text{SNR})$, average token error $P_e(\text{CV}_i, \text{SNR})$, confusion patterns, and entropy. For example, the *average token error* is

$$P_e(\text{CV}_i, \text{SNR}) = 1 - P_{ii} = \sum_{j \neq i} P\{\text{heard } \text{CV}_j | \text{spoken } \text{CV}_i\}, \quad (4.1)$$

where CV_i is the i^{th} token, $i = 1, 2, \dots, 24$ and P_{ii} is the diagonal of the 14x14 CM. Each diagonal element represents a correct response. The off-diagonal elements represent confusions relative to the 14 response (heard) consonants (i.e., CV_j).

From the average token error

$$\bar{P}_e(\text{SNR}) = \frac{1}{24} \sum_{i=1}^{24} P_e(\text{CV}_i, \text{SNR}) \quad (4.2)$$

one may directly graph the confusion patterns (CP), or compute the *average token error* (Eq. 4.1) vs. SNR

4.1.2 Average Probability of Error and Token Errors Analysis for Subject CC

The average probability of error across the 14 natural English spoken consonants for insertion gains FG and NALR is shown Fig. 4.1 for subject CC using the Han (2011) protocol. As discussed in the methods section, Han (2011) studied the impact of hearing aid gain prescriptions on HI ears, for 8 HI native Amer-

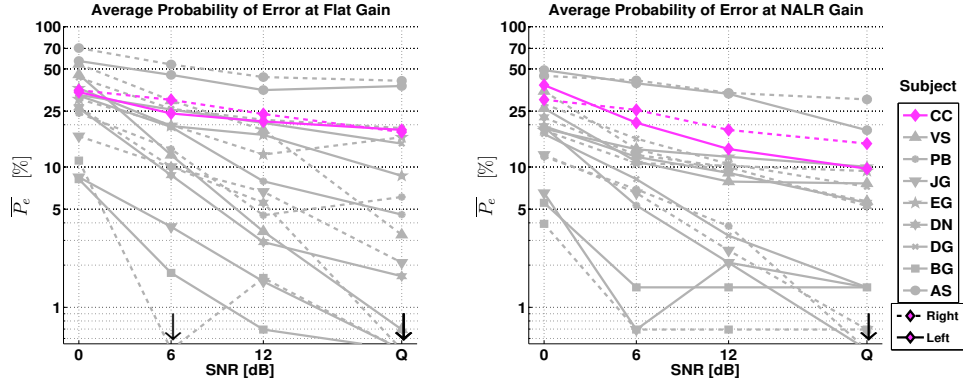


Figure 4.1: Average probability error (Eq. 4.2) for the 8 subjects from Han (2011). Subject CC has been added as the 9th subject in the Han (2011) study. The hearing thresholds of Subject CC are shown in Fig. 3.1. The down arrow (\downarrow) indicates where $\bar{P}_e(\text{SNR}) = 0$.

ican (US) listeners. The two experimental protocols were the same, except for the insertion gain. The first experiment used FG while the second experiment’s gain used NALR, as shown in the left and right panels of Fig. 4.1, respectively. Each chart displays the average probability of error, $P_e(\text{SNR})$, as a function of the wideband SNR. A ninth subject CC (the author) has been added to the Han (2011) experiment. These added curves are the focus of the present analysis abbreviated as ‘CC’, as shown at the top of the legend.

The average probability of error for CC for NALR is slightly lower than the FG condition, when the $\text{SNR} \geq 6$ [dB], which is the general trend for the 8 subjects of Han (2011). However, when the $\text{SNR} = 0$ [dB], the probability of error for the right ear for the NALR case is lower than the right ear for the FG case. For the FG at the same SNR condition (i.e., $\text{SNR} = 0$ [dB]), the right and left ears are approximately the same.

The average token error $P_e(CV_i, \text{SNR})$ (Eq. 4.2) is a more useful measure. Fig. 4.2 shows the token error for /ka/, subject CC, talker f103 for FG and NALR gain conditions. The red (i.e. ‘x’) and blue curves (i.e. ‘+’) represent FG and NALR, respectively. Figure 4.2 shows the huge differences for SNRs of 6, 12, and 18 [dB]. NALR outperformed FG in the error difference of $\approx 0.5, 0.85,$ and 0.5 at 6, 12, and 18 [dB], respectively. The error under the NALR condition increased dramatically from when the SNR decreased from 6 [dB] to 0 [dB]. The maximum error for both conditions is when the SNR is 0 [dB]. Thus we see that individual consonants can deviate dramatically from the average probability error $P_e(\text{SNR})$, a point previously made by Trevino and Allen (2013). This is concrete evidence

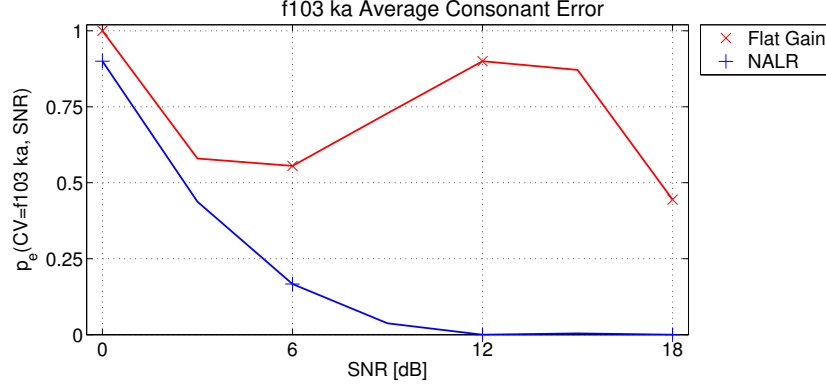


Figure 4.2: Subject CC’s left ear error result for token f103 for the two gain conditions, FG vs. NALR. The primary cue for consonant /k/ is located around 1-2 [kHz], where the HL threshold for Subject CC is 90 [dB] (i.e., profound). The biggest improvement in terms of error decreased is when the NALR gain is 12 and 18 [dB], where it goes from 90 and 44 [%] error, to zero.

that $P_e(\text{SNR})$ does a poor job in characterizing subject performance, mainly due to averaging out larger difference in a small number of tokens.

4.1.3 Entropy for Subject CC

The entropy (\mathcal{H}) is a measure of consistency, uncertainty, or randomness defined as

$$\mathcal{H}_k = \mathcal{E} \log_2 \mathcal{I}(P) = \sum_{k=1}^{24} p(x_k) \log_2 \left(\frac{1}{p(x_k)} \right), \quad (4.3)$$

where the “information density” is defined as $\mathcal{I}_k = \frac{1}{P_k}$, and the information in bits as $\log_2 \mathcal{I}(P)$, \mathcal{E} is the “expected value”, and $p(x_k)$ is the probability of random variable x_k , where x_k represents the 24 tokens.

We plotted $\mathcal{H}(P_e, \text{SNR})$ to understand the pattern of the left ear for Subject CC’s responses, as illustrated in Fig. 4.3. This plot shows the entropy of FG versus NALR for different SNRs. The ordinate and abscissa represent the entropy and percent error, respectively. Each symbol represents the entropy versus percent error for a specific SNR for f103 /ka/. A table of the results from Fig. 4.3 is shown in Table 4.17.

When the NALR was applied for Subject CC’s left ear at 18 and 12 [dB] SNRs, both the errors and entropies decreased to zero. At 0 [dB] SNR, with 90% error, the confusion for f103 /ka/ is distributed with 3 other tokens. As a result, the entropy is 1.571, which shows that at 0 [dB], Subject CC has serious difficulty

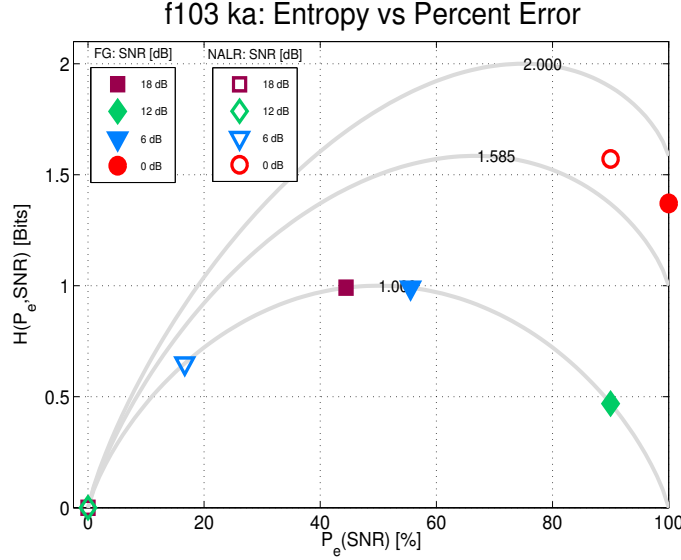


Figure 4.3: The entropy results for Subject CC’s left ear identify a /ka/ spoken by f103. For the FG condition, Subject CC has 44% error and the entropy is .991 at 18 [dB] SNR, and has 90% error and the entropy is 0.469 at 12 [dB] SNR. Under the NALR condition at 18 and 12 [dB], the error and the entropy decreased to zero. As a result, Subject CC consistently performs better when the gain is NALR.

recognizing /ka/.

4.1.4 Sorted Error for Subject CC

This section presents a technique called sorted error method (Trevino and Allen, 2013; Abavisani and Allen). The sorted error graph for Subject CC’s left and right ears is depicted in Fig. 4.4. There are six panels. Top and bottom rows represent the left and right ears, respectively. Column 1 shows the raw data along

Table 4.17: This shows how \mathcal{P}_E and \mathcal{H} vary with SNR for the two gain conditions. These relations are non-monotonic and depend strongly on the gain conditions. We shall explain why this is happening later in the analysis. Note how $\mathcal{P}_E = 0$ for NALR at 18 and 12 [dB] SNR. Then at 0 [dB], \mathcal{P}_E jumps to 90%. For the FG case $\mathcal{P}_E \geq 44\%$ at all SNRs.

SNR [dB]	FG		NALR	
	\mathcal{P}_E [%]	\mathcal{H} [Bits]	\mathcal{P}_E [%]	\mathcal{H} [Bits]
18	44	0.991	0	0
12	90	0.469	0	0
6	55.56	0.991	16.67	.65
0	100	1.371	90	1.571

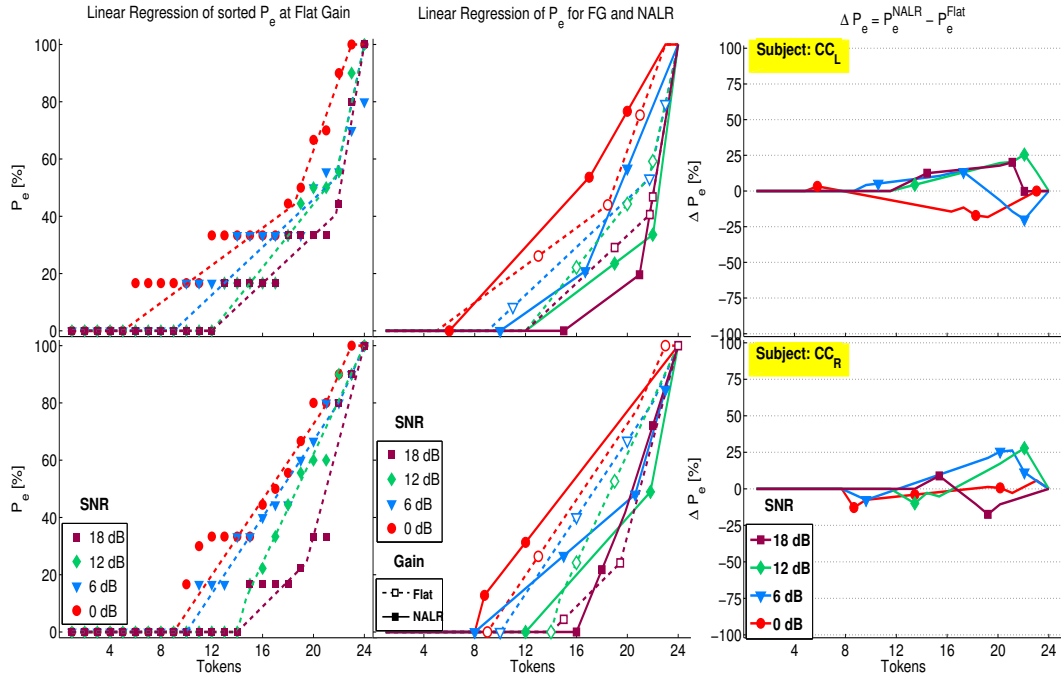


Figure 4.4: The sorted error regression (adapted from Abavisani and Allen) for Subject CC. The top row is the left ear and the bottom row is the right ear, for Subject CC. The legends at the bottom row represent the data for both top and bottom row. The first column shows the token error P_e as a function of token after sorting the token order in ascending order of error. The dashed and solid lines in the second column are regression fits to the sorted errors for the FG and NALR, respectively. The error difference (ΔP_e) between FG and NALR is shown in Column 3. The tokens are improved when $\Delta P_e \geq 0$, and are degraded when $\Delta P_e < 0$.

with the regressions (i.e. dashed lines) for the FG condition. The ordinate shows the probability of error in percent while the abscissa is the 24 tokens, sorted from the minimum to maximum probability error. Each symbol represents a token at a particular SNR coded by symbols (i.e. $\square = 18$ [dB], $\diamond = 12$ [dB], $\nabla = 6$ [dB], $\circ = 0$ [dB]). Each set of symbols used a linear regression fitting method to generate each dashed curve. At the highest SNR (purple square symbols), half of the tokens had zero error (ZE). Five tokens had one error or low error (1/6 or 16%). When two curves of a different SNR intersect, recognition of that token is independent of SNR.

Column 2 shows eight curves that represent one of the four SNRs approximated by a linear regression fitting method, for both FG (dashed lines) and NALR (solid lines). When the NALR curve is to the right of the FG curve, those tokens are improved. When the curve of NALR is to the left of the FG curve, those tokens are degraded. For example, at 18 [dB], when the NALR was applied for the left ear, the curves shifted right. Thus, NALR reduced the error by a total of 3 or

4 zero error tokens. Column 3 is ΔP_e between FG and NALR. A set of tokens improved when $\Delta P_e \geq 0$, otherwise the tokens are degraded.

This technique automatically identifies zero, low, and high error differences when NALR is applied. As the result of this analysis $\Delta P_e(\text{FG-NALR}, \text{SNR})$ characterizes the effectiveness of the prescribed gain, at the token level.

4.1.5 Confusion Patterns for Subject CC

As shown by Trevino and Allen (2013), it is important to evaluate the distribution of error responses at the token level. The confusion pattern, shown in Fig. 4.5, is the tool that allows us to do this. The confusion pattern is defined as one row of the normalized CM ($P_{h|s}(\text{SNR})$), as a function of SNR. The ordinate and abscissa for each panel are, $P_{h|s}(\text{SNR})$, the probability of being heard given spoken vs. SNR. There are 5 plosive consonants, /p, t, k, d, g/, as shown in Fig. 4.5. The red line divides the female talkers from the male talkers. Each row and column represents one of the 5 plosive's consonants and one of the two prescribed gain conditions (i.e., FG vs. NALR), respectively. There are a total of 20 tokens. Figure 3.1 shows that Subject CC has a 90 dB hearing loss at 1 [kHz], a most important speech frequency. The consonants in Fig. 4.5 that are associated with those important speech frequencies are unvoiced /k/ and voiced /g/. The location of the primary cue for /k/ and /g/ is usually between 1 and 3 [kHz], where Subject CC has the greatest HL.

Next we examine the error for consonant /k/. When the NALR was applied at 0 [dB] SNR, Subject CC performed very poorly in identifying f103 /ka/. Note that the confusions are spread to as many as 3 other consonants. Significantly in the presence of noise, the confusions belong in the same group as the primary cue. On the other hand, this was not the case for m111 /ka/, where the probability of being correct for both gain conditions is much higher than f103 /ka/.

For f103 /pa/ for the FG case at 0 [dB] SNR, Subject CC has 33.33% error. The source of error is the conflicting cues /ta/ and /ka/. For the same scenario when the NALR was applied, for f103 /pa/, Subject CC has zero error.

For the FG case for f109 /ga/ at 0 [dB] SNR, Subject CC has 90% error. The error is spread across 3 other consonants. Two of the three consonants are conflicting cues, /ba/ and /da/. The error is reduced to 55.56% when the NALR is applied at 0 [dB] SNR for f109 /ga/. However, the number of consonants with

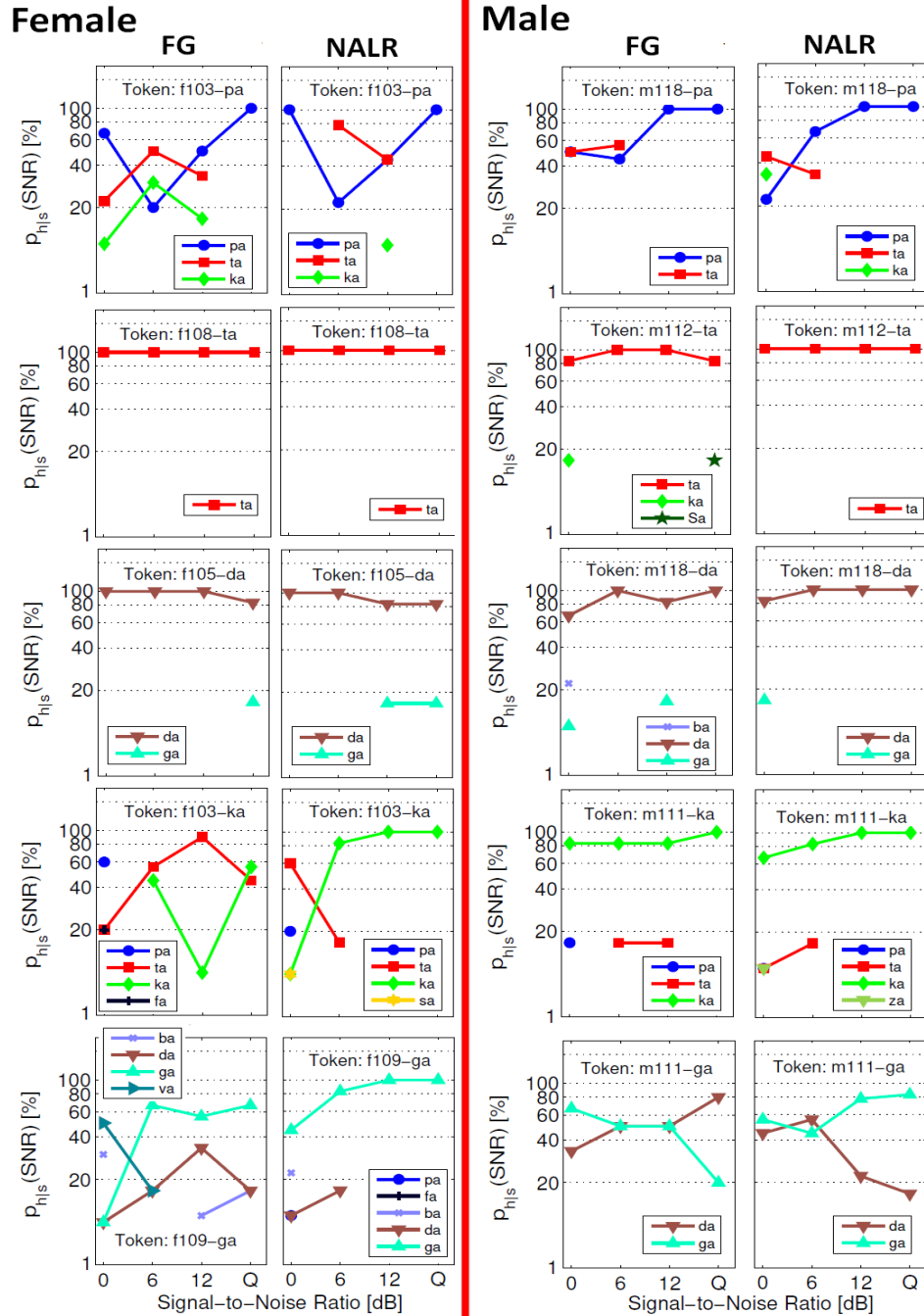


Figure 4.5: Subject CC's left ear confusion pattern results for 5 plosive CVs, /p, t, k, b, d, g/, using FG and NALR conditions. The red line splits the two token genders, female vs. male. For each talker, the results show the comparison at each gain condition, FG vs. NALR. Notice the source of error for mid-frequency sounds f103 /ka/ and f109 /ga/ are determined by conflicting cues. The conflicting cues for consonant /k/ are /t/ and /p/, and for consonant /g/ are /d/ and /b/. Hence this explains the /p, t, k/ and /b, d, g/ confusion groups (Miller and Nicely, 1955).

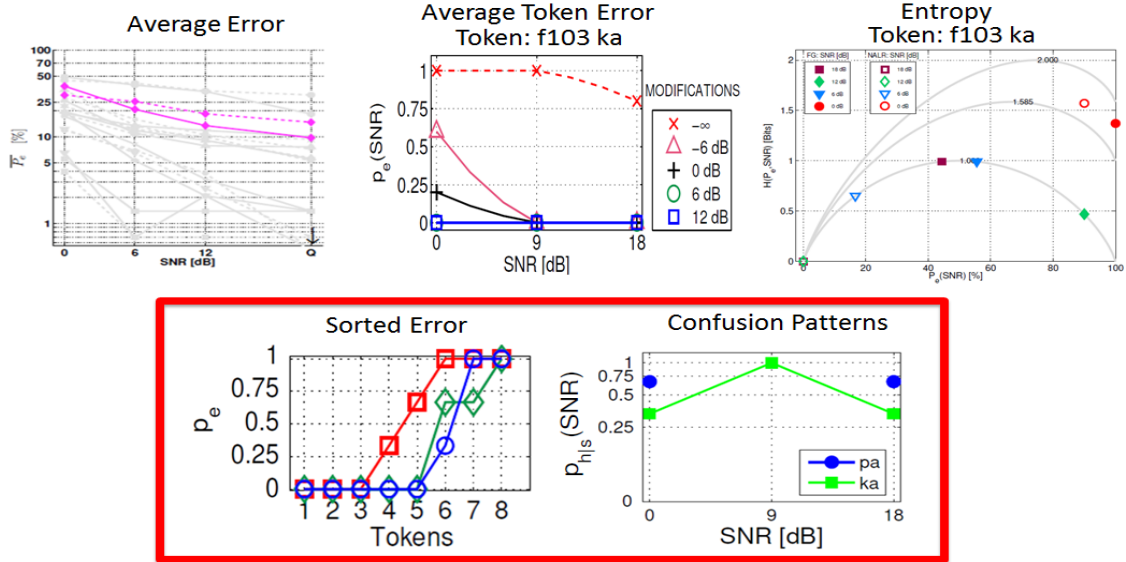


Figure 4.6: Five techniques for analyzing HI and NH psycho-acoustic data. This figure summarizes the various tools that were explored in the Subject CC’s case study. It is clear we must use the sorted error and $P_{h|s}(SNR)$ methods.

error increased from 3 (/ba, da, va/) at 0 [dB] SNR for the FG, to 4 (/pa, fa, ba, da/) at 0 [dB] SNR for NALR.

For the analysis of this research, of the five techniques we analyze from the case study of CC, we chose the sorted error and confusion patterns, as outlined in the red box of Fig. 4.6. These two techniques provided the greatest utility for answering our research questions.

4.2 Experiment I: Primary Cue is Varied

In Exp. I we collect confusions from 15 subjects, 5 NH and 10 HI. We averaged the 5 NH listeners together, to form a fictitious “average normal hearing” (ANH) control subject. Once the data was collected from each subject, we compute two statistical measures: 1) sorted errors (Fig. 4.7, 4.9) and 2) confusion patterns (Sect. 4.2.3).

4.2.1 Sorted Error

In Fig. 4.7 we review the results of the token errors sorted by error. There are 18 panels. The top row represents the average NH subject (‘AVG. NH’). They

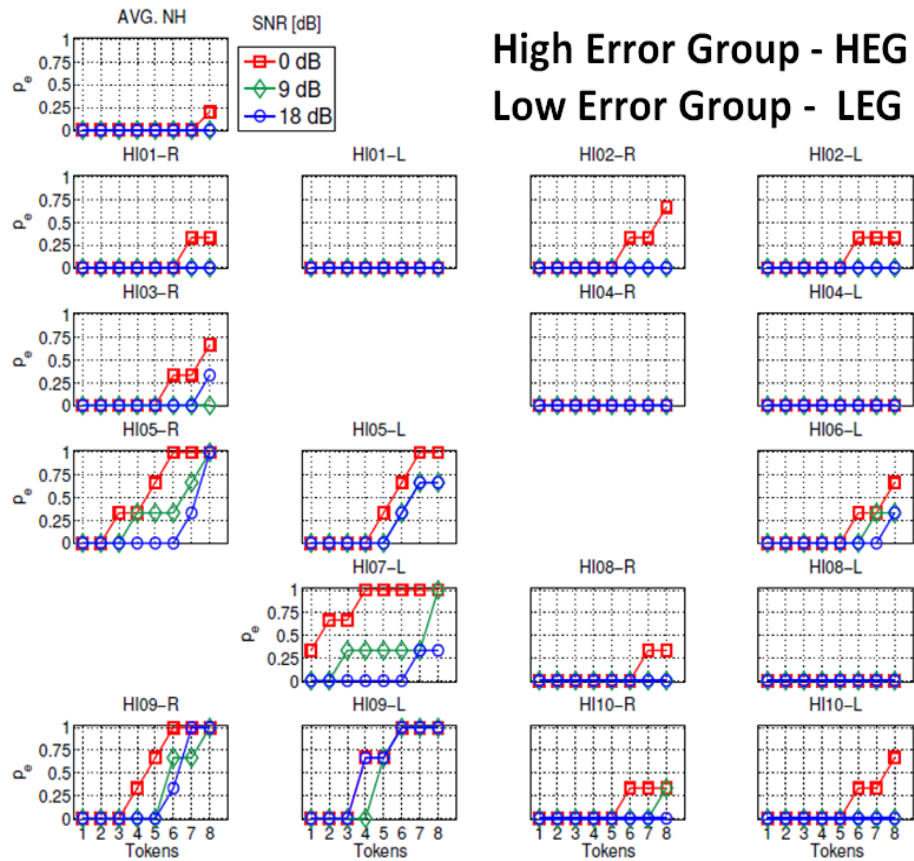


Figure 4.7: This figure shows the raw data at 3 SNRs for the average of the 5 average NH subjects (Row 1, labeled AVG NH) and 10 HI subjects (Rows 2-6). The tokens are sorted in error from 0 (all subject have some 0 error tokens) to their maximum, thus every curve is monotone. Row 2, columns 1, 2 shows HI subject 01, right and left ears. The right ear has a maximum of 30% error, while the left ear has 0 error at all SNRs. Columns 3, 4 show HI subject 02, right and left ears. This subject has a maximum of 60% (right ear) and 30% (left ear) error at 0 [dB] SNR, and no error at the other two SNRs. Row 3, column 1 is HI subject 3, where only the right ear was measured, leaving column 2 blank. Columns 3, 4 correspond to HI subject 4, left and right ears, with 0 loss in both ears, at all SNRs. Rows 4, 5 and 6 contain subjects 5R,L, 6L, 7L, 8R,L, 9R,L and 10R,L. Note that Ear HI07-L has the largest error, at all SNRs. At 0 [dB] 1 token has 30% error and 5 have 100% error.

basically have no error. The sorted error plots for subject HI01 are shown in panel 2 and panel 3, labeled ‘HI01-R’ and ‘HI01-L’, respectively. The other 9 HI subjects are shown in panels 4-18 (missing panels are not counted). There are 3 HI subjects for which only one ear was tested because the other ear is normal (HI03-R, HI06-L, HI07-L). All the other figures for sorted error, average token error and confusion patterns in Appendices A and B are organized in the same way.

The ordinate of each panel of Fig. 4.7 represents the probability of error and the abscissa represents P_e for the 8 tokens, sorted from minimum to maximum error. For each panel, three SNR conditions were used. The red square symbols represent 0 [dB], the green diamond symbols represent 9 [dB], and blue circle symbols represent 18 [dB], the quiet condition.

Unmodified primary cue results for NH and HI ears: The sorted error method and the closely related histograms of P_e are useful for classifying the type of token, and as described in the next section, to compute the difference between Exp. I and Exp. II. Therefore, from Fig. 4.7, an easier and more efficient tool to count the number of ZE, NZE, and ME is to use the histogram as shown in Fig. 4.8. The NH group (top left) has ZE tokens across all SNRs (except for 1 token). Four HI ears (HI01-L, HI04-R, HI04-L, and HI08-L) have ZE tokens across all SNRs. Nine of seventeen HI ears (HI01-R, HI01-L, HI02-R, HI02-L, HI04-R, HI04-L, HI08-R, HI08-L, HI10-L) have all ZE tokens for both 9 [dB] and 18 [dB] SNR. At quiet or 18 [dB] SNR, 3 HI ears (HI05-R, HI09-R, HI09-L) have probability of error of 1 or they were not able to recognize at least one token. When the SNR is decreased to the lowest SNR for varying the primary cue, 5 HI ears (HI05-R, HI05-L, HI07-L, HI09-R, HI09-L) have ME for at least 2 tokens.

The analysis of the unmodified primary cue for HI ears: An observation from Fig. 4.8 is that the 10 HI subjects can be organized into two groups, low error group (LEG) and high error group (HEG). The LEG ears have results similar to the average NH ears. The results from 6 HI subjects (HI01, HI02, HI03, HI04, HI08, and HI10) are quite similar to the results from the average NH subjects, while the other 4 HI subjects (HI05, HI06, HI07, HI09) results are not similar. The LEG ears have more ZE tokens than the HEG ears.

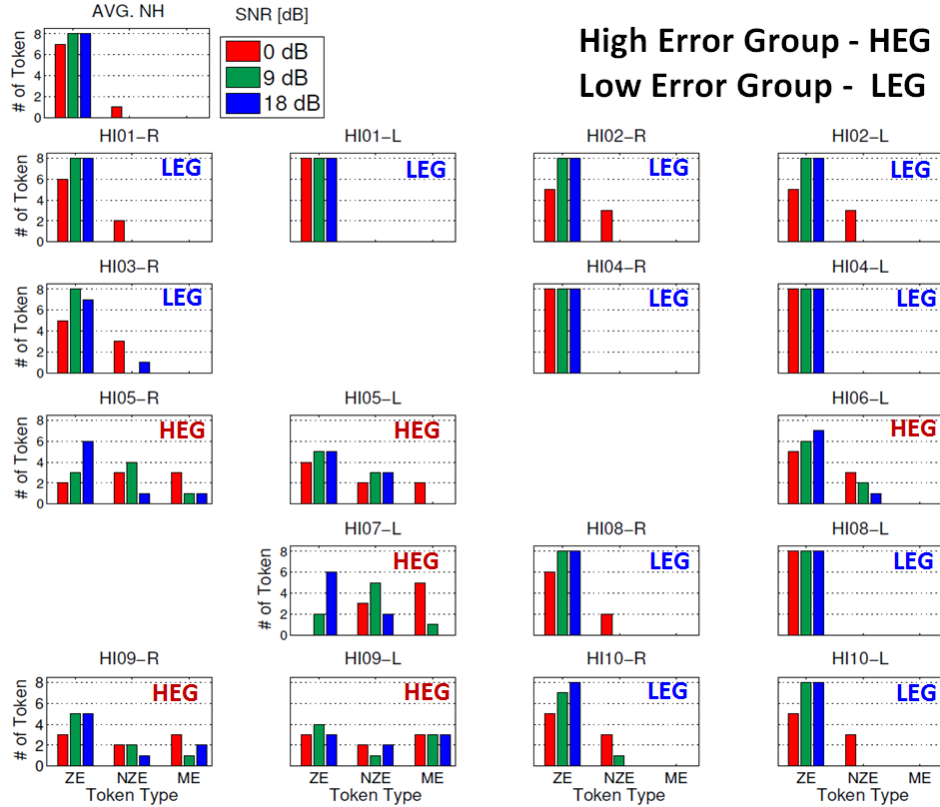
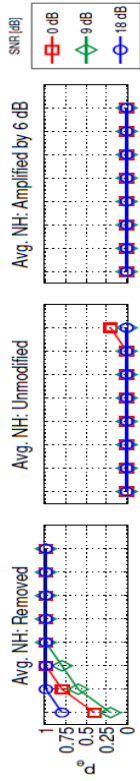


Figure 4.8: This figure is organized the same way as Fig. 4.7, showing histograms of all the token, at 3 SNRs, for the AVG NH group (Row 1) and the 10 HI subjects (Rows 2-6). This analysis assumes three token groups, as indicated on the abscissa: Tokens that have no error are labeled zero-error (ZE); tokens with error between 0 and 1 are labeled non-zero-error (NZE), and tokens that have error of 1 are called maximum-error (ME). The SNR is color coded (0, 9 and 18 [dB]). In each case the number of tokens always sum to 8. Ignoring the perfect ZE tokens, the remaining ears with error are divided into two subgroups: the *low error group* (LEG) with mostly ZE tokens (thus fewer NZE and ME tokens) and a *high error group* (HEG) with fewer ZE sounds, and therefore increased NZE and ME sounds.

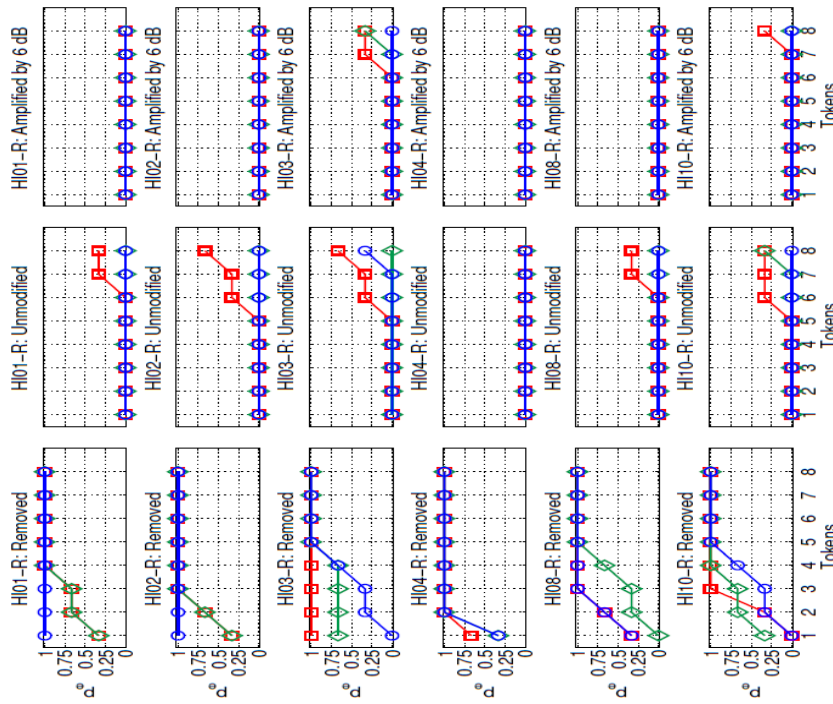
4.2.2 Analysis of Exp. I

The analysis of Exp. I is divided into three parts. First is the presentation of the raw data, Fig. 4.9. Second are the corresponding histograms in Fig. 4.10, which characterize the raw data of Fig. 4.9. These two figures include the results of Figs. 4.7, 4.8 as the two middle columns of rows 2-7 of Figs. 4.9 and 4.10, which replicate (are identical to) the results of Fig. 4.7 and 4.8. Finally, Sect. 4.2.3 provides a discussion of the confusion pattern (CP) data of Figs. 4.11 and 4.12.

Primary Cue is Varied Sorted Error Plots



Low Error Group



High Error Group

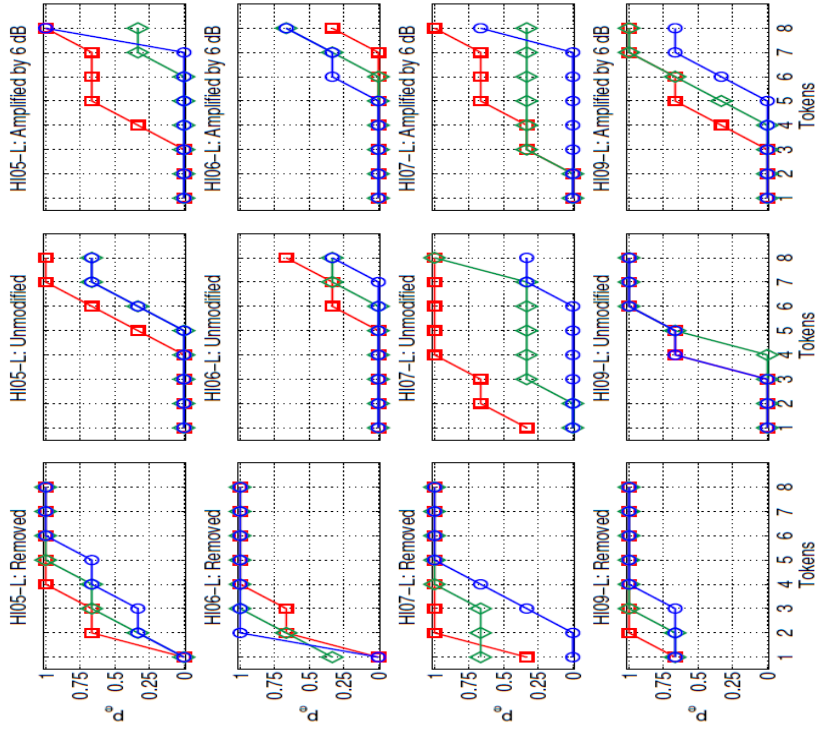


Figure 4-9: The comparison of varying the masking of the primary cue using the sorted error method. The red line divides the figure into 3 groups of ears: the average of the 5 NH ears represent the top, the LEG ears on the right, and the HEG ears on the left. Each group of ears has 3 columns: the left column represents primary cue removed ($\times -\infty$), the primary cue unmodified ($\times 0$) in the middle, and the primary cue amplified by 6 [dB] ($\times 2$). Most errors happen when the primary cue is removed for all NH and HI ears (both LEG and HEG ears). A 6 [dB] amplification, the primary cue for all ears shows a decrease in errors.

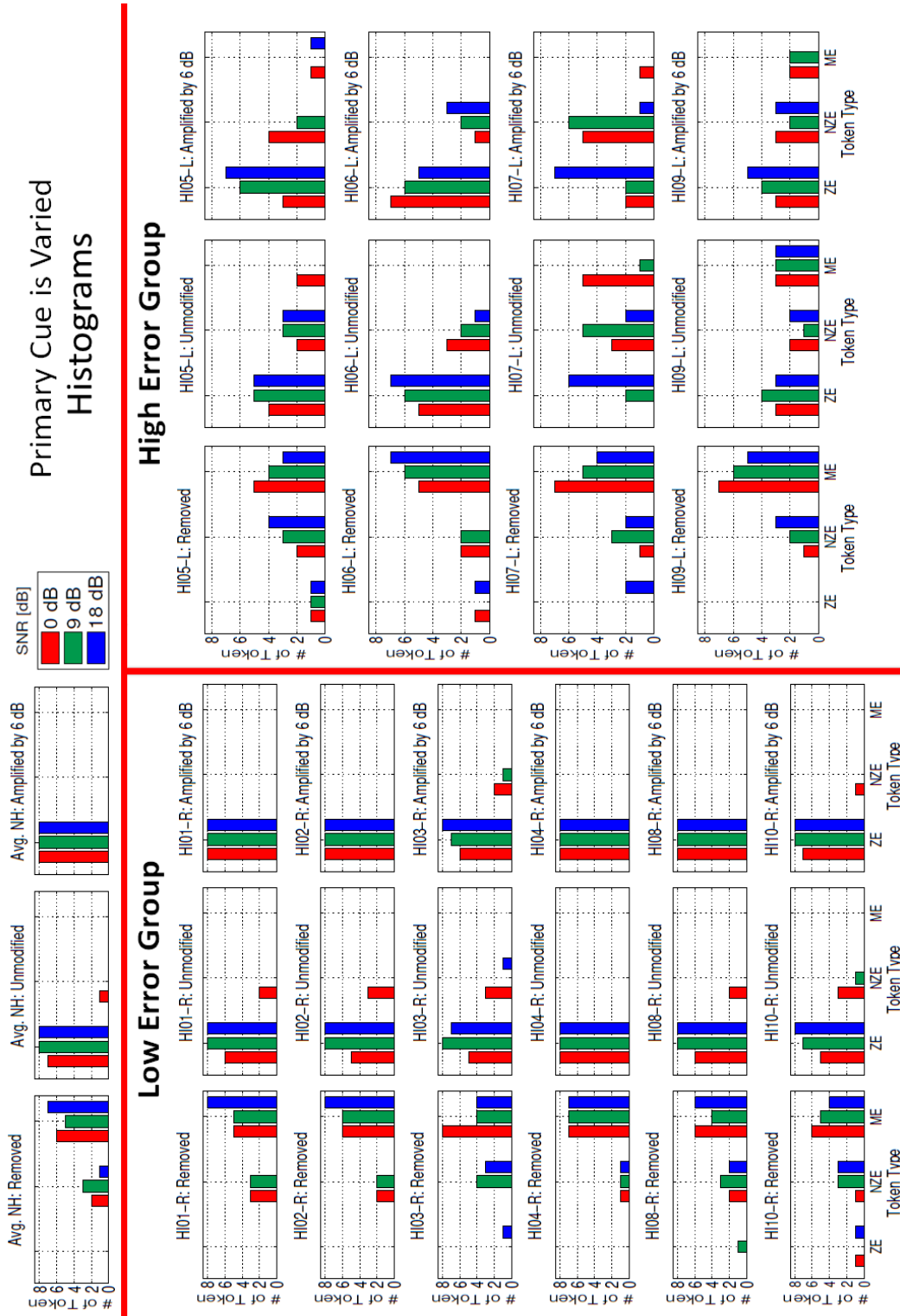


Figure 4.10: The histograms of the number of tokens for the average NH, HEG, and LEG ears. Each result from the previous figure (Fig. 4.9) is classified as ZE, NZE, or ME. The results when the primary cue is removed, which is mostly ME, are similar for the 3 ear groups. On the other hand, when the primary cue is unmodified or amplified, the number of ZE tokens shows a huge difference between the average NH and LEG ears versus HEG ears.

Raw data: Figure 4.9 provides a major reorganization of all the measured data, displayed in sorted error format, with solid-red lines separating the three listener groups: the ANH listener is at the top, the LEG on the left, and the HEG on the right. Each of the three listener-groups is further subdivided into three experimental conditions, indicating the treatment of the primary cue: *Removed* (left column), *Unmodified* (center column), *amplified by 6 dB* (right column). For example, in Fig. 4.9 at the top, the ANH group has three columns. Likewise the LEG and HEG are each divided into the same three columns. The results for the other two experimental conditions, attenuated by -6 [dB] and amplified by 12 [dB], can be found in Appendices A and B.

This figure includes the data of Figs. 4.7, 4.8, as the two middle columns labeled “Unmodified.” The two left columns are labeled “Removed” and the right columns are labeled “Amplified by 6 dB,” indicating the type of modification to the primary cue in the presented tokens.

The expected result is best explained based on the ANH listener group. The unmodified and +6 [dB] errors are zero error (ZE) (one error occurred for one listener for one token (center column), but was removed in the +6 [dB] condition (right column)).

The LEG (Fig. 4.9, lower left) was similar to the ANH group in that the unmodified showed only a few errors, which typically went down, or even to zero, with the +6 [dB] boost of the primary cue.

The HEG however (Fig. 4.9, lower right) was a very different story. First the errors were much greater in the unmodified condition, and only one ear (HI06-L) improved when the primary cue was boosted by +6 [dB].

When the primary cue was removed (column 1) all three listener groups showed a dramatic degradation in performance. The ANH group went from zero error, to nearly 100% error. Between 1 and three tokens showed some error, depending on the SNR, with lower SNR (more noise) showing more error. This result is consistent with the design of the experiment, and with the results of Kapoor and Allen (2012). The LEG followed similar trends as the ANH listeners in that removing the primary cue dramatically increased the error. This proves that HE listeners having the more moderate loss are using the primary cues, like the ANH group.

The largest differences are with the HEG. The error increases dramatically with the removal of the primary cue in this case as well, as with the ANH and LEG. However since the error is much higher with no modification, the change is only for those tokens with small or zero error.

Both the LEG and HEG have lower error, on average, compared to the ANH. It is likely that this is due to the HI getting fewer errors, due to guessing. The way to test this would be to look at the entropy.

Histograms of raw data: The results of Fig. 4.9 are then summarized in Fig. 4.10 as error histograms, by binning the errors into three groups: zero error (ZE), non-zero error (NZE) and maximum error (ME). The format of Fig. 4.10 is identical to that of 4.9 (three listener groups each divided into three conditions). In this figure the main conclusions of Fig. 4.10 are clearly displayed. For the ANH (top) the unmodified conditions are all (but one token-listener) ZE. Boosting the primary cue results in 100% ZE tokens, and removing the cue (column 1) results in mostly ME along with a few NZE tokens.

The corresponding histograms for the LEG are very similar to those for the ANH listeners, but with a slight increase in the number of token errors. However it is notable that a +6 [dB] boost of the primary cue nearly removes all the few errors.

Finally, the HEG is very different from the LEG in that the number of NZE tokens is much higher, and therefore the number of ZE sounds must be lower. Boosting the primary cue by +6 [dB] has only a modest effect in driving down the errors. Removing the primary cue (column 1 of the left panel) is successful in removing most of the ZE sounds. Again this shows that even the HEG is using the primary cue when they get the sounds right.

Confusion patterns: Confusion patterns ($p_{h|s}(\text{SNR})$) are shown for two different tokens, a mid-frequency token f103 /ka/ (Fig. 4.11), and a high-frequency token m115 /ta/ (Fig. 4.12), as discussed in Sect. 4.2.3.

From the study of Kapoor and Allen (2012) we know what to expect for the ANH subjects: There is basically zero error at all SNRs unless the primary cue is removed, in which case the conflicting cues dominate the errors. (It is because of the important role of the conflicting cues that Experiment II was performed, to look at subject errors with conflicting cues removed.)

Briefly the results shown in the confusion patterns reflect what was observed in the sorted error plots and histogram summary, but with the CPs we can see the nature of the errors in greater detail, at the token level. Here the LEG error was high with the primary cue removed, however the errors were not zero as they are with the ANH group. With one LEG subject (HI04-R), the subject reported the

target /k/ correctly 66.67% of the time. However most of the time, when they were correct, the score was closer to 33.33%.

As with the raw data (and summary histograms) the HEG started with higher errors with unmodified sounds, but obtained high scores when the primary cue was boosted by +6 [dB]. Again this proves that the HEG can hear the /ka/ cue correctly, once amplified.

The confusion patterns for the high-frequency /ta/ sounds follow a similar pattern: The ANH group reports the sounds correctly 100% of the time for the unmodified and boosted cases, and has 100% error for the case of the primary feature removed. This follows the expected findings of Kapoor and Allen (2012). As with /ka/, the recognition of /ta/ for the LEG is perfect for the unmodified and boosted cases, and the primary cue removed confusions are similar, if not nearly identical, to those of the ANH ears.

The HEG however is the most interesting case, especially for subject HI05-L, which we must further discuss in Section 4.4.

Removed primary cue: As expected (Kapoor and Allen 2012), when removing the primary cue, there are no ZE tokens across all SNRs for the average NH ears. For the HI ears, 3 out of 6 HI ears in LEG (HI01-R, HI02-R, and HI04-R) have no ZE tokens across all SNRs. All other 3 HI ears in LEG have one ZE token. HI10-R have one ZE token at 0 [dB] and 18 [dB] SNR while the other two HI ears have one ZE token for one SNR condition. The results are similar when removing the primary cue for both average NH ears and HI ears in LEG. When the primary cue is removed, for the HI ears in HEG, three-fourths of the HI ears have either one or two ZE tokens. Therefore, given the removal of the primary cue, the HI ears in HEG are affected similarly to the average NH ears and the HI ears in LEG.

Amplified primary cue (+6 [dB]) for NH and LEG ears: By definition, the average NH ears and 4 HI ears (HI01-R, HI02-R, HI04-R, and HI08-R) in the LEG have ZE for all 8 tokens across all SNRs. Three of four HI ears (HI01-R, HI02-R, HI08-R) make errors when the primary cue is unmodified. Five of the six ears (HI01-R, HI02-R, HI03-R, HI08-R, HI10R) show reduced error when the primary cue is amplified by 6 [dB]. At 0 [dB] SNR, under the unmodified condition, HI10-R has 3 tokens with errors, reducing to a single error when the primary cue is amplified by 6 [dB]. When comparing the primary cue unmodified versus amplified by 6 [dB], the average NH ears and 5 of 6 HI ears increased their

number of ZE tokens.

Amplified primary cue (+6 [dB]) for HEG ears: The results of amplification of the primary cue by 6 [dB], suggest that the HI ears in HEG significantly improve their ability to recognize the primary cue. For instance, under the most extreme noise, for HI07-L, the number of ME tokens went from 5, when the primary cue is unmodified, to one token, for the amplification conditions. Furthermore, at 18 [dB] SNR, the number of ZE tokens increased by two for HI05-L. Results for when the primary cue was amplified by 6 [dB] are shown to be different when comparing the average NH ears and LEG versus HEG.

4.2.3 Confusion Patterns

Next we study the impact of varying the masking of the primary cue on the confusion patterns. The confusion patterns are a graphical interpretation of the response distributions. Figure 4.11 shows the confusion patterns for f103 /ka/ for the average NH ear, each of the LEG ears, and each of the HEG ears. The organization and the panels in the figure are set up the same way as Fig. 4.11. The ordinate and abscissa for each panel are the probability of being heard given that f103 /ka/ was spoken and SNR, respectively.

Removed primary cue for f103 /ka/: Column one of each ear group represents the removal of the primary cue for f103 /ka/. Both the average NH ears and every HI ear from LEG reported the conflicting cue /p/ at 0 [dB] SNR when the primary cue was removed. The average NH ears and two-thirds of HI ears (HI01-R, HI02-R, HI04-R, HI08-R) from the LEG reported a /t/ at 18 [dB] SNR. The average NH ears and every HI ear in the LEG, except HI01-R, reported /k/ (correct) for at least one SNR. Every HI ear from HEG reported a /p/ (incorrect) at least one time across all SNRs, except HI05-L at 18 [dB], for the case of the removal of the primary cue.

Unmodified primary cue for f103 /ka/: For the unmodified case (column 2), both NH ears and HI ears from the LEG reported a /k/ (correct) for 9 and 18 [dB] SNR. However, when the SNR decreased to 0 [dB], the average NH ears and half of LEG HI ears reported /p/ (incorrect) or ‘other’. For the unmodified f103 /ka/ at

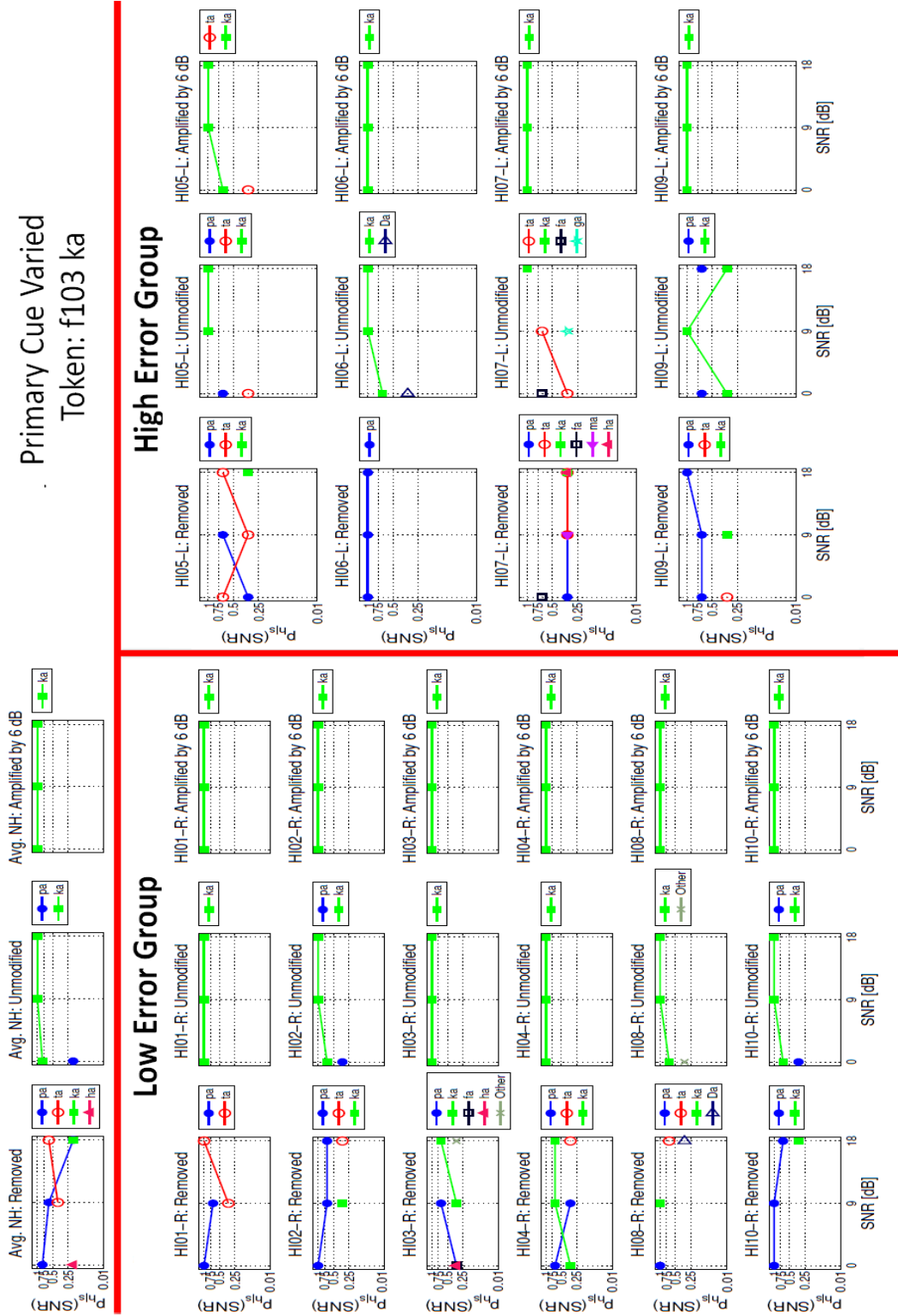


Figure 4.11: Confusion patterns for f103 /ka/ when varying the primary cue. The token f103 /ka/ is a mid-frequency sound in which the primary cue is located between 1-2 [kHz]. When the primary cue is removed, the error increases for all ears (average NH ears, and LEG and HEG ears), and the dominant confusion is reported for /p/. The error decreases for all ears when the primary cue for f103 /ka/ is amplified by 6 [dB].

Primary Cue Varied Confusion Patterns

Token: m115 ta

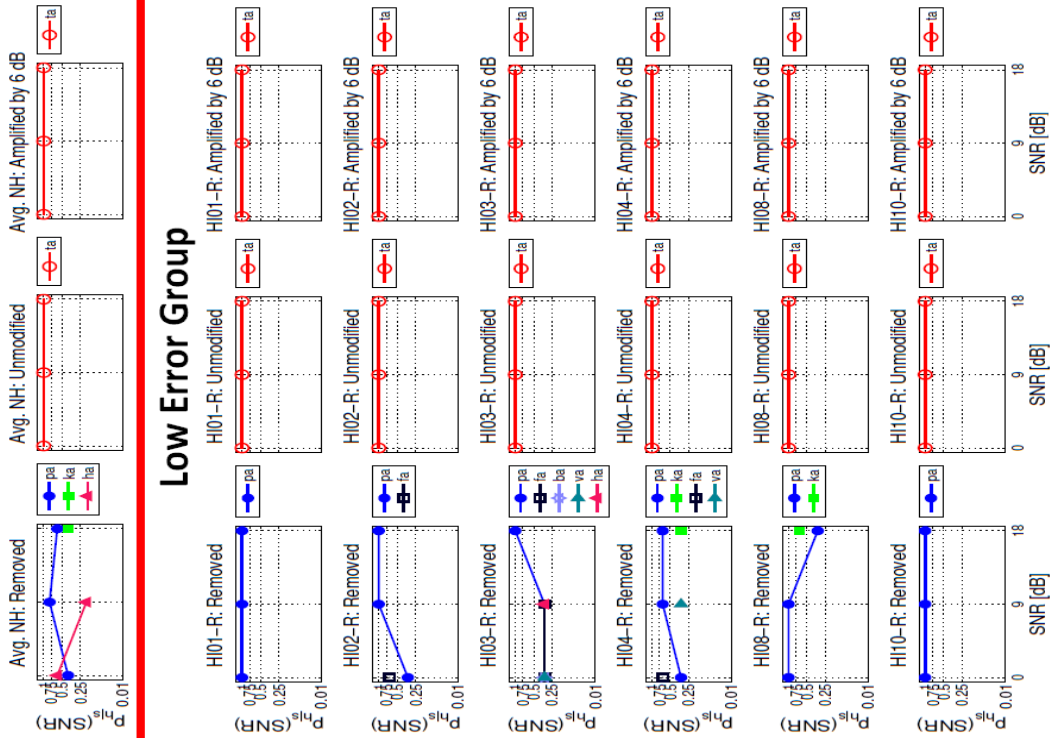


Figure 4.12: m115 /ta/ comparison of varying the primary cue using the confusion pattern method. The token m115 /ta/ is a high-frequency sound in which the primary cue is located above 4 [kHz]. The audiometry profile for the HEG ears shows severe to profound HL for frequencies above 4 [kHz](Table 3.1). Due to HL, the HEG ears have more errors than both the average NH and LEG ears.

18 [dB] SNR, every HI ear from HEG reported /k/ with a probability of 1, except for HI09-L.

Amplified primary cue for f103 /ka/: For the amplification condition (column 3) for the average NH ears and HI ears from LEG, no ears made any errors across all SNRs. However from the HEG, only one error was observed: a /t/ was reported by HI05-L at 0 [dB] SNR. Thus /t/ is very robust once the primary cue was boosted by 6 [dB].

Removed primary cue for m115 /ta/: The confusion patterns for m115 /ta/ are shown in Fig. 4.12. The average NH ears and the 6 HI ears of LEG have the same error and reported /p/ when the primary cue of m115 /ta/ was removed. Likewise, the error /p/ appears at every SNR except for HI03-R (at 0 [dB] SNR). Not a single ear from the average NH ears or the HI ears from LEG reported a /t/ across all SNRs. Unlike the average NH ears and HI ears from LEG, three-fourths of HEG ears (HI05-L, HI07-L, HI09-L) correctly heard consonant /t/ at 18 [dB] SNR. Like the average NH ears and HI ears from LEG, every HI ear from HEG reported a /p/ at 0 [dB] SNR.

Unmodified and amplified primary cue for m115 /ta/: Unlike the average NH ears and HI ears from LEG, when the primary cues were unmodified, and amplified by 6 [dB], no HEG ears were able to consistently recognize /t/ with a probability score of 100% across all SNRs. Three-fourths of the HI ears have some trials that were confused with /p/ for at least two SNR when m115 /ta/ was not modified. The same three-fourths of HI ears were still confused with /p/ even when a 6 [dB] amplification was applied to the primary cue.

4.3 Experiment II: The Removal of the Conflicting Cues

The purpose of Exp. II is to investigate the impact of the removal of the conflicting cues on HI ears, with and without the presence of the primary cues. The question being addressed is “Will the error change due to the removal of a token’s conflicting cue?” In cases where the conflicting cues are the dominant source of error (e.g., when the primary cues have been removed), then removing conflicting

cues should reduce the errors, or at least modify them in some significant way, since conflicting cues compose the confusion groups for NH listeners (Miller and Nicely, 1955).

In this section, we summarize the results of Exp. II to analyze the effects of removing conflicting cues, with and without the presence of the primary cues. The hope is that the errors will reduce in the HI ear, when the conflicting cues are removed, showing that the HI ear is influenced by these conflicting cues, namely that they are one of the primary sources of errors in HI ears.

4.3.1 Analysis of Exp. II

In the first experiment, by varying the masking of the primary cue, we demonstrate that the errors made by the listeners of our study (average NH ears, HI ears in the LEG, HI ears in the HEG) are generally confused with the consonants cues above and/or below the primary cue (Kapoor and Allen, 2012), that is, with the conflicting cues. We have shown that for HI ears, the strength of the primary cue is a significant source of error, especially in the presence of noise. For example, when we increased the magnitude of the primary cue, many of the errors go to zero. The strength of the primary cue for some tokens is too weak for an HI ear to recognize, especially if the frequency of the primary cue falls in the range where the level of HL [dB] is severe. Therefore, the error rate decreased for many listeners when we disambiguate the tokens by boosting the primary cue (Fig. 4.9).

The purpose of Experiment II is to investigate the removal of the conflicting cues for both NH ears and HI ears. If we remove the token's conflicting cues, will the error rate decrease more when comparing to the experiment for unmodified conflicting cues (Exp. I)? It makes sense that if the conflicting cues are the dominant source of errors, then removing these cues should benefit and improve speech perception, for in both NH and HI ears.

The research question is: Does the removal of the conflicting cues reduce the error when the primary cues are manipulated? This question is answered by using the raw data alone, and it is necessary to directly compare tokens from the two experiments. The practical way is to directly compare the two sets of the raw data (sorted error) for Exps. I and II.

Effect of Conflicting Cues for Average NH Ears

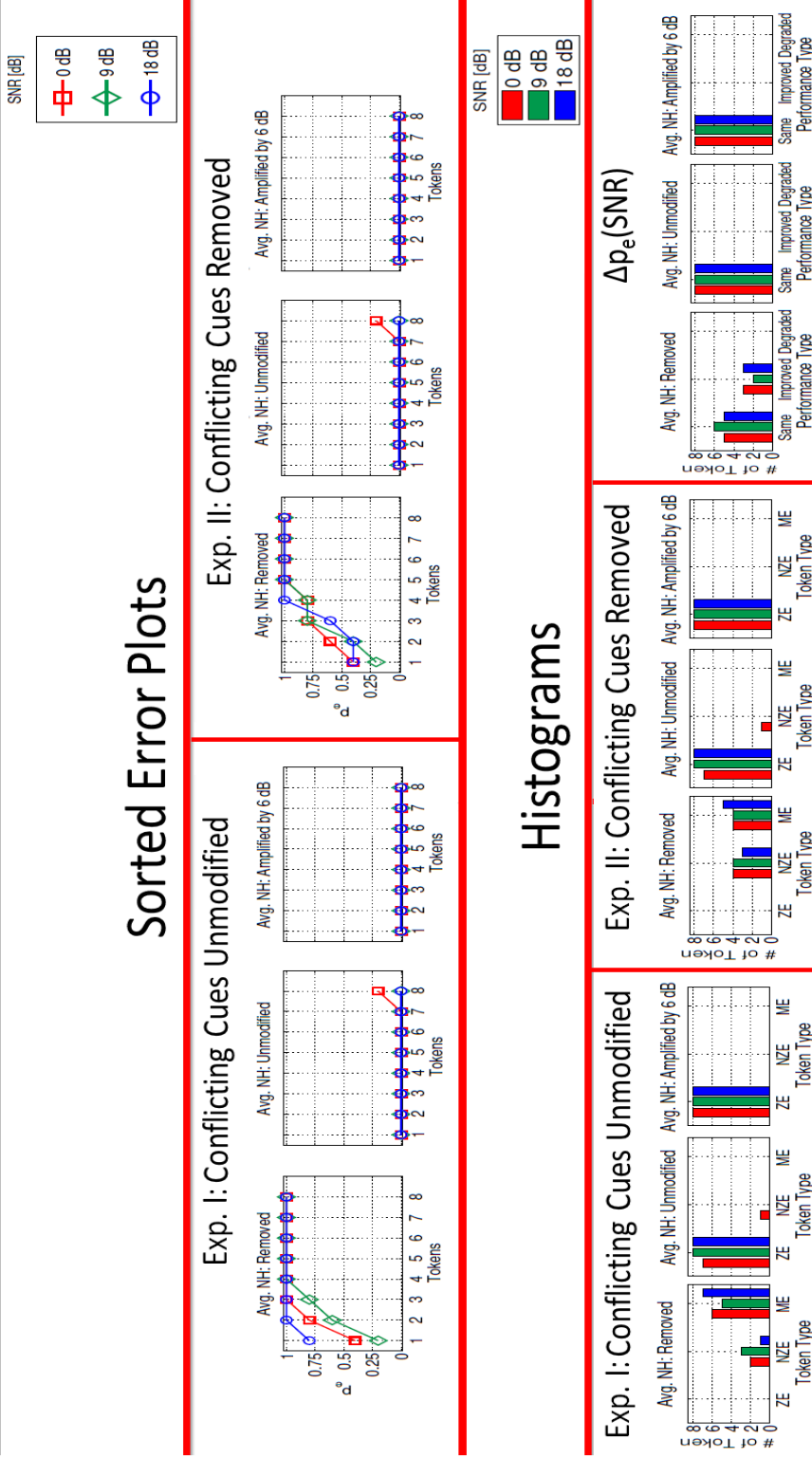


Figure 4.13: **Top:** Sorted error plots of Exp. I and II for the ANH ears. On the left are the results of Exp. I, *conflicting cues unmodified*. On the right are the results of Exp. II *conflicting cues removed*. For both experiments, the primary cue is removed, unmodified, or amplified by 6 [dB] (left, center, right columns). When the primary cue is removed, results show small difference between the two experiments. When the primary and conflicting cues are removed, the errors are much larger. **Bottom:** When the primary and conflicting cues are removed (for all SNRs), the number of ME tokens is decreased by one or two tokens, indicating a small effect of the cues for ANH ears. **Bottom-right:** The histogram of the sorted error difference between Exp. I,II is denoted as $\Delta P(\text{SNR})$ (p. 73). It is defined as $\Delta p_e(\text{SNR}) = p_e(\text{Exp. II,SNR}) - p_e(\text{Exp. I,SNR})$. The token is the same, improved, and degraded when $\Delta p_e(\text{SNR}) = 0$, $\Delta p_e(\text{SNR}) > 0$, and $\Delta p_e(\text{SNR}) < 0$, respectively. Removing the conflicting cues shows improvement only when the primary cue is removed, by three, two, and three tokens at 18, 9, and 0 [dB], as indicated by the SNR color-code in the legend.

Sorted Error Plots: Low Error Group

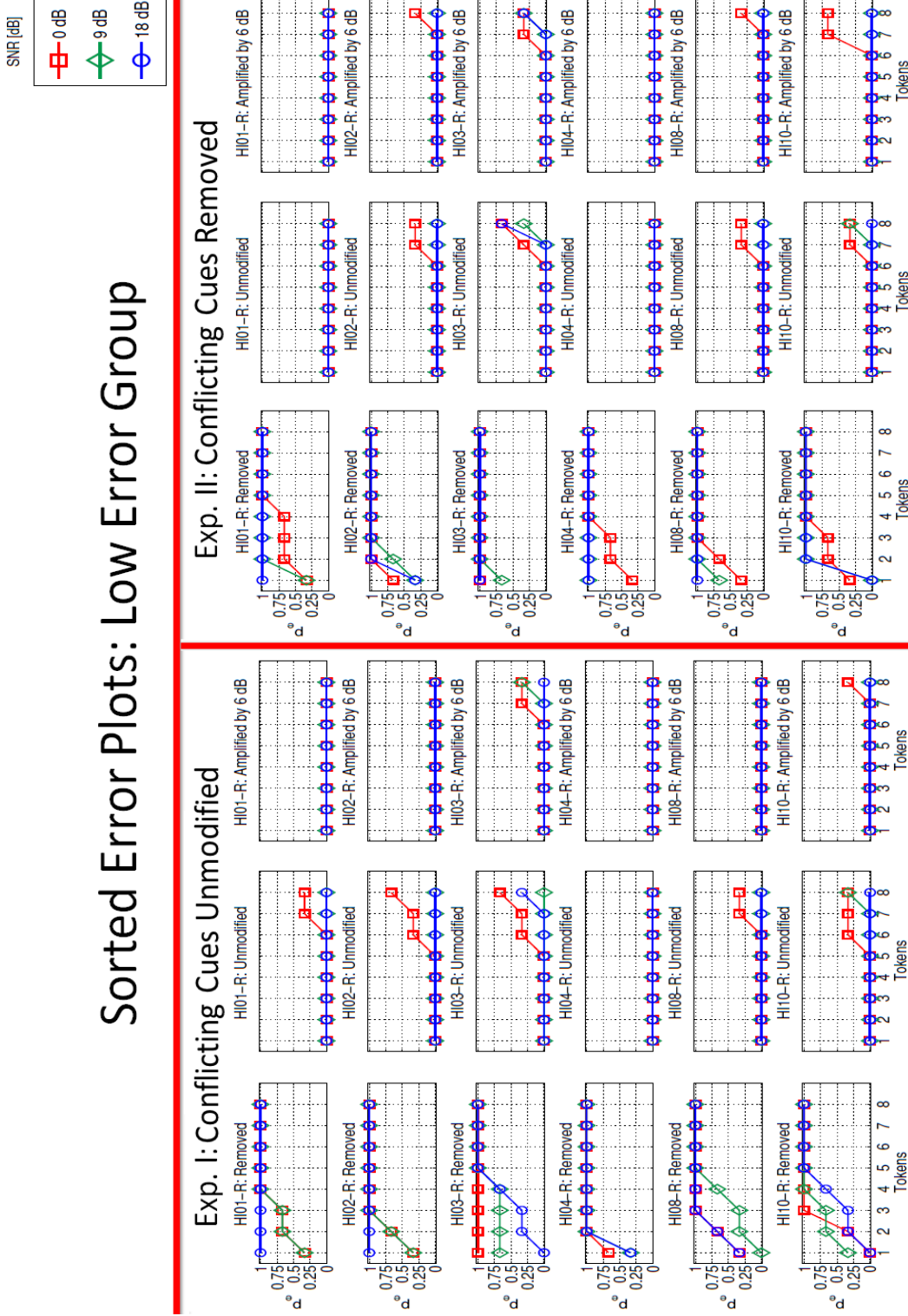


Figure 4.14: Sorted error plots for the LEG group. In each experiment, the primary cue was *Removed* (left column), *Unmodified* (middle column) and *Amplified by +6 dB* (right column). **Left:** Conflicting cues unmodified (Exp. I). **Right:** Conflicting cues removed (Exp. II). **Summary:** The LEG behaves similarly to the ANH group. Removing the primary cue increases the errors. The role of the secondary cues is small.

Direct comparisons of Exps. I and II raw data: This direct comparisons of the raw data from Exp. I and Exp. II is shown in Figs. 4.13 (ANH), 4.14 (LEG), and 4.16 (HEG). The figures are organized to provide a direct comparison between Exps. I (left) and II (right) for the three modification conditions, where the primary cue has been *Removed* (left column), *Unmodified* (middle column) and *Amplified by +6 [dB]* (right column).

The upper half of Fig. 4.13 shows the sorted error plots, while the lower half shows the corresponding histograms, with its three natural groups *ZE*, *NZE*, *ME*, as described in Fig. 4.8 (p. 54).

Histograms of the raw data from Exp. II: The sorted error plots from Exp. II are shown in Figs. 4.13 (ANH, top), 4.14 (LEG), and 4.16 (HEG). The corresponding histograms for the *ZE*, *NZE* and *ME* tokens are shown in Fig. 4.13 (ANH, bottom), Fig. 4.15 (LEG), and Fig. 4.17 (HEG). As may be seen from these three figures, the histograms for the ANH (Fig. 4.13) and LEG (Fig. 4.15) groups have mostly *ZE* tokens for both Exp. I and Exp. II, when the primary cue was unmodified (center column) and boosted by +6 [dB] (right column).

The most interesting results are concentrated in Fig. 4.17 (p. 72), for the HEG tokens, which are mostly spread between ZE and NZE, even when the primary cue was unmodified and boosted by 6 [dB]. There are a few cases where some HEG subjects have tokens spread across all three token types (ZE, NZE, ME).

Since the most tokens in both Exp. I and Exp. II have *ZE* when the primary cue is unmodified and boosted by +6 [dB], the removal of the conflicting cues does not improve the results for the ANH and LEG. Most of the tokens when computing the difference are the same.

The results from Column III for the HEG differ significantly from the ANH and LEG. The error in Exp. II when the primary cue is unmodified and boosted by +6 [dB] has either increased or remained the same. Therefore in most cases, the tokens degraded or stayed the same. Again as mentioned above, not all subjects have tokens that degraded, some have tokens that improved. Depending on the subjects from the HEG, the removal of the conflicting cues can enhance or diminish speech perception.

Summary: When we computed the histogram sorted error difference (Fig. 4.15, right column) we conclude that when the primary cue is removed, the LEG can hear the conflicting cues. In fact the LEG ears are similar to the ANH ears in this

Effect of Conflicting Cues for LEG Ears

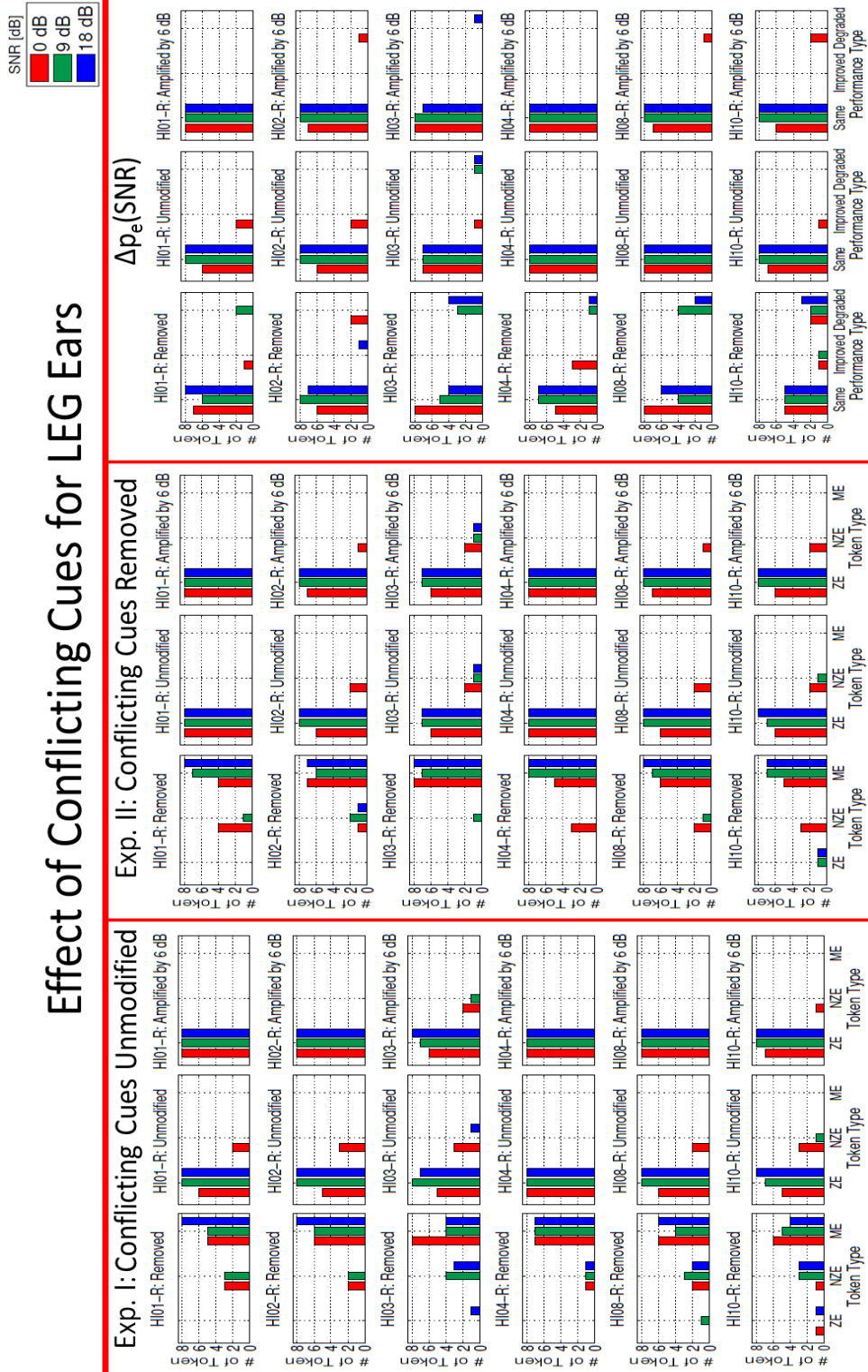


Figure 4.15: Histogram plots for the LEG ears. The raw data of Fig. 4.14 has been converted into histograms, making the results more accessible. **Left:** Conflicting cues unmodified (Exp. I). **Middle:** Conflicting cues removed (Exp. II). **Right:** Error difference, Δp_e (SNR). **Summary:** When the primary cue is removed all the sounds show dramatic improvements, and are zero for two ears. At 0 [dB] SNR, HI04-R has 3 tokens that improved when the primary cue is removed. When the primary cue is unmodified or amplified by +6 [dB], almost all the errors disappear. When the primary cue is unmodified at 0 [dB] SNR, HI01-R and HI02-R benefit the most when the conflicting cues are removed. They both have two tokens that improved. These results are similar to ANH ears. This figure proves that LEG HI ears can hear conflicting cues.

respect.

But the most dramatic results were those of the HEG, who were very sensitive to the presence of the conflicting cues. *This result is what we set out to show, and is possibly the most important finding of this study.*

Confusion patterns: As mentioned in summary section of Exp. I, we review the confusion patterns ($p_{h|s}(\text{SNR})$) for two tokens, a mid-frequency f103 /ka/ and a high-frequency m115 /ta/. The CPs are shown in Fig. 4.18 for ANH, Figs. 4.19-4.20 for LEG, and Figs. 4.21, 4.23 for the HEG. From Exp. I, we know the major source of error are the conflicting cues. The conflicting cues for f103 /ka/ are /ta/ and /pa/, and for m115 /ta/ are /ka/ and /pa/. We hope in Exp. II, the removal of the conflicting cues for f103 /ka/ and m115 /ta/ would enhance speech perception.

The $p_{h|s}(\text{SNR})$ results in Exp. II for the ANH and LEG are similar to the results in Exp. I when the primary cue is unmodified and boosted by +6 [dB]. If we compute and compare the difference between Exp. I and Exp. II for the ANH and LEG, we would find that the results have zero or small change in probability. Both ANH and LEG performed well in both experiments. The important takeaway from the results of Exp. I or Exp. II is that the primary cue of the token is the most important speech feature that the ANH and LEG relies on for speech perception.

In Fig. 4.21, the CPs results from Exp. I and Exp. II for HEG are notably different when the primary cue is unmodified and are the same when the primary cue is boosted by +6 [dB]. An additional analysis is used to break down the CPs into histograms as shown in Fig. 4.22. The figure is separated similar to the histogram of the raw data, the left column is Exp. I, the middle column is Exp. II, and the right column is the probability difference $\Delta p_{h|s}(\text{SNR})$. Each response from the CPs is classified as primary cue (PC), conflicting cues (CC), secondary cues (SC), noise (N), and other (O). The secondary cues are sounds that are not the PC, CC, N, or O. An example of sounds that pertained to 'O' are not listed in stimuli corpus as shown in Table 3.4 - 3.5.

Column III of Fig. 4.22 shows the effect of the removal of the conflicting cues. The effect of the removal of the conflicting cues is defined as the probability difference between Exp. I and Exp. II, $\Delta p_{h|s}(\text{SNR})$. The response for the primary cue improved in Exp. II when the $\Delta p_{h|s}(\text{SNR})$ increased and decreased for the other responses (CC, SC, N, O). Likewise, the response for the primary cue degraded in Exp. II when the $\Delta p_{h|s}(\text{SNR})$ decreased and increased for the other responses.

The results from Column III of Fig. 4.22 show the response of the primary

Sorted Error Plots: High Error Group

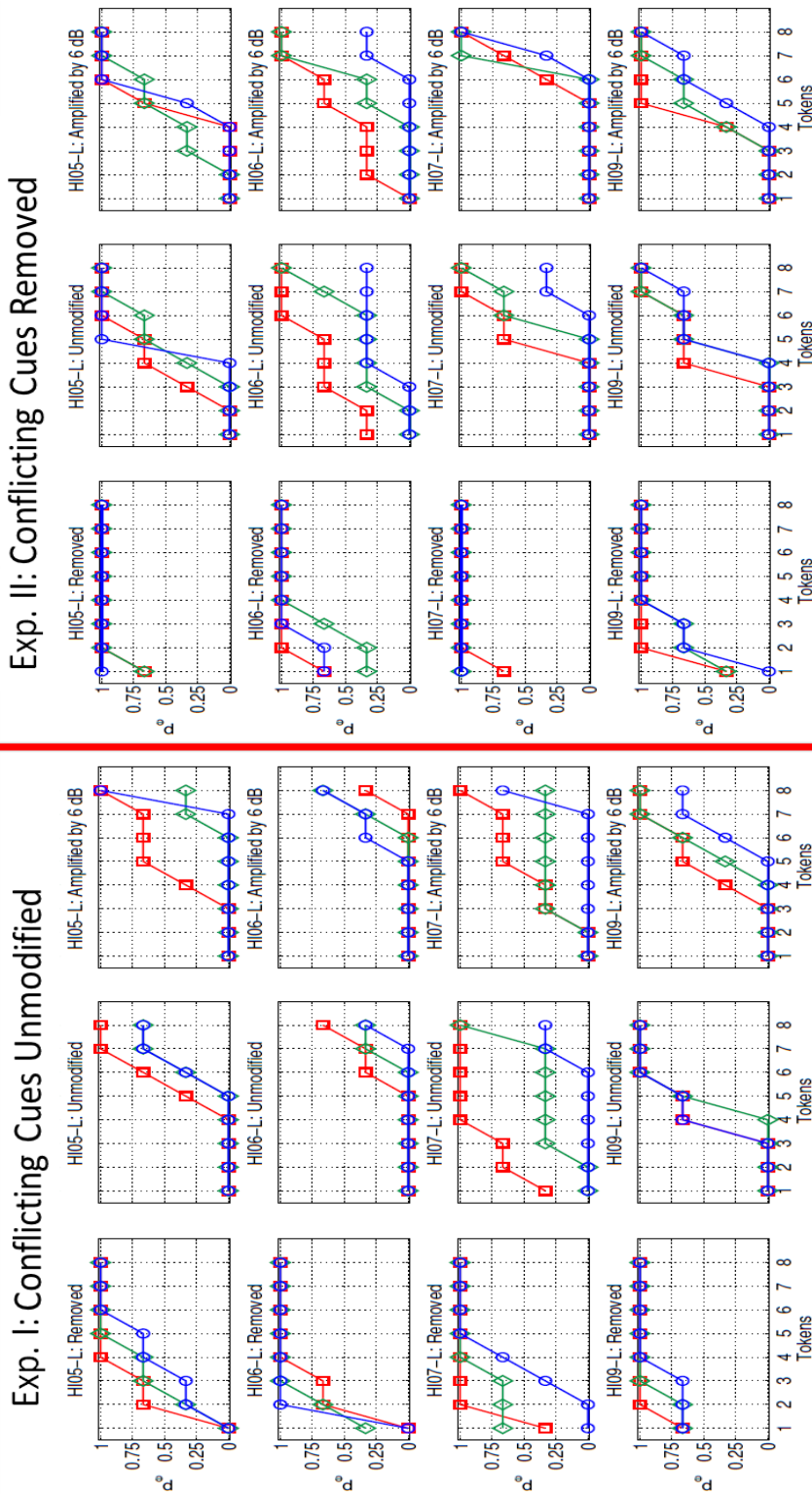
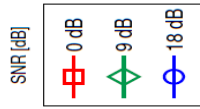


Figure 4.16: The sorted error plots for the HEG ears: conflicting cues unmodified (Exp. I) versus the conflicting cues removed (Exp. II). When the primary cue is removed in both experiments, only two ears, HI06-L and HI09-L, have tokens that decrease in error in Exp. II. For both experiments when the primary cue is unmodified, HI07-L and HI09-L are the only ears in HEG to have tokens with decreased error in Exp. II. HI07-L has the biggest decrease in error in Exp. II when comparing both experiments when the primary cue is amplified by 6 [dB].

cue in Exp. II when the primary cue is unmodified, degraded for half of the ears from the HEG, which means that the $\Delta p_{h|s}(\text{SNR})$ for the primary cue decreased. Since the difference for the primary cue decreased, the conflicting cues increase the error, and $\Delta p_{h|s}(\text{SNR})$ for conflicting cues increased. For the +6 [dB] case in Exp. II, the $\Delta p_{h|s}(\text{SNR})$ for the primary cue has zero or small effects. In other words, the difference between Exp. I and Exp. II has zero or small probabilities.

Figure 4.23 and 4.24 show the CPs and histograms of m115 /ta/ for the HEG. Using the histograms to understand the patterns of error from the CPs, shows the $\Delta p_{h|s}(\text{SNR})$ for primary cue decreased for half of HEG ears in Exp. II when the primary cue is unmodified. Furthermore, the +6 [dB] boost increased the number of HEG ears that have $\Delta p_{h|s}(\text{SNR})$ for primary cue decreased. The increased errors are closely distributed equally between the conflicting and secondary cues. The $\Delta p_{h|s}(\text{SNR})$ for both conflicting and secondary cues increased about the same. The removal of the conflicting cues for m115 /ta/ demonstrates that *conflicting cues* are important speech features for some ears from HEG for correct recognition. The *secondary cues* are also vital features in speech perception. Section 4.4 provides results and a brief discussion of an example of a case study that shows the importance of the secondary cues for a particular subject who relies on both primary and secondary cues to identify m115 /ta/.

4.3.2 Effect of Conflicting Cues for the Average NH Ears

Sorted Error: The comparison between Exp. I (conflicting cues unmodified) versus Exp. II (conflicting cues removed) for the average NH ears, using the sorted error and histogram plots, is shown in Fig. 4.13. The difference between the two experiments is the removal of the conflicting cues. The tokens for both experiments have their primary cue varied. In Exp. II, the errors in column I decrease when both the primary and conflicting cues are removed. In general, the errors went down and became independent of SNR.

Histograms: In columns I and II of the histograms for Exp. I versus Exp. II, the number of tokens shows differences for the NZE and ME. For the NZE token in Exp. II, the number of tokens increased by two, one, and two at 18, 9, and 0 [dB], respectively. The number of tokens that increased for NZE decreased proportionally for ME by two, one, and two at 18, 9, 0 [dB], respectively.

Effect of Conflicting Cues for HEG Ears

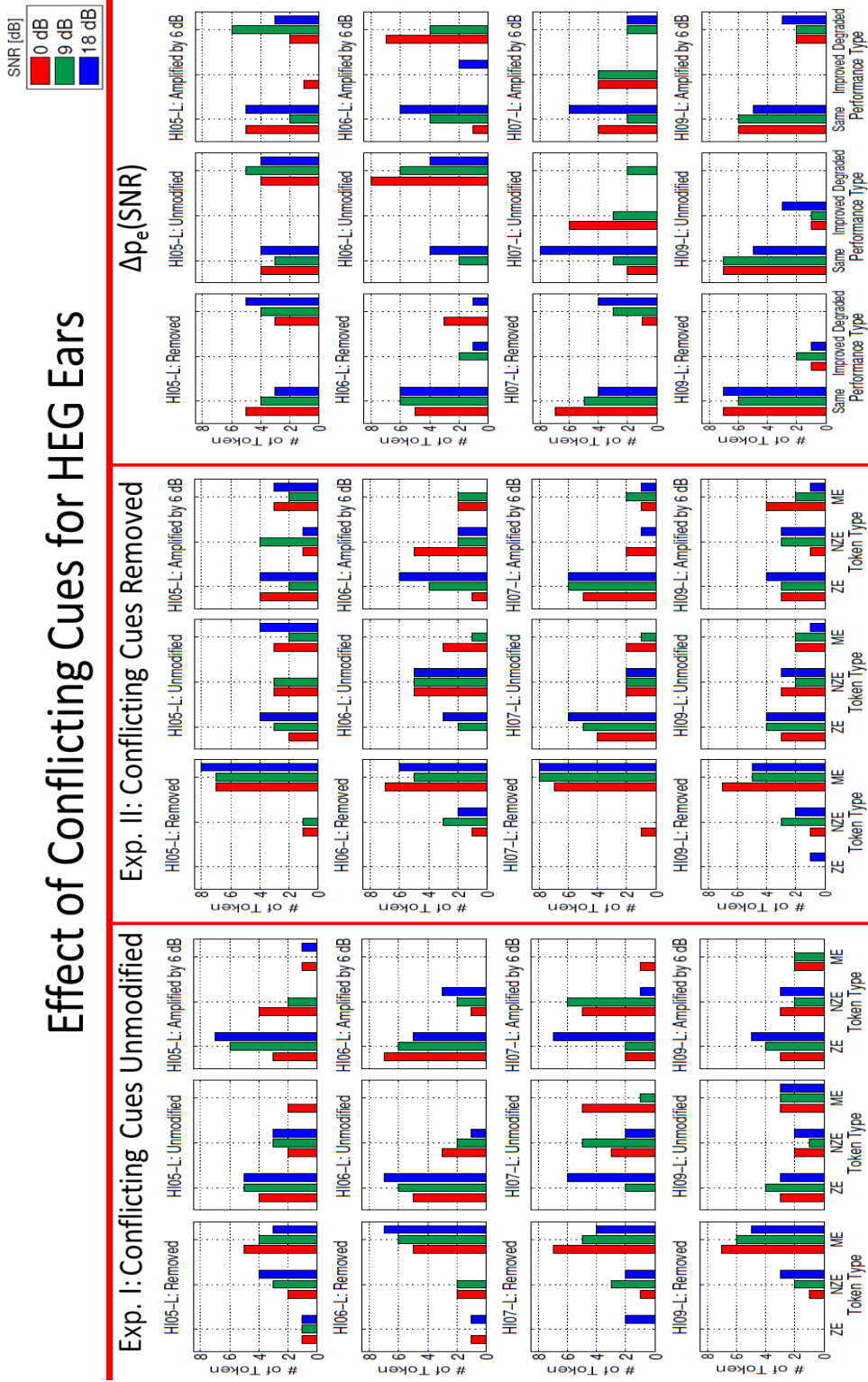


Figure 4.17: The effect of conflicting cues for the HEG ears using the histogram analysis. For the removal of the primary cue in both experiments at 0 [dB] SNR, two tokens improved for HI06-L and HI09-L in Exp. II. For Exp. I versus Exp. II, when the primary cue is unmodified at 0 [dB] SNR, HI07-L has the maximum number of tokens that improved in Exp. II, while HI06-L has the maximum number of tokens that degraded in Exp. II. In addition to the same modification condition for all SNRs, HI05-L and HI06-L have at least half of the set of tokens that degraded in Exp. II. The conflicting cues have an important influence on HI05-L and HI06-L.

$\Delta p_e(\text{SNR})$: Column III of the histogram shows the sorted error difference between Exp. I and II defined as $\Delta p_e(\text{SNR}) = p_e(\text{Exp. II, SNR}) - p_e(\text{Exp. I, SNR})$. The histogram results for $\Delta p_e(\text{SNR})$ show the number of tokens that are the same, improved, and degraded. The number of tokens that degraded for ANH is zero and is the same for all modification types. The token is the same, improved, and degraded when $\Delta p_e(\text{SNR}) = 0$, $\Delta p_e(\text{SNR}) > 0$, and $\Delta p_e(\text{SNR}) < 0$, respectively. The average NH ears show improvement for 3, 2, and 3 tokens when the primary cue is removed. Therefore, removing the conflicting cues only has an impact on the average NH ears when the primary cue is removed.

4.3.3 Effect of Conflicting Cues for LEG Ears

Sorted Error: Sorted error method is shown in Fig. 4.14. When comparing Exp. I and II for token 1 through 3 at 0 [dB] SNR, HI04-L has the biggest error decreases at 0 [dB] in Exp. II. The error decreased from $\frac{2}{3}$, 1, and 1 to $\frac{1}{3}$, $\frac{2}{3}$, and $\frac{2}{3}$, respectively. The removal of the primary cue in both experiments at 18 [dB] SNR, for token 1 through 4, HI03-R has increased error for 4 tokens in Exp. II. The error increased from 0, $\frac{1}{3}$, $\frac{1}{3}$, $\frac{2}{3}$, to 1, 1, 1, and 1, respectively. The unmodified primary cue for both Exp. I and II, HI01-R, HI02-R, and HI03-R, have tokens with error decreased at 0 [dB] SNR. The removal of the conflicting cues benefits HI01-R, HI02-R, and HI03-R. On the other hand, HI10-R has token with error increased when the primary cue was unmodified and amplified in Exp. II.

Histograms: We believe Fig. 4.15 demonstrates that the primary cue is the most important speech perceptual cue in human speech recognition for both NH ear and HI ears. The removal of the conflicting cues benefits most of the HI ears in the LEG as for example, subject HI01-R. In the first experiment at 0 [dB] SNR, the subject has 6 ZE tokens when the primary cue unmodified. For the case of the removal of the conflicting cues with primary cue is unmodified, the number of ZE tokens increases from 6 to 8. HI02-R and HI03-R have one ZE token that increases when the primary cue is unmodified in Exp. II at 0 [dB] SNR. This demonstrates that conflicting cues play a significant role in the errors made by HI ears.

$\Delta p_e(\text{SNR})$: Column III in Fig. 4.15 shows the error difference, $\Delta p_e(\text{SNR})$, on the LEG ears. All ears in the LEG have at least one token that degrades when the primary and conflicting cues are removed. HI03-R, HI08-R and HI10-R have at

least 3 tokens that degrade when both cues are removed at a specific SNR. HI01-R and HI02-R have the most tokens, which is two, that improved at 0 [dB]. When the primary cue is amplified, HI10-R has two tokens that degrade.

4.3.4 Effect of Conflicting Cues for HEG Ears

Sorted Error: Figure 4.16 shows the comparison of Exp. I versus Exp. II, for HEG ears. This figure is very important because the HEG is the most sensitive to the conflicting cues, as shown in the figure.

Removed primary cue: When the primary cue is removed in Exp. II, the error increases from Exp. I by several tokens for HI05-L and HI07-L. For HI05-L, at 18 [dB] SNR, 5 tokens with error increase from 0, $\frac{1}{3}$, $\frac{1}{3}$, $\frac{2}{3}$, and $\frac{2}{3}$ in Exp. I, to $\frac{2}{3}$, 1, 1, 1, 1, and 1 in Exp. II, respectively. At the same SNR for HI07-L, the error increases from 0, 0, $\frac{1}{3}$, $\frac{2}{3}$ in Exp. I, to $\frac{2}{3}$, 1, 1, and in Exp. II, respectively.

Unmodified primary cue: For the unmodified primary cue at 0 [dB] SNR in Exp. II, the error increases for HI05-L. The error increases from 0, 0, $\frac{1}{3}$, and $\frac{2}{3}$ in Exp. I, to $\frac{1}{3}$, $\frac{2}{3}$, $\frac{2}{3}$, and 1 in Exp. II, respectively. At the same SNR for HI07-L, the error decreases from $\frac{1}{3}$, $\frac{2}{3}$, $\frac{2}{3}$, 1, 1, and 1 in Exp. I, to 0, 0, 0, 0, $\frac{2}{3}$, and $\frac{2}{3}$ in Exp. II, respectively.

Amplified primary cue: When comparing both experiments for the amplified primary cue at 0 [dB] SNR, 7 tokens with error increased for HI06-L, and 4 tokens with error decreased for HI07-L.

Histograms: The histograms for Exp. I,II for the HEG ears are shown in columns I and II of Fig. 4.17, respectively.

Removed primary cue: In Exp. I with primary cue removed, three HI ears have at least one token with ZE, for at least one SNR. Following the removal of the conflicting cues, the number of HI ears reduced to one HI ear (HI09-L) who had a token with ZE at 18 [dB] SNR. When the primary cue was removed in Exp. I, all HEG ears for all SNRs had at least four tokens with ME, and five tokens when the conflicting cues were removed.

Unmodified primary cue: Half of the HI ears in HEG (HI07-L, HI09-L) improve when the conflicting cues were removed, when the primary cue was unmodified. On the other hand, the other two HI ears (HI05-L, HI06-L) have more tokens with ME and therefore did not improve. In Exp. II with the primary cue amplified by 6 [dB], three-fourths of the HI ears improve when compared to the unmodified primary cue. Specifically three HI ears (HI05-L, HI06-L, HI07-L) have fewer tokens with ME.

Amplified primary cue: The results for 6 [dB] amplification for Exp. I versus Exp. II, show (except for HI07-L) three HI ears (HI05-L, HI06-L, HI09-L) performed better when the conflicting cues were present. Some of these tokens may have critical secondary cues near the primary cue that HI ears in the HEG use for correct identification. A demonstration of this the role of secondary cues on HI05 will be shown in Sect. 4.4. The parameters for removing the conflicting cues may need to be spread further from the primary cue, because the cues that were removed might be the related or part of the primary cue.

$\Delta p_e(\text{SNR})$: The histogram of the error difference, $\Delta p_e(\text{SNR})$, on the HEG ears in illustrated in Column III in Fig. 4.17. We will show that the $\Delta p_e(\text{SNR})$ is a great tool that provides evidence that HI ears do depend on conflicting cues.

Removed primary cue: HI05-L has the most tokens that degraded at each SNR. The number of tokens that degraded for HI05-L is 5, 4, and 3 at 18, 9, and 0 [dB] SNR, respectively. As you can see, the masking effect reduces the number of tokens that degraded for HI05-L. When both primary and conflicting cues are removed, HI09-L is the only ear that has improvement at each SNR. The number of tokens that improved for HI09-L is 1, 2, and 1 at 18, 9, and 0 [dB] SNR, respectively.

Unmodified primary cue: HI05-L and HI06-L have the most tokens that degraded. At 9 and 0 [dB] SNR, HI06-L has 6 and 8 tokens that degraded, respectively. HI05-L and HI06-L have 4 tokens that degraded at 18 [dB] SNR. The removal of the conflicting cues when the primary cue is unmodified does not improve speech perception for HI05-L and HI06-L. Results show that they depend on the conflicting cues. Unlike HI05-L and HI06-L, the removal of the conflicting

cues does improve speech perception for HI07-L. HI07-L has 3 and 6 tokens that improved at 9 and 0 [dB] respectively.

Amplified primary cue: HI05-L and HI06-L have more tokens degraded than improved for the error difference, $\Delta p_e(\text{SNR})$. HI05-L has 3 tokens that degraded at 18 [dB] SNR, and HI06-L has 4 and 6 tokens that degraded at 9 and 0 [dB] SNR, respectively. HI07-L has 4 tokens that improved at 9 and 0 [dB] SNR.

4.3.5 Effect of Conflicting Cues for f103 /ka/ and m115 /ta/ on the Average NH Ears

We analyze the confusion patterns for f103 /ka/ and m115 /ta/ for Exp. 1 versus Exp. II. We will study the patterns of errors made by each ear, and study the influence of conflicting cues. We wish to answer the key question: Can the removal of the conflicting cues improve speech intelligibility for HI ears as it did for NH ears (Li and Allen 2011)?

Confusion patterns for f103 /ka/ on the average NH ears (Fig. 4.18): Our analysis begins with Fig. 4.18, which shows the average NH ears. When the primary cue was removed for f103 /ka/, there are more confusions for conflicting cues unmodified (Exp. I). In the first column of Exp. I, the entropy increased when the conflicting cues were unmodified. The number of confusions are reduced from 3 when the conflicting cues were unmodified, to 2 when the conflicting cues were removed. In the first column for both experiments, when the noise was added, the consonant /pa/ is the dominant confusion. In the second column for both experiments, when the primary cue was unmodified, one subject from the average NH ear made an error (0 [dB] SNR), and no error was made when the primary cue was amplified by 6 [dB].

Confusion patterns for m115 /ta/ on the average NH ears (Fig. 4.18): When the primary cue was removed for m115 /ta/, the number of confusions for both experiments is the same. Like f103 /ka/, consonant /pa/ is the dominant confusion. For both experiments, the second and third columns for m115 /ta/, the NH ears have zero errors.

Confusion Patterns: Average NH Ears

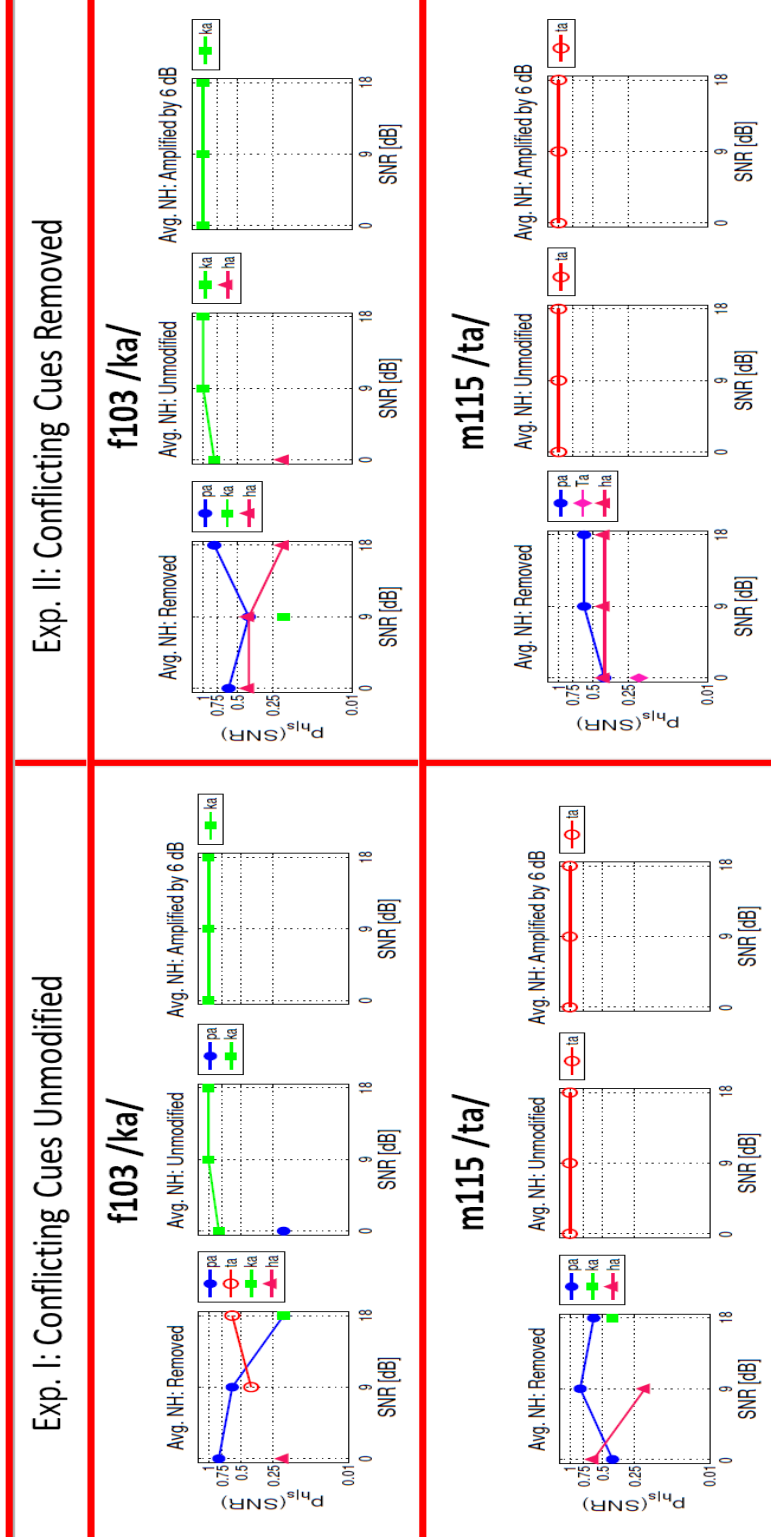


Figure 4.18: Exp. I versus Exp. II confusion patterns for f103 /ka/ and m115 /ta/. Results are for the two sounds. The error is high when the primary cue is removed and zero or very low when the primary is unmodified. The errors are similar, independent of the conflicting cues, as expected. For the case when the primary cue is amplified by 6 [dB], the errors are zero.

Confusion Patterns for f103 /ka/: Low Error Group

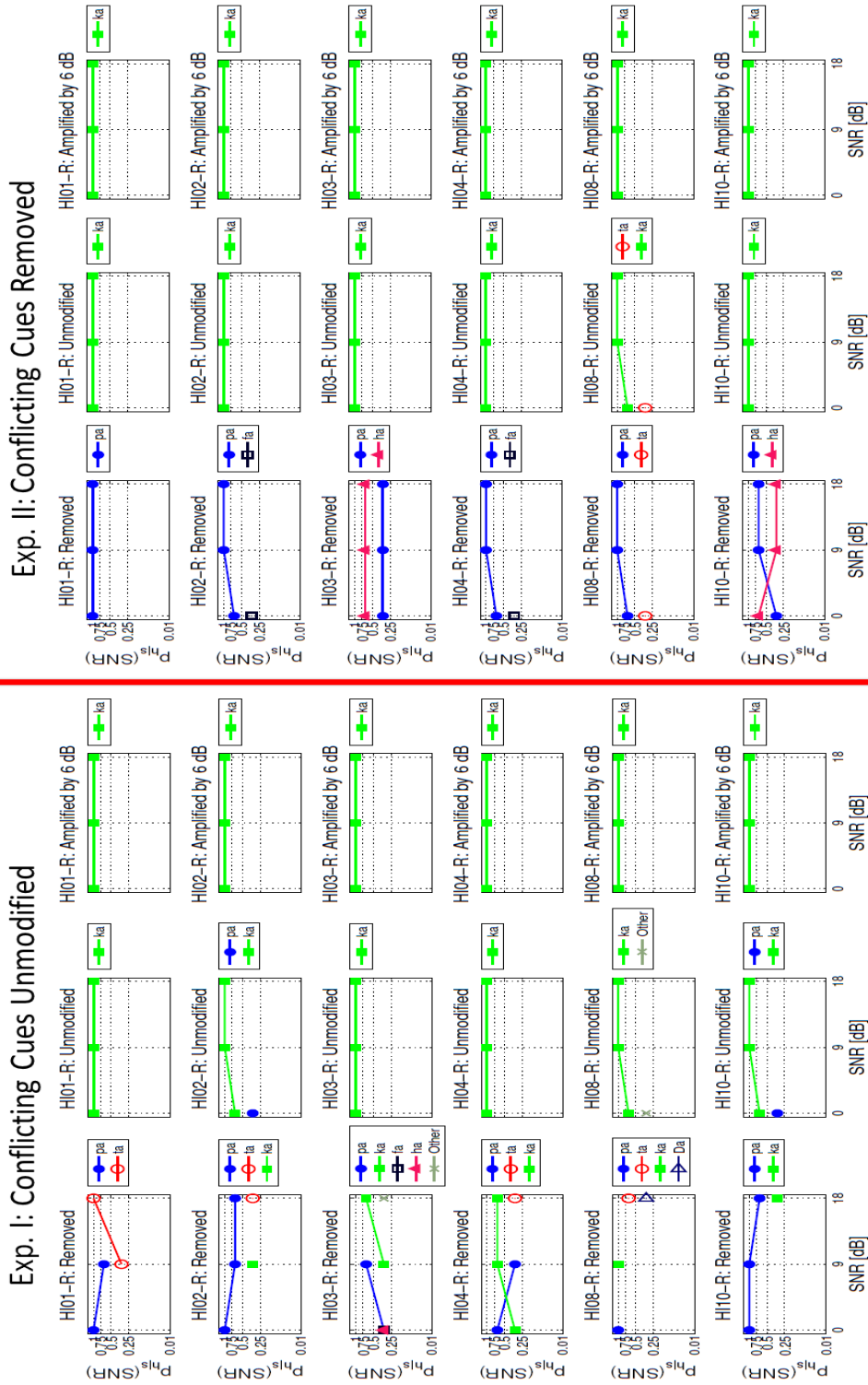


Figure 4.19: LEG confusion results for Exp. I,II for f103 /ka/. The results are very similar to the average NH ears, showing no significant differences. There are differences when the primary cue is removed, showing that the LEG is sensitive to conflicting cues. Note this shows that primary and conflicting cues are audible at MCL for the LEG.

Confusion Patterns for m115 /ta/: Low Error Group

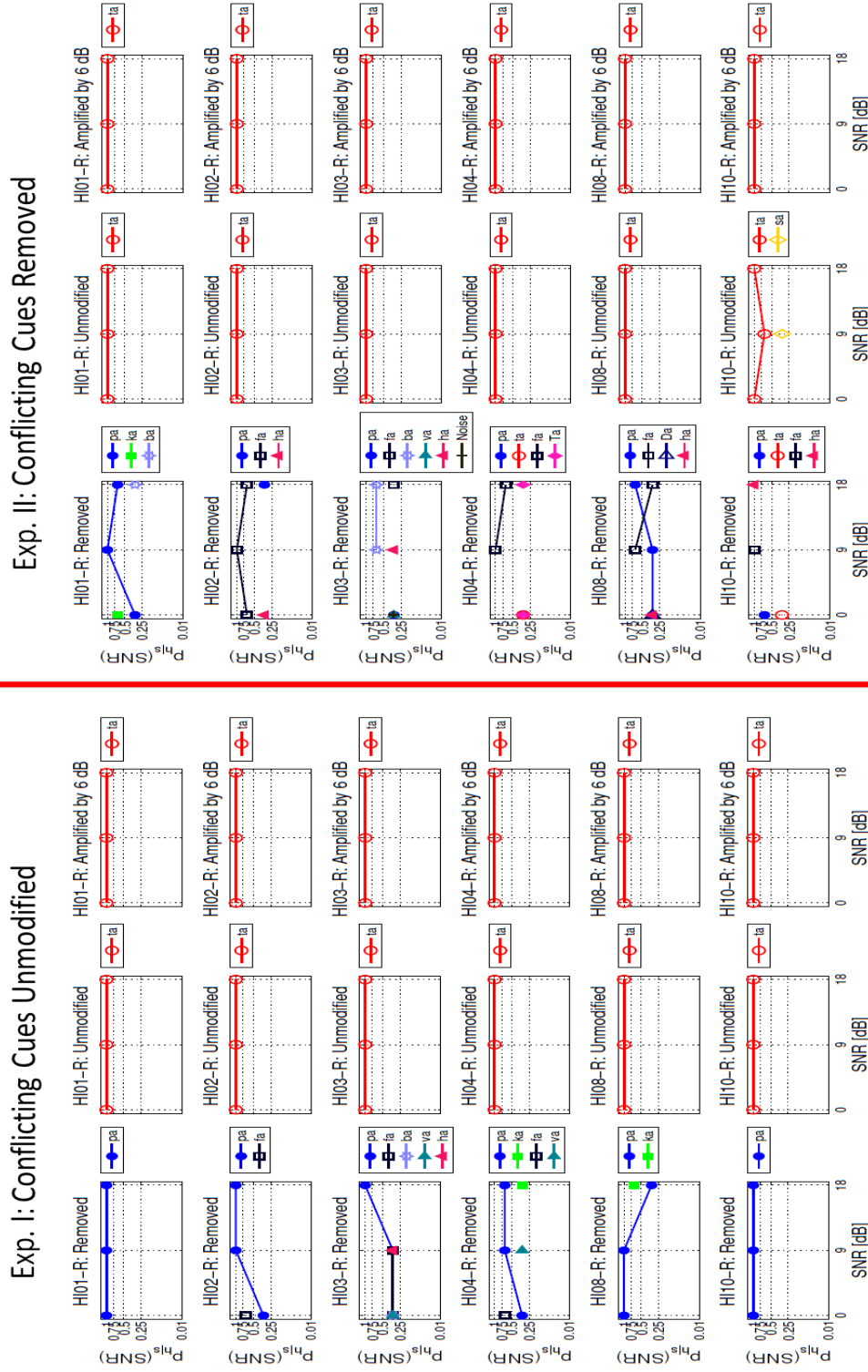


Figure 4.20: The LEG ears' confusion patterns for token m115 /ta/. This shows negligible error for unmodified and 6 [dB] cases. The error is the same independent of the conflicting cues that were unmodified or removed unless the primary cue was removed, then error is similar to the ANH ears..

4.3.6 Effect of Conflicting Cues for f103 /ka/ and m115 /ta/ on the LEG Ears

Confusion patterns for f103 /ka/ on the LEG ears (Fig. 4.19): The confusion patterns for f103 /ka/ for LEG, are shown in Fig. 4.19. The number of confusions is either very small or zero, when the conflicting cues were removed unless the primary cue was removed. The confusion /pa/ was the most common error when the conflicting cues were removed. No ears reported a /ka/, once the conflicting cues were removed. Given the unmodified tokens, two HI ears (HI02-R, HI10-R) reported a confusion /pa/, at 0 [dB] SNR, when the conflicting cues were not removed. As a result of the removal of the conflicting cues, the same two ears that reported a confusion /pa/, have no confusion at any SNR. We concluded that the errors made by HI02-R and HI10-R were due to the presence of the conflicting cues. HI02-R and HI10-R reported /ka/ once the conflicting cues were removed.

Confusion patterns for m115 /ta/ for the LEG ears (Fig. 4.20): When the primary cue was removed for Exp. I and Exp. II, the number of confusions increased once the conflicting cues were removed, with entropy between 0-1 [bits]. For two HI ears (HI08-R and HI10-R) the number of confusions increased from 1, for Exp. I, to 3, for Exp. II. The removal of both primary cue and conflicting cues caused the entropy to increase for m115 /ta/, especially in the presence of noise. Four HI ears (HI02-R, HI03-R, HI08-R, HI10-R) and the average NH ears have the same confusions, /pa/ and /ha/. The errors for all HI ears from the LEG are the same for both experiments (Exp. I vs. Exp. II), when the primary cue was unmodified and amplified by 6 [dB], except for HI10-R, who made an error when the primary cue was unmodified and the conflicting cues were removed.

4.3.7 Effect of Conflicting Cues for f103 /ka/ and m115 /ta/ on the HEG Ears

Confusion patterns for f103 /ka/ (Fig. 4.21): The confusion patterns for f103 /ka/ are shown in Fig. 4.21. The /ka/ sound is a mid-frequency burst. The associated conflicting cues for /ka/ are the /ta/ sound in the high-frequency and /pa/ in the low-frequency. Any sounds that are not associated as primary or conflicting cues are the other secondary cues. In this section, we will provide more evidence to demonstrate the importance of the primary cue for correct recognition. We will

Confusion Patterns for f103 /ka/: High Error Group

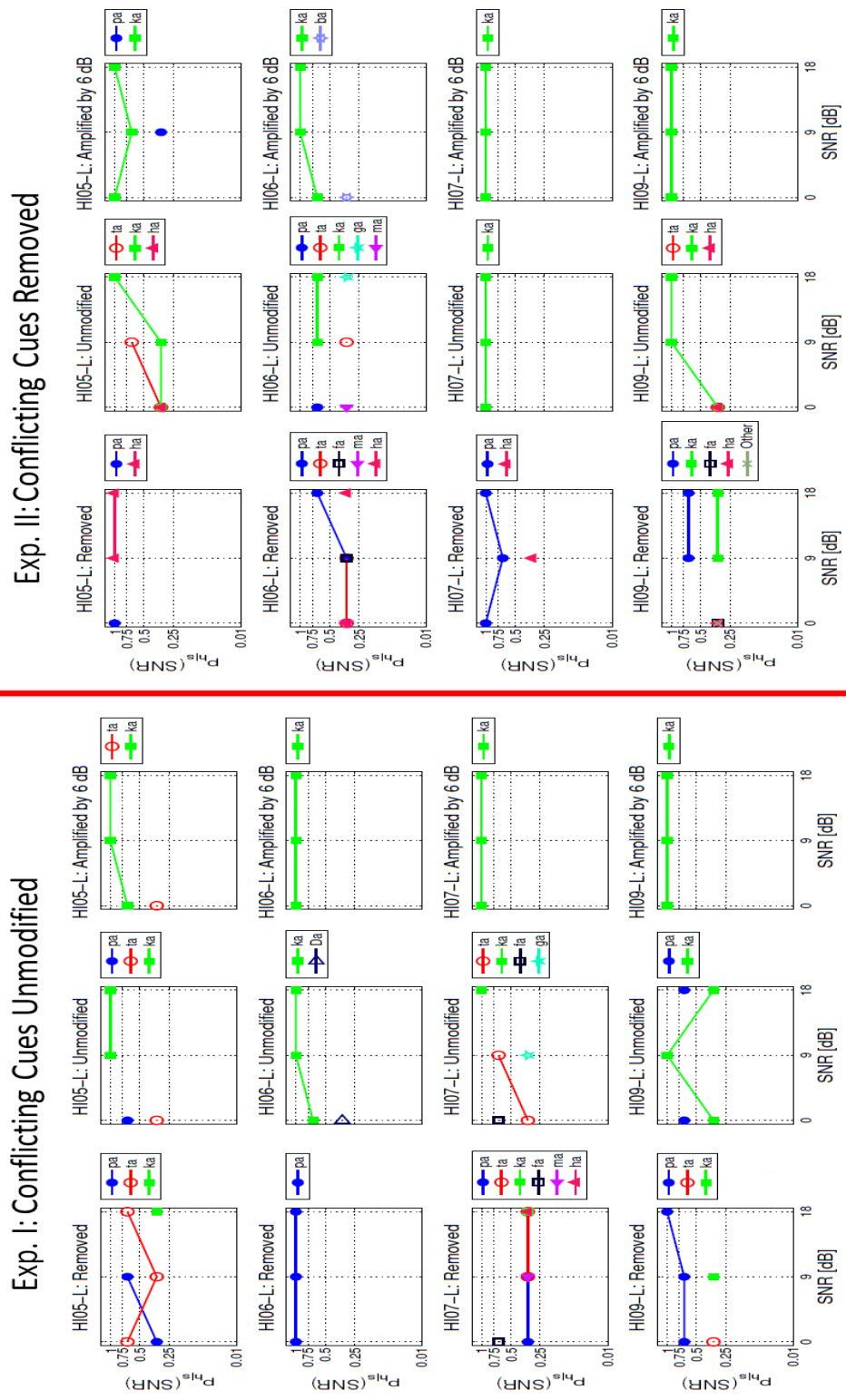


Figure 4-21: Mid-frequency sound f103 /ka/ confusion patterns results for HEG ears. For the removal of the primary cue in both Exp. I, II, all ears hear the confusion /pa/. The error decreased for HI07-L when the primary cue is unmodified in Exp. II, while the error increased for HI05-L and HI06-L. The number confusions for HI07-L decreased from 3 in Exp. I, to zero in Exp. II, while for HI06-L, it increased from 1 in Exp. I, to 4 in Exp. II. The effect of the removal of the conflicting cues has both positive and negative impact on the HI ears in the HEG.

also provide evidence that conflicting cues for some HEG ears are also important for correct recognition.

Removed primary cue: For both experiments, the error is high when the primary cue is removed. When comparing Exp. I and Exp. II, the number of confusions for three-fourths of the HI ears (HI05-L, HI06-L, HI09-L) have increased or remained the same when the primary cue is removed in Exp. II. The number of confusions for HI07-L has decreased from 5 in Exp. I to one in Exp. II. For both experiments, the confusion /pa/ was the dominant error for most HI ears. In Exp. I, three HI ears reported confusion /ka/ at some SNRs when the primary cue is removed. However, when the primary and conflicting cues are removed, the number of HI ears who reported a /ta/ confusion was reduced to 1.

Unmodified primary cue: Independent of the conflicting cues, the HI ears in the HEG depend on the primary cue. For example, for both experiments, compare the cases where the primary cue is unmodified. The removal of the conflicting cues when the primary cue is unmodified benefits two HI ears (HI07-L, HI09-L). At 9 and 0 [dB] SNR, the error decreased for HI07-L, from 1 in Exp. I, to 0 in Exp. II. HI06-L did not improve when the primary cue is unmodified in Exp. II. At 18, 9, and 0 [dB] SNR the error increased from 0, 0, and $\frac{1}{3}$ in Exp. I, respectively, to $\frac{1}{3}$, $\frac{1}{3}$, and 0, respectively.

Amplified primary cue: For both experiments, the amplification of the primary cue reduced the error almost to zero HEG ears. HI07-L and HI09-L have no error for both experiments. At 0 [dB] SNR, the error increased for HI06-L, from 0 in Exp. I to $\frac{1}{3}$ in Exp. II, while at the same SNR, the error decreased for HI05-L from $\frac{1}{3}$ in Exp. I, to 0 in Exp. II.

Histogram for f103 /ka/ (Fig. 4.22): The histograms for f103 /ka/ are shown in columns I, II in Fig. 4.22. Exp. I and Exp. II are shown in column I and column II, respectively. The x-label and y-label are type of response and probability of heard given spoken, $p_{h|s}(\text{SNR})$. 'PC', 'CC', 'SC', 'N', and 'O' on the x-axis label are primary cue, conflicting cues, secondary cues, noise, and other sounds, respectively. 'SC' are responses that are not 'PC', 'CC', 'N', and 'O'. A subject may choose 'N' or 'O' if they did not hear 'PC', 'CC', and 'SC'. The 'PC' for f103 /ka/ is /ka/. The 'CC' for f103 /ka/ are /ta/ and /pa/.

Effect of Conflicting cues for f103/ka/ on HEG Ears

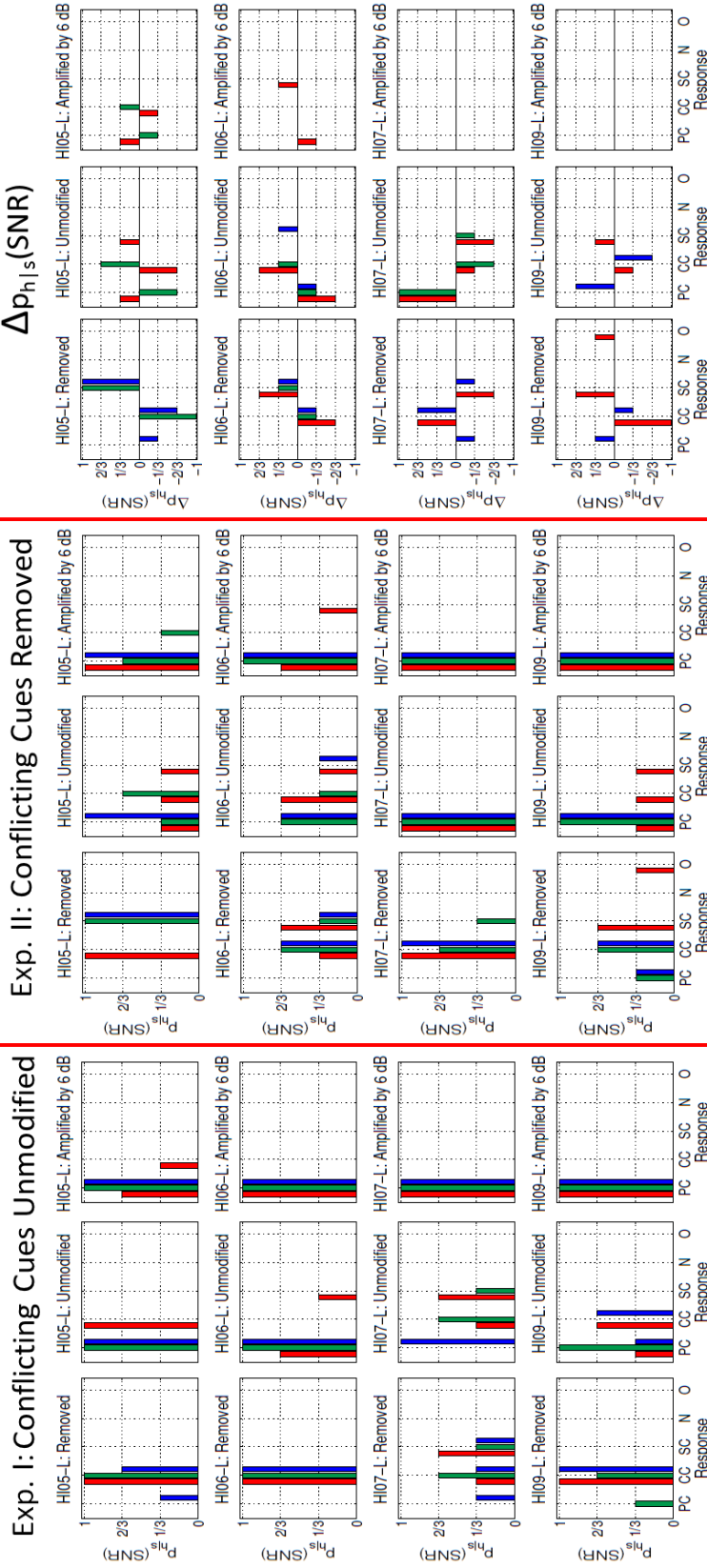
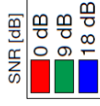


Figure 4.22: The histograms for token f103 /ka/ for Exp. I and Exp. II on HEG ears are examined in Columns I and II. The x-axis labels contain 5 different types of subject responses. Three of the five responses are the type of cues for a sound: primary cues ('PC'), conflicting cues ('CC'), and secondary cues ('SC'). The other two responses are noise ('N') and other sounds ('O'). A subject may choose to report 'N' or 'O' if he/she does not hear a sound that contains one of the 20 sounds available on the MATLAB GUI. The 'PC' is /ka/ and the 'CC' is /ta/ and/or /pa/. The 'SC' are responses that are not 'PC', 'CC', 'N', and 'O'. Note in Columns I and II, at each SNR the probability must sum to one. The probability difference for Exp. I, II is ΔP_{his} . The ΔP_{his} increase is proportional to the ΔP_{his} decrease. For example, for HI05-L, when the primary cue is removed in Exp. I, II at 18 [dB] SNR, the response for the secondary cues increased by 1 in Exp. II, while the responses for the primary and conflicting cues decreased by $\frac{1}{3}$ and $\frac{2}{3}$, respectively. Therefore, with the removal of the conflicting cues for HI05-L, the error for f103 /ka/ increased by $\frac{1}{3}$ and the response for the secondary cues increased to 100%.

Removed primary cue: All ears have errors in Exp. I when the primary cue is removed. At least two-thirds of the response at all SNRs in Exp. I for HI05-L, HI06-L, and HI09-L are the conflicting cues. HI07-L is the only ear that reported a sound with secondary cues at all SNRs in Exp. I. In Exp. II at 18 and 9 [dB] SNR, 100% of the response for HI05-L is the secondary cues, while at 0 [dB] SNR, 100% of the response is the conflicting cues. The response for HI05-L was substantially impacted by the masking noise. In Exp. II at all SNRs, the probability that the conflicting cues was reported is at least $\frac{2}{3}$ for HI07-L.

Unmodified: When the primary cue is unmodified in Exp. I, no ears had reported a primary cue at a rate of 100% at all SNRs. One-hundred percent of the response for HI05-L and HI06-L at 18 and 9 [dB] SNR is the primary cue. When the SNR decreased for HI07-L, the response for the primary cue decreased from 1 at 18 [dB] SNR to 0 at SNR below 18 [dB]. The error is 1 for HI07-L when the SNR is at 9 and 0 [dB] SNR. One-hundred percent of the response for HI07-L when the primary cue is unmodified at all SNRs in Exp. II, is the primary cue. A probability of 1 that the primary cue was reported at 18 and 9 [dB] SNR is 1 for HI09-L. The responses for HI05-L and HI06-L are spread between the primary, conflicting, and secondary cues.

Amplified primary cue: In Exp. I when the primary cue is amplified by 6 [dB] at all SNRs, 100% of the response for HI06-L, HI07-L, and HI09-L is the primary cue. In both experiments when the primary cue is unmodified, the $p_{h|s}$ is the same for HI07-L and HI09-L. In Exp. II, the probability that the primary cue was not reported is low for HI05-L and HI06-L. One-third of the response for HI05-L is the conflicting cues at 9 [dB] SNR, and for HI06-L, is the secondary cues at 0 [dB] SNR.

$\Delta p_{h|s}(\text{SNR})$ for f103 /ka/ (Fig. 4.22): The probability difference for Exp. I and Exp. II as a function of SNR is defined and denoted as $\Delta p_{h|s}(\text{SNR}) = p_{h|s}(\text{Exp. II}, \text{SNR}) - p_{h|s}(\text{Exp. I}, \text{SNR})$ as shown on column III in Fig. 4.22. The $\Delta p_{h|s}(\text{SNR})$ for a response type (i.e. ‘PC’, ‘CC’, ‘SC’, ‘N’, ‘O’) is the same, increased, or decreased when $\Delta p_{h|s}(\text{SNR}) = 0$, $\Delta p_{h|s}(\text{SNR}) > 0$, $\Delta p_{h|s}(\text{SNR}) < 0$, respectively. Another way to describe $\Delta p_{h|s}(\text{SNR})$ is the sum of $\Delta p_{h|s}(\text{SNR})$ that increased plus the sum of $\Delta p_{h|s}(\text{SNR})$ that decreased plus the

sum of $\Delta p_{h|s}(\text{SNR})$ that is equal, in which it is expressed mathematically as

$$\Delta p_{h|s}(\text{SNR}) = \sum_{x=1}^3 (\Delta p_{h|s}(\text{SNR}_x) > 0) + (\Delta p_{h|s}(\text{SNR}_x) < 0) + (\Delta p_{h|s}(\text{SNR}_x) = 0). \quad (4.4)$$

The effect of the conflicting cues improved or degraded if the response type is the primary cue and $\Delta p_{h|s}(\text{SNR}) > 0$ or $\Delta p_{h|s}(\text{SNR}) < 0$, respectively. Vice versa, if the response type is not the primary cue, then $\Delta p_{h|s}(\text{SNR}) > 0$ and $\Delta p_{h|s}(\text{SNR}) < 0$ means the effect of the conflicting cues degraded and improved, respectively.

Removed primary cue: At 18 [dB] SNR, the error for HI05-L was increased from $\frac{2}{3}$ in Exp. I to 1 in Exp. II. The conflicting cues were decreased by $\frac{2}{3}$, while the secondary cues increased to 1. At 9 [dB] SNR for HI05-L, the removal of the conflicting cues increased the probability of response for the secondary cues to 1. The effect of the conflicting cues for HI06-L at 18, 9, and 0 [dB] SNR decreased the conflicting cues response by $\frac{1}{3}$, $\frac{1}{3}$, and $\frac{2}{3}$, respectively, and increased the secondary cues response to $\frac{1}{3}$, $\frac{1}{3}$, $\frac{2}{3}$, respectively. This show the removal of the conflicting cues for HI05-L and HI06-L increased the response for the secondary when the primary cue is removed.

Unmodified: The removal of the conflicting cues when the primary cue is unmodified has a positive impact for HI07-L and HI09-L, while it has a negative impact on HI05-L and HI06-L. When comparing Exp. I, II for the unmodified primary cue at 9 and 0 [dB] SNR, the primary cue response for HI07-L increased to 1 in Exp. II, while the conflicting and secondary cues decreased to $\frac{2}{3}$ and $\frac{1}{3}$ at 9 [dB], respectively, and $\frac{1}{3}$ and $\frac{2}{3}$ at 0 [dB] SNR, respectively. At 9 [dB] SNR for HI05-L, the effect of the removal of the conflicting cues decreased the primary cue response by $\frac{2}{3}$ and increased the secondary response by $\frac{2}{3}$. At 0 [dB] SNR, the conflicting cues response for HI05-L decreased by $\frac{2}{3}$ and increased the primary and secondary cues by $\frac{1}{3}$. The effect of the conflicting cues for HI05-L decreased the conflicting cues response and increased the secondary cues response.

Amplified primary cue: The effect of the conflicting cues for f103 /ka/ when the primary cue is amplified by 6 [dB] shows zero or low effect for the HEG ears. The response for HI07-L and HI09-L are the same when comparing both

experiments. At 9 [dB] SNR for HI05-L, the conflicting cue response decreased by $\frac{1}{3}$ and increased by the secondary cues response by $\frac{1}{3}$. At 0 [dB] SNR, the removal of the conflicting cues decreased the conflicting cues by $\frac{1}{3}$ and increased the primary cue by $\frac{1}{3}$.

Confusion patterns for /ta/ (Fig. 4.23): The confusion pattern results for m115 /ta/ for HI ears from the HEG is shown in Fig. 4.23. There are serious challenges for improving speech perception of high frequency sounds such as /t/ and /d/. All HI ears from HEG have severe to profound HL at high frequencies. Since m115 /ta/ is a high frequency sound, we can expect the HI ears from the HEG are going to make significant errors. The conflicting cues for m115 /ta/ are /ka/ and /pa/.

Removed primary cue: With the removal of the primary cue at 18 [dB] SNR, the error increased from Exp. I to Exp. II for HI05-L, HI07-L, and HI09-L. The error increased from 0 in Exp. I to 1 in Exp. II for HI05-L, from $\frac{1}{3}$ in Exp. I to 1 in Exp. II for HI07-L, and from $\frac{2}{3}$ in Exp. I to 1 in Exp. II for HI09-L. When comparing both experiments for the removal of the primary cue at 0 [dB] SNR, the confusion /pa/ increased from $\frac{1}{3}$ in Exp. I to $\frac{2}{3}$ in Exp. II for HI05-L and HI07-L, and $\frac{2}{3}$ in Exp. I to 1 in Exp. II for HI09-L.

Unmodified primary cue: At 18 [dB] SNR when the primary cue is unmodified for both experiments, the error decreased in Exp. II for HI06-L and HI09-L, while it increased in Exp. II for HI05-L. Comparing both experiments at 9 [dB] SNR, the error decreased by $\frac{1}{3}$ in Exp. II for HI06-L, and increased by 1 in Exp. II for HI07-L. When comparing both experiments at 0 [dB] SNR, the error decreased by $\frac{1}{3}$ in Exp. II for HI06- and increased by 1 in Exp. II for HI05-L. Since the error increased for HI05-L when the conflicting cues were removed at 18 and 0 [dB] SNR, HI05-L is depending on the conflicting cues for correct recognition for m115 /ta/.

Amplified primary cue: When the primary cue is amplified by 6 [dB] for both experiments at 18 [dB] SNR, the error decreased in Exp. II by $\frac{1}{3}$ for HI06-L and increased in Exp. II by 1 and $\frac{1}{3}$ for HI05-L and HI09-L, respectively. In Exp. II at 9 [dB] SNR, the error decreased by $\frac{1}{3}$ for HI06-L and increased by $\frac{1}{3}$, $\frac{2}{3}$, and $\frac{1}{3}$ for HI05-L, HI07-L, and HI09-L, respectively. Comparing both experiments when the primary cue is amplified by 6 [dB], no ears improved at 0 [dB] SNR.

Confusion Patterns for m115 /ta/: High Error Group

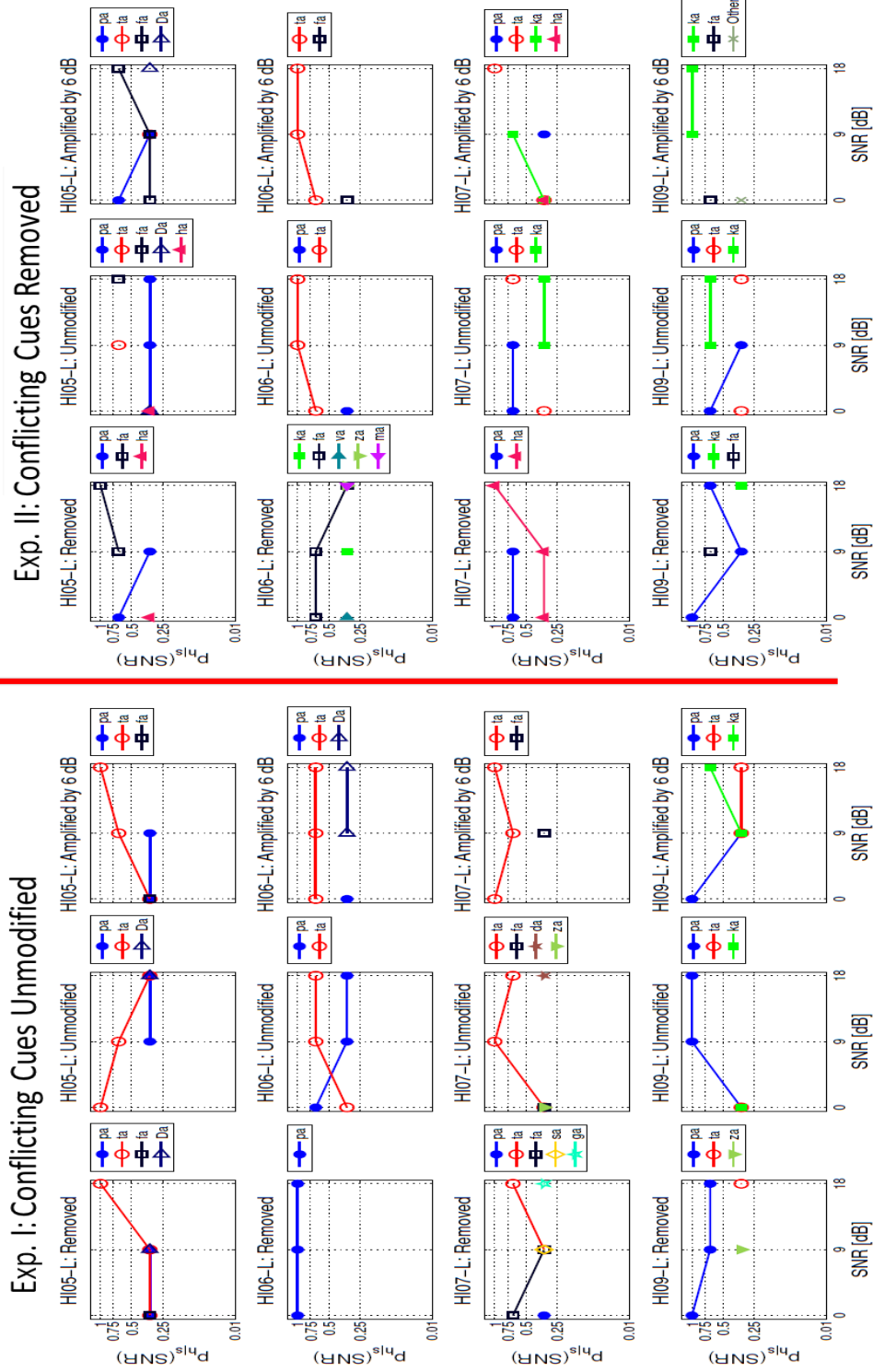


Figure 4.23: For the comparison for the removal of the primary cue in Exp. I, II, the error increased in Exp. II for HI05-L and HI07-L. Out of all the ears in HEG, HI06-L shows improvement when the primary cue is unmodified in Exp. II. Also, HI06 is the only ear that has error decreased in Exp. II when the primary cue is amplified by 6 [dB]. HI05-L and HI07-L depend on the conflicting cues.

The error for HI05-L and HI07-L had decreased by $\frac{1}{3}$ and 1, respectively. Overall, the amplification of 6 [dB] on the primary cue for m115 /ta/ enhanced recognition for HI06-L, while it degraded recognition score for HI05-L, HI07-L, HI09-L. For better recognition score for HI05-L, HI07-L, and HI09-L, it is better not to remove the conflicting cues.

Histogram for m115 /ta/ (Fig. 4.24): The histogram of the response for m115 /ta/ is shown in column I (Exp. I) and column II (Exp. II) on Fig. 4.24.

Removed primary cue: One would expected the error to be 1 when the primary cue for m115 /ta/ is removed for the HEG ears. However, that was not the case for two HI ears, HI05-L and HI07-L. In Exp. I at 18 [dB] SNR, the error for the primary cue for HI05-L and HI07-L is 0 and $\frac{1}{3}$, respectively. But in Exp. II, the error for the PC increased to 1 for HI05-L and HI07-L.

As expected in Exp. I when the primary cue was removed for HI06-L, at all SNRs, the probability of that the conflicting cues was reported is 1, while in Exp. II, the probability of the secondary was reported is 1, $\frac{2}{3}$, and 1 at 18, 9, 0 [dB], respectively.

Unmodified primary cue: The removal of the conflicting cues when the primary cue is unmodified has a positive impact on HI06-L. The probability that the primary cue was reported for HI06-L in Exp. I, is $\frac{2}{3}$, $\frac{2}{3}$, and $\frac{1}{3}$ at 18, 9, and 0 [dB] SNR, while the probability that the conflicting cues were reported is $\frac{1}{3}$, $\frac{1}{3}$, and $\frac{2}{3}$ at 18, 9, and 0 [dB] SNR. The conflicting cues are the main source of error for HI06-L. The probability that the conflicting cues were reported for HI06-L is reduced in Exp. II to 0, 0, and $\frac{1}{3}$ at 18, 9, and 0 [dB] SNR, respectively, and the probability that the primary cue was reported in Exp. II increased to 1, 1, and $\frac{1}{3}$ at 18, 9, and 0 [dB] SNR, respectively.

The removal of the conflicting cues when the primary cue is unmodified has a negative impact on HI05-L. The probability that the primary cue was reported in Exp. I is $\frac{1}{3}$, $\frac{2}{3}$, and 1 at 18, 9, and 0 [dB] SNR, respectively. The probability that the conflicting cues were reported in Exp. I is $\frac{1}{3}$, $\frac{1}{3}$, and 0 at 18, 9, and 0 [dB] SNR, respectively, and for the secondary cues is $\frac{1}{3}$ at 18 [dB] SNR. The removal of the conflicting cues decreased the probability that the primary cue to 0 at 18 and 0 [dB] SNR, while it increased the probability of the secondary cues to $\frac{2}{3}$ at

Effect of Conflicting cues for m115/ta/ on HEG Ears

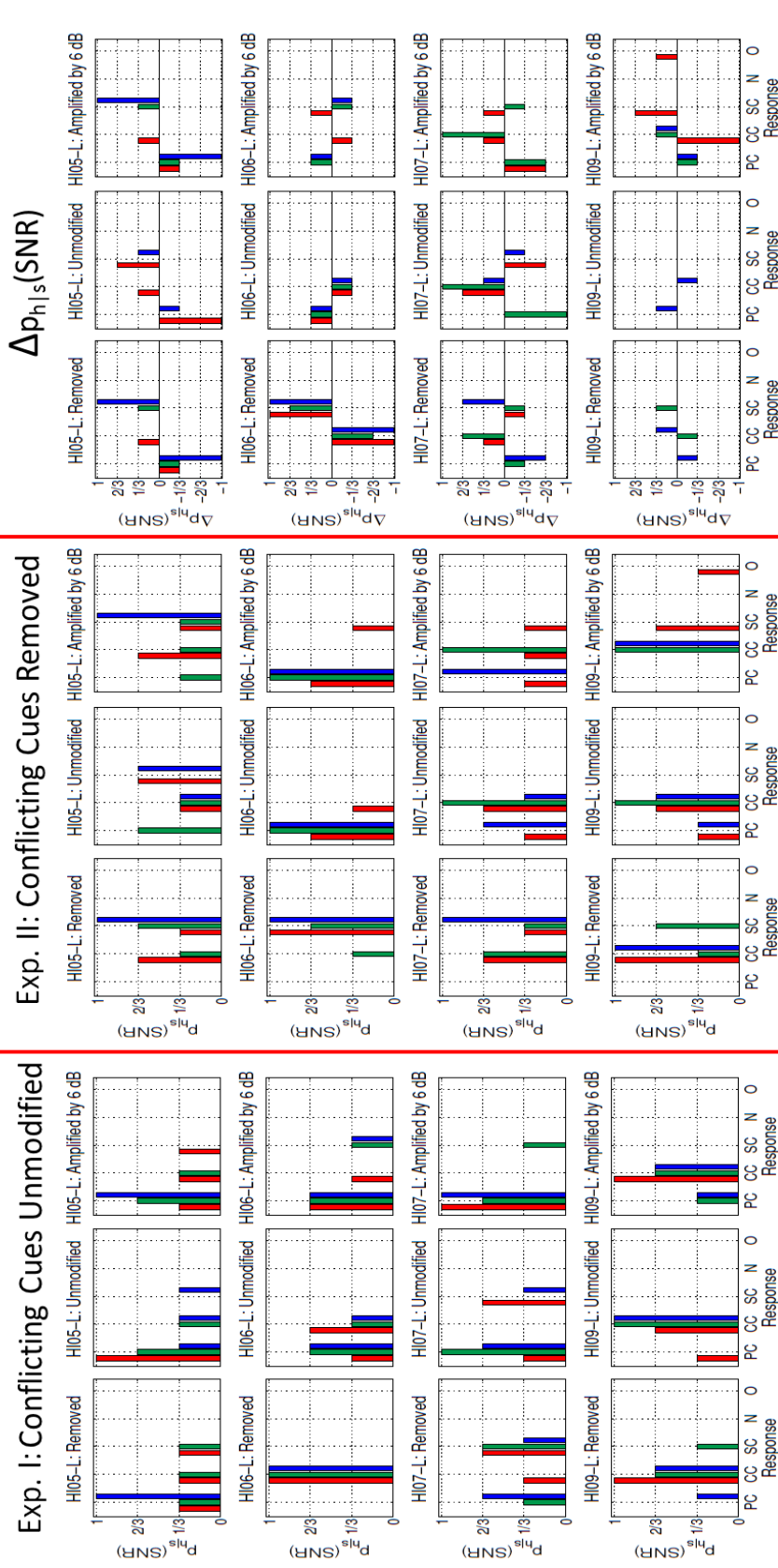
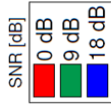


Figure 4.24: The effect of the conflicting cues for m115 /ta on HEG ears shows the importance of conflicting cues. When the primary cue is unmodified at 0 [dB] SNR for Exp. I, II, the primary cue response for HI05-L decreased to 1, while the conflicting and secondary responses increased to one-third and two-thirds, respectively. Under the same modification condition at 9 [dB] SNR, the HI07-L response for the primary cue decreased to 1, while the conflicting cues increased to 1. HI06-L is the only ear that benefits in Exp. II. The primary response for HI06-L increased by one-third at all SNRs when the primary cue is unmodified, and one-third at 9 and 18 [dB] SNR when the primary cue is amplified. The other ears show no improvement in Exp. II for when the primary cue is unmodified or amplified.

18 and 0 [dB] SNR. The importance of the conflicting cues for recognizing m115 /ta/ is vital for HI05-L.

Amplified primary cue: The removal of the conflicting cues had greatly affected HI06-L when the primary cue was unmodified by 6 [dB]. In Exp. I at all SNRs for HI06-L, the probability that the primary cue was reported is $\frac{2}{3}$, the probability that the conflicting cues were reported is $\frac{1}{3}$ at 0 [dB] SNR, and the probability that the secondary cues were reported is $\frac{1}{3}$ at 18 and 9 [dB] SNR. In Exp. II the probability of the primary increased from HI06-L to 1 at 18 and 9 [dB] SNR, while the probability of secondary cues decreased to 0 at 18 and 9 [dB] SNR.

At 9 and 0 [dB] SNR, the removal of the conflicting cues had severely effected HI07-L when the primary cue was unmodified by 6 [dB] SNR. The probability of the primary cue in Exp. I for HI07-L is $\frac{2}{3}$ and 1 at 9 and 0 [dB] SNR, respectively. In Exp. II, the probability of primary cue is reduced to 0 and $\frac{1}{3}$ at 9 and 0 [dB] SNR.

$\Delta p_{h|s}$ for m115 /ta/ (Fig. 4.24)

Removed primary cue: The effect of the removal of the primary cue in Exp. II had highly effected HI05-L. As a result of Exp. II, the difference for the primary cue decreased by 1, $\frac{1}{3}$, and $\frac{1}{3}$ at 18, 9, and 0 [dB] SNR, respectively. The difference for the secondary cues had increased by 1 and $\frac{1}{3}$ at 18 and 9 [dB], respectively, and increased for the conflicting cues by $\frac{1}{3}$ at 0 [dB] SNR.

Due to the effect of the removal of the conflicting cues for HI06-L, the difference for the conflicting cues was reduced to 1, $\frac{2}{3}$, and 1 at 18, 9, and 0 [dB] SNR, respectively. The difference for the secondary cues was increased relatively by the same amount as the difference for the conflicting cues.

Unmodified primary cue: The difference for the primary cue when the primary cue is unmodified in Exp. II for HI05-L, was decreased by $\frac{1}{3}$ and 1 at 18 and 0 [dB] SNR. As a result of the negative impact of the removal of the conflicting cues, the difference for the secondary cues increased by $\frac{1}{3}$, and $\frac{2}{3}$ at 18 and 0 [dB] SNR, respectively, and increased the difference for the conflicting cues by $\frac{1}{3}$ at 0 [dB] SNR. The results of the removal of conflicting cues for m115 /ta/ degraded HI05-L performance.

HI06-L was highly effective when the primary cue was unmodified in Exp. II. The difference for the primary cue was increased by $\frac{1}{3}$ at all SNRs, while the difference for the conflicting cues was increased by $\frac{1}{3}$ at all SNRs. The results of the removal of the conflicting cues for m115 /ta/ improved HI06-L performance.

Amplified primary cue: HI06-L is the only ear that improved due to the effect of the removal of the conflicting cues when the primary cue was amplified by 6 [dB]. The difference for the secondary decreased by $\frac{1}{3}$ at 18 and 9 [dB] SNR, and increased the difference for the primary cue by $\frac{1}{3}$ at 18 and 9 [dB] SNR.

The performance of the removal of the conflicting cues when the primary cue was amplified by 6 [dB] greatly degraded HI07-L. The difference for the primary cue was decreased by $\frac{2}{3}$ at 9 and 0 [dB] SNR and the difference for the secondary was decreased by $\frac{1}{3}$ at 9 [dB] SNR. As a result of the difference for the primary and secondary cues that were reduced, the difference for the conflicting cues increased by 1 and $\frac{1}{3}$ at 9 and 0 [dB] SNR, and the difference for the secondary cues increased by $\frac{1}{3}$ at 0 [dB] SNR.

4.4 Decoding Strategy for m115 /ta/

In Fig. 4.25 we saw complex patterns for HI05-L. In an attempt to explain these strange results, we provide the AI-gram of m115 /ta/. The green highlighted region is the modification of the primary cue. The blue outline is the possible secondary cues that HI05-L seems to be using for recognizing m115 /ta/. Additionally, the time and frequency location of the blue outline share similar perceptual cues as the /ðɑ/ sound. The middle and right panels show the confusion patterns from Exp. I, when the primary cue is removed and unmodified, respectively.

HI05-L has at least a 90 [dB] HL at 4 kHz and above, which is a profound HL, thus it cannot detect high frequency sound. However, due to the influence of the noise masking effects, and the removal of the primary cue at the high frequency end, other delayed mid frequency to high frequency perceptual cues seem to be audible, and therefore could play a role in the choices of consonants selected by HI05-L. Notice the confusion pattern that results when the primary cue is removed at 18 [dB]. The ear was able to identify the target /ta/ with an error of 0. When a 9 [dB] SNR was added to the token, the error increased to two-thirds ($1 - P_{h|s} = \frac{2}{3}$). However, the ear reported /ðɑ/ one-third of the time. As shown in Fig. 2. 9(b), the

HI05-L Decoding Strategy for Detecting m115 ta

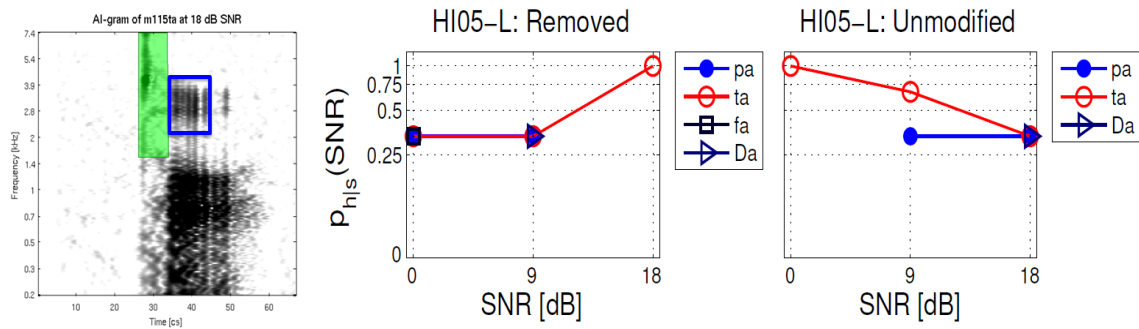


Figure 4.25: HI05-L Decoding Strategy for Recognizing m115 /ta/.

3DDS finding for the /ð̃a/ cue lies at the end of the target’s primary cue. The offset time for /ta/ ends at 20 [cs], while the onset time for /ð̃a/ starts at 20 [cs]. Next, look at the unmodified condition at 18 [dB] SNR, where the ear has a two-thirds error. This is the same confusion /ð̃a/. Finally, as the SNR decreased, the noise behaves like a low pass filter, since it masks the high frequency components. As the noise increased, the error decreased as a function of the SNR, so that, at 0 [dB] SNR, the error is 0. The error behaves inversely like the removal of the primary cue. The /ð̃a/ sound is a mid to high-frequency sound. When the noise masked the high-frequency components near the burst region, the delayed perceptual cue at the mid-frequency which normal ears report as /ð̃a/, seems to be mapped to /ta/.

We hypothesize that for HI05-L, when the primary cue of /ta/ was removed, or masked for the case when the primary cue was not removed, the /ð̃a/ was perceived as /ta/. It seems likely this occurs for two reasons: 1) Since HI05-L has a profound HL at the high frequency end, the primary cue for /t/ is inaudible; therefore, the perceptual cues that are associated with /ð̃a/ (i.e., blue outline in Fig. 4.25) are secondary cues that HI05-L uses for recognizing the /t/ sound. 2) Due to the masking effects of the noise, the primary cue and conflicting cues were masked, and the token behaves more like m115 /ta/ once the primary cue was removed.

When the strength of the primary cue, and/or conflicting cues, are reduced due to the noise masking effects, the strength of the /ð̃a/ secondary cues at the end of the /ta/ primary cue play a role for HI05-L. This case study demonstrates that the HI ears use both the primary cues and the secondary cues.

CHAPTER 5

DISCUSSION

Experiment I varied the plosive's primary cue for eight tokens, using two statistical tools to examine the errors: sorted error and confusion patterns. From the results of varying the masking of the primary cue, we seek to determine the impact of the primary cue in an HI ear. Next, we review the results from Exp. II, in which we both removed the conflicting cues and varied the plosive's primary cue for eight tokens. We perform the same statistical analysis to evaluate the errors. We compare Exp. I and Exp. II to see if there is any improvement when the conflicting cues are removed. By removing the conflicting cues, we hope to determine the role of the conflicting cues in an HI ear. Finally, we review a case study of a particular HI ear with a profound hearing loss (HL) at the high frequencies, and demonstrate a possible decoding strategy for a high frequency sound.

5.1 Discussion of Case Study CC

The author has profound hearing loss, up to 90 [dB] HL at the most important speech frequencies (i.e. 1000 Hz, 1500 Hz, and 2000 Hz). As a demonstrative case study, therefore, Section 4.1 analyzed the confusion and sorted error patterns of the author, hereafter called Subject CC. The objective of this analysis is to study the errors made by Subject CC's ears when applying FG and NALR. Furthermore, we evaluate five different statistical analysis techniques, and determine which techniques provide a better explanation of the errors made by Subject CC's ears. We hope that the results in Section 4.1 could lead us to a better understanding of the errors made by Subject CC's ears and can suggest methods of amplification that would reduce errors for the hearing impaired listeners. In summary, in a comparison between two gain conditions, Subject CC slightly benefited from the NALR gain, particularly at SNR of 6 [dB] or higher. The question remains open: Is there enough evidence to draw strong conclusions about the impact of

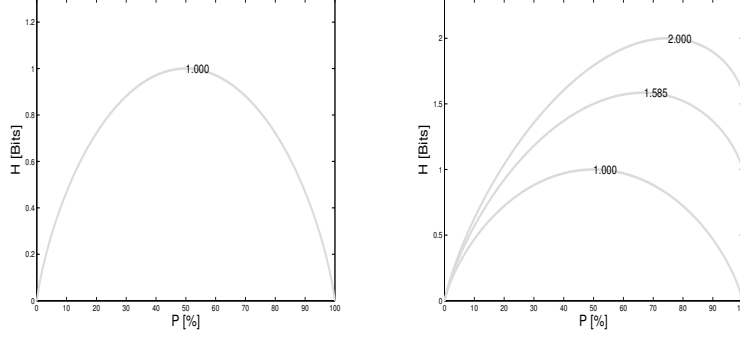


Figure 5.1: Plots of \mathcal{H} as function of probability. On the left is the case of two outcomes, $[P, 1 - P]$, with a maximum 1 bits. The plot to the right has 3 curves. The lower curve is the same plot as on the left. The middle curve is for the case of 3 outcomes, $\left[P, \frac{1-P}{2}, \frac{1-P}{2}\right]$, with a maximum bits of 1.585. Finally, the curve above both curves is the case of 4 outcomes, $\left[P, \frac{1-P}{3}, \frac{1-P}{3}, \frac{1-P}{3}\right]$, with a maximum of 2 bits.

gain prescriptions, using the average probability of error?

Average probability of error is an insufficient metric of the performance. The average probability of error can be a misrepresentation of the results due to the error of each token that can increase or decrease the overall average probability of error (Trevino and Allen (2013)). What is needed is the individual consonant or token (utterance) error. As we shall see, Subject CC performs poorly in recognizing some CVs but often does well on different CVs, at the same SNR. Therefore, when analyzing the response, it is necessary to use more powerful techniques (e.g. less averaging) that take the token confusion into account.

A more effective measure is to take an average per token at each SNR, which is called the token error (i.e., corresponding to the target token on the row normalized CM).

Another more effective measure is entropy (\mathcal{H}). The entropy is a measure of consistency, uncertainty, or randomness defined as

$$\mathcal{H}_k = \mathcal{E} \log_2 \mathcal{I}(P) = \sum_{k=1}^{24} p(x_k) \log_2 \left(\frac{1}{p(x_k)} \right), \quad (5.1)$$

where the “information density” is defined as $\mathcal{I}_k = \frac{1}{P_k}$, and the information in bits as $\log_2 \mathcal{I}(P)$, \mathcal{E} is the “expected value”, and $p(x_k)$ is the probability of x_k . The unit of \mathcal{H} is bits. The entropy of a token tells us something about the number of conflicting cues and/or about the ambiguity of the primary cue. Examples of

plots of \mathcal{H} as a function of probability are shown in Fig. 5.1. The ordinate and abscissa for each panel represent the entropy and probability of a correct outcome, respectively. On the left is a plot of \mathcal{H} with two outcomes having probabilities $[P, 1-P]$. An unbiased coin is an example of an experiment with two equally probable outcomes; heads and tails are each 50% likely. The maximum entropy occurs when the probability is equal to 50%. The higher the entropy, the higher the inconsistency, uncertainty, or randomness. The lower the entropy, the lower the inconsistency, uncertainty or randomness. The figure to the right shows 3 curves, where each curve represents a different number of outcomes: starting from the bottom, the curves show the entropies of experiments with two, three, and four outcomes, respectively.

Entropy of the case study responses was plotted in Fig. 4.3, in order to understand the pattern of responses. The probability of error is low at 6 [dB] SNR; therefore, Subject CC is consistent. However, that is not the case at 0 [dB] SNR, when there is a 90% error rate for f103 /ka/ and the error is distributed with 3 other tokens. As a result, the entropy is high: at 0 [dB], Subject CC has difficulty recognizing /k/.

Further analysis of the case study was performed in Section 4.1.5 using confusion matrices as the analysis tool. This is informative as to an example why we should not average across tokens. One of the conclusions is that the strength of the primary cue for f103 /ka/, compared to m111 /ka/, is weaker and much more sensitive to the masking noise. In other words, the noise masks the primary cue for f103 /ka/, making it ambiguous, once noise was added. Therefore, the location of a weak primary cue of a token near 1 [kHz], where Subject CC has the largest HL, leads to errors (confusion) with other conflicting cues that are in the same group as the primary cue. An alternative to the current amplification strategy (i.e. FG vs. NALR) is to strengthen the weak primary cue via amplification. This is the key point of the case study.

5.2 Discussion of Experiment I

In Exp. I, the weak primary cue was strengthened via amplification. Perceptual responses of NH and HI ears were analyzed at different SNRs. HI ears were divided into two groups, based on the pattern of their responses: a low-error group (LEG), and a high-error group (HEG).

5.2.1 Sorted Error

Discussion of the unmodified, removed and amplified primary cue for the LEG and HEG ears: To find the answer to the research questions of this study, we only need to analyze when the primary cue is unmodified, removed, and amplified by 6 [dB]. When removing the primary cue, there are no ZE tokens across all SNRs for the average NH ears, and 3 out of 6 HI ears in LEG (HI01-R, HI02-R, and HI04-R) have no ZE tokens across all SNRs. All 3 other HI ears in LEG have one ZE token. HI10-R has one ZE token at 0 [dB] and 18 [dB] SNR while the other two HI ears have one ZE token for one SNR condition. For the most part, the results are similar when removing the primary cue for both average NH ears and HI ears in LEG. When the primary cue is removed, of the HI ears in HEG, three-fourths have either one or two ZE tokens. Therefore the removal of the primary cues affects the HI ears in HEG similarly to NH ears, and to the HI ears in LEG. The results for amplification of the primary cue by 6 [dB] confirmed that the HI ears in HEG significantly improve their ability to recognize the primary cue. For instance, under the most extreme noisy condition in this experiment, HI07-L was able to reduce the number of tokens with ME from 5 tokens when the primary cue is unmodified to one token for the 6 [dB] amplification condition. Another example: at 18 [dB] SNR, the number of ZE tokens increased by two tokens for HI05-L. However, comparing the average NH ears with LEG and the average NH ears with HEG shows that the NH ears are very similar with LEG, while they are notably different from HEG. Nonetheless, there is sufficient evidence in Exp. I that HI ears in both groups are relying on the cues that the NH ears are using for speech perception. Among the relevant evidence: no ears in Exp. I could recognize at least three-fourths of the tokens when the primary cue was removed. But after presenting the token's primary cue of its original form, the error rate dramatically decreased for all SNRs. The HI ears in both groups had lower error rates when the primary cue was strengthened. When the primary cue is strengthened, the energy level is increased for all signal components whose time and frequency are in the primary cue region; the strengthened primary cue is able to mask the surrounding secondary cues that belong to other competing sounds.

5.2.2 Confusion Patterns

Discussion of the unmodified, removed and amplified primary cue for m115 /ta/ for the LEG and HEG ears: The final analysis for studying the impact of varying the masking of the primary cue is the confusion patterns. Earlier we stated that the confusion patterns illustrate a graphical interpretation of the response distribution. Though the groups NH, LEG, and HEG show different error frequencies, their confusion patterns in Exp. I are similar. For example, Fig. 4.12 showed confusion patterns when the token was m115 /ta/. Even after the removal of the primary cue for HEG, three-fourths of HI ears (HI05-L, HI07-L, HI09-L) reported that they heard a consonant /t/ at 18 [dB] SNR. Like the average NH ears and HI ears from LEG, every HI ear from HEG reported a /p/ for at least one attempt at 0 [dB] SNR. Unlike the average NH ears and HI ears from LEG, when the primary cues were unmodified, or amplified by 6 [dB], no HI ears from HEG were able to consistently recognize /t/ for all attempts across all SNRs. Three-fourths of HI ears have some trials that were confused with /p/ for at least two SNRs when m115 /ta/ was not modified. The same three-fourths of HI ears were still confused with /p/ even when a 6 [dB] amplification was applied to the primary cue.

The results of varying the masking of the primary cue provide an insight into the importance of the primary cue for HI ears. Without the primary cue, results show significant error increase for all ears.

There are a small number of cases where an HI ear could recognize some tokens (i.e. m115 /ta/) when the primary cue was removed. The last section of Chapter 4 attempts to explain for this particular case. With this exception noted, the results are consistent in showing that the primary cue is the main feature used by HI ears for correct consonant recognition.

5.3 Discussion of Experiment II

Without the primary cue, results of Exp. I show significant error rate increase for all ears. In this section, we review the results from the second experiment of our study, the removal of the conflicting cues. In the first experiment (conflicting cues unmodified), varying the masking of the primary cue, we demonstrate that the errors made by the listeners of our study (average NH ears, HI ears in the LEG, HI ears in the HEG) are generally confused with the consonants whose cues lie

above and/or below the primary cue. The primary cue for some tokens is too weak for an HI ear to recognize, especially if it falls in the frequency range where the level of HL [dB] is severe. Therefore, the error rate decreased for many listeners when we disambiguated the tokens by boosting the primary cue.

5.3.1 Sorted Error

Weakness of the primary cue is a significant source of error, especially in the presence of significant noise. Therefore, the error rate decreased for many listeners when we disambiguated the tokens by boosting the primary cue. This demonstrates that the primary cue is the most important speech perceptual cue in human speech recognition.

Discussion of the unmodified conflicting cues vs. removed conflicting cues for the LEG and HEG ears: As expected, the removal of the conflicting cues benefits most of the HI ears in the LEG, especially when the primary cue is unmodified. By contrast, several ears in the HEG suffered error increases after the removal of the conflicting cues. For example, in the first experiment, unmodified conflicting cues, three HI ears have at least one token with ZE for at least one SNR. In the second experiment, by contrast, after the removal of the conflicting cues, the number of HI ears reduced to one HI ear (HI09-L) that has a token with ZE. When the primary cue was removed in Exp. I, for all SNRs, all HI ears in HEG had at least four tokens with ME, but five tokens when the conflicting cues were removed in Exp. II. When the primary cue was removed, the removal of the conflicting cues made the tokens more ambiguous across all SNRs for half of the HI ears in HEG, while the other half of the HI ears in HEG (i.e. HI07-L and HI09-L) improved when the conflicting cues were removed and the primary cue was unmodified.

5.3.2 Confusion Patterns

This section examines the effects of the removal of the conflicting cues on the average NH ears, HI ears from the LEG, and HI ears from the HEG. We want to study the patterns of errors made by an ear, and determine if the errors are due to the influence of the conflicting cues. Also, we want to find an answer to the key unanswered question: Can the removal of the conflicting cues improve speech

intelligibility for HI ears as it did for NH ears in the Li and Allen study (Li and Allen 2011)?

Discussion of f103 /ka/ for the LEG ears: Confusion patterns of the LEG were studied in response to token f103 /ka/ in Fig. 4.19. When comparing both experiments for the removal of the primary cue, for every HI ear from LEG, except HI10-R, the number of confusions is reduced or unchanged after the conflicting cues are removed. The confusion /pa/ was the dominant error of choice, when the conflicting cues were removed. No HI ears reported a /ka/ when the conflicting cues were removed. When the primary cue was restored to its original form (i.e., unmodified token, conflicting cues not removed), only two HI ears (HI02-R, HI10-R) reported a confusion /pa/ at 0 [dB] SNR. After removal of the conflicting cues, the same two HI ears that reported confusion /pa/ have no remaining confusion at any SNR. The errors made by HI02-R and HI10-R were due to the impact of the conflicting cues. The removal of the conflicting cues demonstrates that HI02-R and HI10-R were able to improve consonant identification for f103 /ka/.

Discussion of m115 /ta/ for the LEG ears: The pattern is different for LEG. Fig. 4.20 shows the confusion pattern results for m115 /ta/ for the HI ears from LEG. When both the primary cue and conflicting cues are removed, the number of confusions significantly increases. Two HI ears, HI08-R and HI10-R, have an increased number of confusions, from 1 when the conflicting cues are not removed, to 3 when the conflicting cues are removed. The removal of both primary cue and conflicting cues caused m115 /ta/ to become more ambiguous, especially in the presence of noise. Four HI ears (HI02-R, HI03-R, HI08-R, HI10-R) and the average NH ears have the same confusions, /pa/ and /ha/.

Discussion of f103 /ka/ for the HEG ears: Confusion patterns of the HEG were studied in response to token f103 /ka/ in Fig. 4.21. The number of confusions for three-fourths of the HI ears (HI05-L, HI06-L, HI09-L) increased or remained the same when both the primary cue and conflicting cues were removed. Regardless of whether the conflicting cues were removed or not, the HI ears from HEG depend on the primary cue, as shown when the primary cue is unmodified and amplified by 6 [dB]. The removal of the conflicting cues when the primary cue is unmodified benefits two HI ears (HI07-L, HI09-L). For the other two HI

ears (HI05-L, HI06-L), the error rate increased when the conflicting cues were removed. These HI ears depend on other secondary cues, including, apparently, the removed conflicting cues, to make their correct choice.

Discussion of m115 /ta/ for the HEG ears: Since m115 /ta/ is a high frequency sound, we can expect the HI ears from the HEG to make significant errors. We expected the error rate to be high when the primary cue was removed and the conflicting cues were not removed. However, that was not the case for two HI ears, HI05-L and HI07-L. They have an error rate below 25% in quiet condition (18 [dB]). But the error rate increased to 100% when the conflicting cues were removed. We will revisit HI05-L in the next section and propose a strategy by which HI05-L may be identifying m115 /ta/.

5.4 Discussion of the Case Study

Earlier we analyzed the confusion patterns for HI05-L. In this section, we revisit the results from HI05-L for when the primary cue of the m115 /ta/ is removed and the conflicting cues are unmodified (Exp. I). In this situation we expect a high error rate, but HI05-L exhibits zero error at 18 [dB] SNR, as shown in Fig. 4.25. Similarly, even if the primary cue is not removed, noise may mask it. As the SNR decreased, the noise behaved like a low pass filter, masking the high frequency components, similar to the removal of the primary cue. As the noise increased, the error rate decreased as a function of the SNR. At 0 [dB] SNR, the error rate that was reported for recognizing the target token was 0, exactly matching the zero error with which this subject recognized this token after explicit removal of the primary cue. It is therefore clear that this subject recognizes this token correctly only when the primary cue is missing; the rest of this paragraph proposes a possible explanation for this unusual finding. The /ðɑ/ sound is a mid- to high-frequency sound, but since HI05-L cannot hear the high-frequency cues for /ta/, he is accustomed to identifying both /ðɑ/ and /ta/ on the basis of these mid-frequency cues; therefore, when the high frequency cues are removed from m115 /ta/, the subject hears /ta/ even more clearly because the mid-frequency cues that he uses to identify /ta/ are still present. In summary, for HI05-L recognizing m115 /ta/, when the primary cue of /ta/ was removed, or masked for the case when the primary cue was not removed, the /ta/ was perceived. This event occurred for

two reasons: 1) Since HI05-L has a profound HL at the high frequency end, the primary cue for /t/ is inaudible; therefore, the perceptual cues that are associated with /ðɑ/ (i.e., blue outline in Fig. 4.25) are secondary cues that HI05-L uses for recognizing the /t/ sound. 2) Due to the masking effects of the noise, the primary cue and conflicting cues were masked, and the token behaved more like m115 /ta/ once the primary cue was removed. When the strength of the primary cue, and/or conflicting cues, is reduced due to the noise masking effects, the strength of the primary offset plays a significant role for HI05-L, thus permitting this ear to avoid reporting sounds that are associated with the conflicting cues. This case study is a perfect example demonstrating that the HI ears use not only the primary cue for correct recognition, but also the secondary cues.

One of the main questions we hope to answer with this research is: What is the strategy of each HI ear in detecting consonants? In this study, we provided an example of an HI ear that uses the secondary cues for correct recognition. The subject HI05-L has profound HL at the high frequency end, and was unable to identify m115 /ta/ unless the primary cue was removed or masked. Since /ta/ is a high frequency sound, HI05-L was unable to detect the primary cue due to high frequency HL. However, when the primary cue was removed, it is possible that the strength of the secondary cues at the offset of the primary cue was detected and used for correct identification of m115 /ta/. For the case when the primary cue was not removed, HI05-L has a maximum allowable error rate at 18 [dB], which is a very high SNR. The maximum error rate for recognizing m115 /ta/ was reduced to ZE at 0 [dB], which is a low SNR. The noise masked the primary cue at the high frequency end, and the conflicting cues below the primary cue were probably too weak for HI05-L to hear. As a result of the masking effects of the noise, the m115 /ta/ was treated as if the primary cue was removed. Therefore, most likely, HI05-L's decoding strategy for recognizing m115 /ta/ probably uses the secondary cues at the offset of the primary cue. Thus, we conclude that HI05-L may or may not have used the secondary cues at the offset of the primary cue for recognizing m115 /ta/, but the ear is definitely using the secondary cues.

Varying the masking of the primary cue and the removal of the conflicting cues, we conclude that the primary cues, secondary cues and conflicting cues play a significant role in speech perception. For most sounds, the primary cue of a target sound is an important feature for correct recognition, while the conflicting cues are the dominant nontarget competing sounds that lead to errors and confusions for an HI ear. As for the secondary cues, for some tokens, they are vital

unique cues on which the HI ears depend for correct recognition.

CHAPTER 6

CONCLUSIONS

In this research, we have two objectives. The first is to examine the effect of the plosive's primary cue in an HI ear. The second is to examine the role of the conflicting cues in an HI ear. To achieve these objectives, we implemented two experiments. To achieve objective 1, we conducted an experiment to study the effects of varying the masking of the primary cue for eight plosive CVs. To achieve objective 2, we conducted another experiment to study the influences of the removal of the conflicting cues for eight plosive CVs. For both experiments, the primary cue was modified four ways: removed by $-\infty$, attenuated by -6 [dB], amplified by 6 [dB], and amplified by 12 [dB]. Additionally for the removal of the conflicting cues, we removed the conflicting cues by $-\infty$. Noise was added to the stimuli, adjusting the signal-to-noise ratio (SNR) to 0, 9, and 18 [dB]. For both experiments, 5 NH and 10 HI subjects took part in this study. We conducted pure tone audiometry (PTA) to obtain an audiogram of each of the 10 HI subjects. Each subject was placed in a sound booth. Before each subject took the speech tests, we set the most comfortable level (MCL) to meet audibility requirements. We generated a count matrix after each subject completed each experiment. We selected two statistical analyses for both experiments: the sorted error and confusion patterns.

In this research, we study the errors of 8 plosive consonants for three groups of ears: NH ears as the control group, HI ears in the low error group (LEG), and HI ears in the high error group (HEG). The two groups of HI ears were divided because their errors were notably different. The HI ears in the LEG tend to make fewer errors than the HI ears in the HEG. Additionally, the audiometric profiles from each HI group show a significantly different hearing loss (HL), mainly in the high frequency region at or above 4 kHz. The HI ears from LEG usually have more zero error (ZE) tokens than HI ears from HEG. Due to the severe high frequency HL for the HI ears in HEG, they tend to struggle identifying high frequency sounds like m115 /ta/.

What do we now know about the impact of varying the masking of the plosive's primary cue that can possibly explain the errors of an HI ear? The HI ears are listening to the same cues as the NH ears. To determine which cues are being used by each HI ear group, we remove the primary cue and study the errors. The results show that most tokens have maximum error (ME) when each group lacks access to the primary cue. The error significantly decreases when the primary cue is in its original form. In many cases, most tokens have ZE or low error (LE). Furthermore, the error rate for all groups decreases even more when the primary cue is boosted by 6 [dB]. The strength of the primary cue is an important variable in speech perception, especially in the presence of significant noise. Results show that some ears made errors due to the masking noise and the conflicting cues. The strength of a primary cue is weakened when noise is added. Therefore, as a result, target tokens end up ambiguous, which causes several HI ears to make errors, especially those with severe HI loss. Nonetheless, the weak primary cue can be strengthened and disambiguated by amplification. It is important to understand that an HI ear with or without severe HL or high frequency loss can perform consonant classification as well as an NH ear. However, this competency depends on the level of the masking noise on the primary cue and the degree of ambiguity of the primary cue. Varying the masking of the primary cue demonstrates the important use of the primary cue by the HI ears. It also shows, in speech perception, that the HI ears depend on the same cues as the NH ears. We conclude, therefore, that the primary cue is critical and vital information for correct identification of plosive consonants, for both NH and HI ears.

What do we now know about the role of conflicting cues in speech perception for an HI ear? It is possible to conclude that conflicting cues cause errors for both NH and HI, especially, in both cases, when noise masks or signal modification removes the primary cue. It was shown, when the strength of the noise increased, that the primary cue was masked, and the conflicting cue explained the pattern of errors for both NH ears and HI ears. However, in some cases, error rate was not reduced when the conflicting cues were removed. The conflicting cues can have negative impact on the HEG ears. Comparing the two experiments shows that the errors increased for a few HI ears when the conflicting cues were removed. The experimental removal of the conflicting cues demonstrates that some HI ears are using not only the primary cue, but also conflicting cues and other secondary cues for correct recognition.

BIBLIOGRAPHY

- A. Abavisani and J.B. Allen. How NALR tuned hearing aids enhance phone recognition in hearing impaired ears. In preparation for submission to the *Journal of the Acoustical Society of America*.
- J.B. Allen. Short time spectral analysis, synthesis, and modification by discrete fourier transform. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-25(4): 235–238, 1977.
- J.B. Allen. Harvey Fletcher’s role in the creation of communication acoustics. *J. Acoust. Soc. Amer.*, 99(4):1825–1839, 1996.
- J.B. Allen. Consonant recognition and the articulation index. *Acoust. Soc. Am.*, 117:2212–2223, 2005.
- J.B. Allen and F. Li. Speech perception and cochlear signal processing. *IEEE Signal Process. Mag.*, 26:73–77, 2009.
- J.B. Allen, J.L. Hall, and P.S. Jeng. Loudness growth in 1/2-octave bands (lgob)- a procedure for the assessment of loudness. *The Journal of the Acoustical Society of America*, 88(2):745–753, 1990.
- G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, and S. M. Williams. *Neuroscience*. Sinauer Associates, Sunderland, MA, second edition, 2001.
- R.C. Bilger and M.D. Wang. Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech and Hearing Research*, 19(4):718, 1976.
- S.E. Blumstein and K.N. Stevens. Perceptual invariance and onset spectra for stop consonants in different vowel environments. *J. Acoust. Soc. Am.*, 67:648–662, 1980.
- S.E. Blumstein, K.N. Stevens, and G.N. Nigro. Property detectors for bursts and transitions in speech perceptions. *J. Acoust. Soc. Am.*, 61:1301–1313, 1977.
- A. Boothroyd. Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech and Hearing Research*, 27(1):134, 1984.

- T.Z. Burkle, D. Kewley-Port, L. Humes, and J.H. Lee. Contribution of consonant versus vowel information to sentence intelligibility by normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 115(5):2601, 2004.
- D. Byrne and H. Dillon. The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, 7(4):257–265, 1986.
- D. Byrne, H. Dillon, T. Ching, R. Katsch, and G. Keidser. The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *The Journal of the American Academy of Audiology*, 12(1):37–51, 2001.
- R.A. Cole and B. Scott. Property detectors for bursts and transitions in speech perceptions. *J. Acoust. Soc. Am.*, 61:1301–1313, 1974.
- R.A. Cole, Y. Yang, B. Mak, M. Fanty, and T. Bailey. The contribution of consonants versus vowels to word recognition in fluent speech. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings*, 2:853–856, 1996.
- F. Cooper, P. Delattre, A. Liberman, J. Borst, and L. Gerstman. Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Am.*, 24:579–606, 1952.
- R.M. Cox. Using ULCL measures to find frequency/gain and SSPL90. *Hearing Instruments*, 34(7):17–21, 1983.
- R.M. Cox. A structured approach to hearing aid selection. *Ear and Hearing*, 6(5):226–239, 1985.
- R.M. Cox. The MSU hearing instrument prescription procedure. *Hearing Instruments*, 39(1):6–10, 1988.
- R. Dewey. Structures of the ear, 2007. URL <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>.
- H. Dillon. *Hearing Aids*. Thieme, 2nd edition, 2012.
- M.F. Dorman, M. Studdert-Kennedy, and L.J. Raphael. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues. *Percept. Psychophys.*, 22:109–122, 1977.
- W.A. Dreschler. Phonemic confusions in quiet and noise for the hearing-impaired. *International Journal of Audiology*, 25(1):19–28, 1986.

- J.R. Dubno and D.D. Dirks. Evaluation of hearing-impaired listeners using a nonsense-syllable test I. test reliability. *Journal of Speech and Hearing Research*, 25(1):135, 1982.
- J.R. Dubno, D.D. Dirks, and D.E. Morgan. Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America*, 76:87, 1984.
- D.A. Fabry and D.J. Van Tasell. Masked and filtered simulation of hearing loss: Effects on consonant recognition. *Journal of Speech and Hearing Research*, 29(2):170, 1986.
- H. Fletcher and R. Galt. Perception of speech and its relation to telephony. *The Journal of the Acoustical Society of America*, 22:89–151, 1950.
- P. Fousek, P. Svojanovsky, F. Grezel, and H. Hermansky. New nonsense syllables database-analyses and preliminary ASR experiments. In *Proceedings of International Conference on Spoken Language Processing*, pages 2749–2752, 2004.
- S. Gordan-Salant. Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *J. Acoust. Soc. Am*, 80:1599–1607, 1986.
- S. Gordan-Salant. Consonant recognition and confusion patterns among elderly hearing-impaired subjects. *Ear and Hearing*, 8(5):270, 1987.
- W. Han. *Methods for Robust Characterization of Consonant Perception in Hearing-Impaired Listeners*. PhD dissertation, Univ. of Illinois, Urbana-Champaign, 2011.
- V. Harzan and A. Simpson. The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Speech Commun.*, 24:211–226, 1998.
- J.M. Heinz and K.N. Stevens. On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America*, 33(5):589–596, 1961.
- J.D. Hood and J.P. Poole. Improving the reliability of speech audiometry. *British Journal of Audiology*, 11(4):93–102, 1977.
- G.W. Hughes and M. Halle. Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America*, 28(2):303–310, 1956.
- A. Kapoor and J.B. Allen. Perceptual effects of plosive feature modification. *The Journal of the Acoustical Society of America*, 131(1):478–491, 2012.
- G. Keidser, H. Dillon, M. Flax, T. Ching, and S. Brewer. The NAL-NL2 prescription procedure. *Audiology Research*, 1(24):88–99, 2011.

- D. Kewley-Port, D. Pisoni, and M. Studdert-Kennedy. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *J. Acoust. Soc. Am.*, 73:1778-1793, 1983.
- D. Kewley-Port, T.Z. Burkle, and J.H. Lee. Contribution of consonant versus vowel information to sentence intelligibility by normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 122:2365, 2007.
- M.C. Killion and G.I. Gudmundsen. Fitting hearing aids using clinical prefitting speech measures: An evidence-based review. *The Journal of the American Academy of Audiology*, 16(7):439-447, 2005.
- D.L. Lawrence and V.W. Byers. Identification of voiceless fricatives by high frequency hearing impaired listeners. *Journal of Speech, Language and Hearing Research*, 12(2):426, 1969.
- F. Li. *Perceptual Cues of Consonant Sounds and Impact of Sensorineural Hearing Loss on Speech Perception*. PhD dissertation, Univ. of Illinois, Urbana-Champaign, 2010.
- F. Li and J.B. Allen. Manipulation of consonants in natural speech. *IEEE Trans. Audio, Speech, Lang. Process.*, 19(3):496-504, 2011.
- F. Li, A. Menon, and J.B. Allen. A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *The Journal of the Acoustical Society of America*, 127:2599-2610, 2010.
- F. Li, A. Trevino, A. Menon, and J.B. Allen. A psychoacoustic method for studying the necessary and sufficient perceptual cues of fricative consonants in noise. *The Journal of the Acoustical Society of America*, 132:2663, 2012.
- A. Liberman. Some results of research on speech perception. *J. Acoust. Soc. Am.*, 29:117, 1957.
- B.E. Lobdell, J.B. Allen, and M.A. Hasegawa-Johnson. Intelligibility predictors and neural representation of speech. *Speech Communication*, 53(2):185-194, 2011.
- G.A. McCandless and S. Lybarger. Prescription of gain/output (POGO) for hearing aids. *Hearing Instruments*, 34(1):16-21, 1983.
- G.A. Miller. *Language and Communication*. McGraw-Hill, 1951.
- G.A. Miller and P.E. Nicely. An analysis of perceptual confusions among some english consonants. *J. Acoust. Soc. Am.*, 27:338-352, 1955.
- M.A. Mines, B.F. Hanson, and J.E. Shoup. Frequency of occurrence of phonemes in conversational English. *Language and Speech*, 21(3):221-241, 1978.

- R.N. Ohde and K.N. Stevens. Effect of burst amplitude on the perception of stop consonant place of articulation. *J. Acoust. Soc. Am.*, 74:706–714, 1983.
- R.N. Ohde, K.L. Haley, H.K. Vorperian, and C.W. McMahon. A developmental study of the perception of onset spectra for stop consonants in different vowel environments. *J. Acoust. Soc. Am.*, 97:3800–3812, 1995.
- E. Owens. Consonant errors and remediation in sensorineural hearing loss. *Journal of Speech and Hearing Disorders*, 43(3):331, 1978.
- S. Phatak and J.B. Allen. Consonant and vowel confusions in speech-weighted noise. *J. Acoust. Soc. Am.*, 121:12121226, 2007.
- S. Phatak, A. Lovitt, and J.B. Allen. Consonant confusions in white noise. *The Journal of the Acoustical Society of America*, 124(2):1220–1233, 2008.
- R.K. Potter, G.A. Kopp, and H.G. Kopp. Consonant confusions in white noise. *Visible Speech* (Dover, New York), 124:22–27, 1966.
- M. Régnier and J.B. Allen. A method to identify noise-robust perceptual features: application for consonant /t/. *J. Acoust. Soc. Am.*, 123(5):2801–2814, 2008.
- C. Scheidiger, J.B. Allen, and T. Dau. Assessing the efficacy of hearing-aid amplification using a phoneme test. *The Journal of the Acoustical Society of America*, 141:1739–1748, 2017.
- C.E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 38: 611–656, 1948.
- R. Singh and J.B. Allen. The influence of stop consonants perceptual features on the articulation index model. *The Journal of the Acoustical Society of America*, 131(4):3051–3068, 2012.
- K.N. Stevens and S.E. Blumstein. Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.*, 64:13581368, 1978.
- A. Trevino and J.B. Allen. Within-consonant perceptual differences in the hearing impaired ear. *The Journal of the Acoustical Society of America*, 134:607–617, 2013.
- D.F. Vysochanskij and Y.I. Petunin. Justification of the 3 rule for unimodal distributions. *Theory of Probability and Mathematical Statistics*, 21:25–36, 1980.
- M.D. Wang, C.M. Reed, and R.C. Bilger. A comparison of the effects of filtering and sensorineural hearing loss on patterns of consonant confusions. *Journal of Speech and Hearing Research*, 21(1):5, 1978.

R.H. Wilson, R.A. McArdle, and S.L. Smith. An evaluation of the BKB-SIN, HINT, QuickSin, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, 50:844–856, 2007.

P.M. Zurek and L.A. Delhorne. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *The Journal of the Acoustical Society of America*, 82(5):1548–1559, 1987.

APPENDIX A

EXPERIMENT I RESULTS

A.1 Modifications

Figure A.1 shows the AI-gram for eight tokens in Experiment I. The modification parameters for the primary cue of each token are highlighted in green.

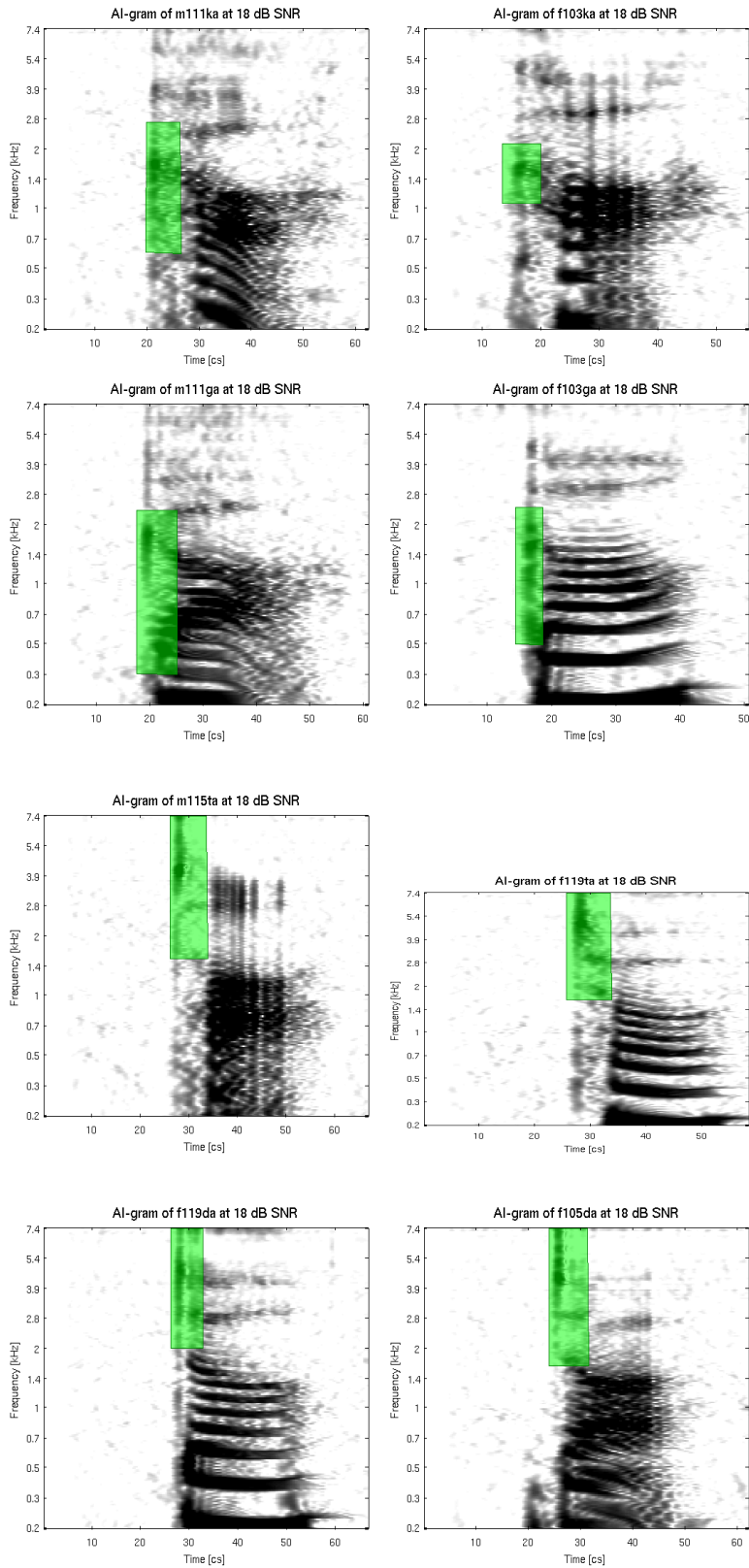


Figure A.1: Modification on the primary cue for 8 tokens. The highlighted green is the modification to the primary cue.

A.2 Sorted Errors

Figures A.2 - A.6 show the sorted error for eight tokens when the primary cue was unmodified, removed, attenuated -6 dB, amplified by 6 dB and 12 dB for the average NH ears and 17 individual HI ears. Each plot is for a specified modification condition (i.e. primary cue was unmodified, removed, attenuated -6 dB, amplified by 6 dB and 12 dB).

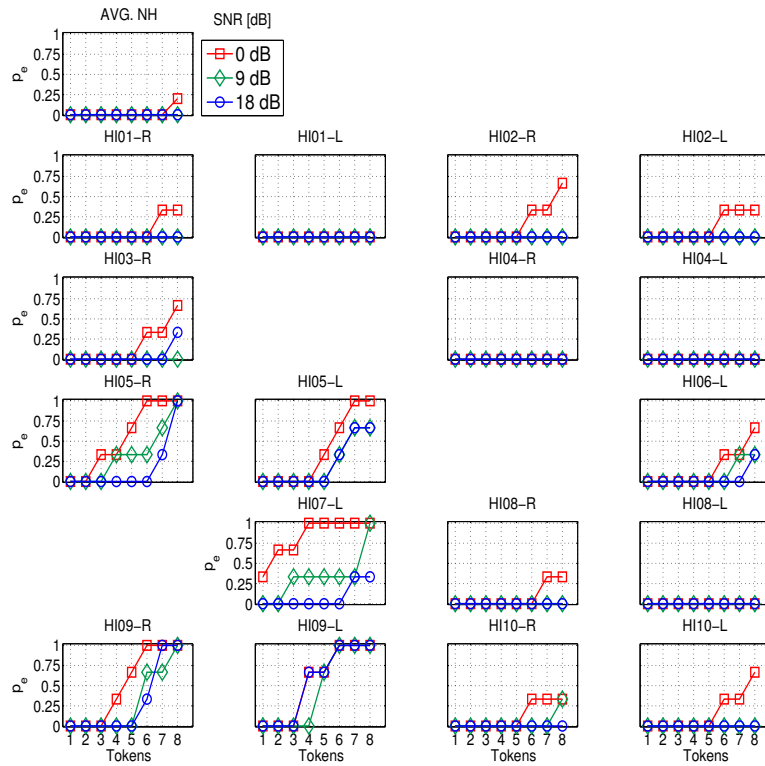


Figure A.2: The sorted errors when primary cues are attenuated are unmodified.

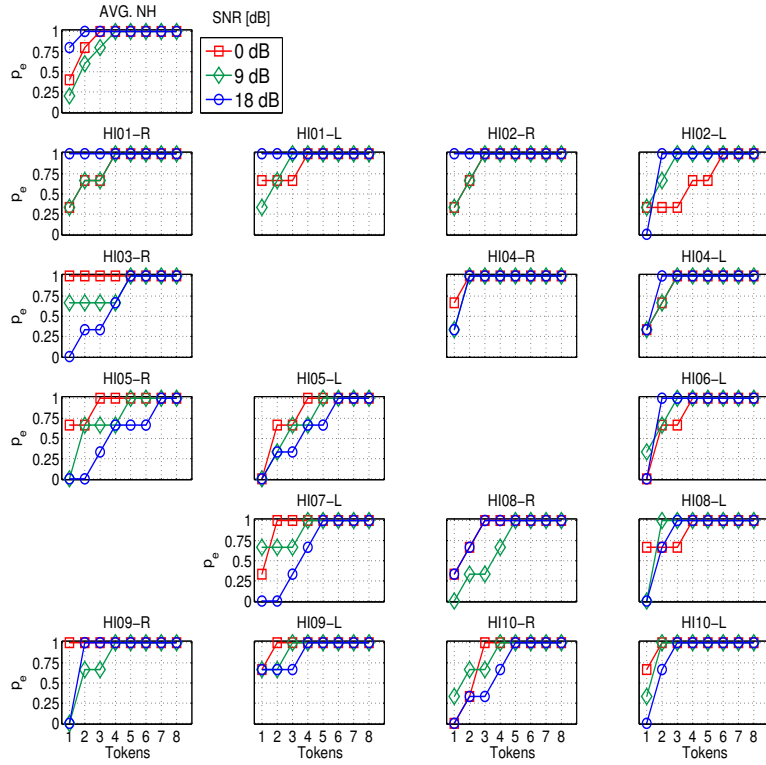


Figure A.3: Experiment I: The sorted errors when primary cues are removed.

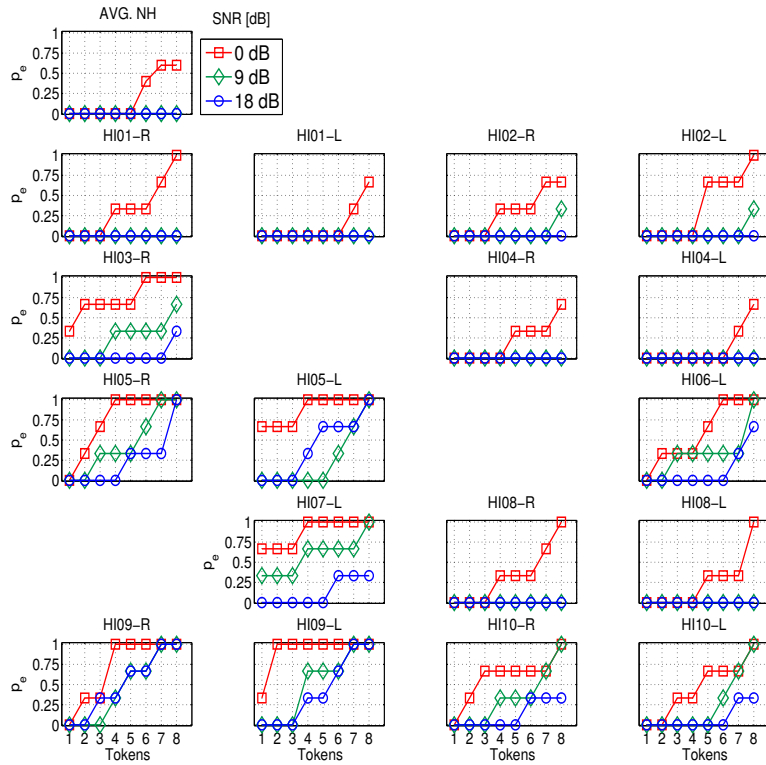


Figure A.4: The sorted errors when primary cues are attenuated by -6[dB].

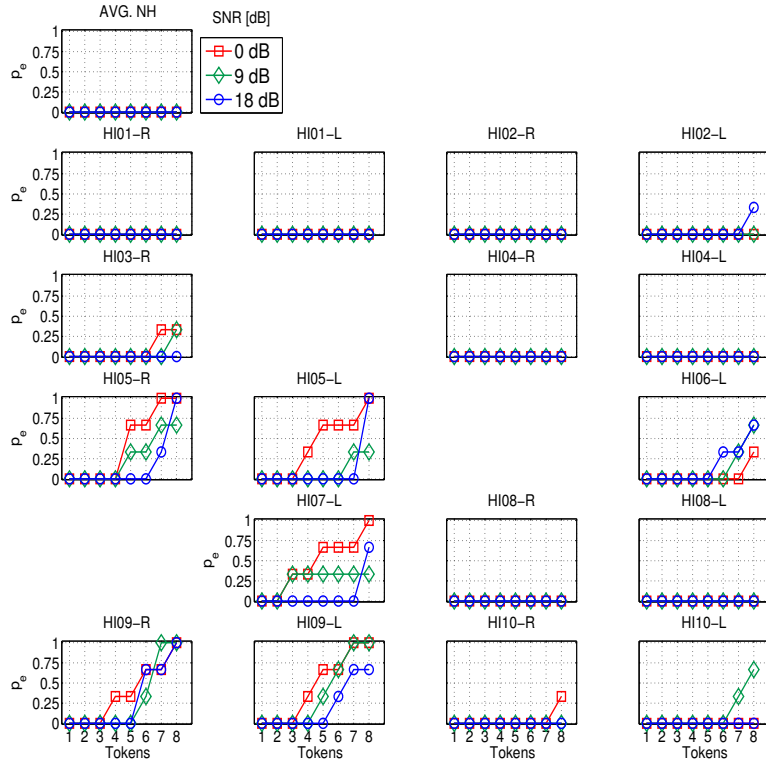


Figure A.5: The sorted errors when primary cues are amplified by 6[dB].

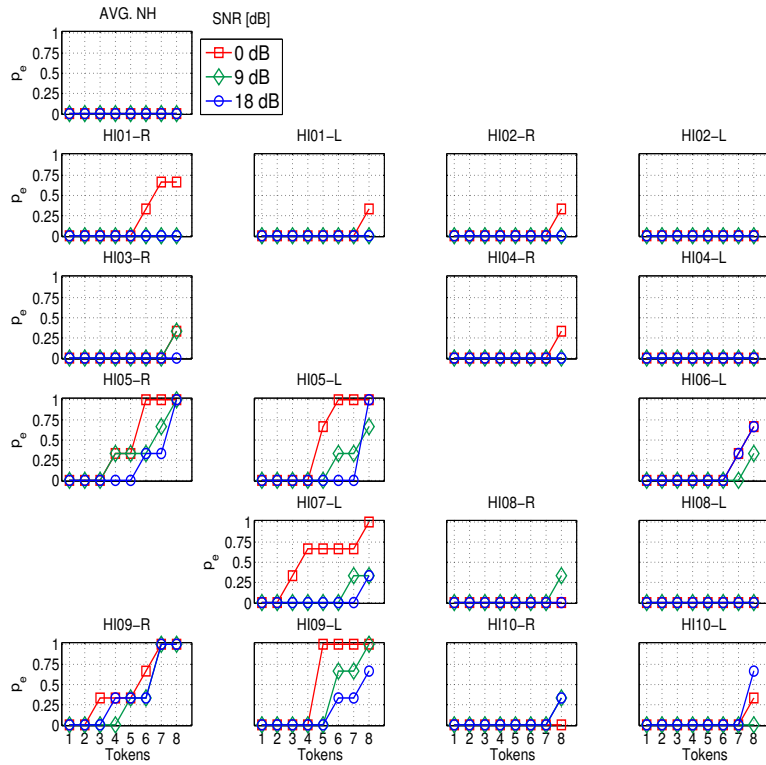


Figure A.6: The sorted errors when primary cues are amplified by 12[dB].

A.3 Average Token Error Analysis

Figures A.7 - A.14 show the average token error for eight tokens when the primary cue was unmodified, removed, attenuated -6 dB, amplified by 6 dB and 12 dB for the average NH ears and 17 individual HI ears. Each plot is for a specified token (i.e. m111 /ka/, f103 /ka/, m111 /ga/, f103 /ga/, m115 /ta/, f119 /ta/ f105 /da/, f119 /da/).

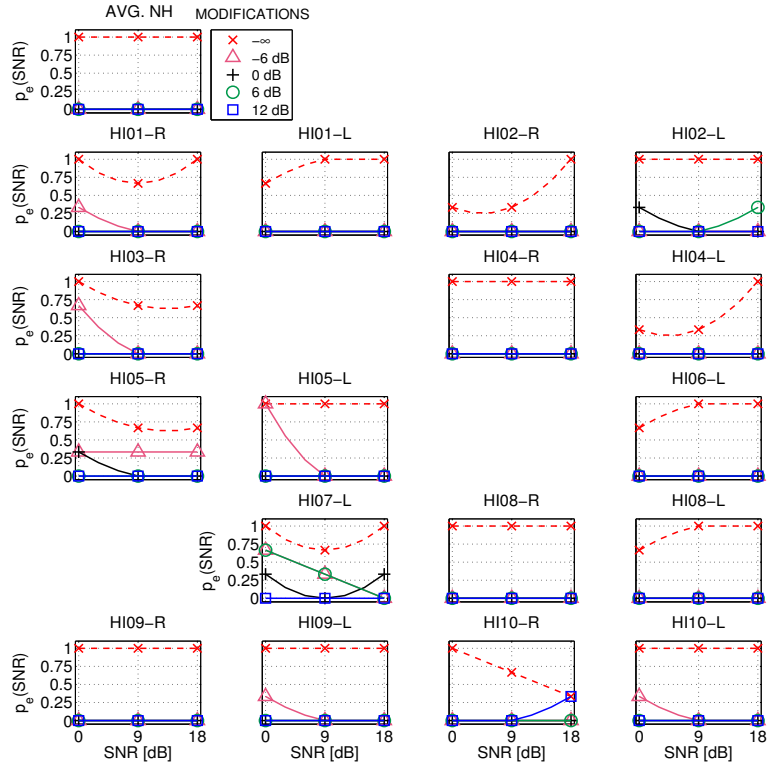


Figure A.7: m111 ka primary cues modifications compared.

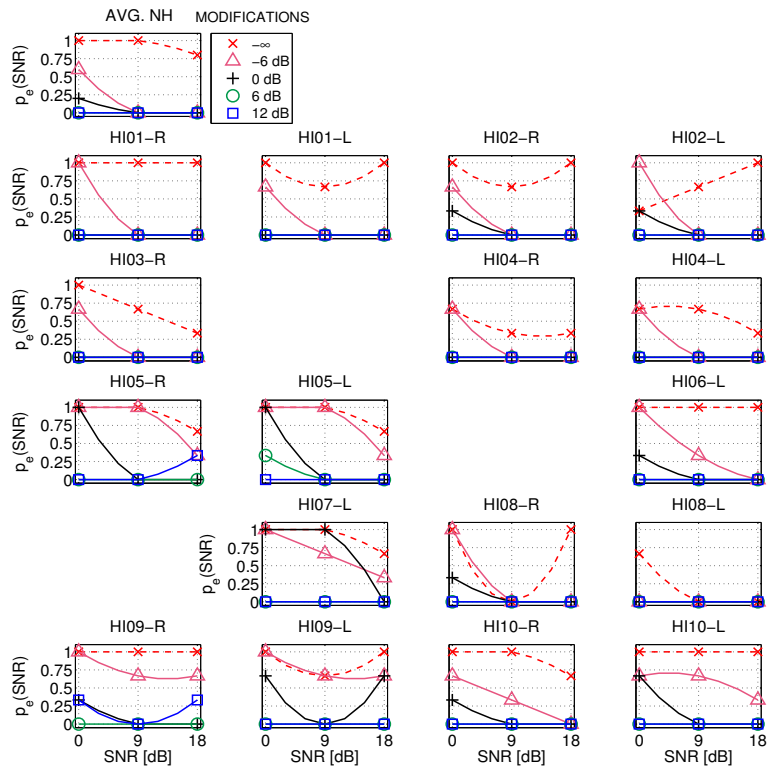


Figure A.8: f103 ka primary cues modifications compared.

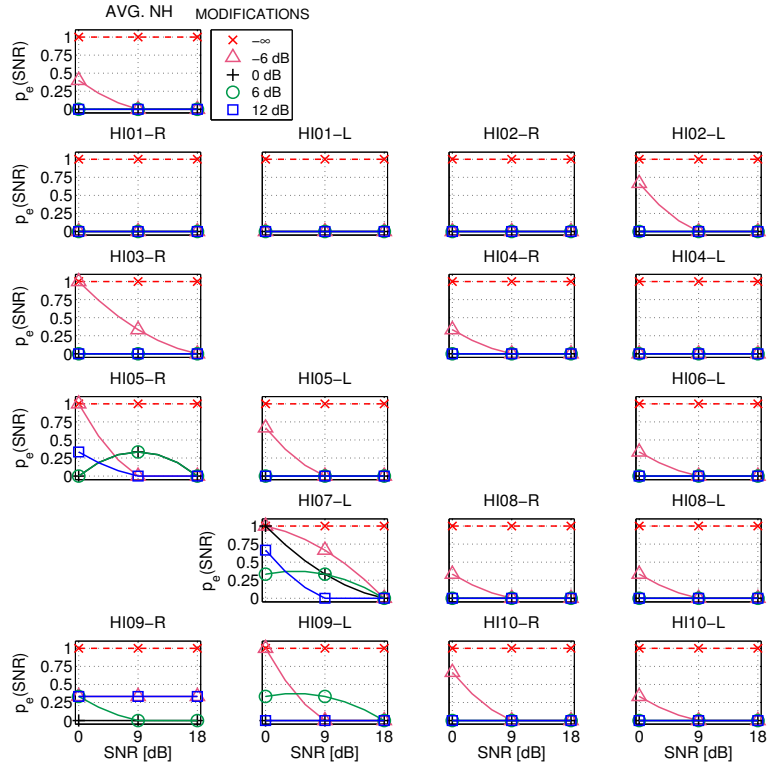


Figure A.9: m111 ga primary cues modifications compared.

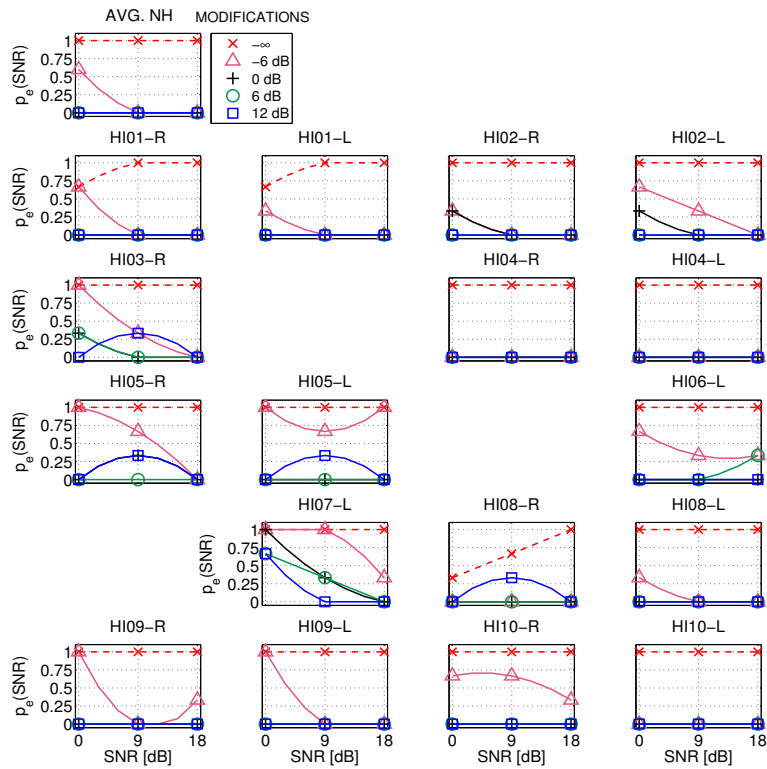


Figure A.10: f103 ga primary cues modifications compared.

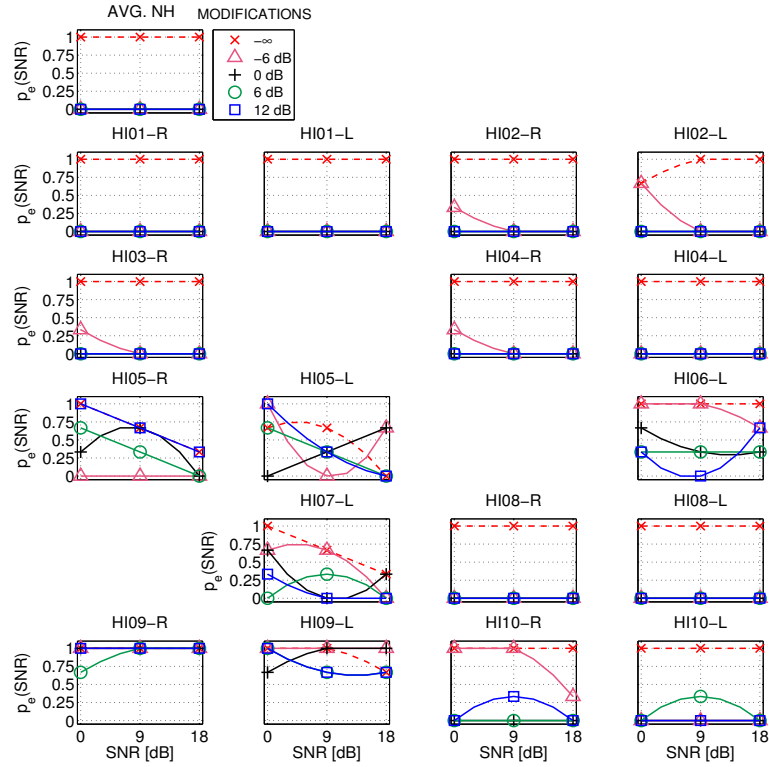


Figure A.11: m115 to primary cues modifications compared.

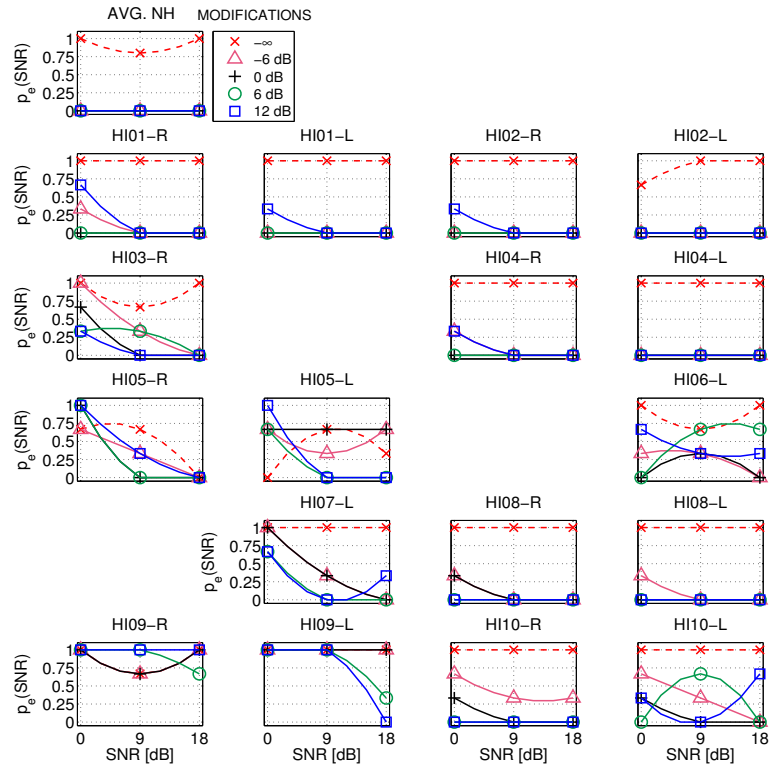


Figure A.12: f119 to primary cues modifications compared.

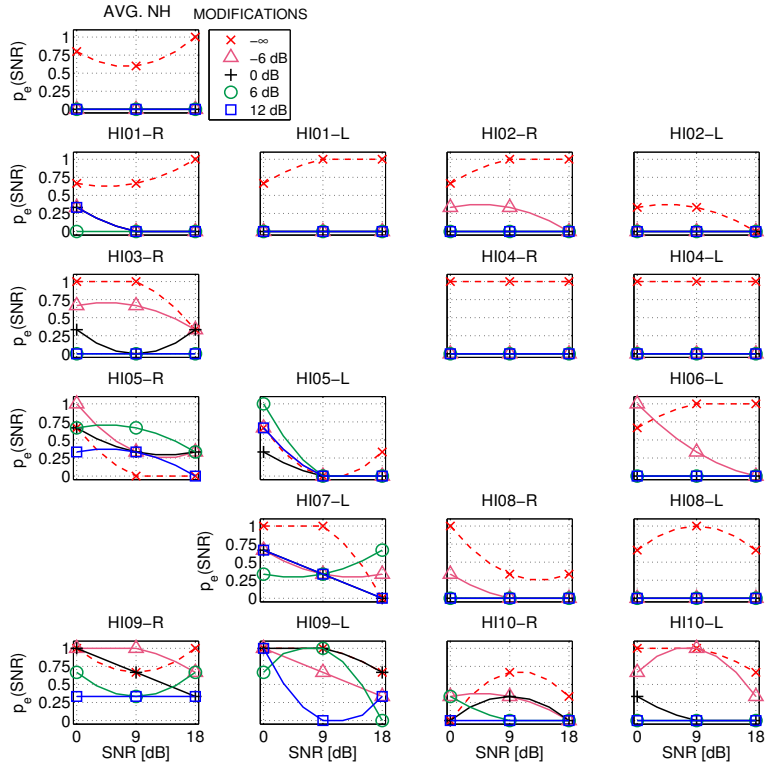


Figure A.13: f105 da primary cues modifications compared.

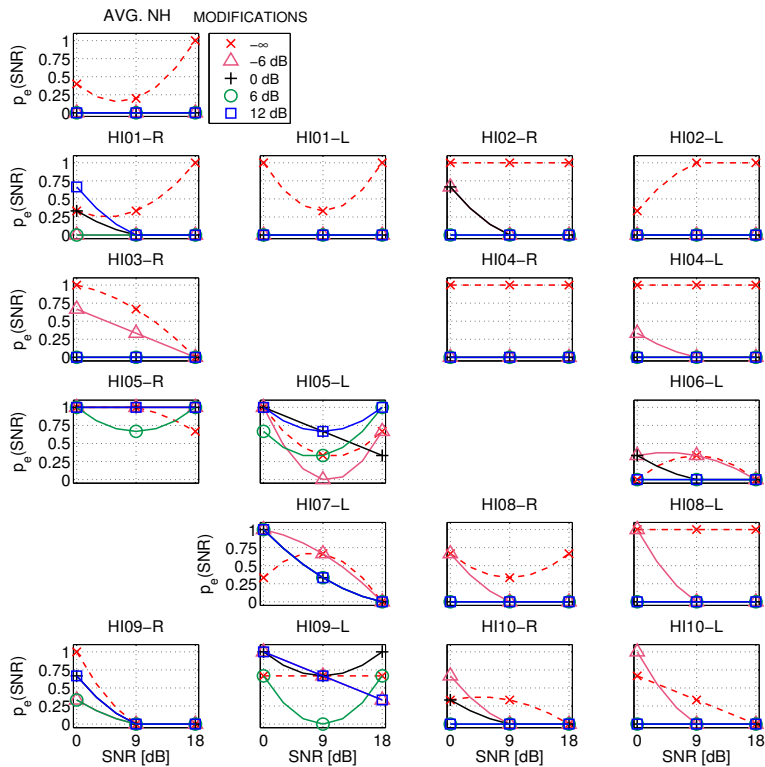


Figure A.14: f119 da primary cues modifications compared.

A.4 Confusion Patterns

Figures A.15 - A.54 show the confusion patterns of a token for the average NH ears and 17 individual HI ears. Each plot is a specified token (i.e. m111 /ka/, f103 /ka/, m111 /ga/, f103 /ga/, m115 /ta/, f119 /ta/ f105 /da/, f119 /kda/) and modification condition (i.e. primary cue was unmodified, removed, attenuated -6 dB, multiplied by 6 dB and 12 dB).

m111ka

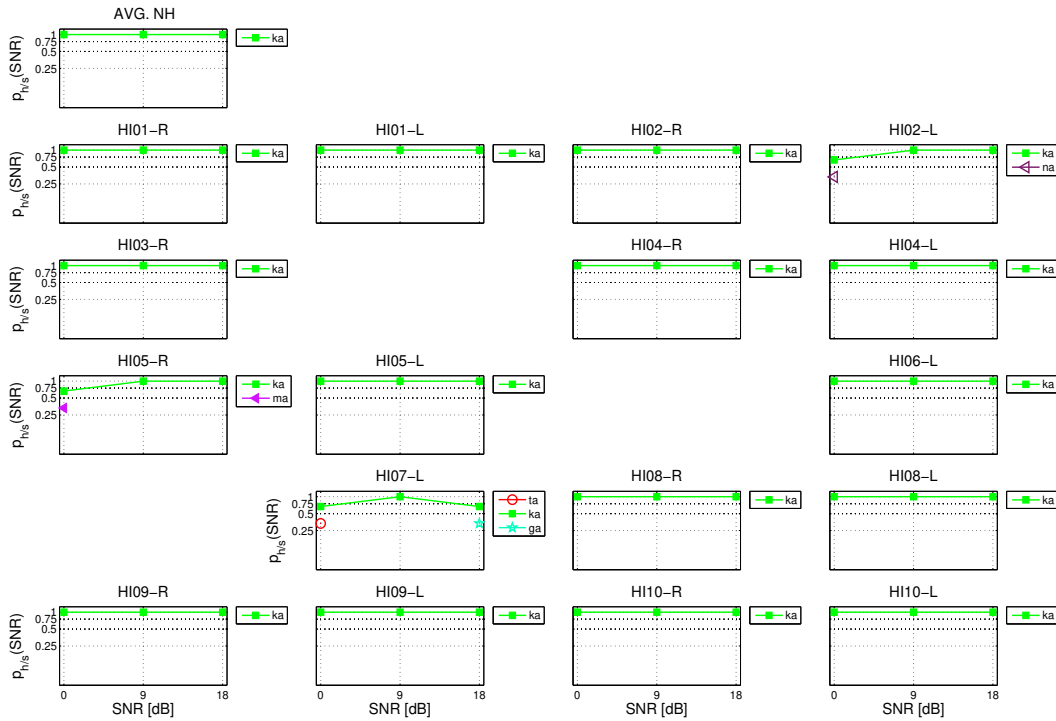


Figure A.15: m111 ka confusions patterns when primary cues are unmodified.

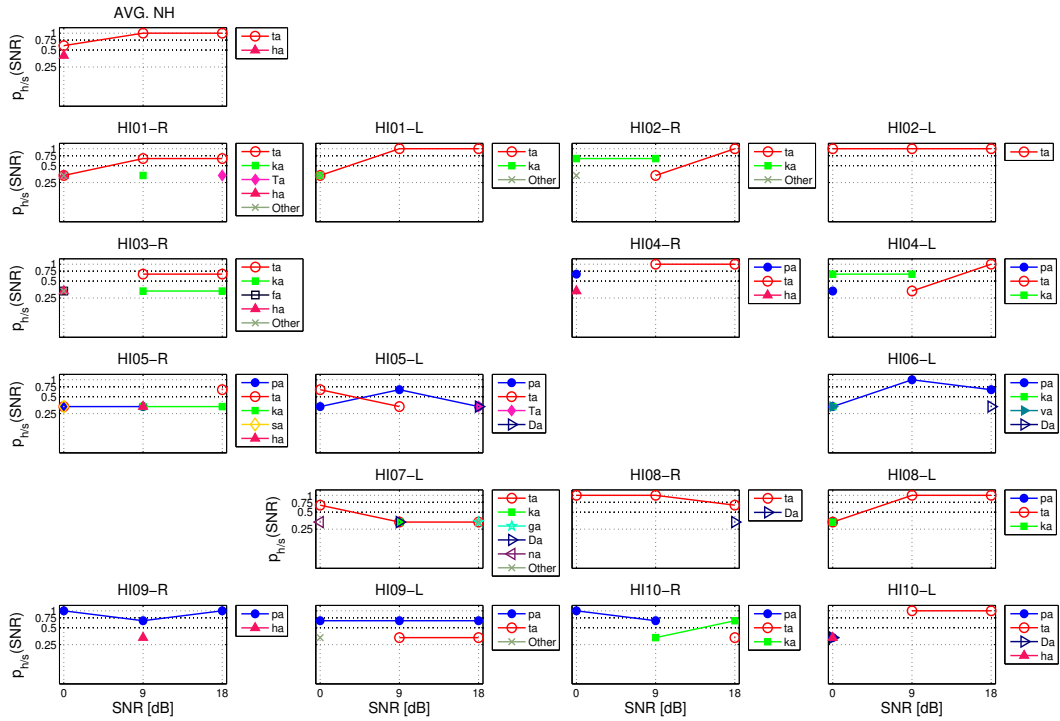


Figure A.16: m111 ka confusions patterns when primary cues are removed.

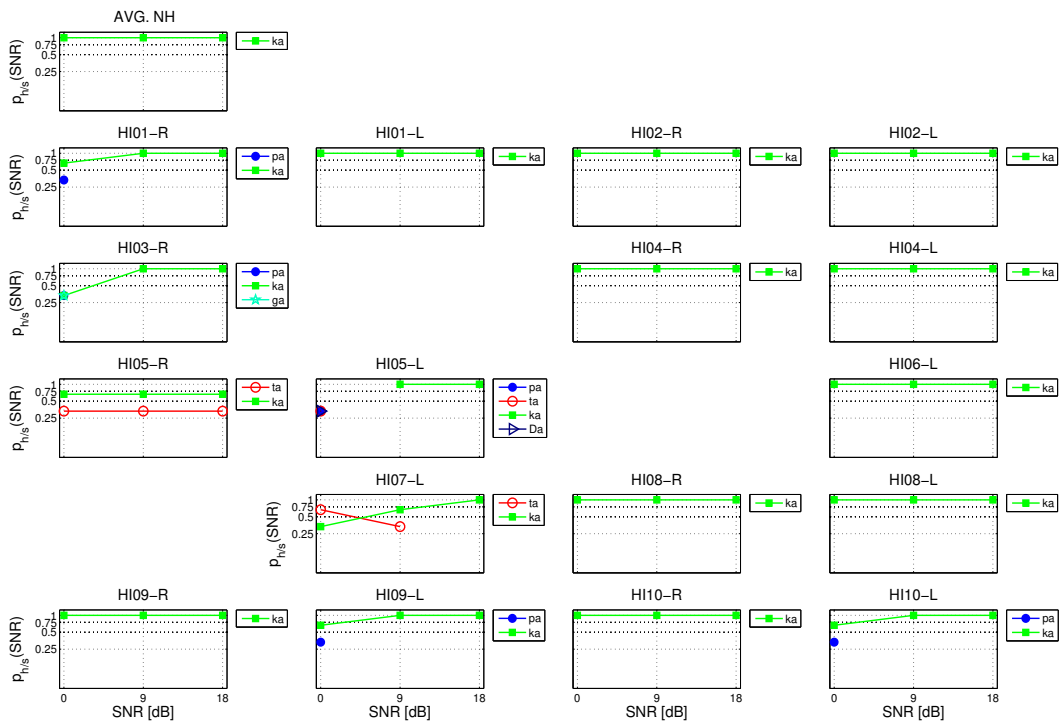


Figure A.17: m111 ka confusions patterns when primary cues are attenuated by -6 [dB].

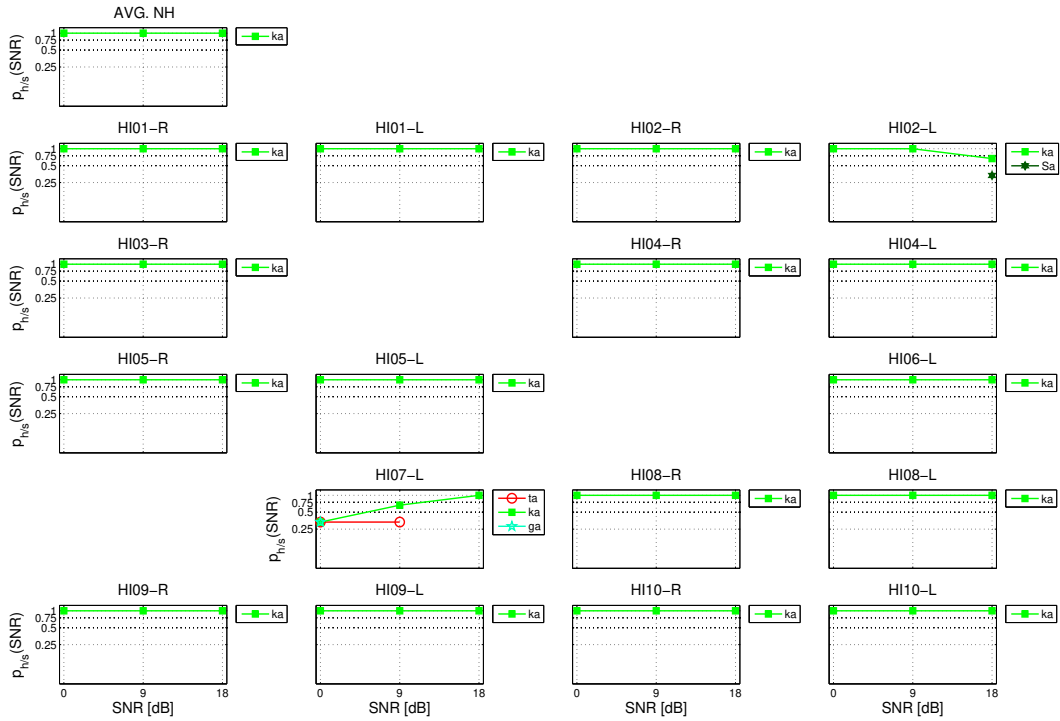


Figure A.18: m111 ka confusions patterns when primary cues are amplified by 6 [dB].

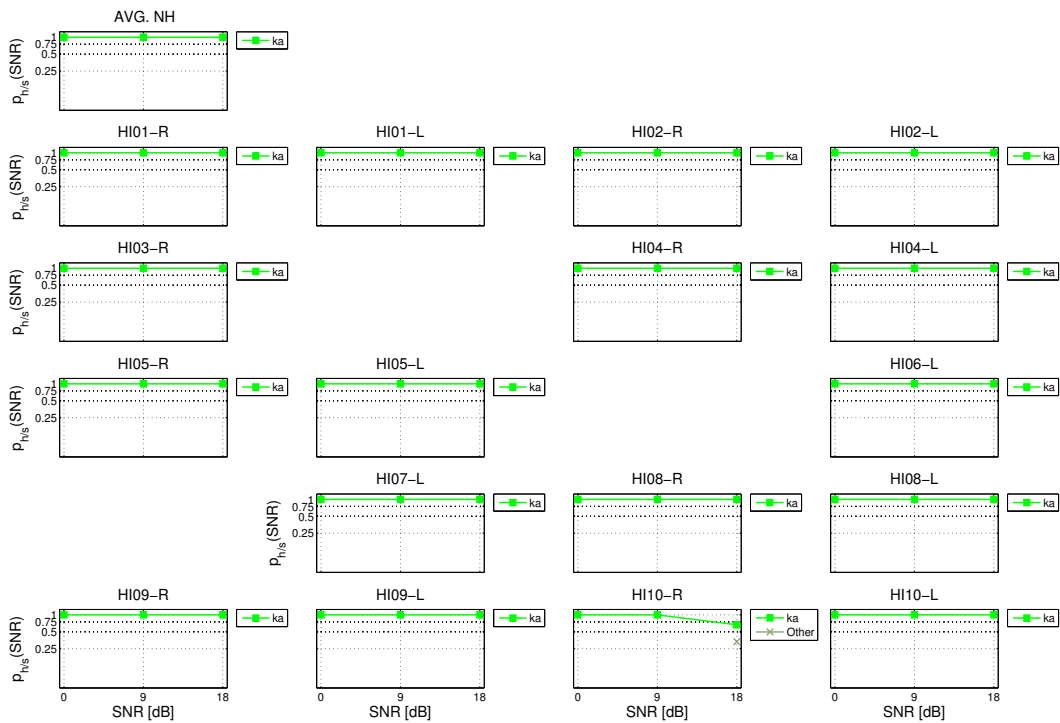


Figure A.19: m111 ka confusions patterns when primary cues are amplified by 12 [dB].

f103ka

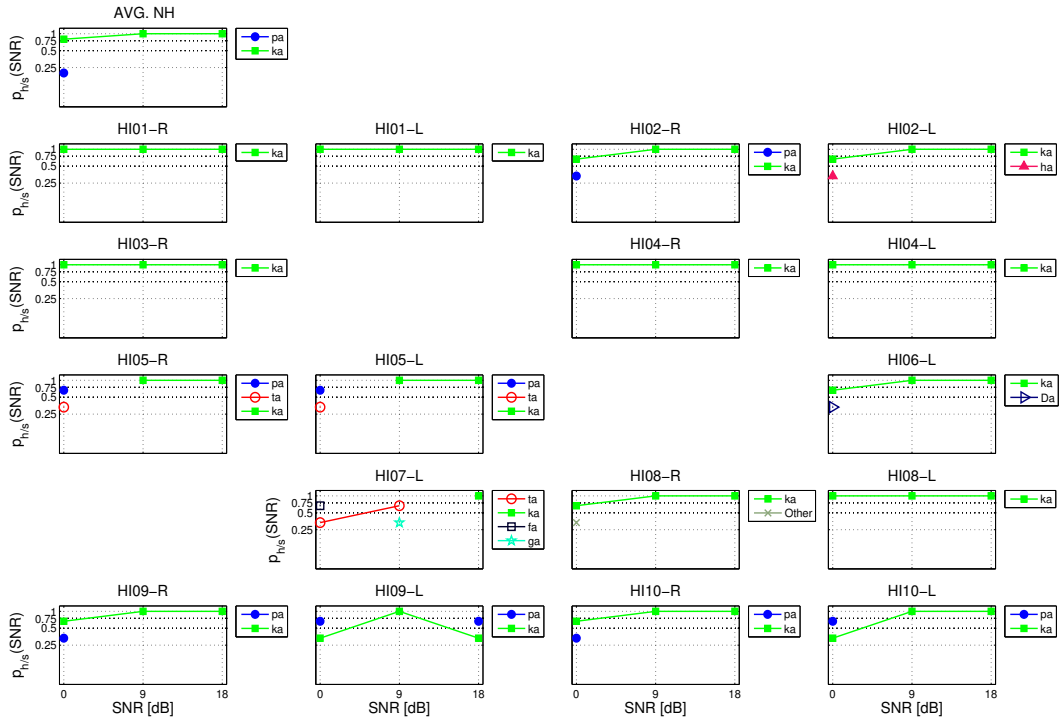


Figure A.20: f103 ka confusions patterns when primary cues are unmodified.

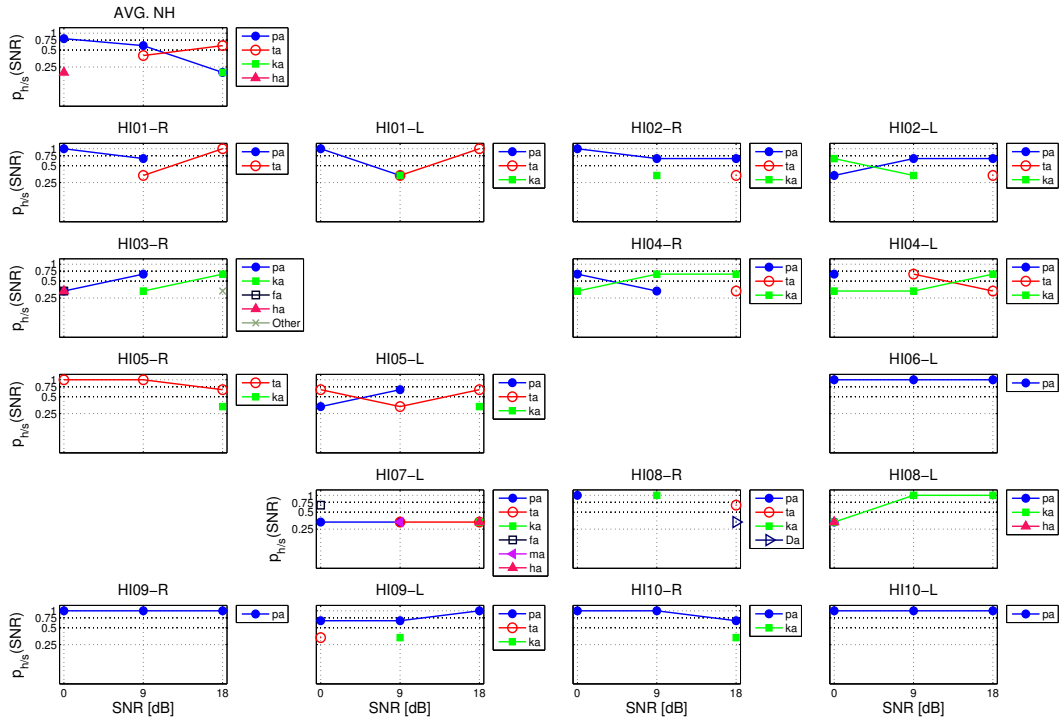


Figure A.21: f103 ka confusions patterns when primary cues are removed.

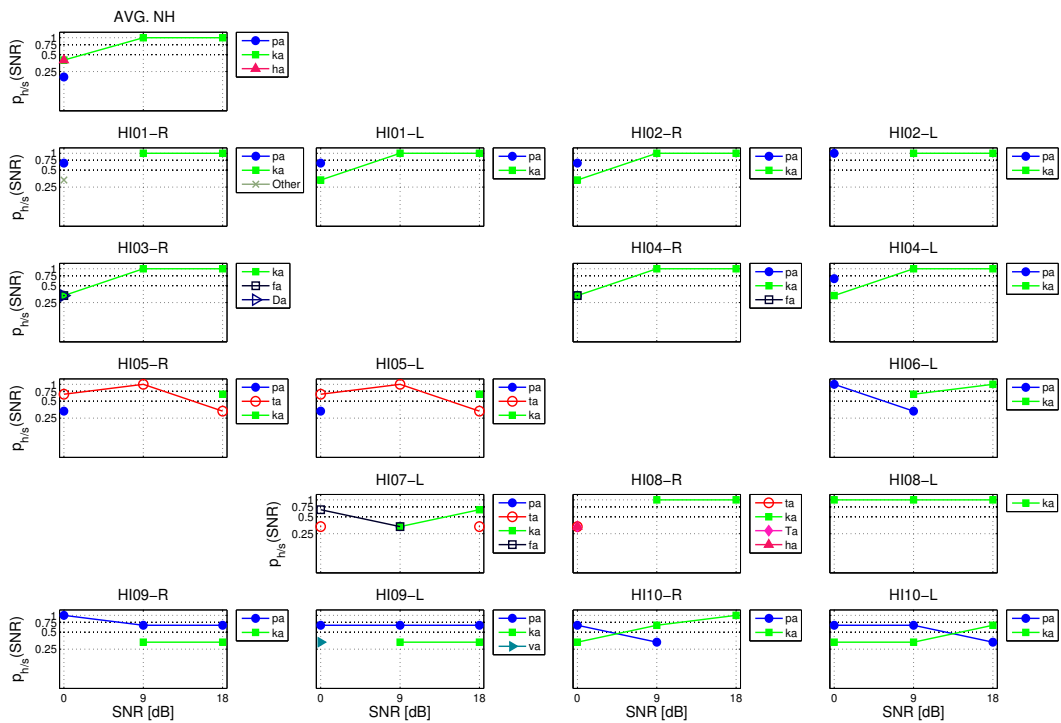


Figure A.22: f103 ka confusions patterns when primary cues are attenuated by -6 [dB].

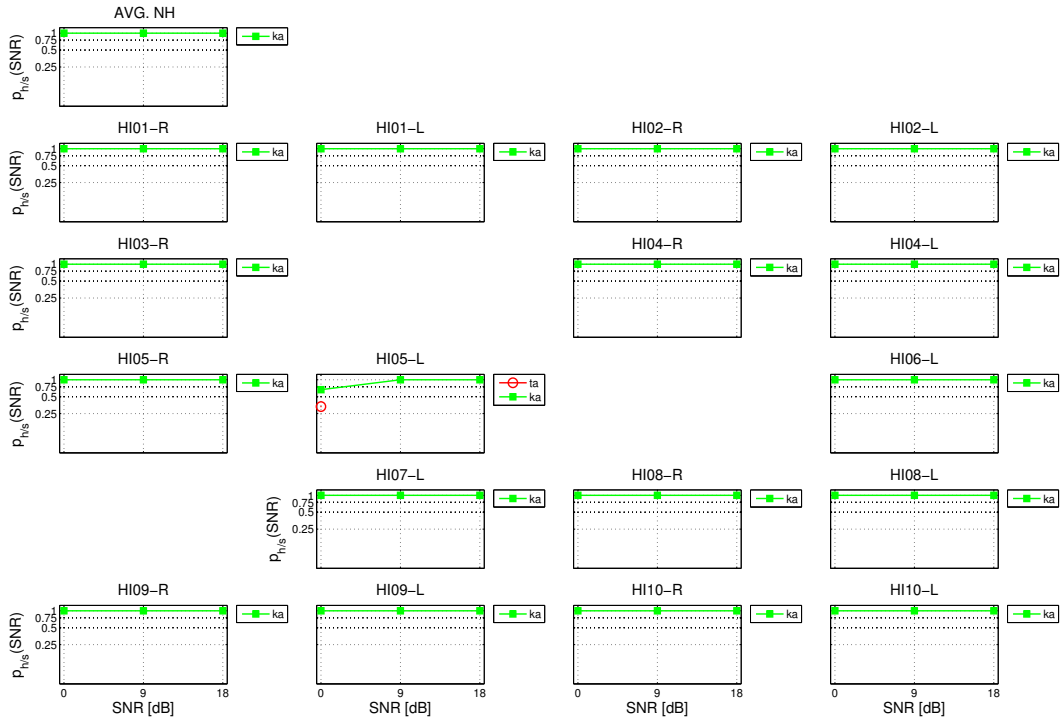


Figure A.23: f103 ka confusions patterns when primary cues are amplified by 6 [dB].

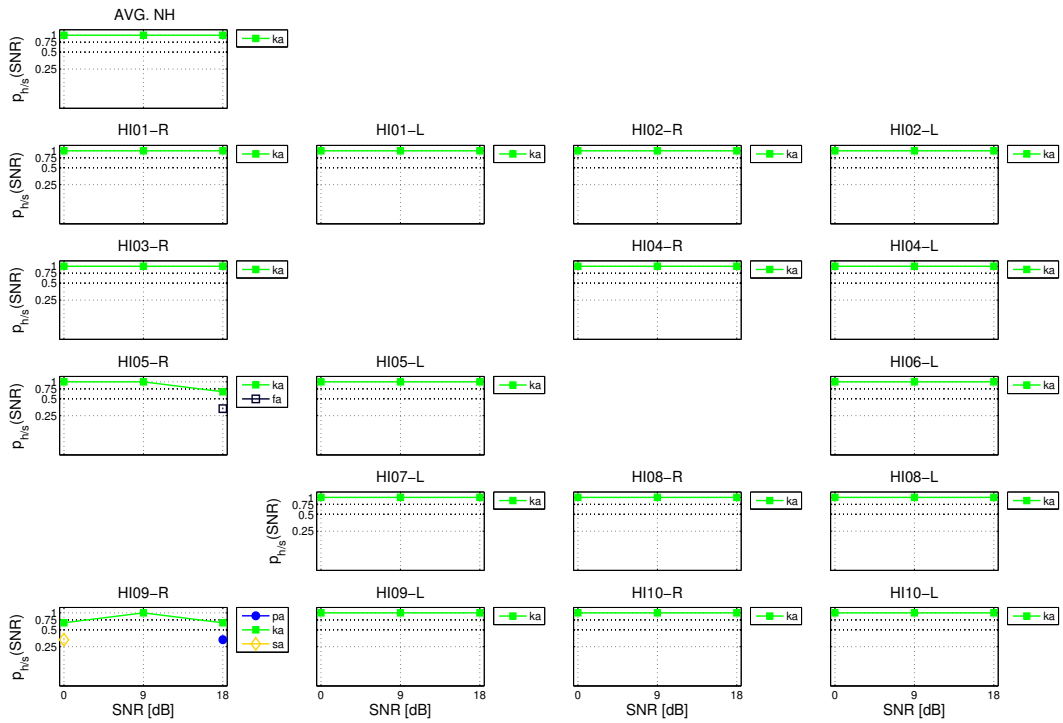


Figure A.24: f103 ka confusions patterns when primary cues are amplified by 12 [dB].

m111ga

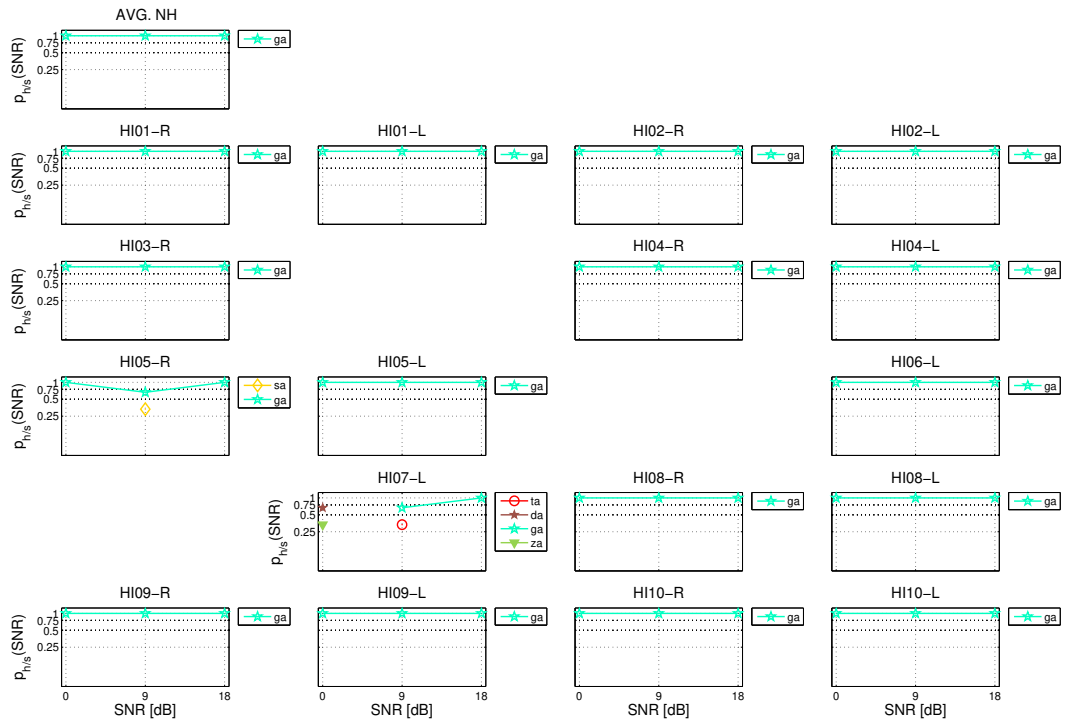


Figure A.25: m111 ga confusions patterns when primary cues are unmodified.

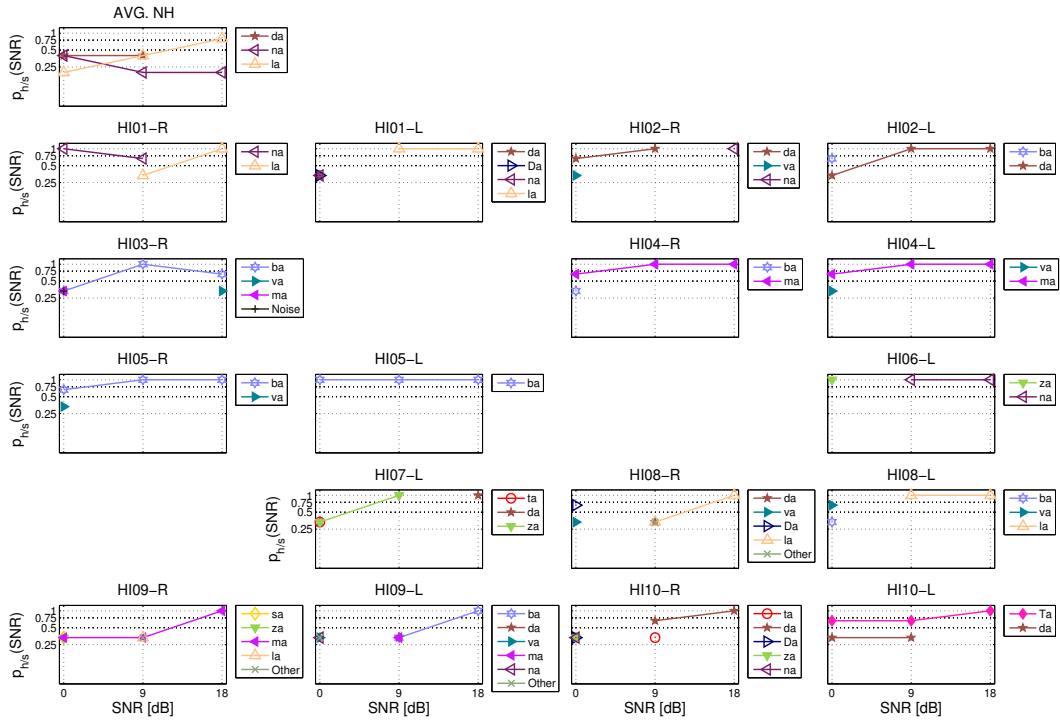


Figure A.26: m111 ga confusions patterns when primary cues are removed.

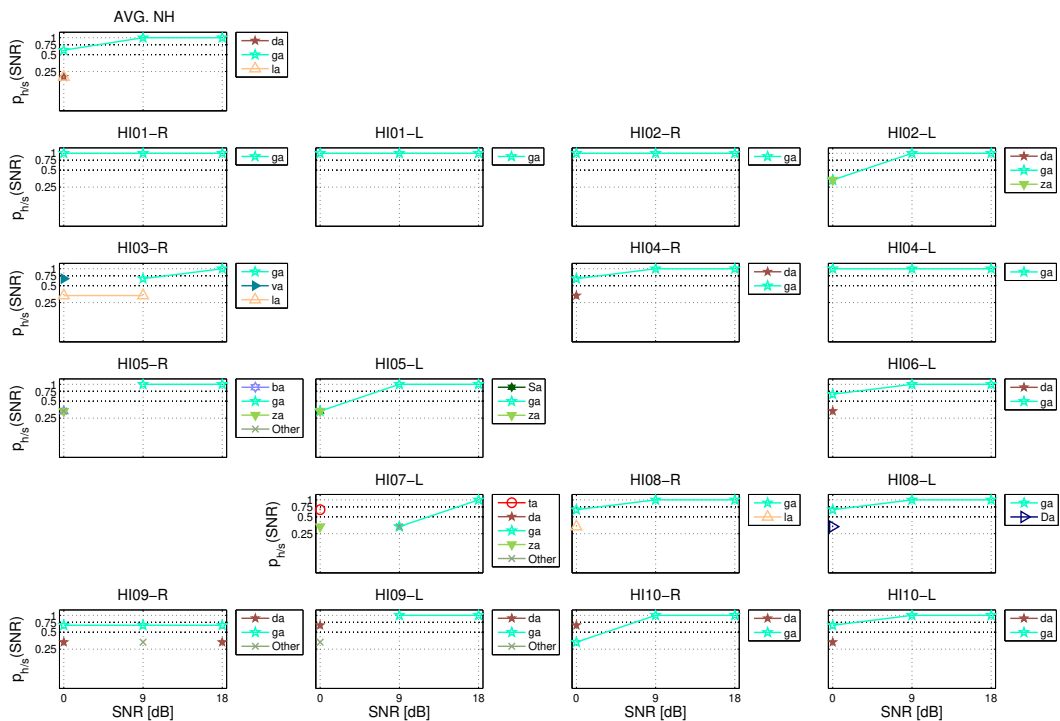


Figure A.27: m111 ga confusions patterns when primary cues are attenuated by -6 [dB].

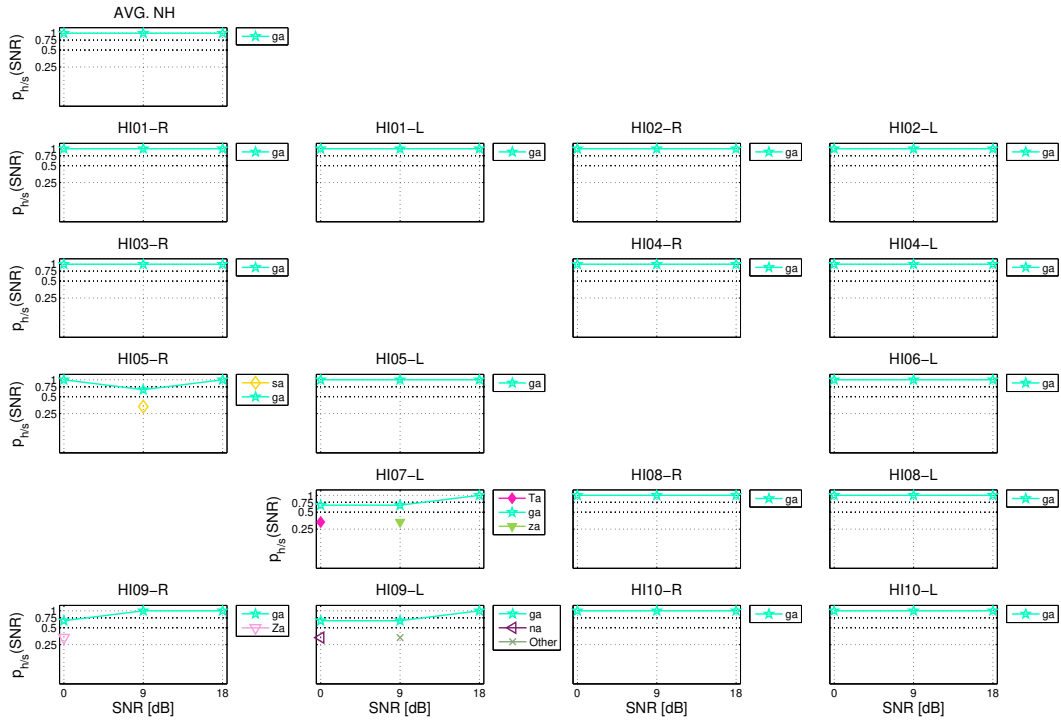


Figure A.28: m11 ga confusions patterns when primary cues are amplified by 6 [dB].

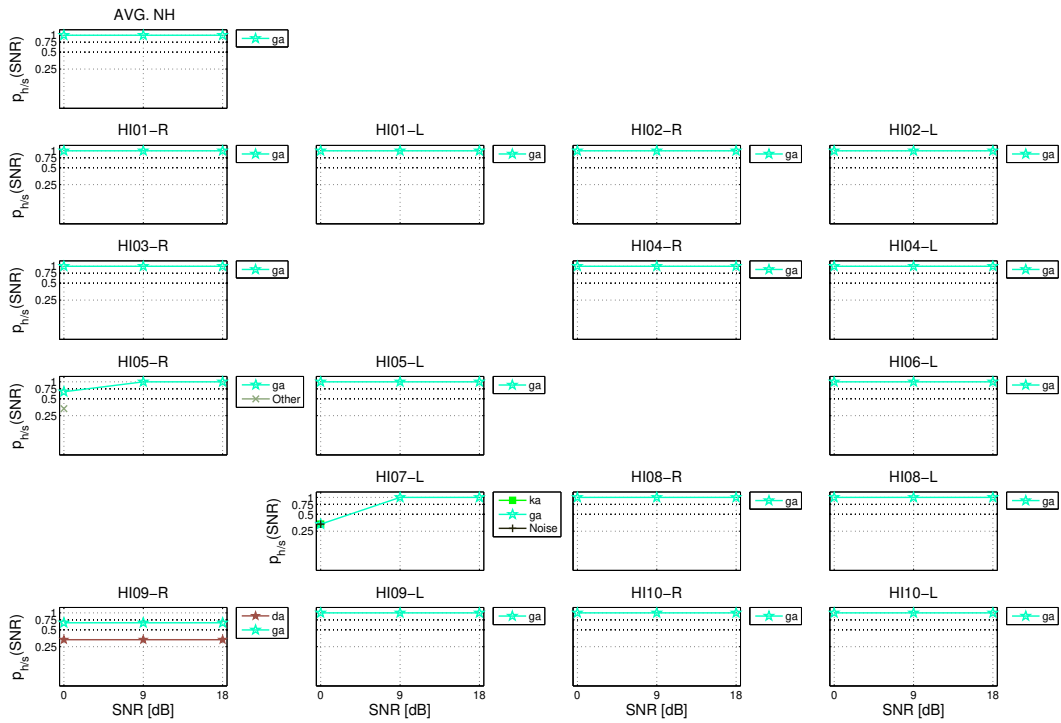


Figure A.29: m11 ga confusions patterns when primary cues are amplified 12 [dB].

f103ga

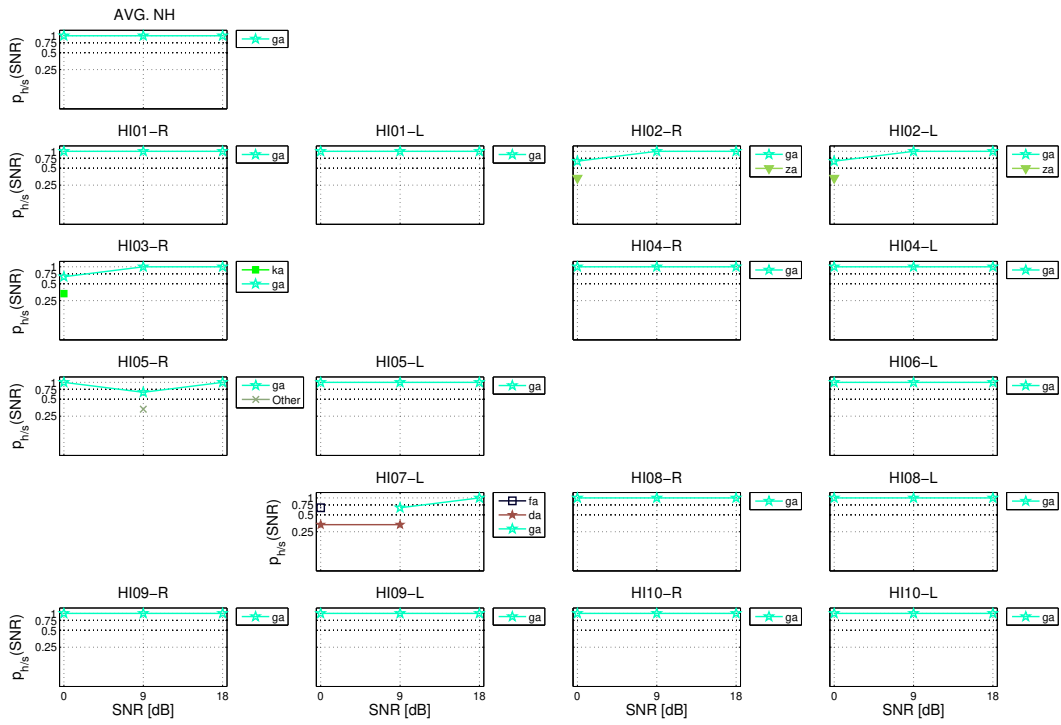


Figure A.30: f103 ga confusions patterns when primary cues are unmodified.

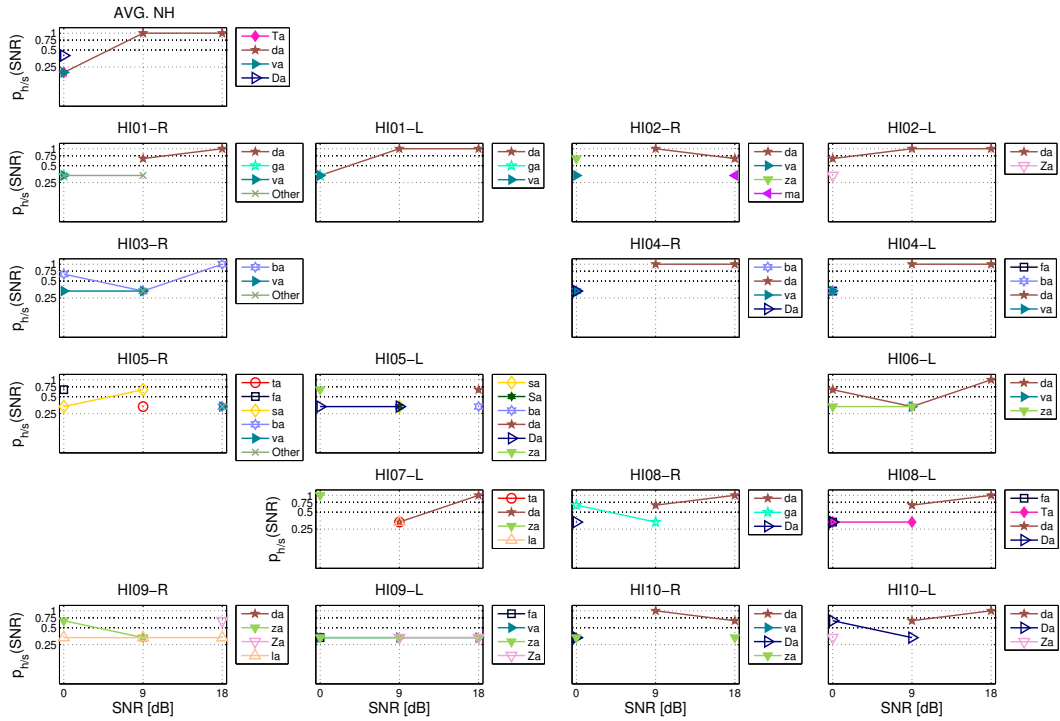


Figure A.31: f103 ga confusions patterns when primary cues are removed.

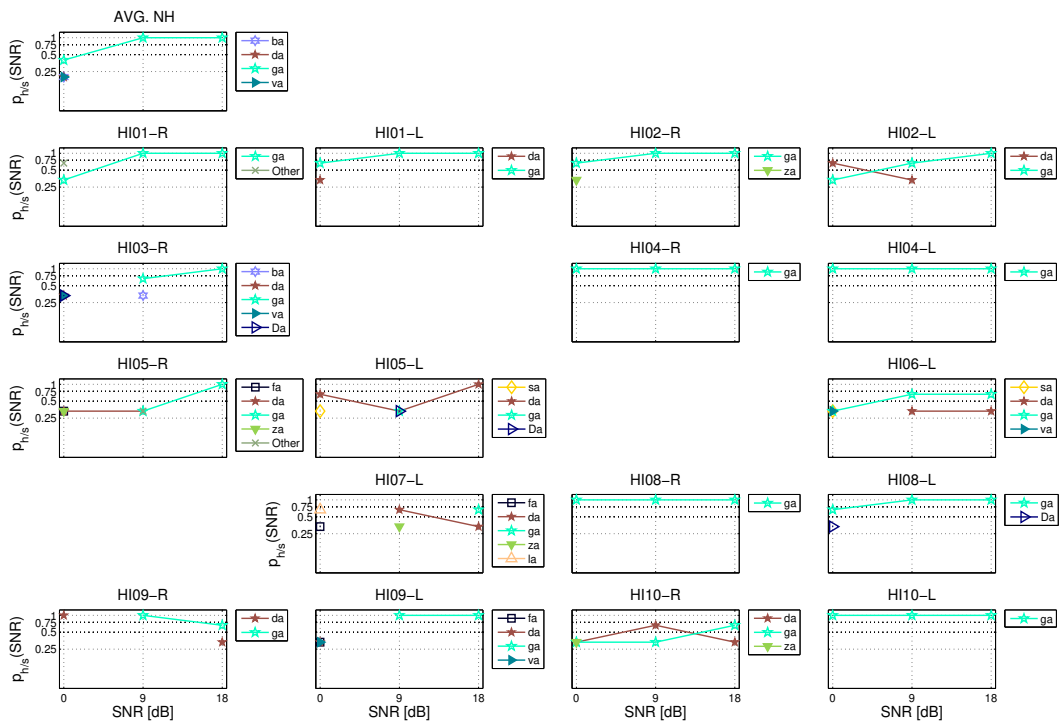


Figure A.32: f103 ga confusions patterns when primary cues are attenuated by -6 [dB].

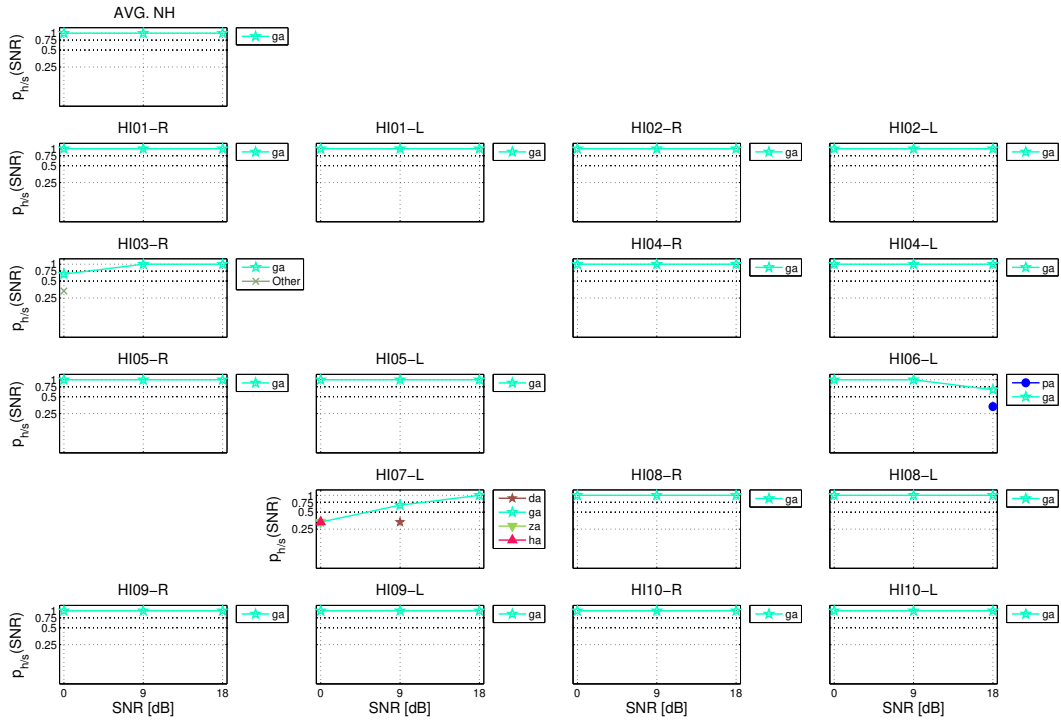


Figure A.33: f103 ga confusions patterns when primary cues are amplified by 6 [dB].

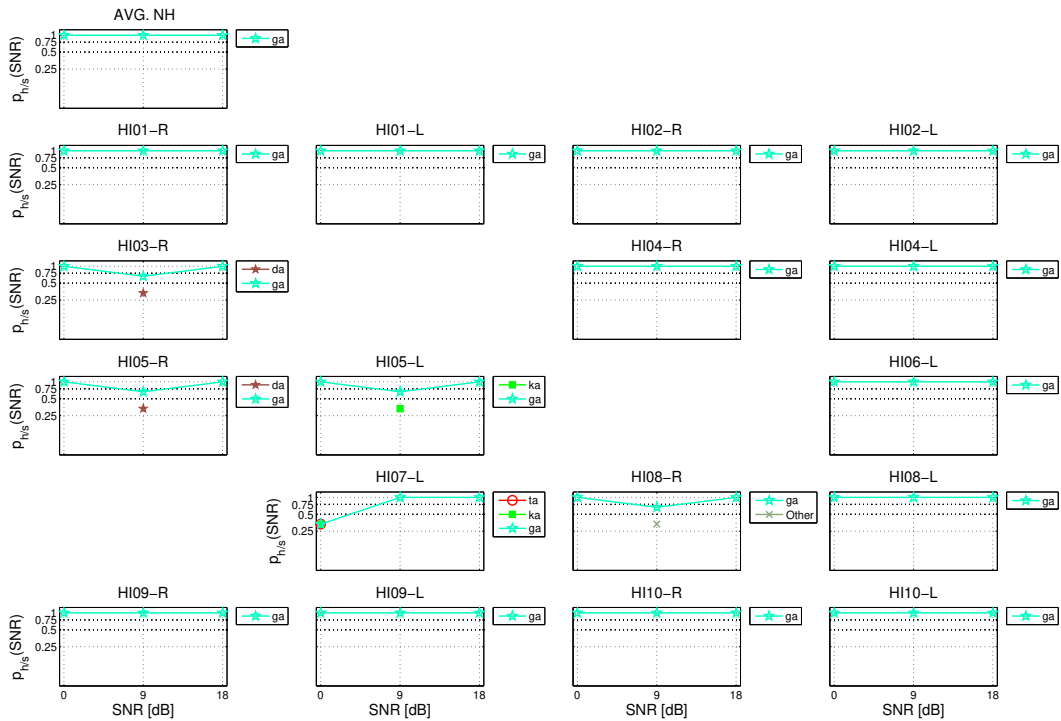


Figure A.34: f103 ga confusions patterns when primary cues are amplified by 12 [dB].

m115ta

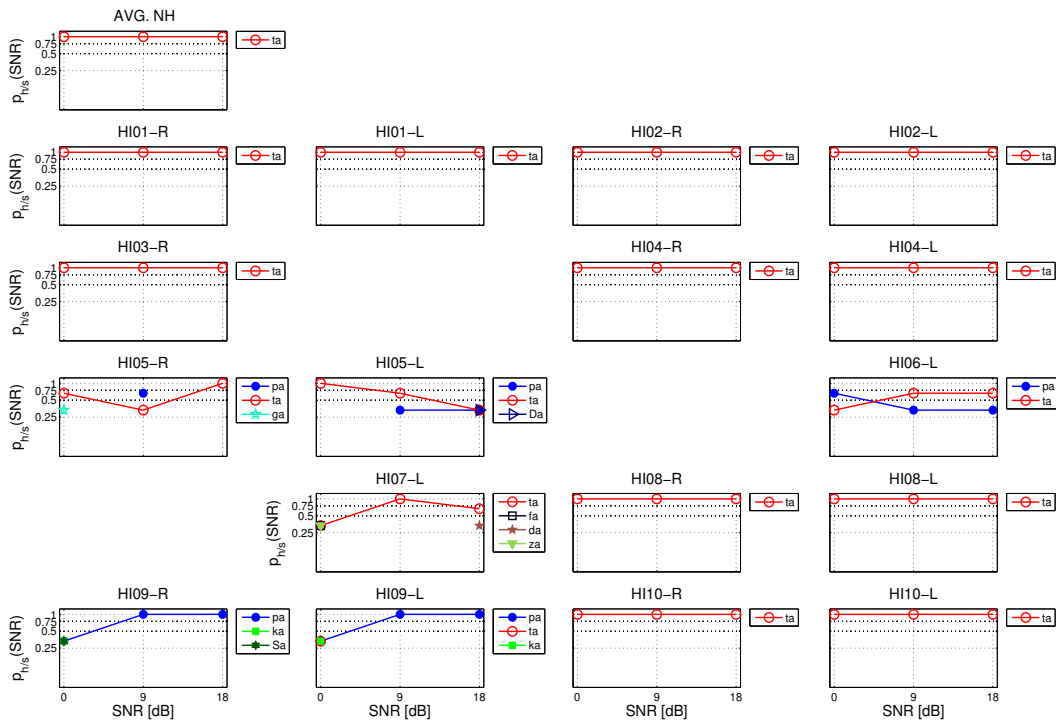


Figure A.35: m115 ta confusions patterns when primary cues are unmodified.

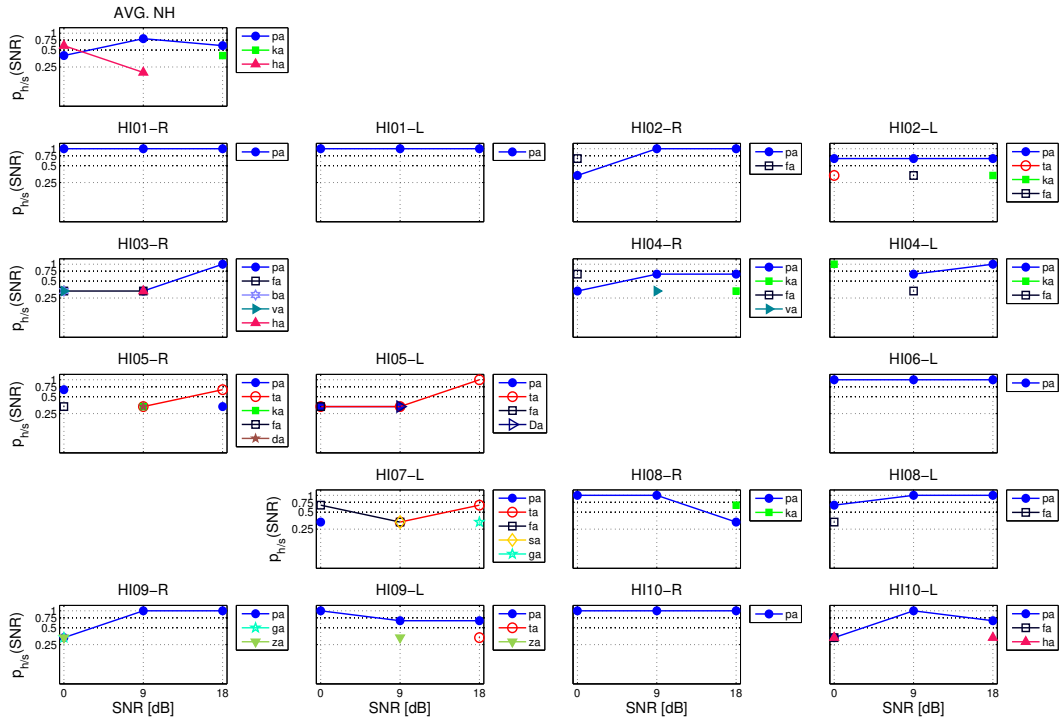


Figure A.36: m115 ta confusions patterns when primary cues are removed.

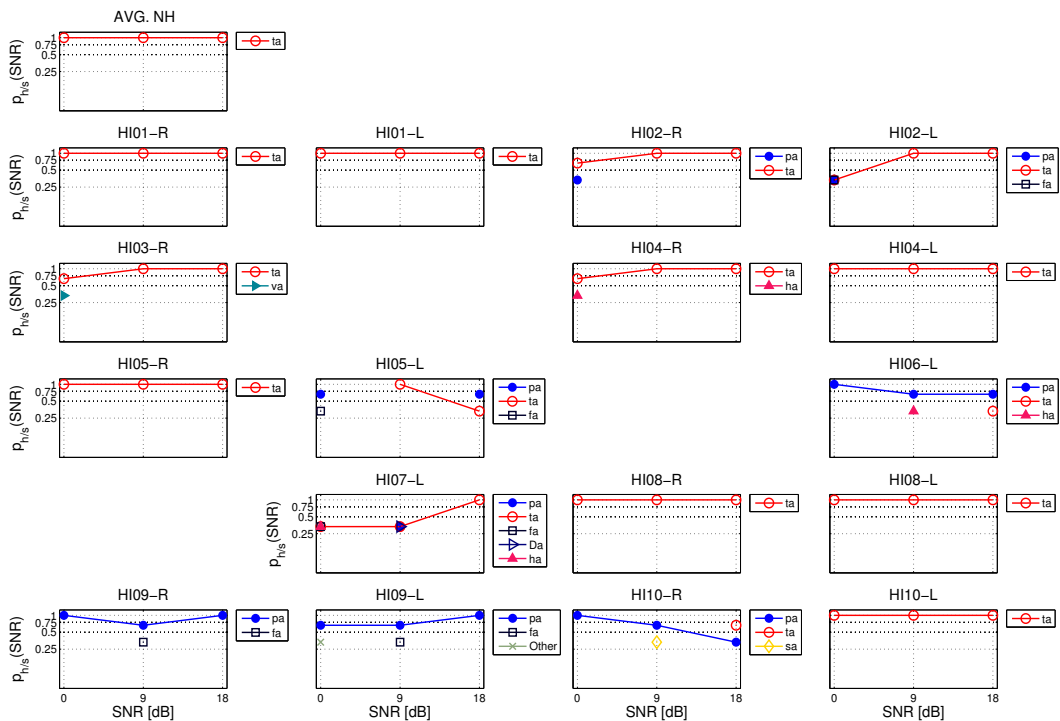


Figure A.37: m115 ta confusions patterns when primary cues are attenuated by -6 [dB].

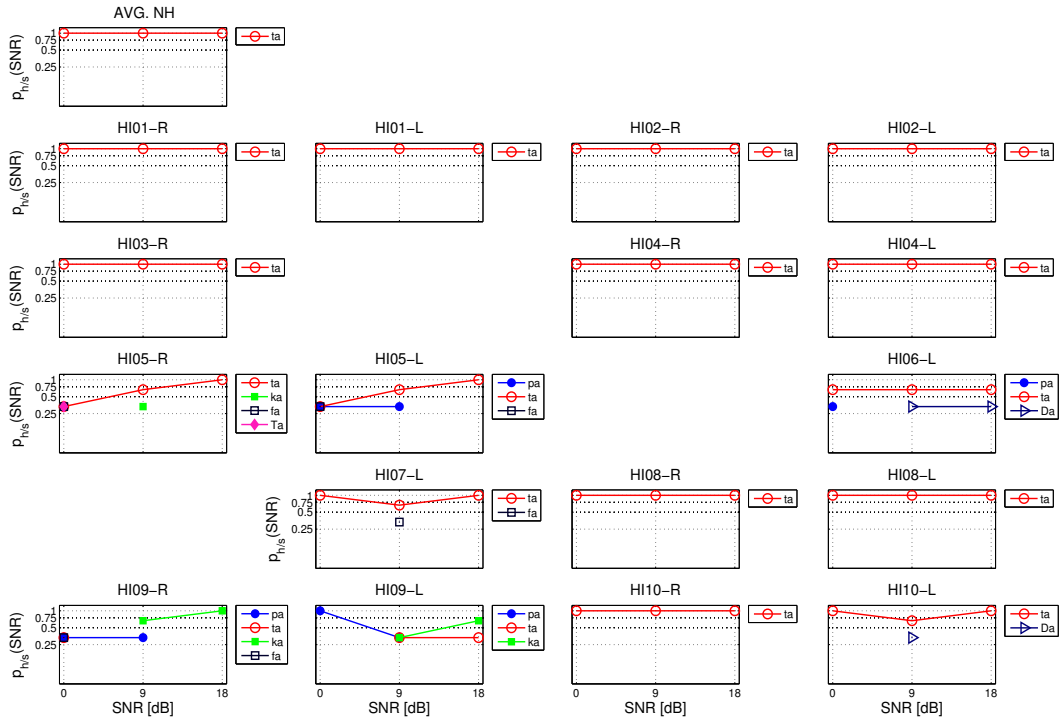


Figure A.38: m15 ta confusions patterns when primary cues are amplified by 6 [dB].

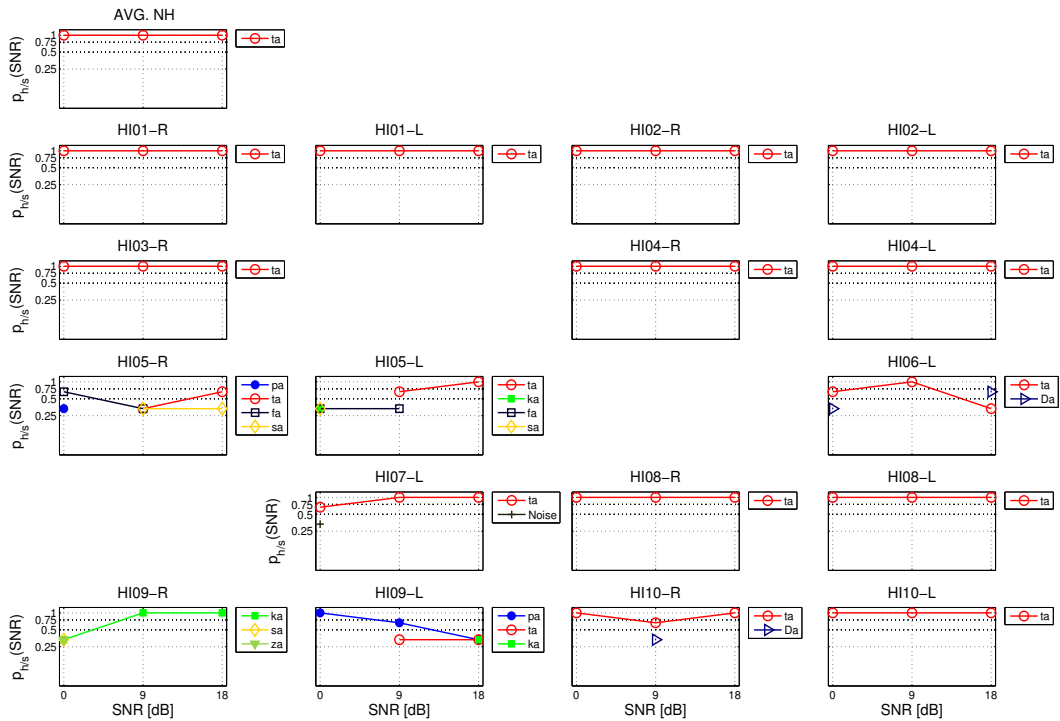


Figure A.39: m15 ta confusions patterns when primary cues are amplified by 12 [dB].

f119ta

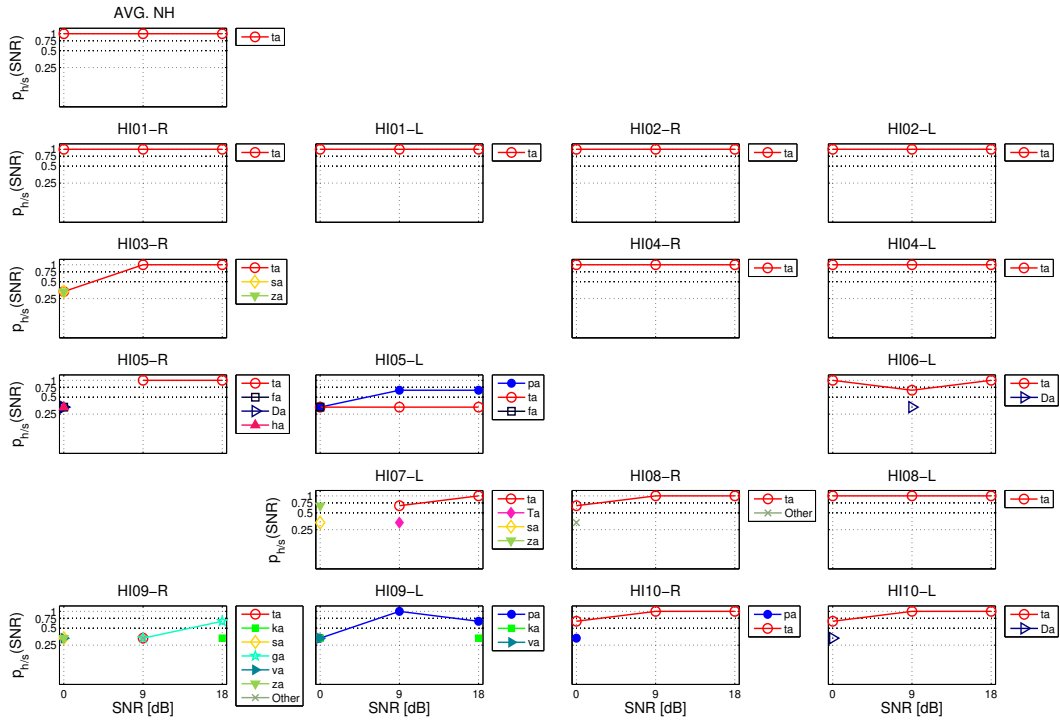


Figure A.40: f119 ta confusions patterns when primary cues are unmodified.

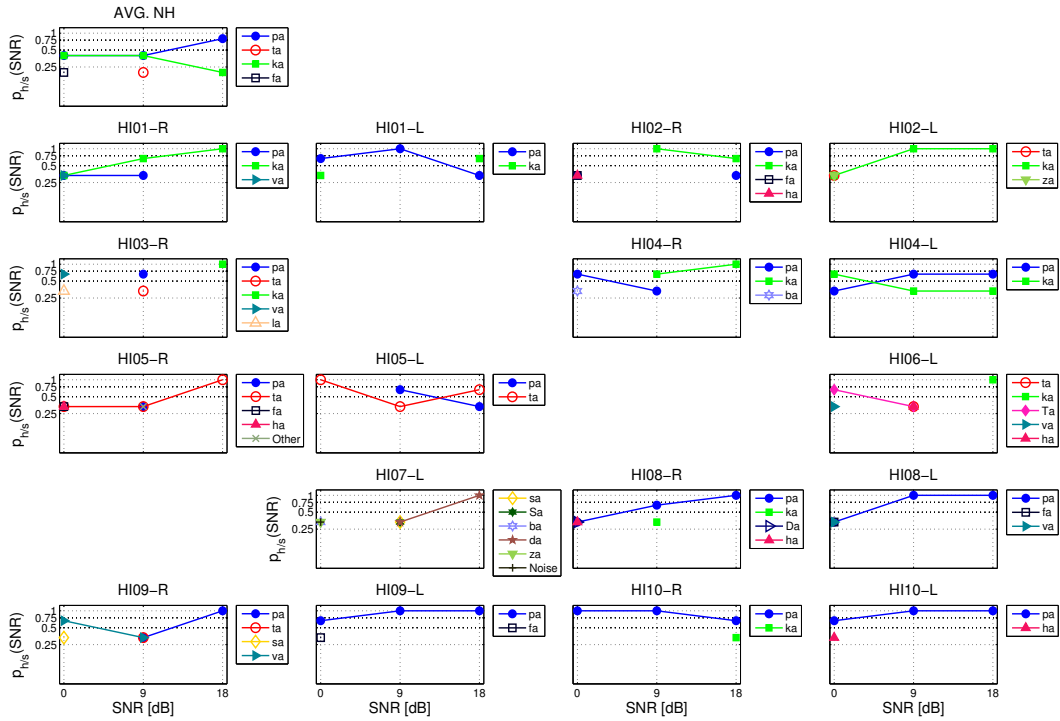


Figure A.41: f119 ta confusions patterns when primary cues are removed.

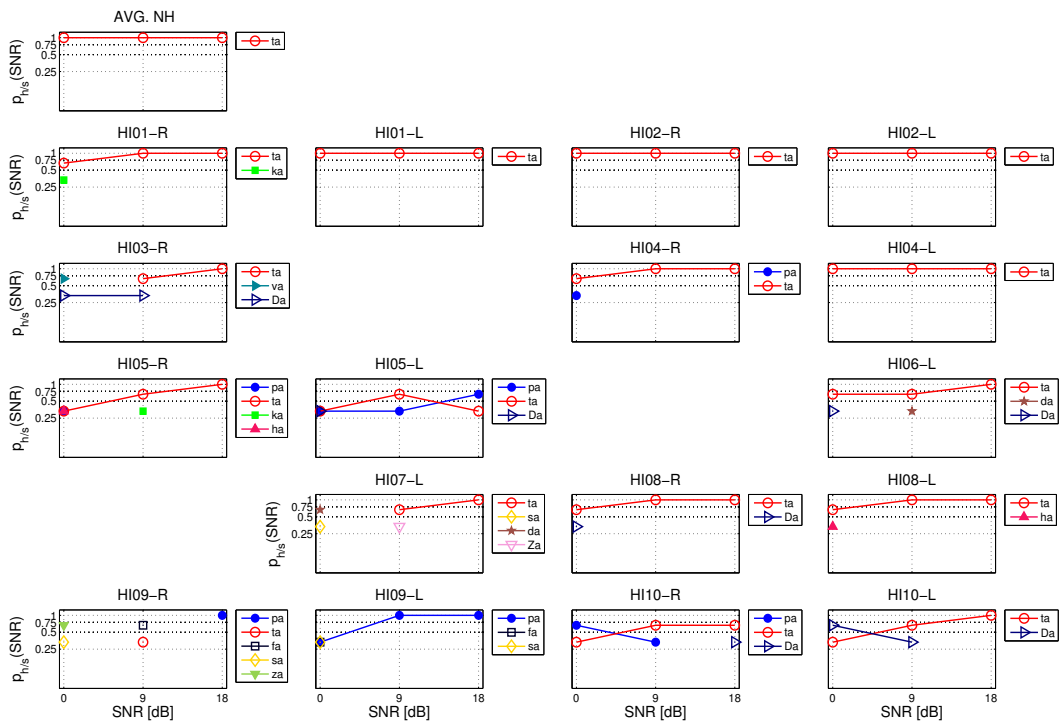


Figure A.42: f119 ta confusions patterns when primary cues are attenuated by -6 [dB].

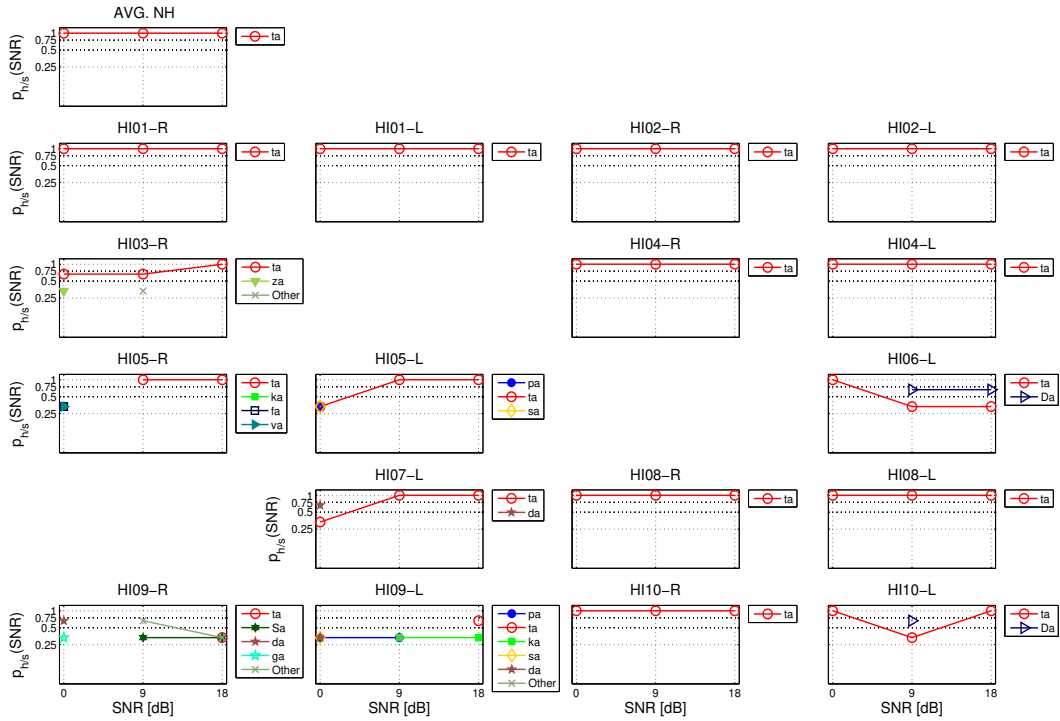


Figure A.43: f119 ta confusions patterns when primary cues are amplified by 6 [dB].

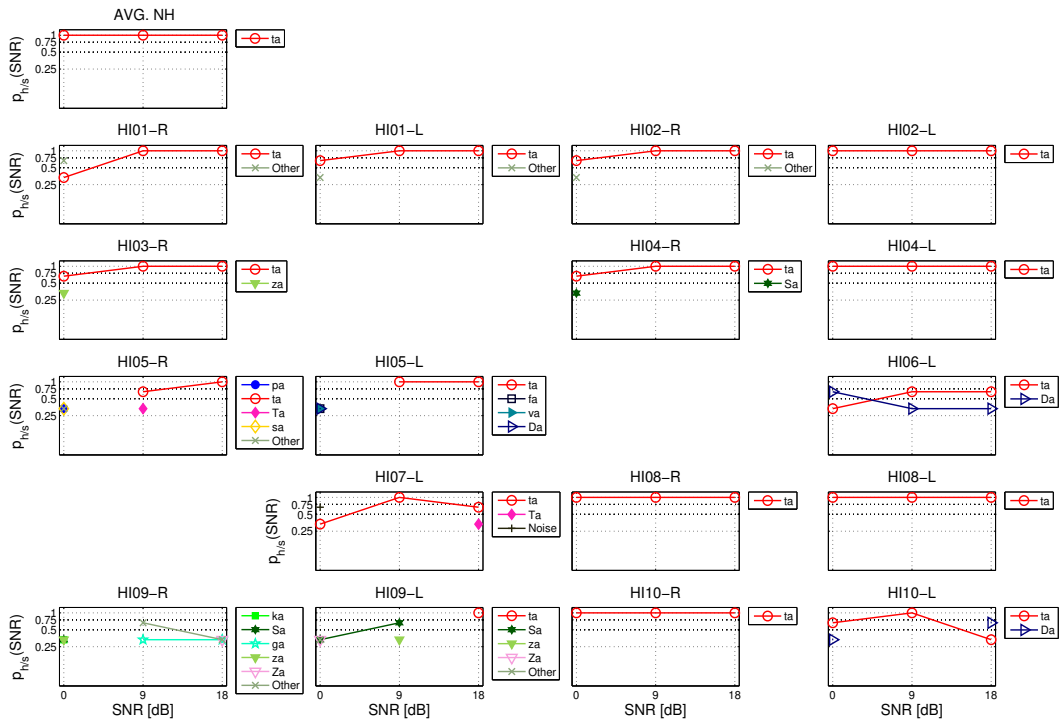


Figure A.44: f119 ta confusions patterns when primary cues are amplified by 12 [dB].

f105da

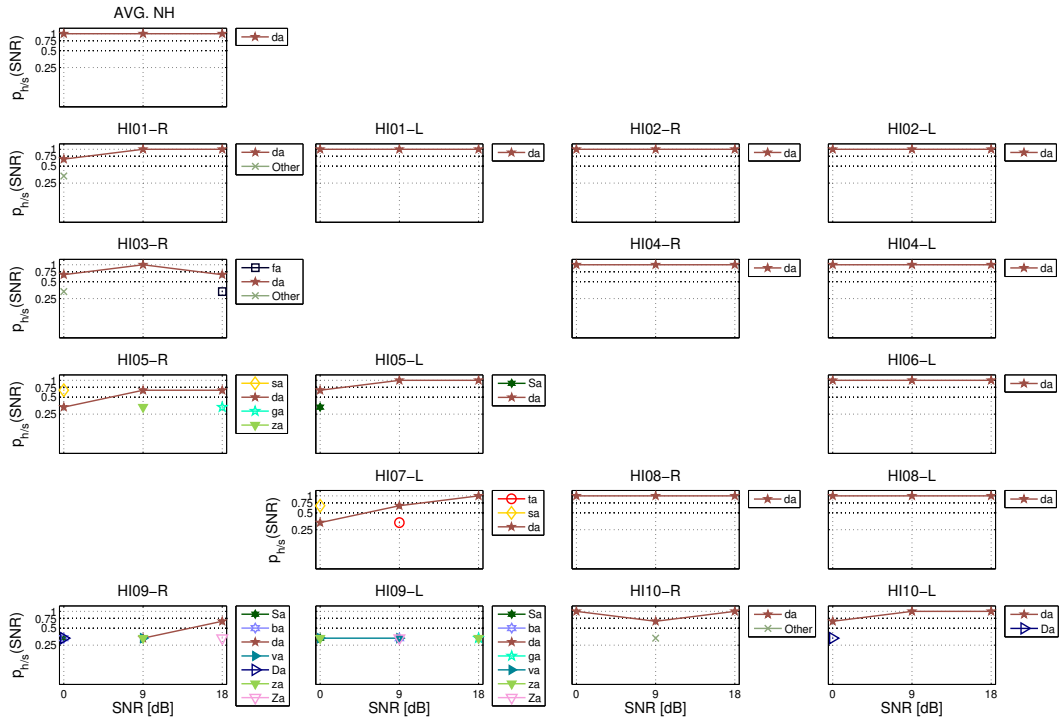


Figure A.45: f105 da confusions patterns when primary cues are unmodified.

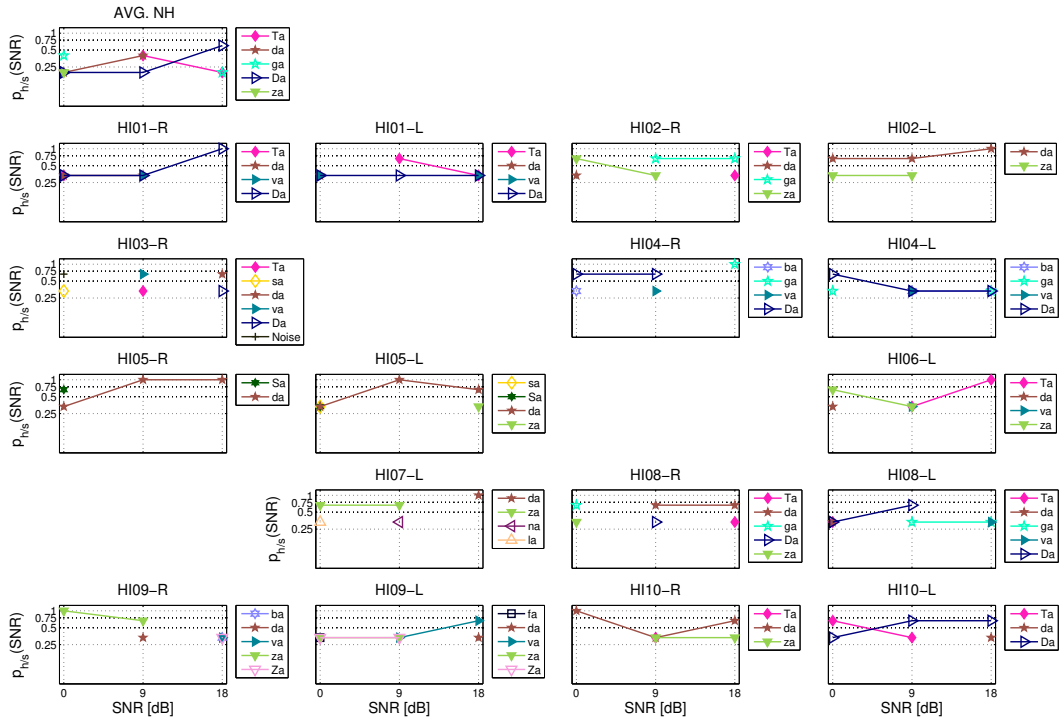


Figure A.46: f105 da confusions patterns when primary cues are removed.

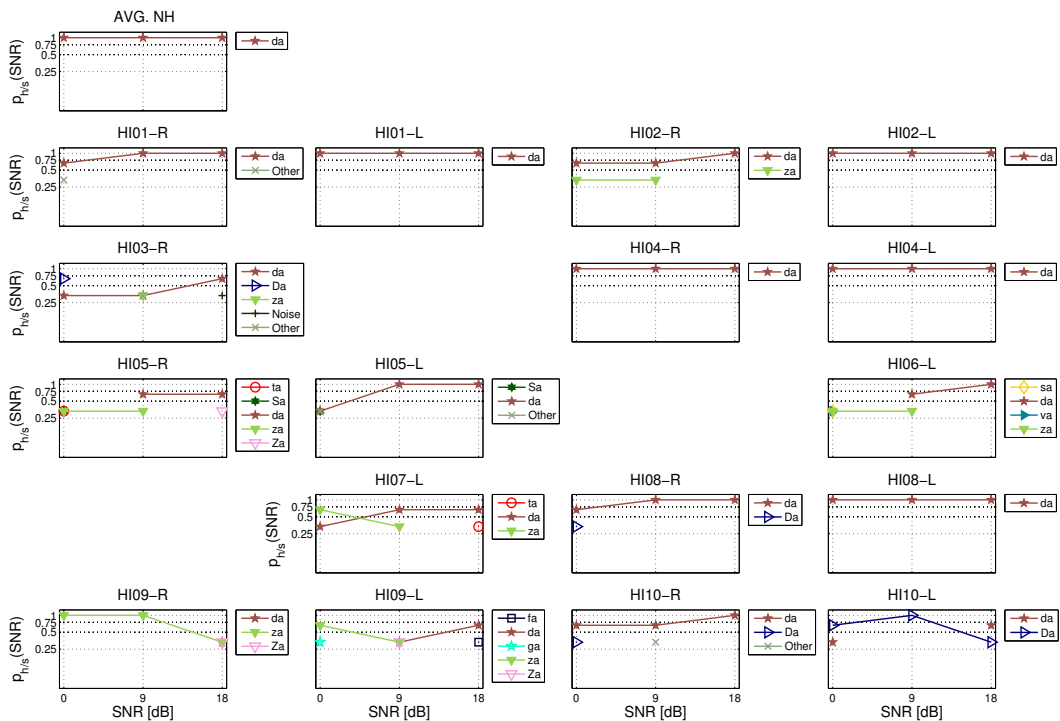


Figure A.47: f105 da confusions patterns when primary cues are attenuated by -6 [dB].

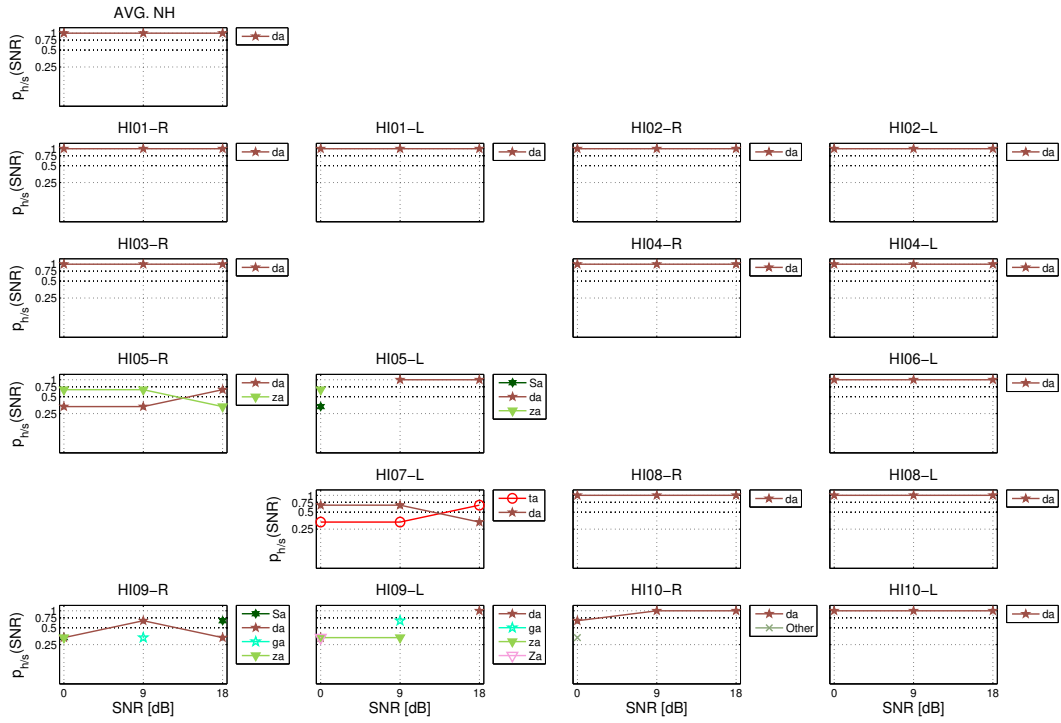


Figure A.48: f105 da confusions patterns when primary cues are amplified by 6 [dB].

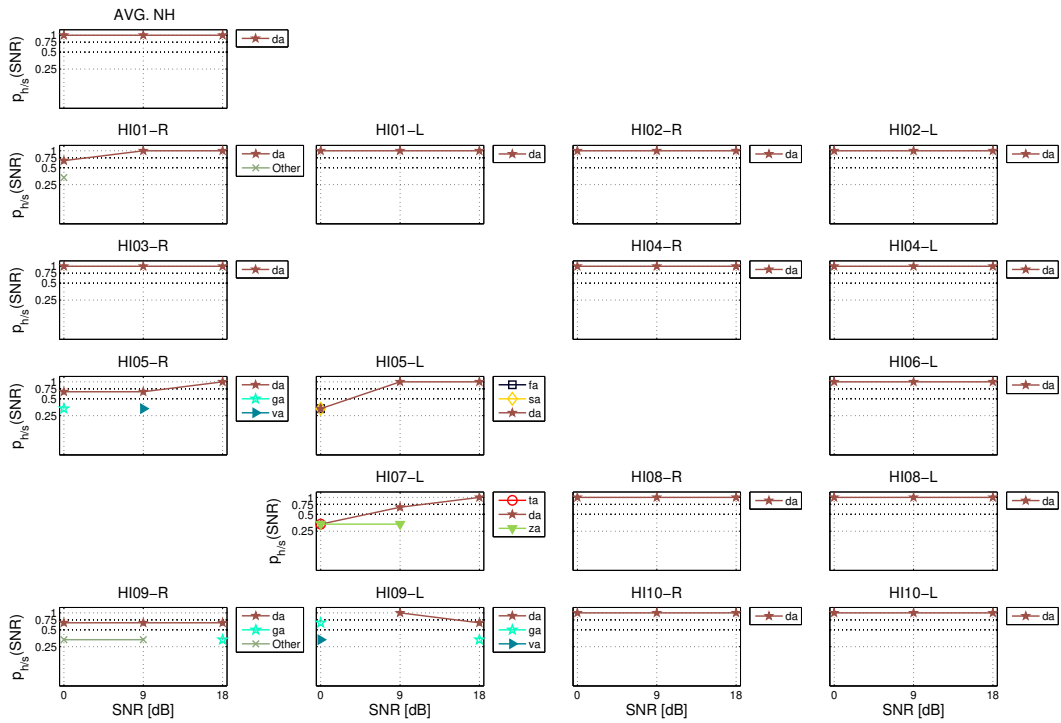


Figure A.49: f105 da confusions patterns when primary cues are amplified by 12 [dB].

f119da

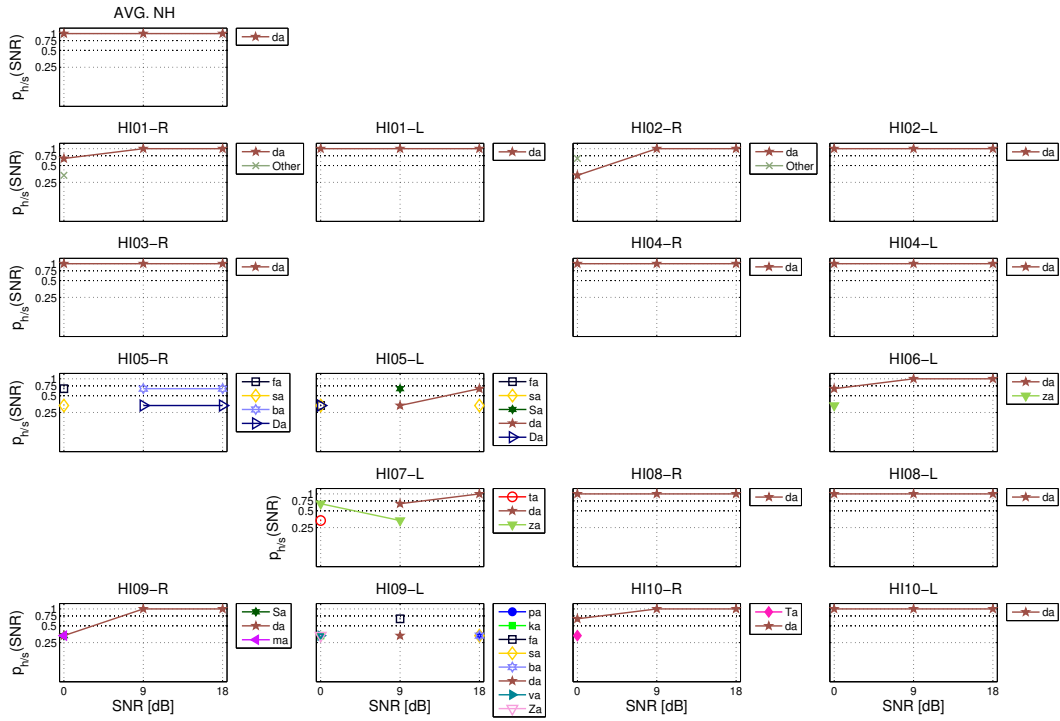


Figure A.50: f119 da confusions patterns when primary cues are unmodified.

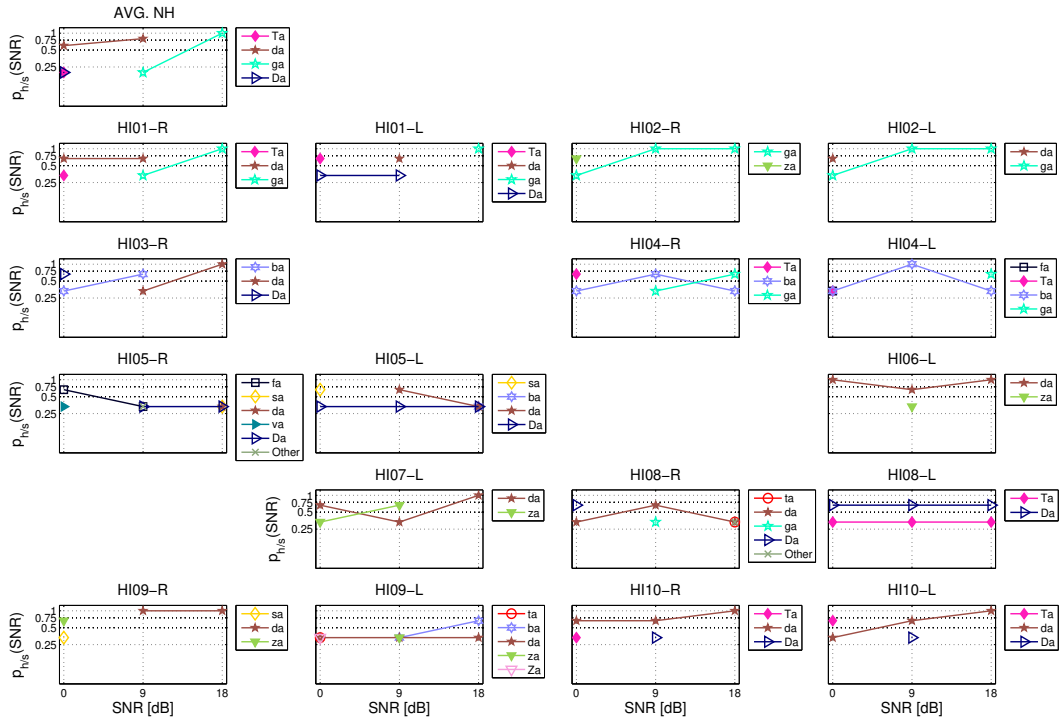


Figure A.51: f119 da confusions patterns when primary cues are removed.

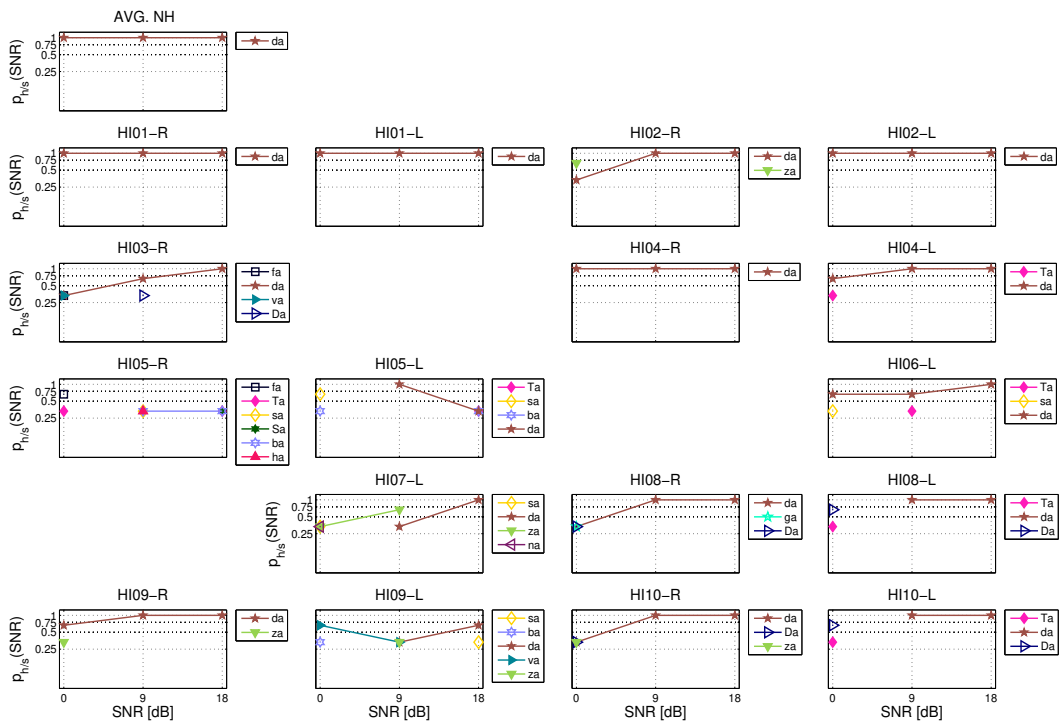


Figure A.52: f119 da confusions patterns when primary cues are attenuated by -6 [dB].

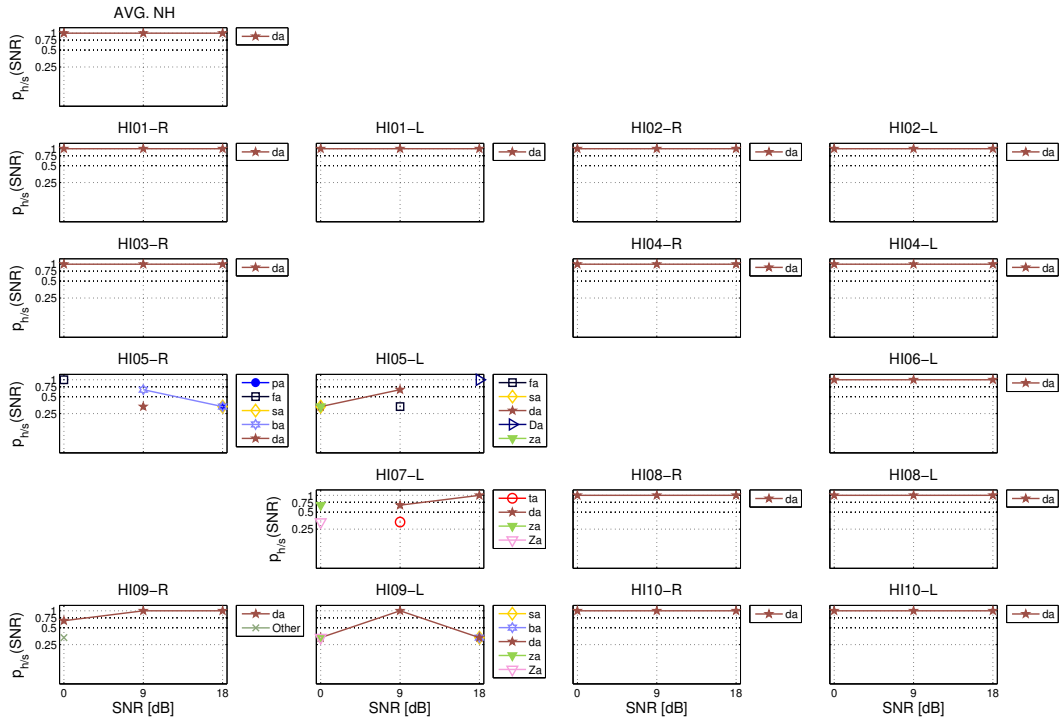


Figure A.53: f119 da confusions patterns when primary cues are amplified by 6 [dB].

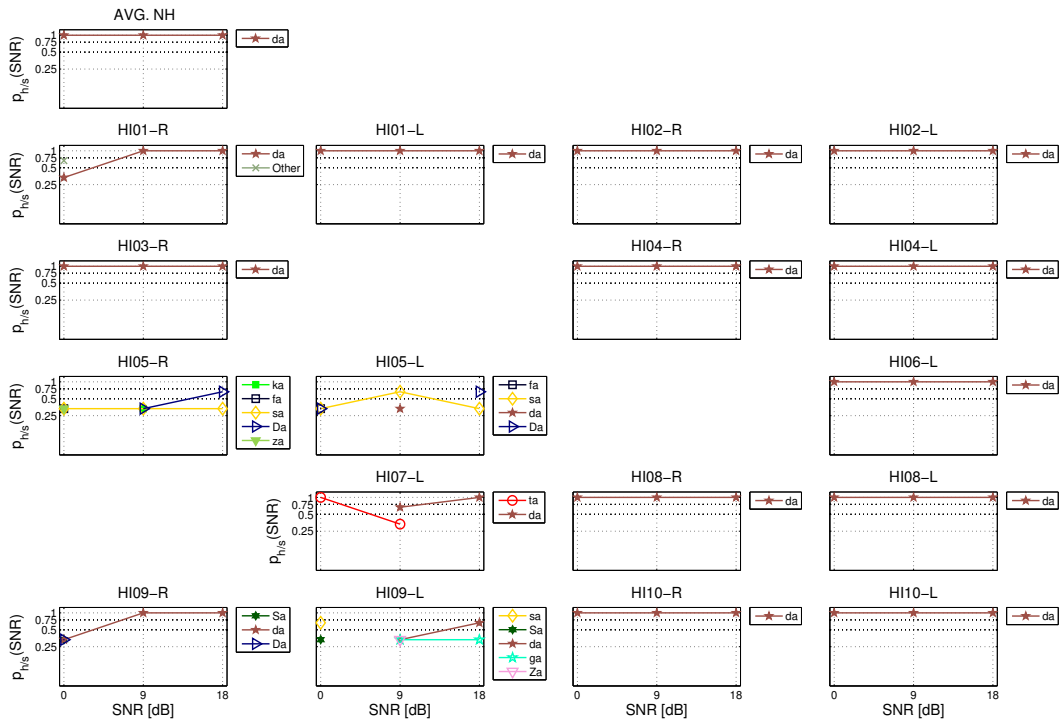


Figure A.54: f119 da confusions patterns when primary cues are amplified by 12 [dB].

APPENDIX B

EXPERIMENT II RESULTS

B.1 Modifications

Figure B.1 shows the AI-gram eight tokens for Experiment II. The modification parameters for the primary cue and conflicting cues of each token are highlighted in yellow and pink, respectively.

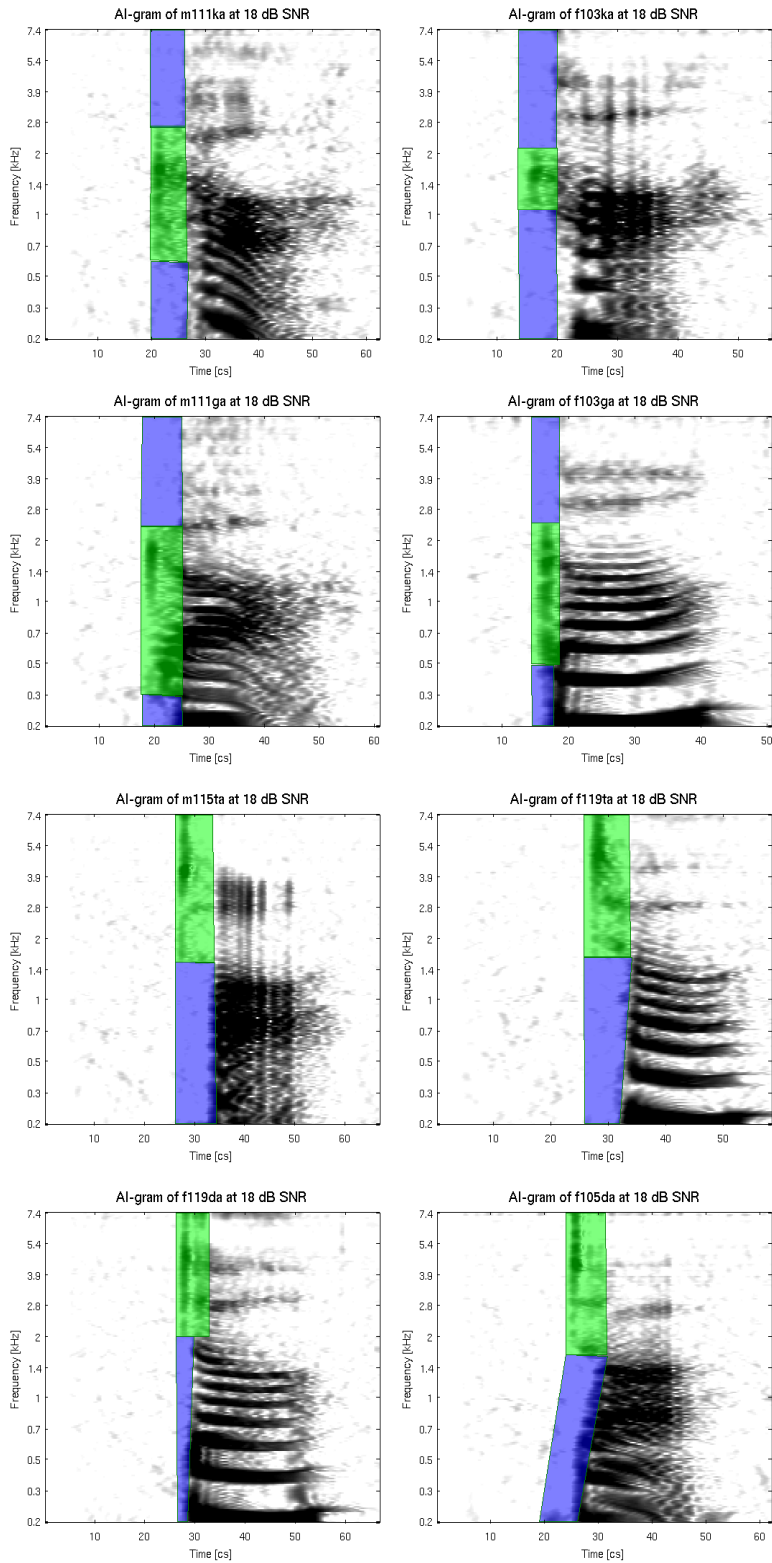


Figure B.1: Modification on the primary cue for 8 tokens. The highlighted green is the modification parameter of the primary cue. The highlighted purple is the parameters for the removal of the conflicting cues.

B.2 Sorted Errors

Figures B.2 - B.6 show the sorted error for eight tokens when the conflicting cues are removed and the primary cue was unmodified, removed, attenuated -6 dB, amplified by 6 dB and 12 dB for the average NH ears and 17 individual HI ears. Each plot is for a specified modification condition.

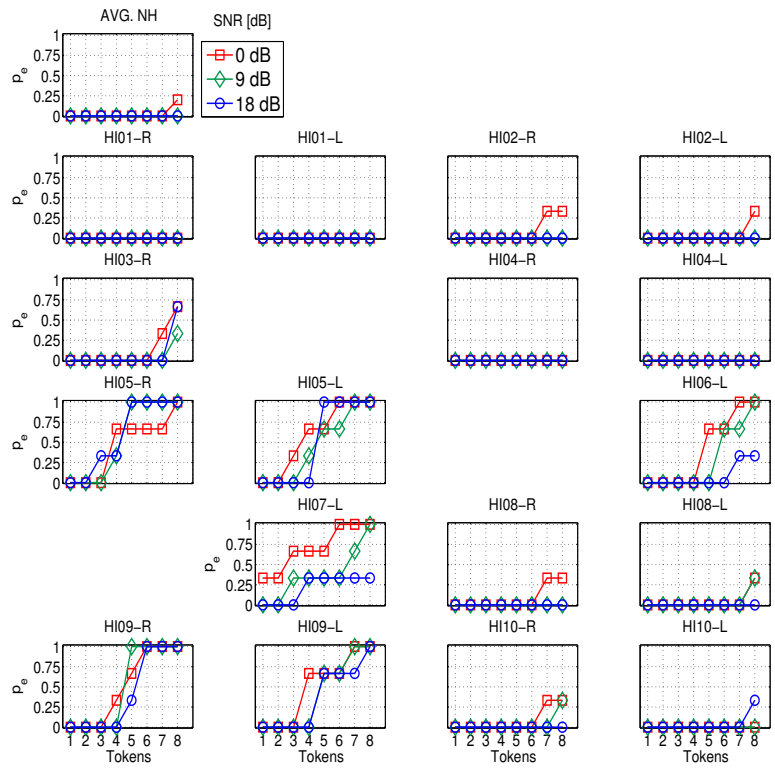


Figure B.2: The sorted errors when primary cues are attenuated are unmodified.

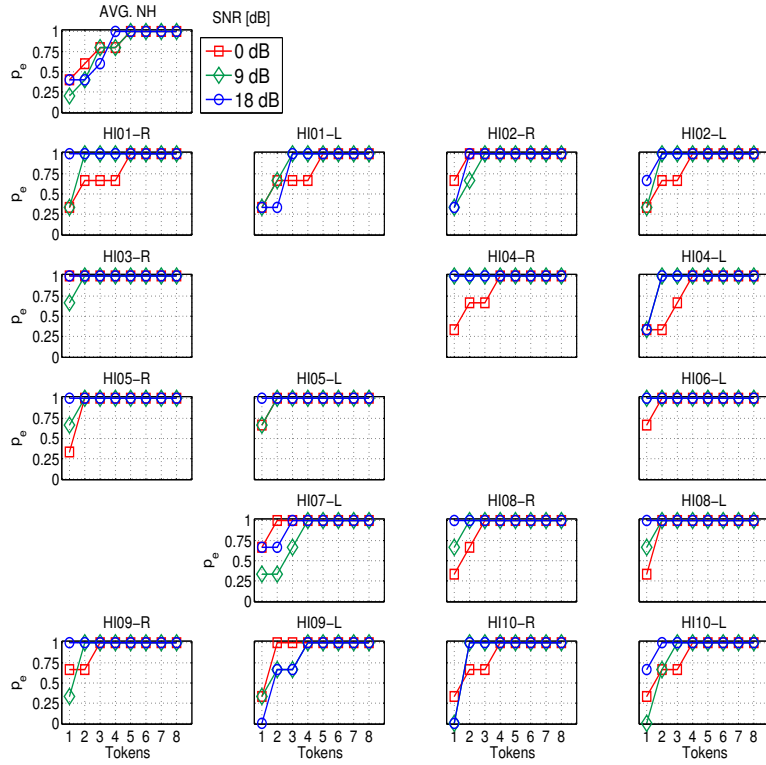


Figure B.3: Experiment I: The sorted errors when primary cues are removed.

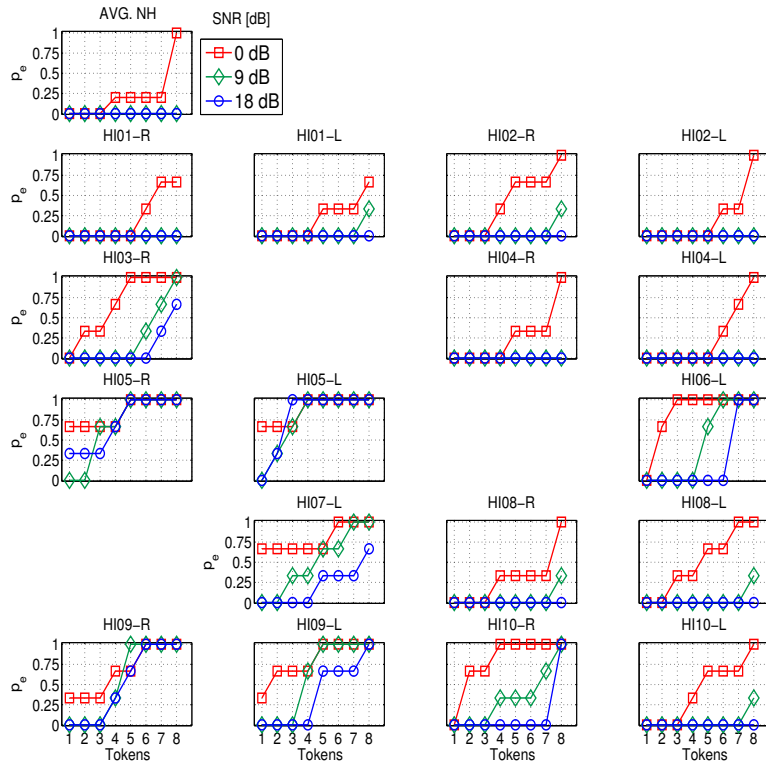


Figure B.4: The sorted errors when primary cues are attenuated by -6[dB].

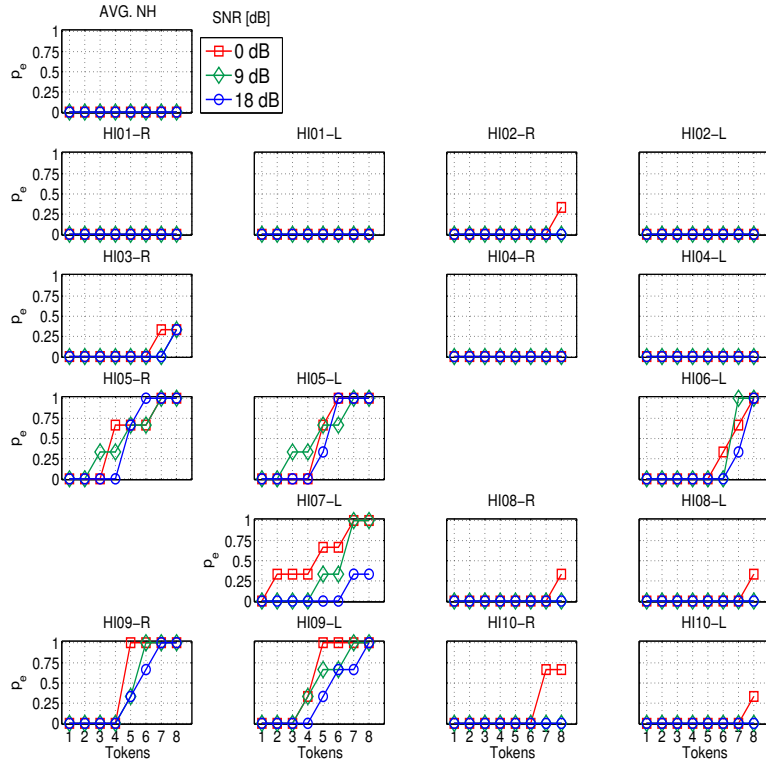


Figure B.5: The sorted errors when primary cues are amplified by 6[dB].

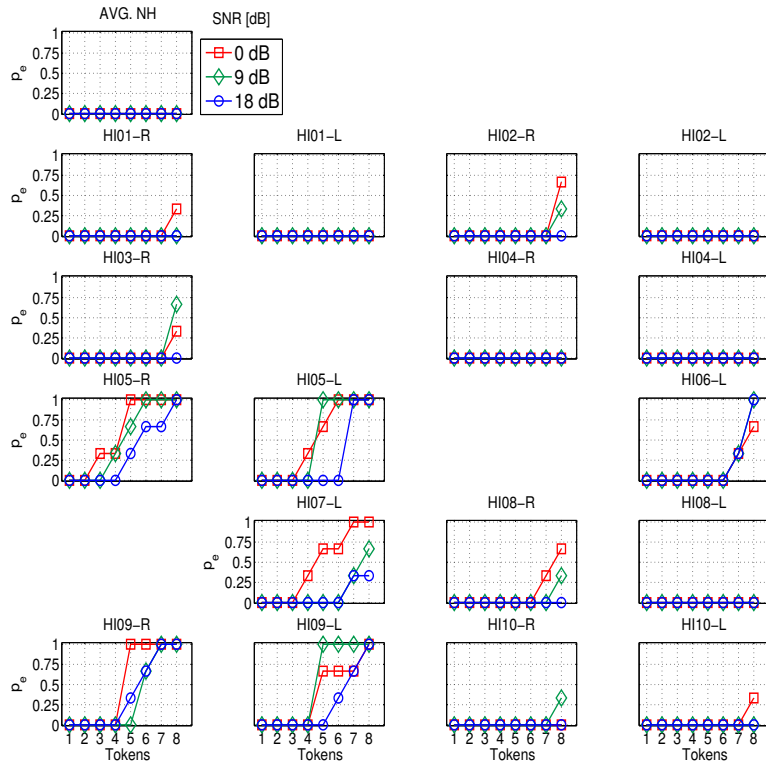


Figure B.6: The sorted errors when primary cues are amplified by 12[dB].

B.3 Average Token Error Analysis

Figures B.7 - B.14 show the average token error for eight tokens when the primary cue was unmodified, removed, attenuated -6 dB, amplified by 6 dB and 12 dB for the average NH ears and 17 individual HI ears. Each plot is for a specified token (i.e. m111 /ka/, f103 /ka/, m111 /ga/, f103 /ga/, m115 /ta/, f119 /ta/ f105 /da/, f119 /da/).

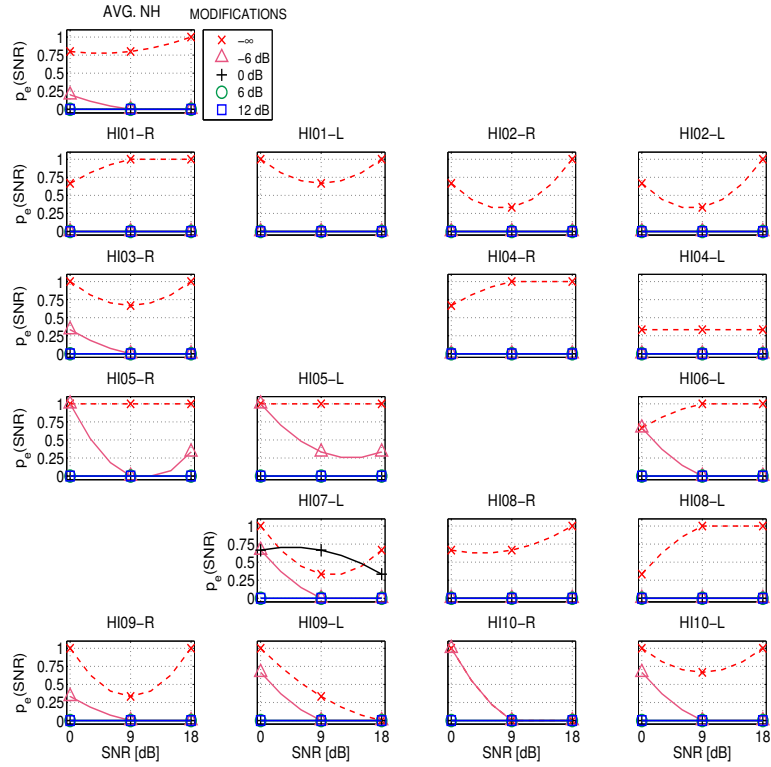


Figure B.7: m111 ka primary cues modifications compared.

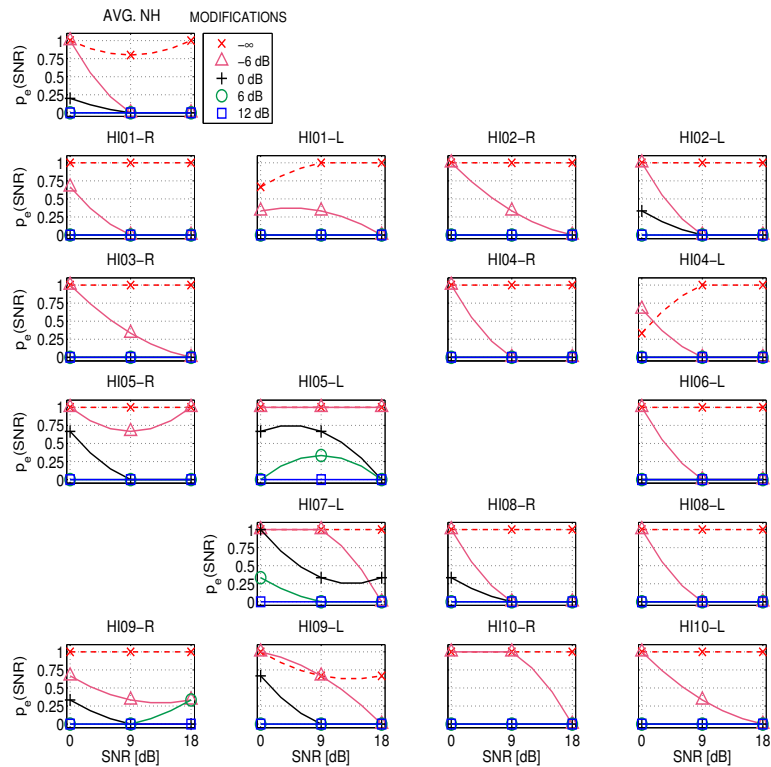


Figure B.8: f103 ka primary cues modifications compared.

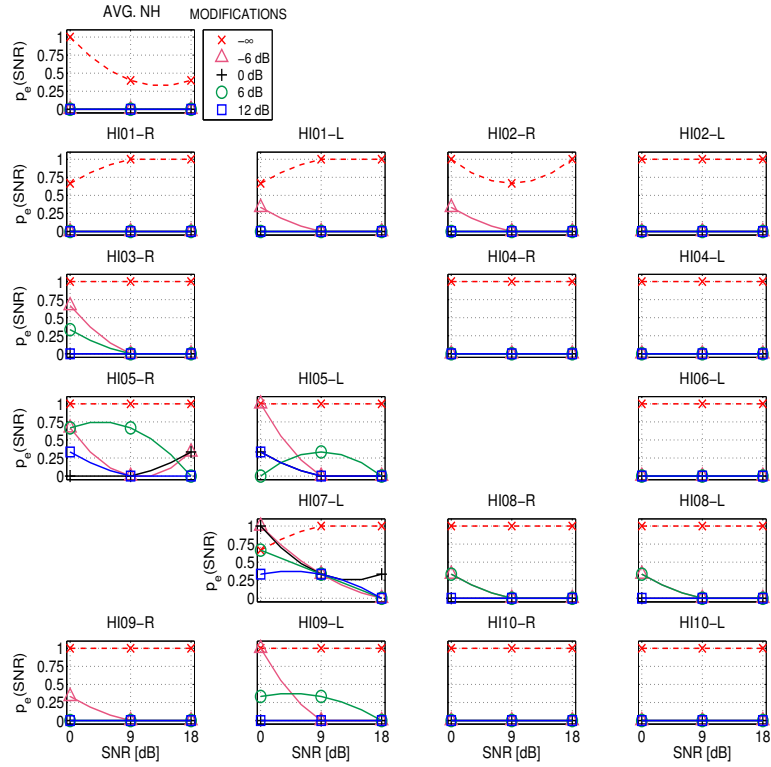


Figure B.9: m111 ga primary cues modifications compared.

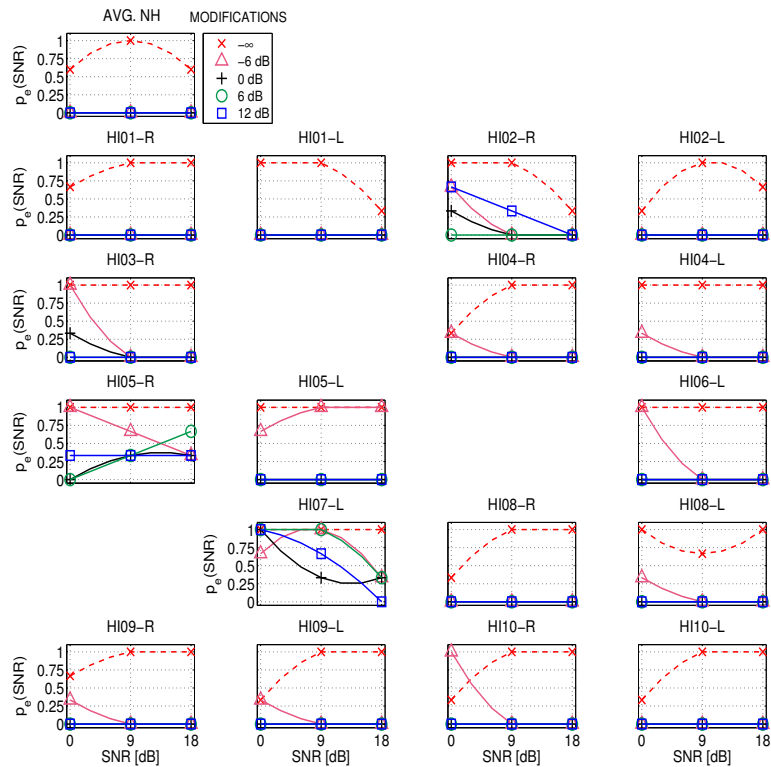


Figure B.10: f103 ga primary cues modifications compared.

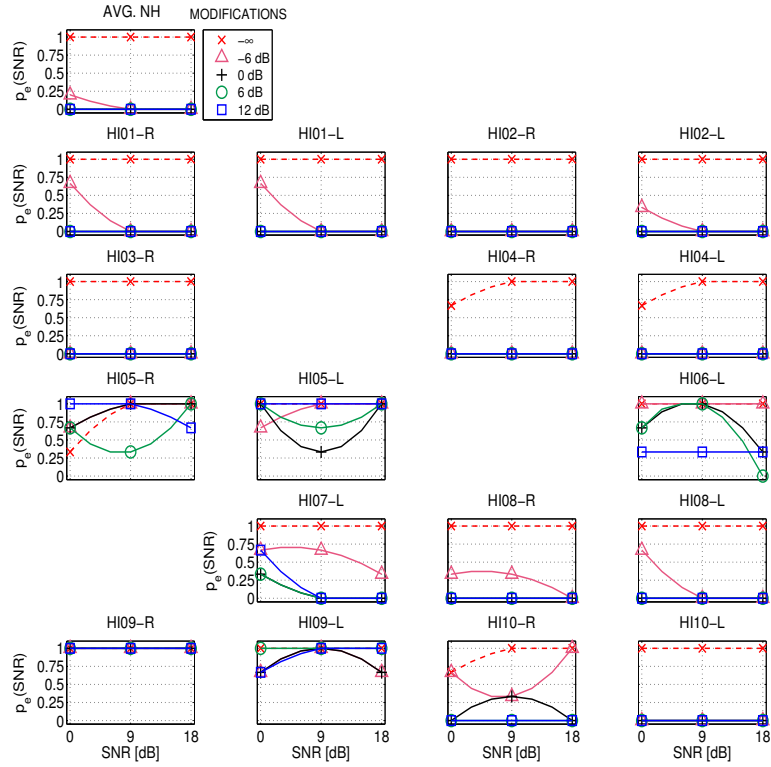


Figure B.11: m115 to primary cues modifications compared.

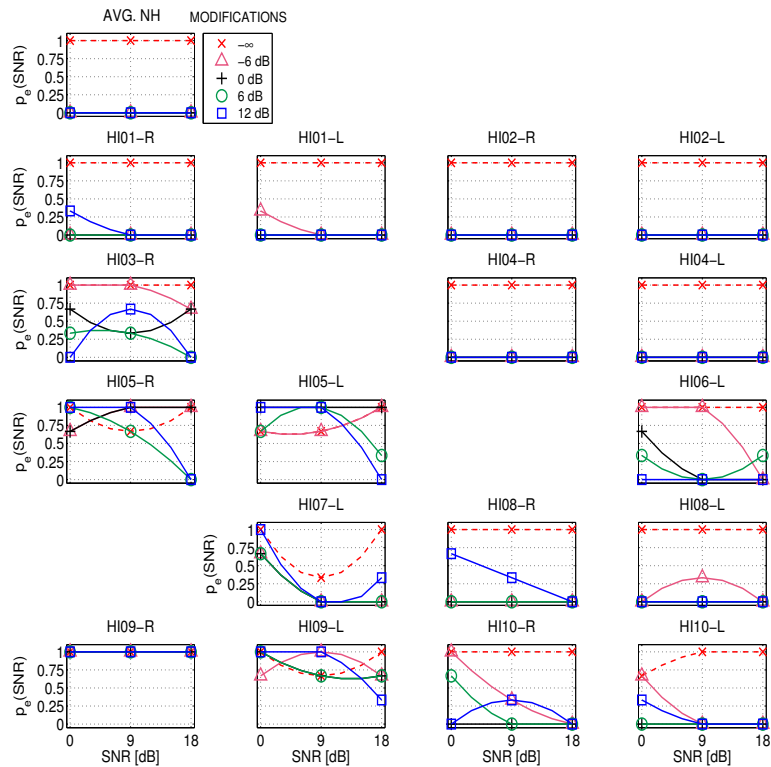


Figure B.12: f119 to primary cues modifications compared.

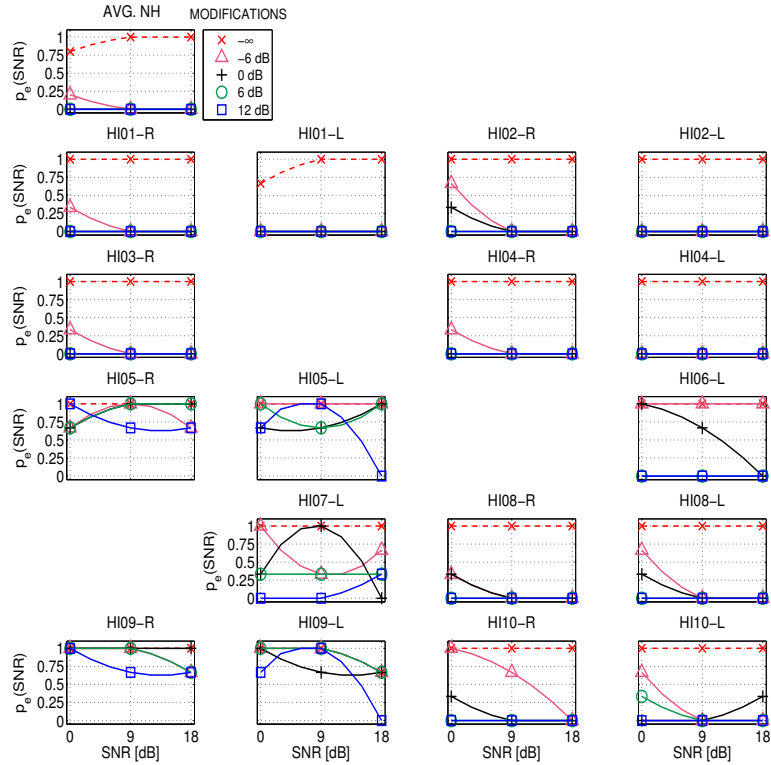


Figure B.13: f105 da primary cues modifications compared.

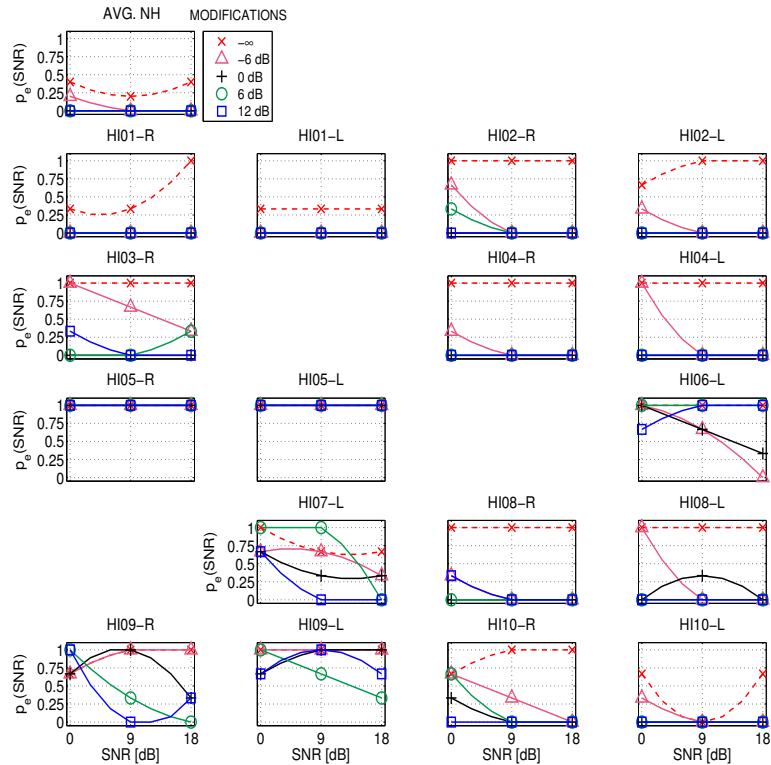


Figure B.14: f119 da primary cues modifications compared.

B.4 Confusion Patterns

Figures B.15 - B.54 show the confusion patterns of a token for the average NH ears and 17 individual HI ears. Each plot is a specified token (i.e. m111 /ka/, f103 /ka/, m111 /ga/, f103 /ga/, m115 /ta/, f119 /ta/ f105 /da/, f119 /kda/) and modification condition (i.e. primary cue was unmodified, removed, attenuated -6 dB, multiplied by 6 dB and 12 dB).

m111ka

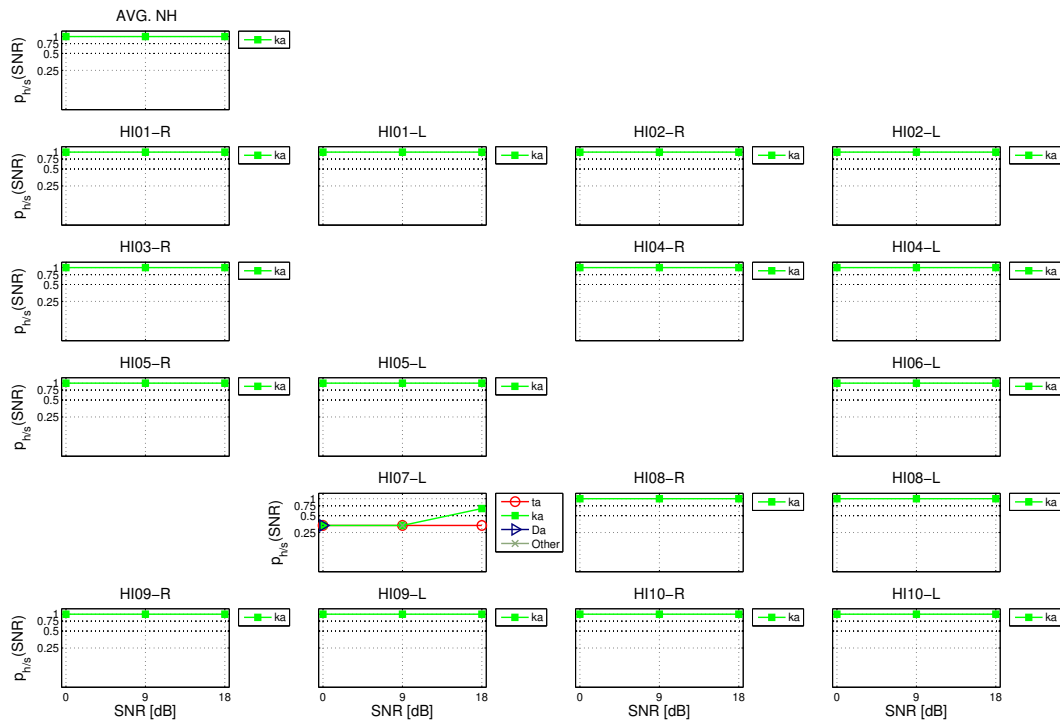


Figure B.15: m111 ka confusions patterns when primary cues are unmodified.

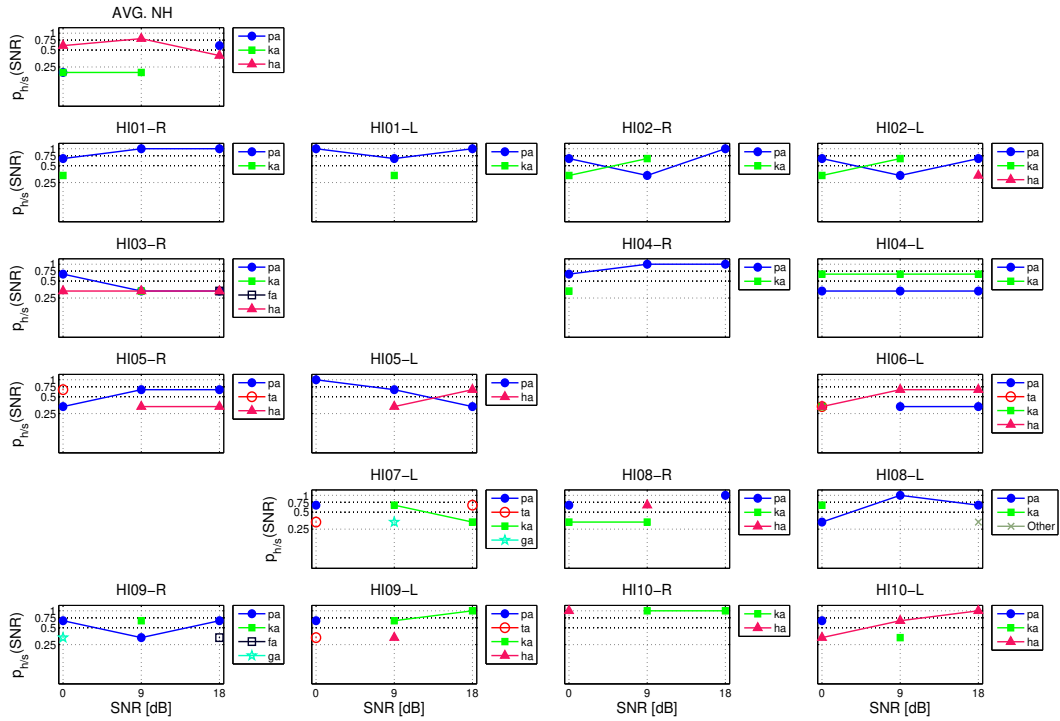


Figure B.16: m111 ka confusions patterns when primary cues are removed.

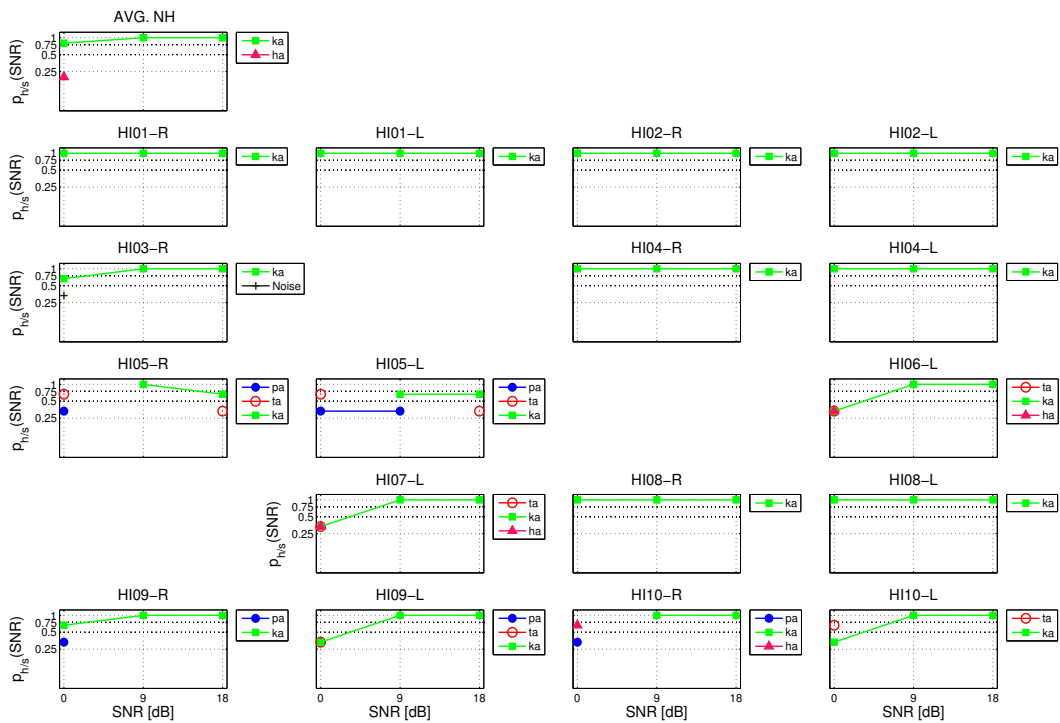


Figure B.17: m111 ka confusions patterns when primary cues are attenuated by -6 [dB].

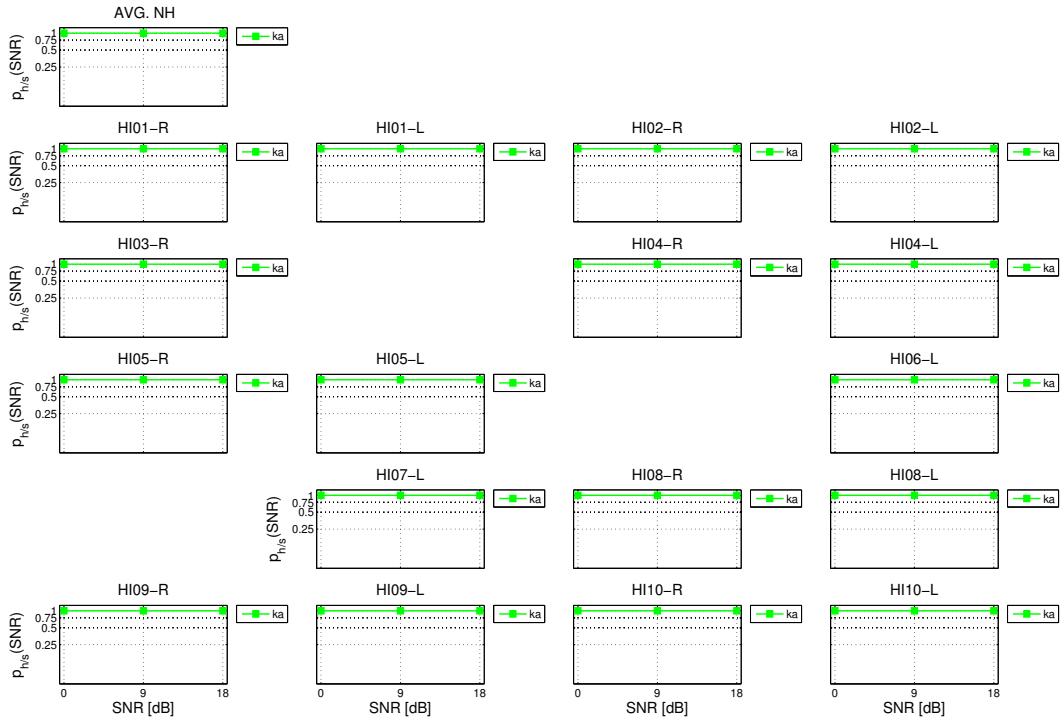


Figure B.18: m111 ka confusions patterns when primary cues are amplified by 6 [dB].

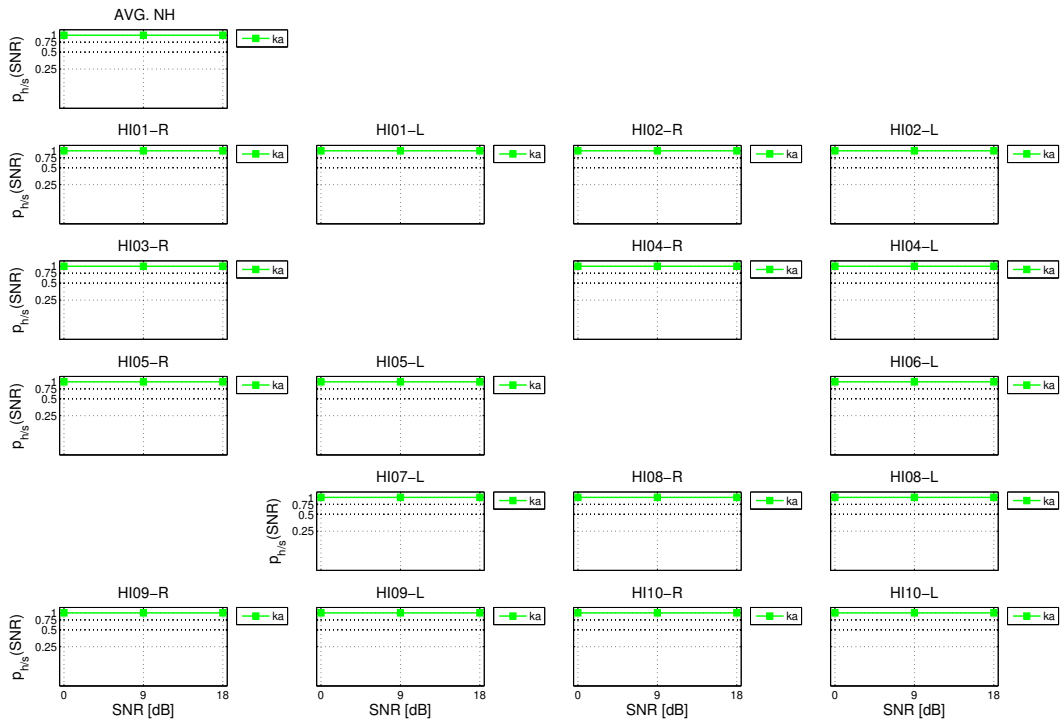


Figure B.19: m111 ka confusions patterns when primary cues are amplified by 12 [dB].

f103ka

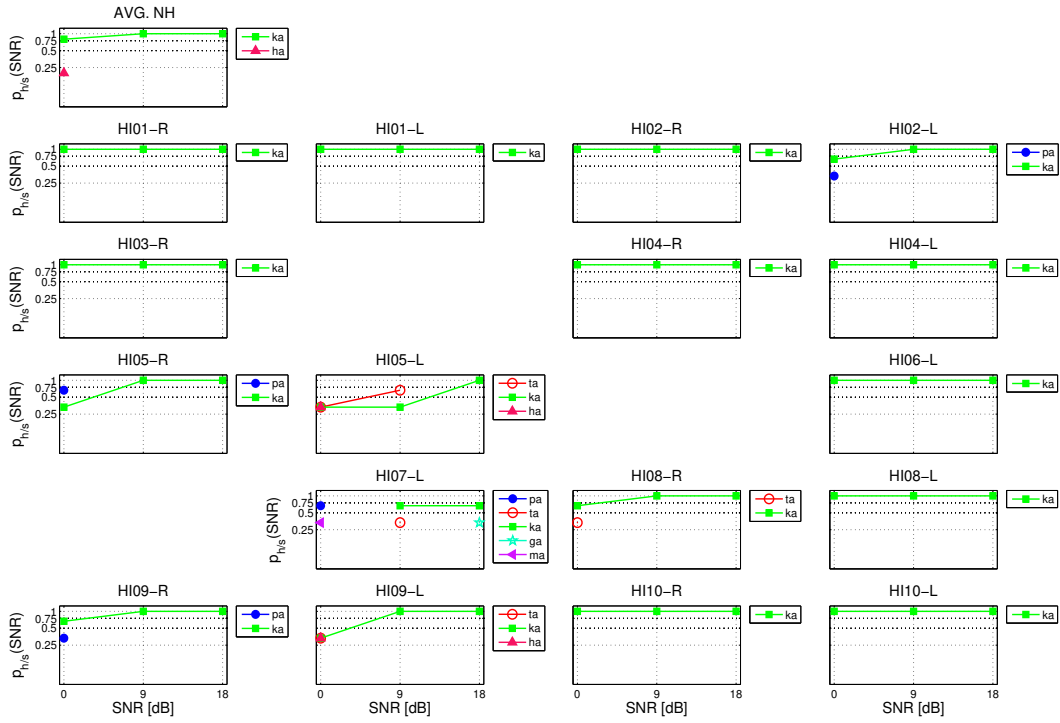


Figure B.20: f103 ka confusions patterns when primary cues are unmodified.

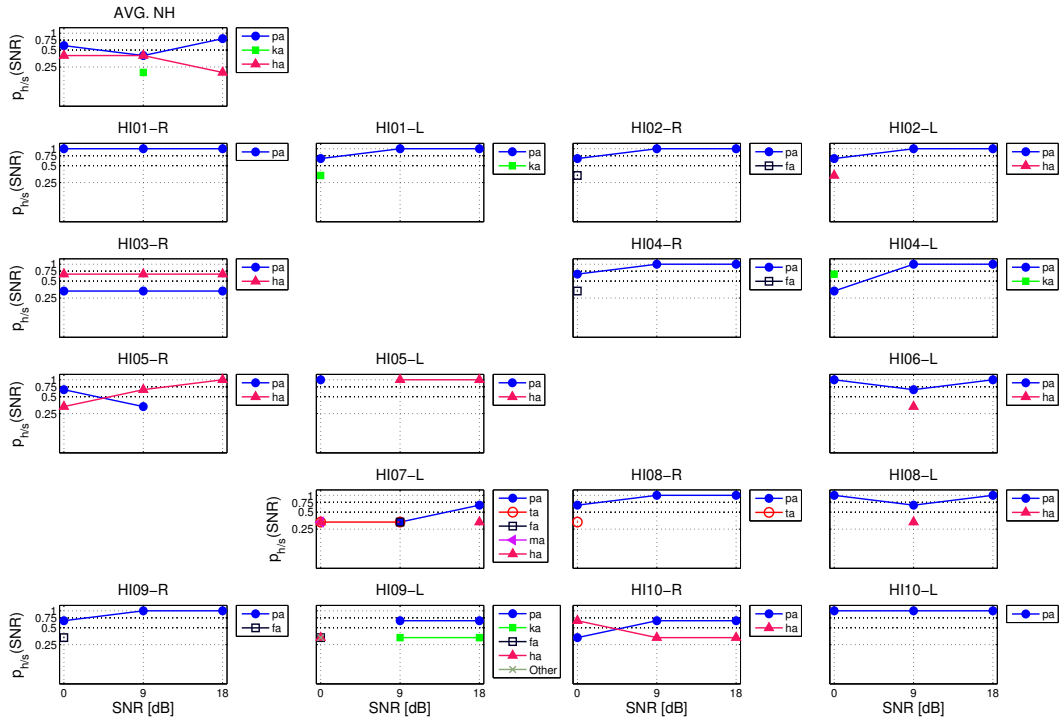


Figure B.21: f103 ka confusions patterns when primary cues are removed.

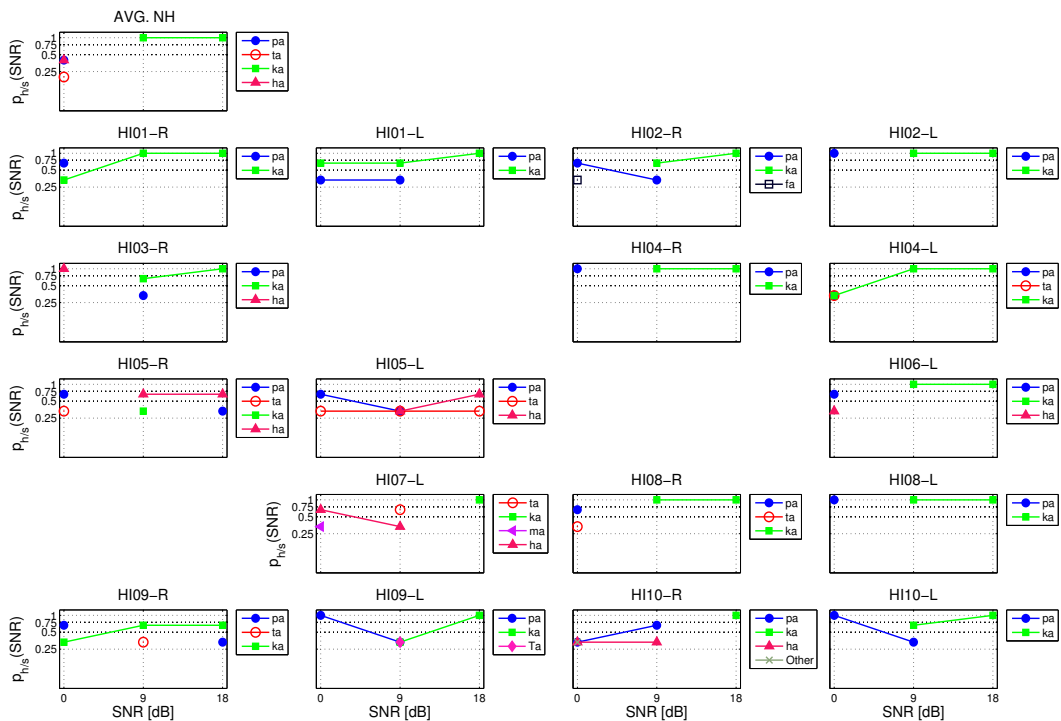


Figure B.22: f103 ka confusions patterns when primary cues are attenuated by -6 [dB].

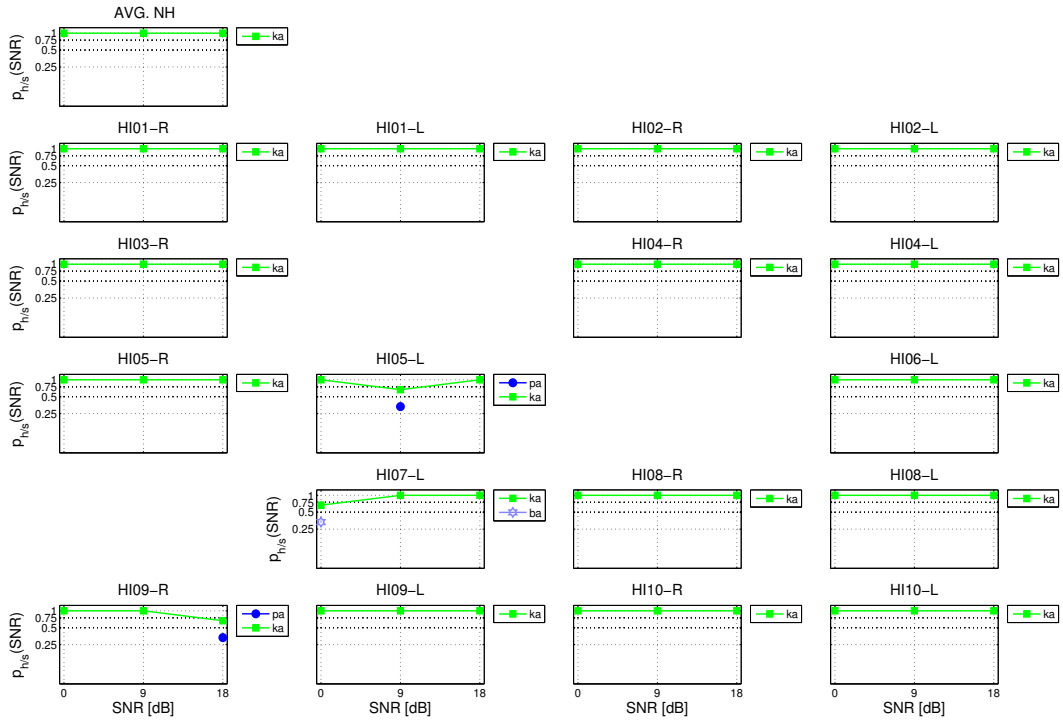


Figure B.23: f103 ka confusions patterns when primary cues are amplified by 6 [dB].

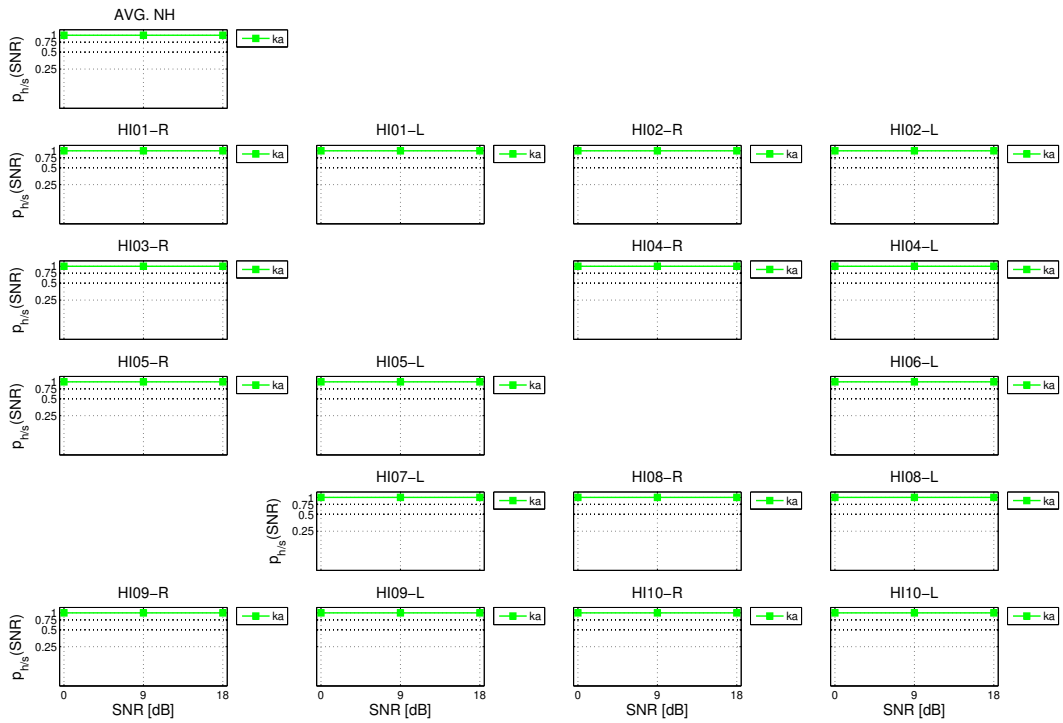


Figure B.24: f103 ka confusions patterns when primary cues are amplified by 12 [dB].

m111ga

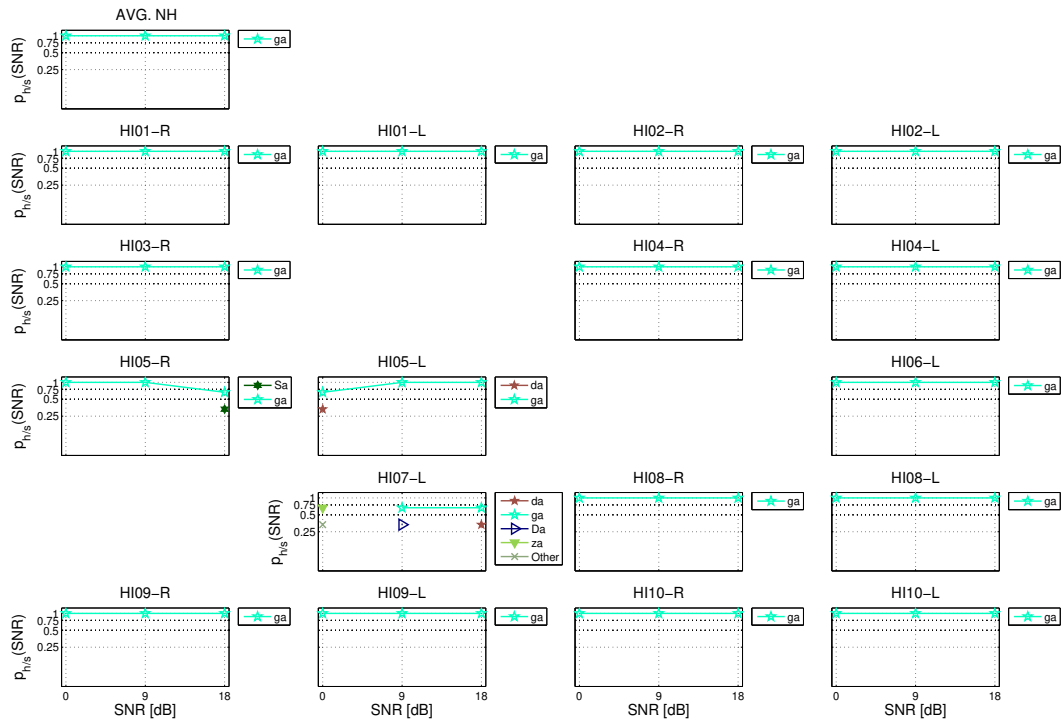


Figure B.25: m111 ga confusions patterns when primary cues are unmodified.

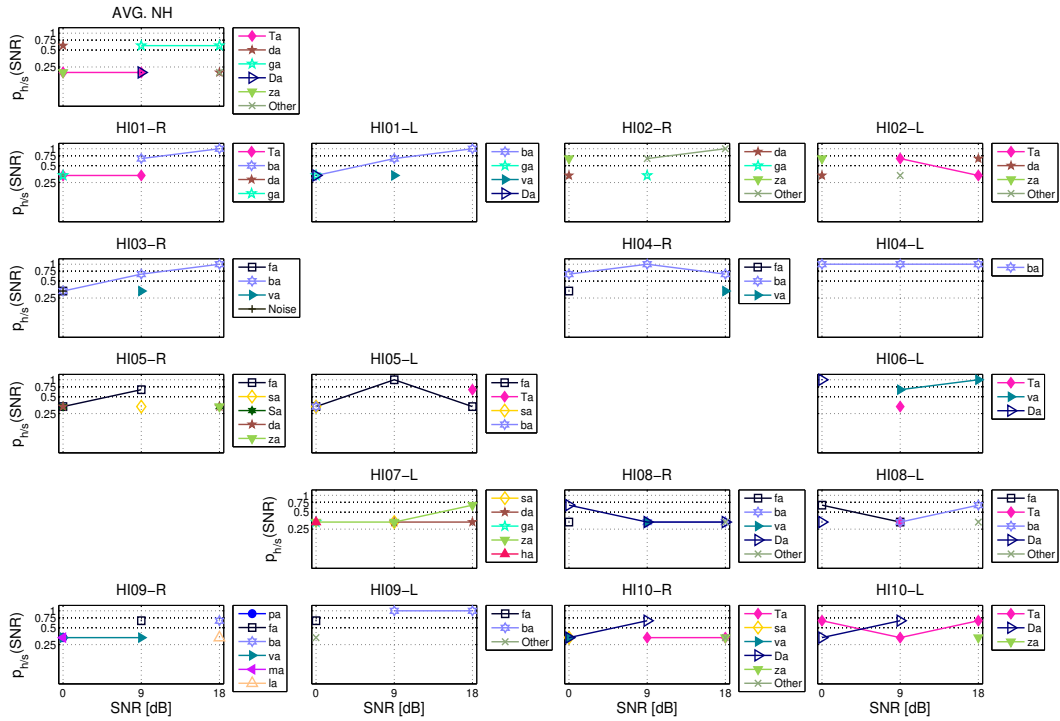


Figure B.26: m111 ga confusions patterns when primary cues are removed.

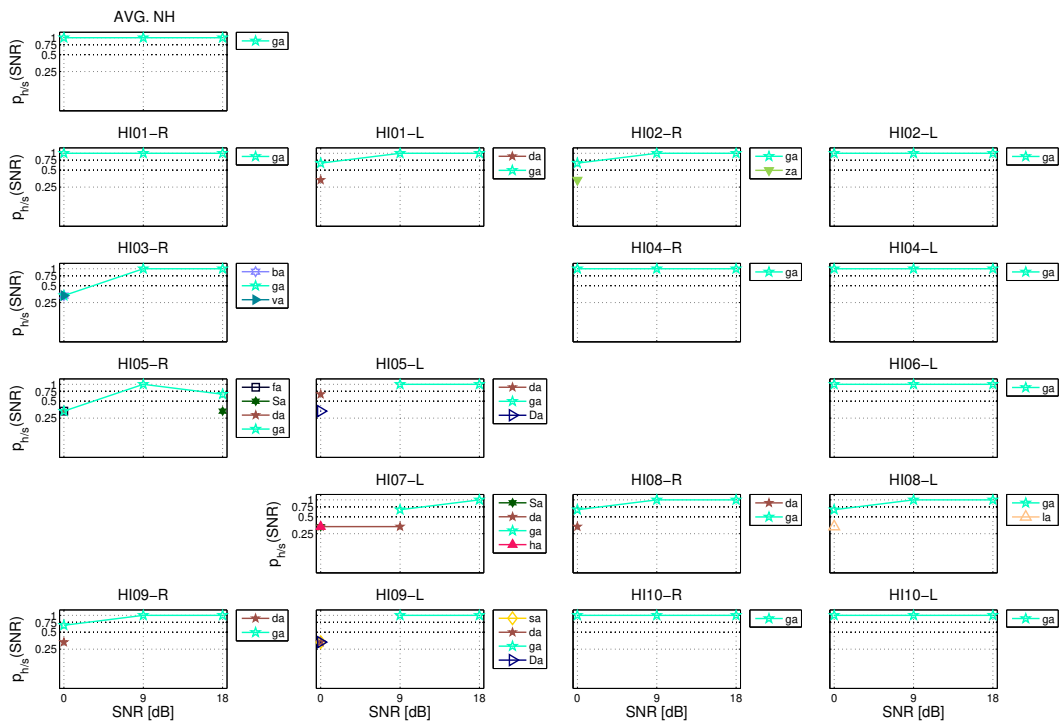


Figure B.27: m111 ga confusions patterns when primary cues are attenuated by -6 [dB].

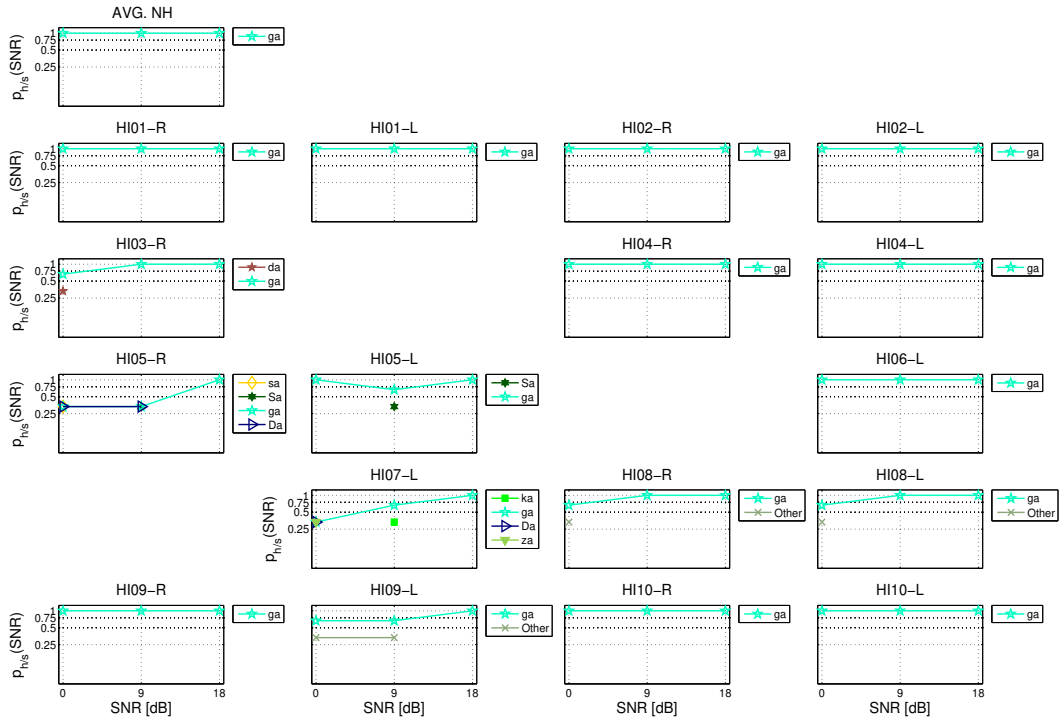


Figure B.28: m111 ga confusions patterns when primary cues are amplified by 6 [dB].

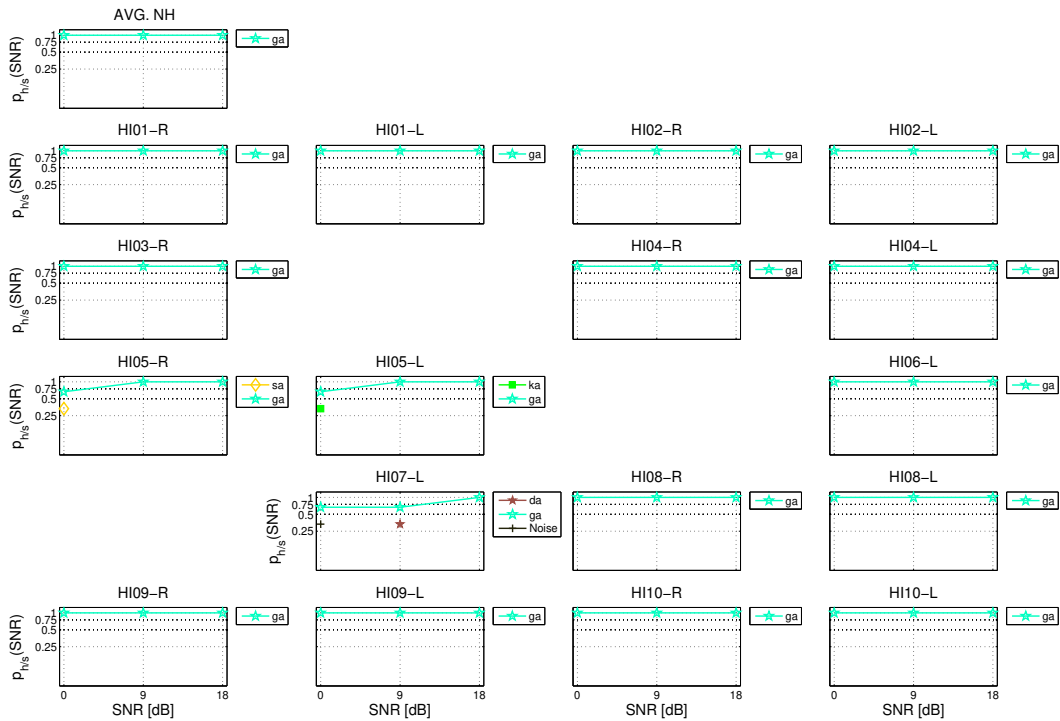


Figure B.29: m111 ga confusions patterns when primary cues are amplified 12 [dB].

f103ga

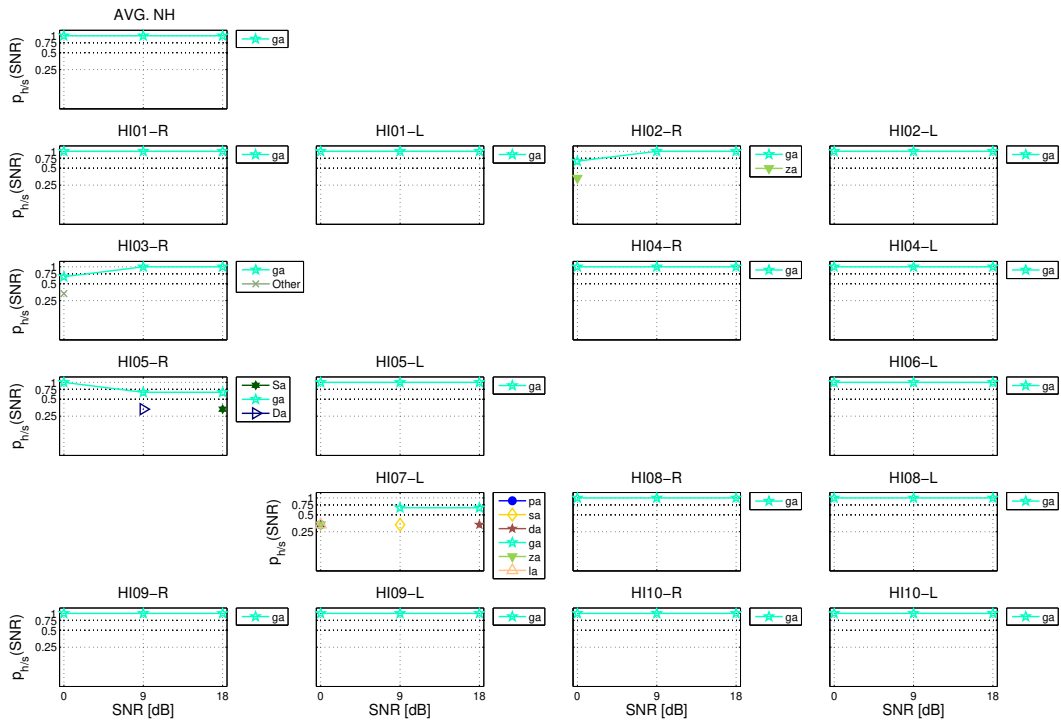


Figure B.30: f103 ga confusions patterns when primary cues are unmodified.

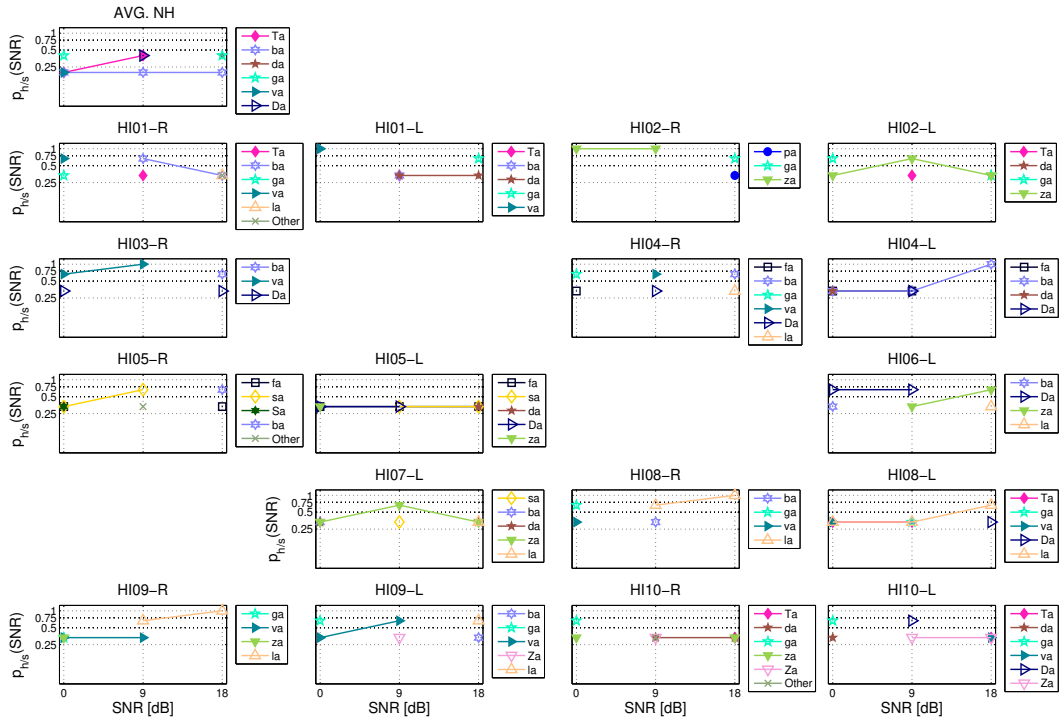


Figure B.31: f103 ga confusions patterns when primary cues are removed.

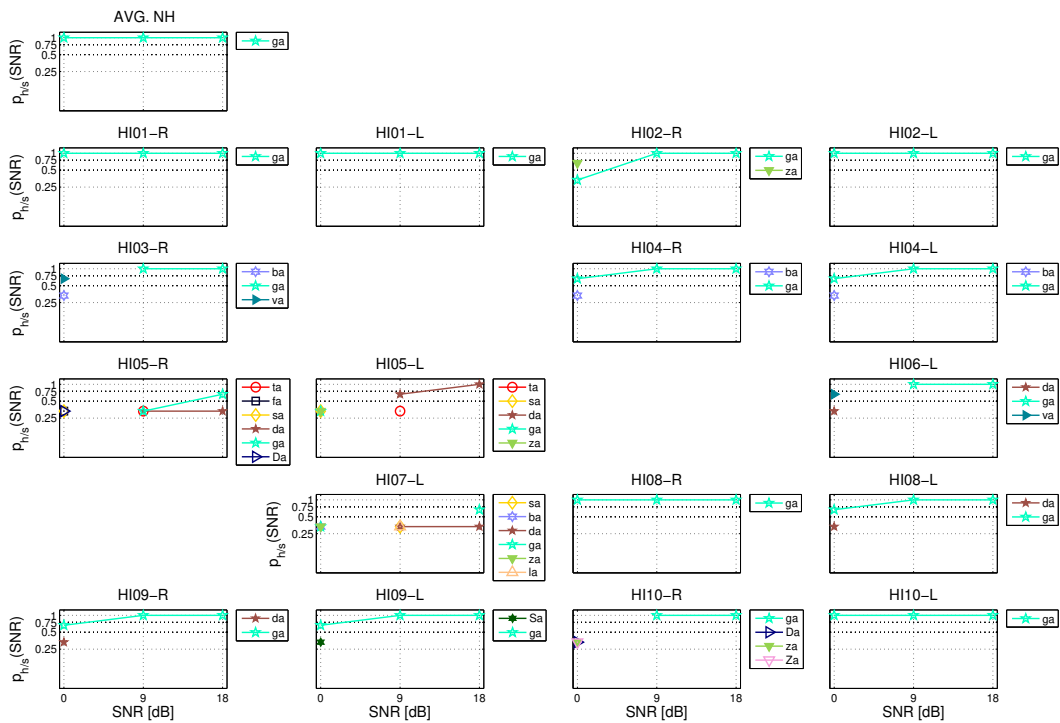


Figure B.32: f103 ga confusions patterns when primary cues are attenuated by -6 [dB].

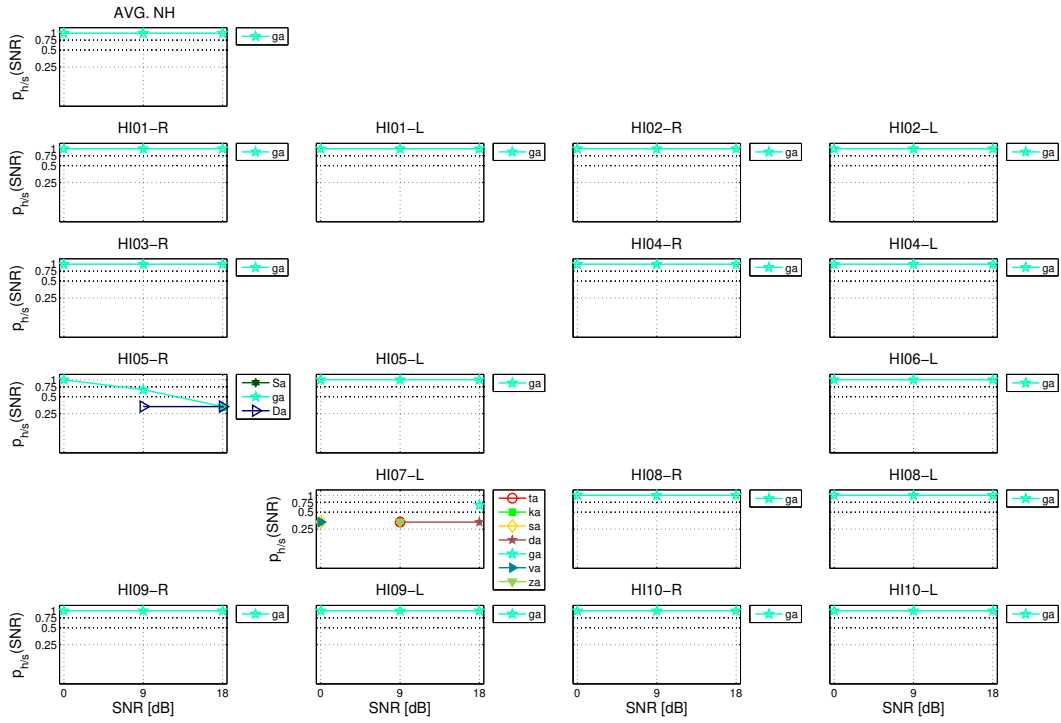


Figure B.33: f103 ga confusions patterns when primary cues are amplified by 6 [dB].

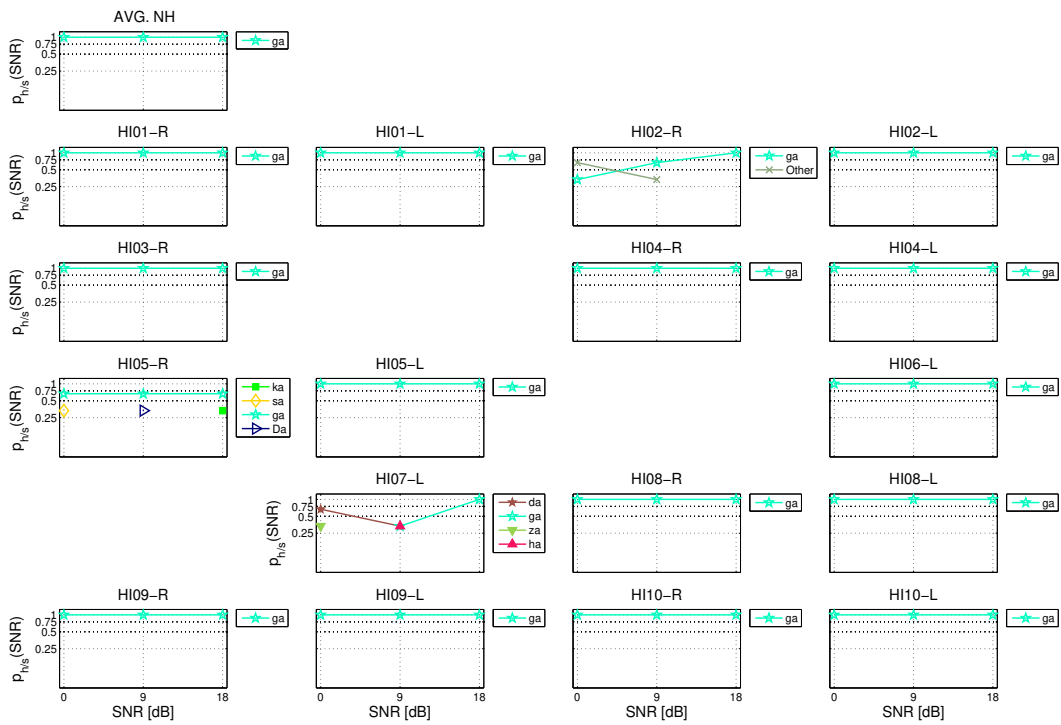


Figure B.34: f103 ga confusions patterns when primary cues are amplified by 12 [dB].

m115ta

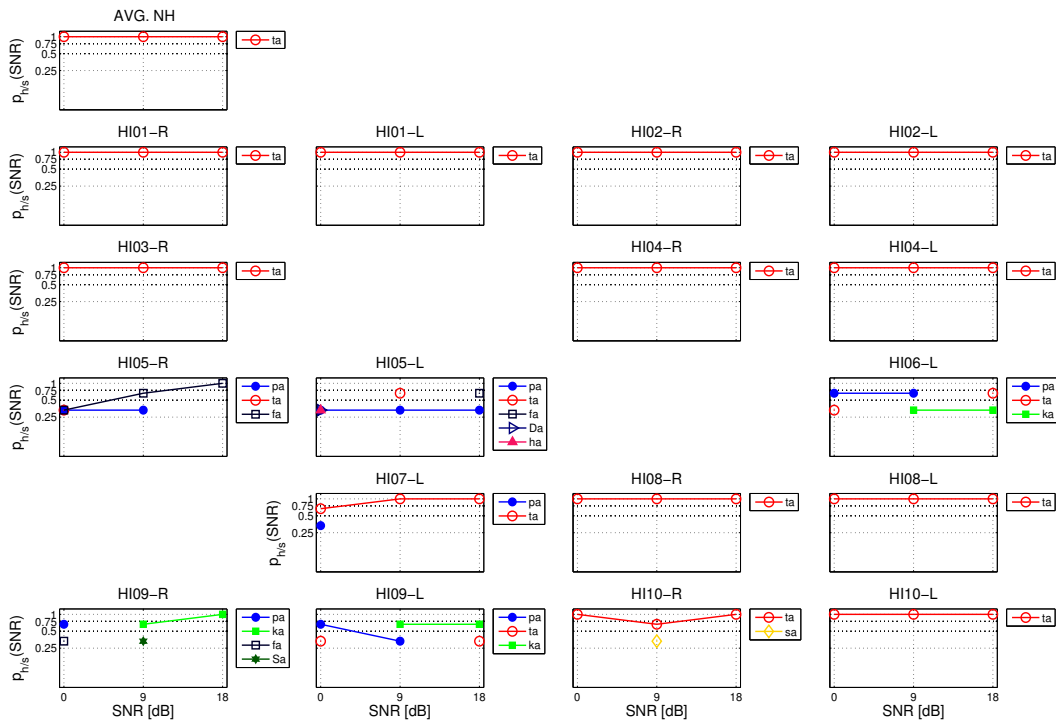


Figure B.35: m115 ta confusions patterns when primary cues are unmodified.

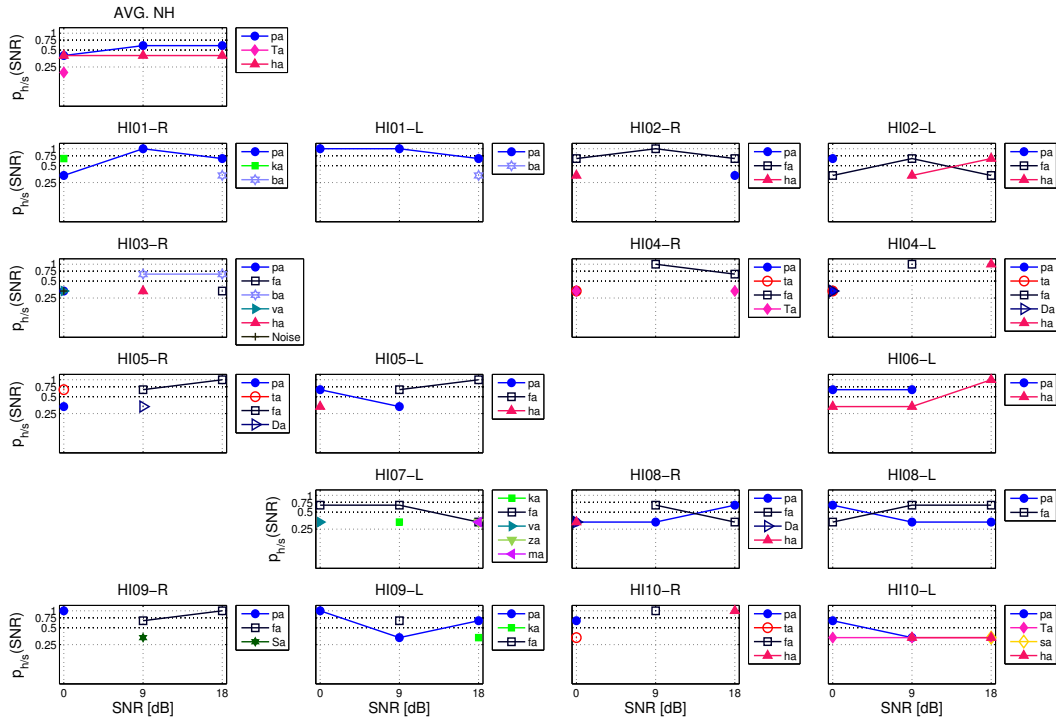


Figure B.36: m115 ta confusions patterns when primary cues are removed.

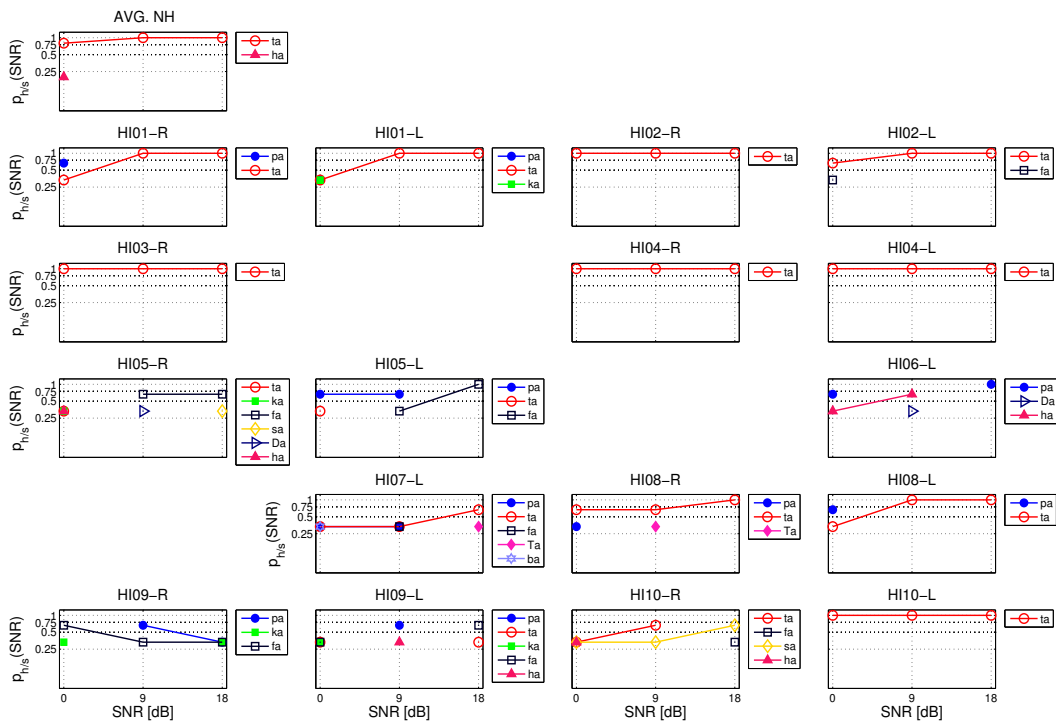


Figure B.37: m115 ta confusions patterns when primary cues are attenuated by -6 [dB].

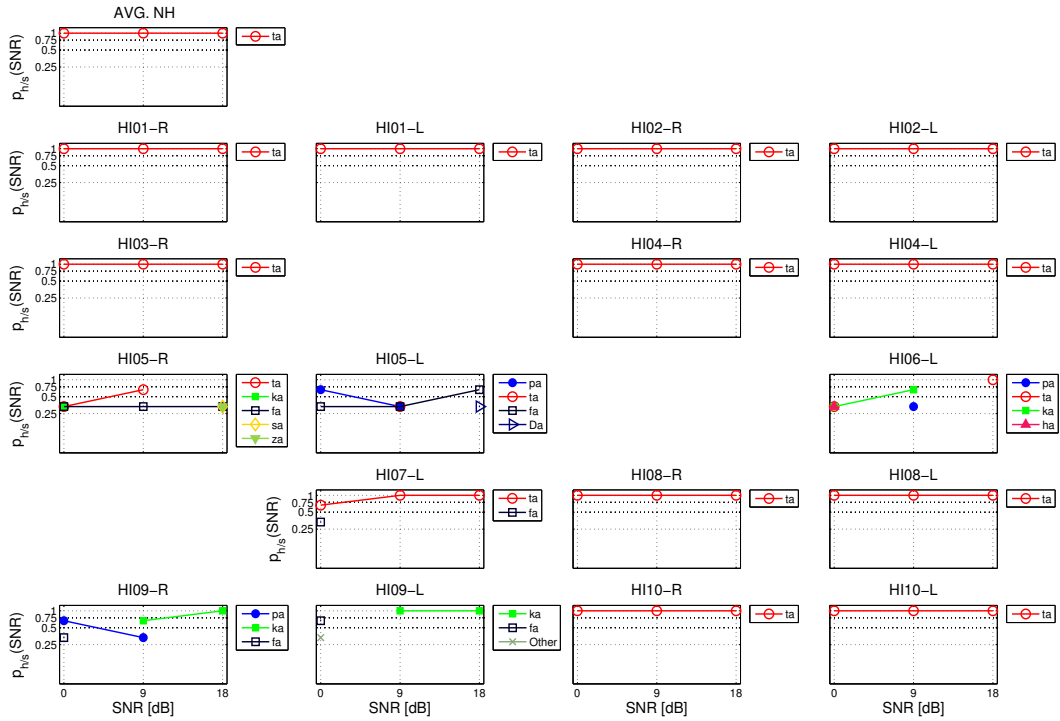


Figure B.38: m15 ta confusions patterns when primary cues are amplified by 6 [dB].

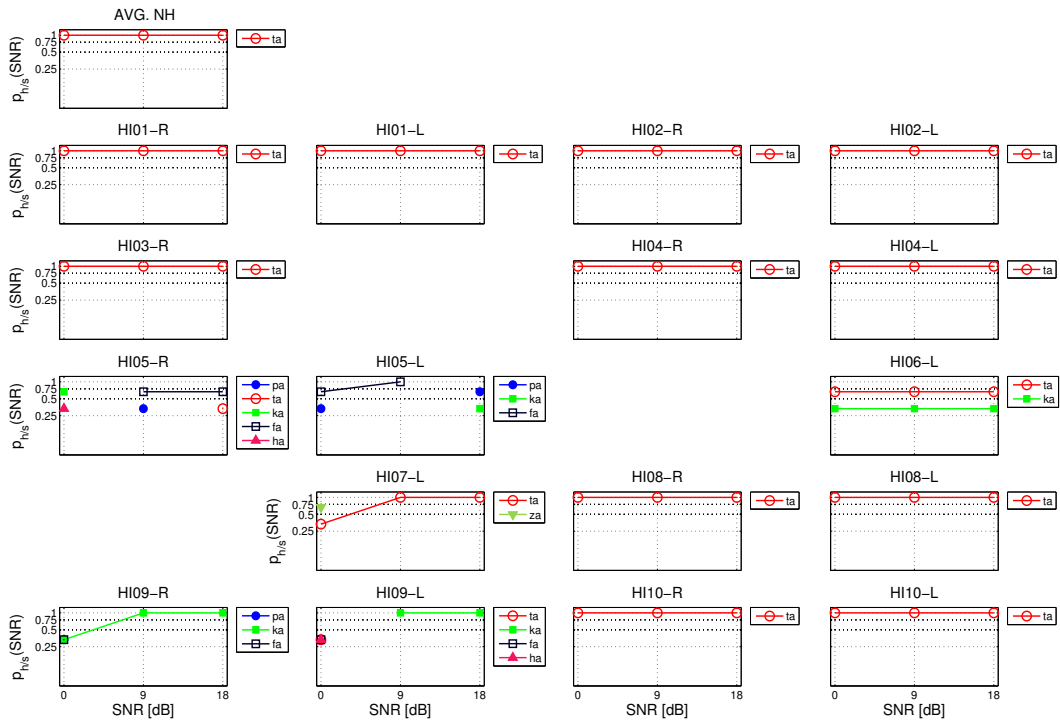


Figure B.39: m15 ta confusions patterns when primary cues are amplified by 12 [dB].

f119ta

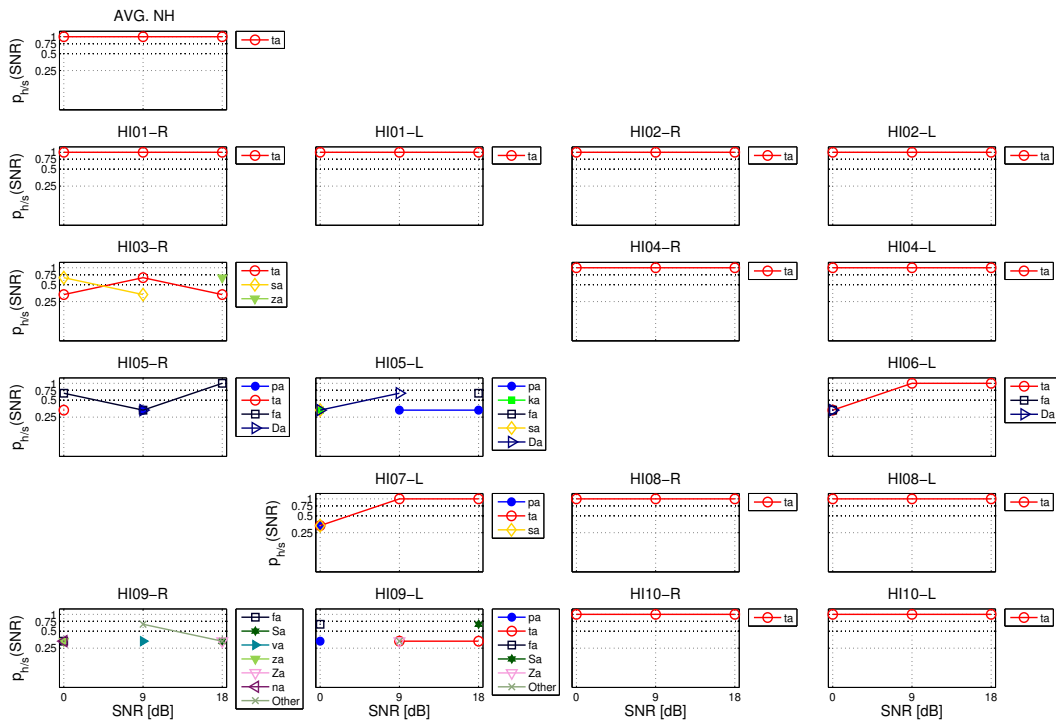


Figure B.40: f119 ta confusions patterns when primary cues are unmodified.

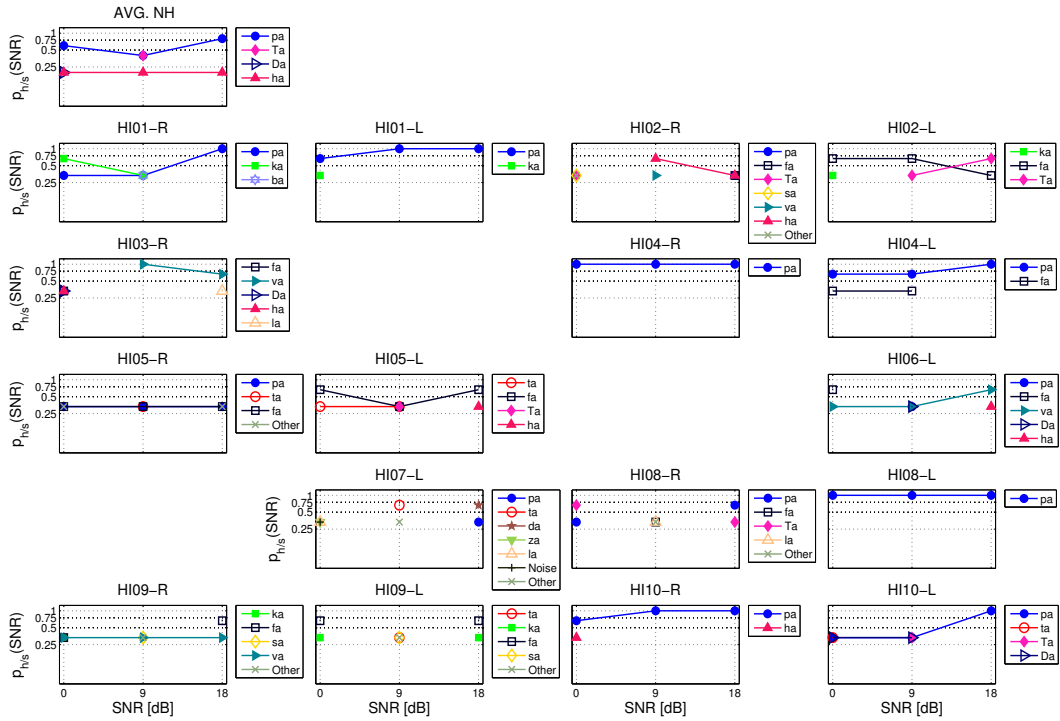


Figure B.41: f119 ta confusions patterns when primary cues are removed.

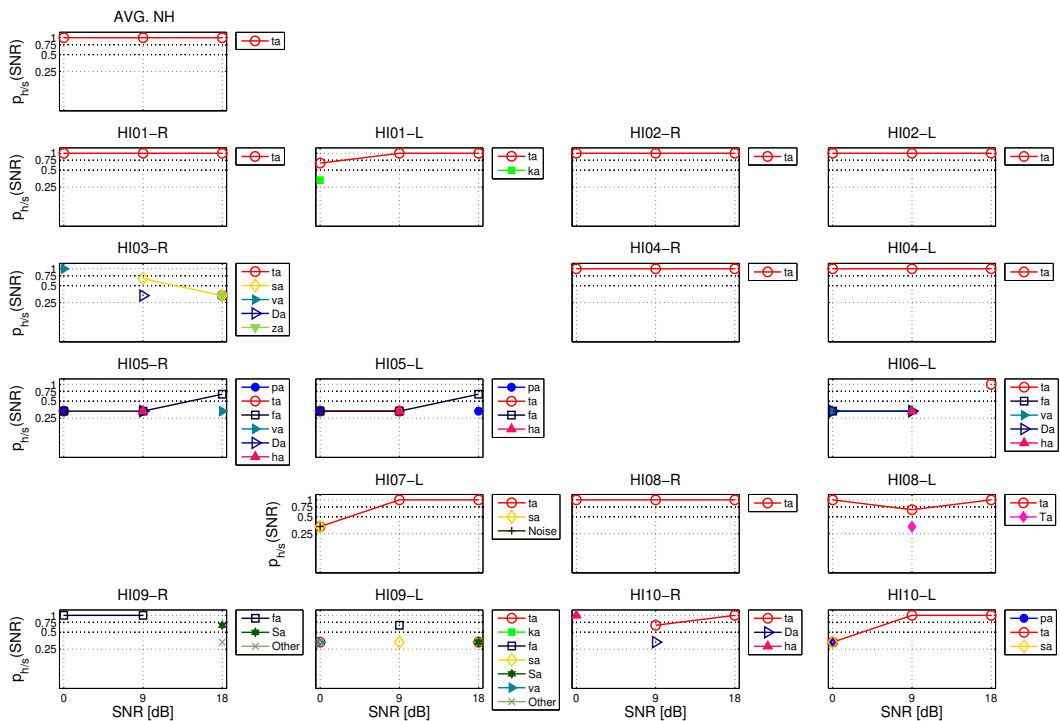


Figure B.42: f119 ta confusions patterns when primary cues are attenuated by -6 [dB].

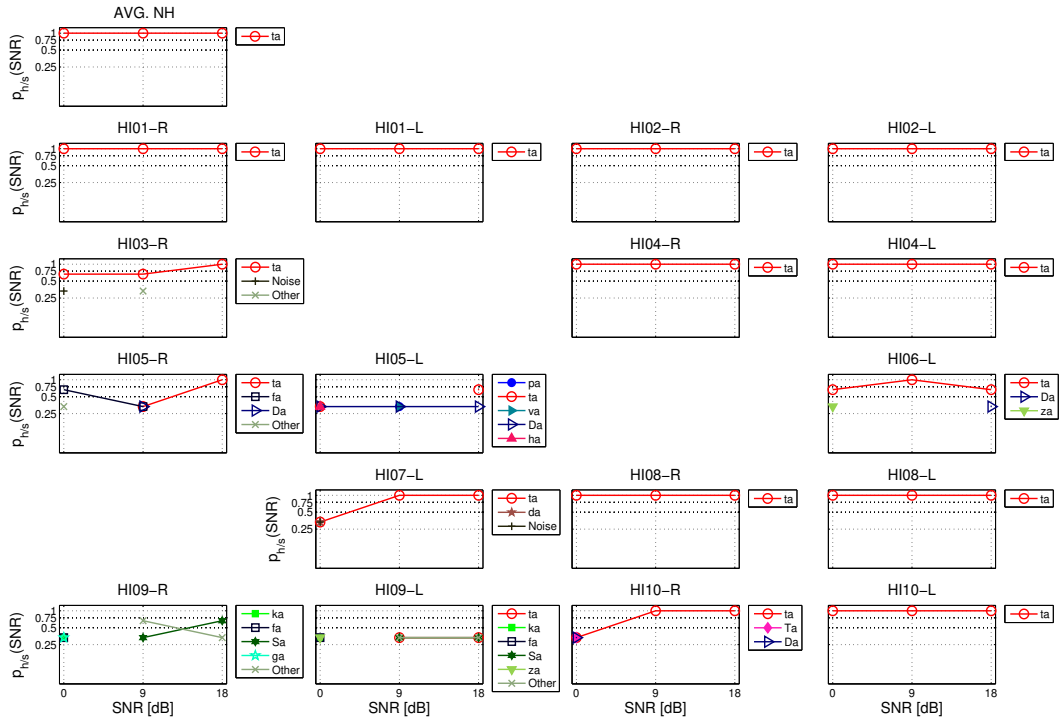


Figure B.43: f119 ta confusions patterns when primary cues are amplified by 6 [dB].

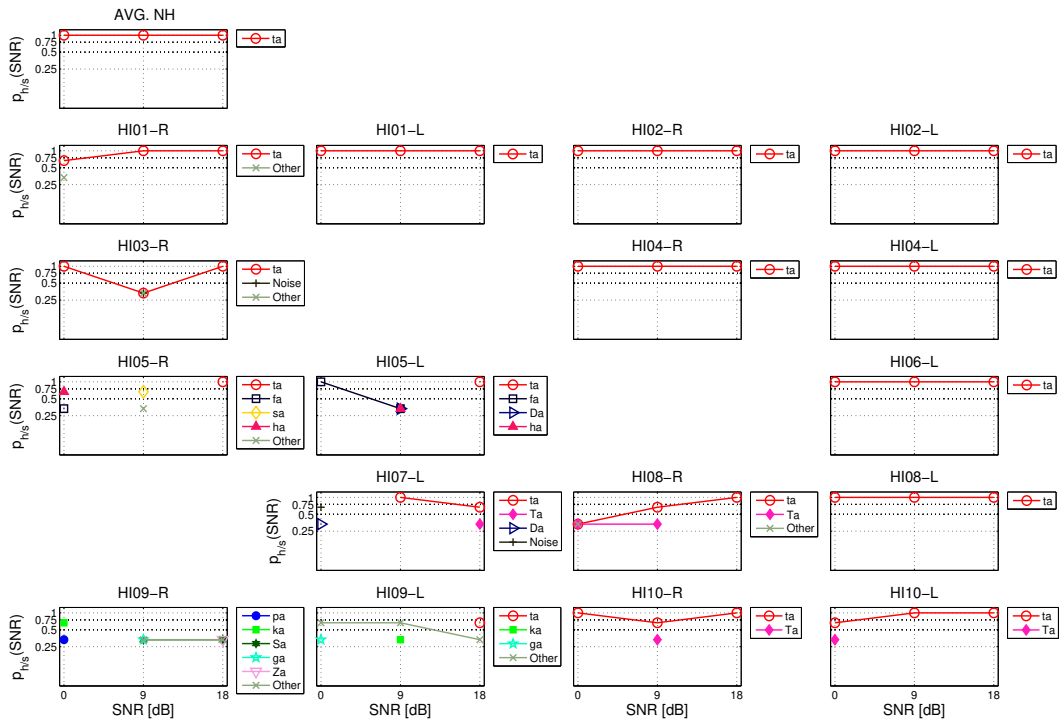


Figure B.44: f119 ta confusions patterns when primary cues are amplified by 12 [dB].

f105da

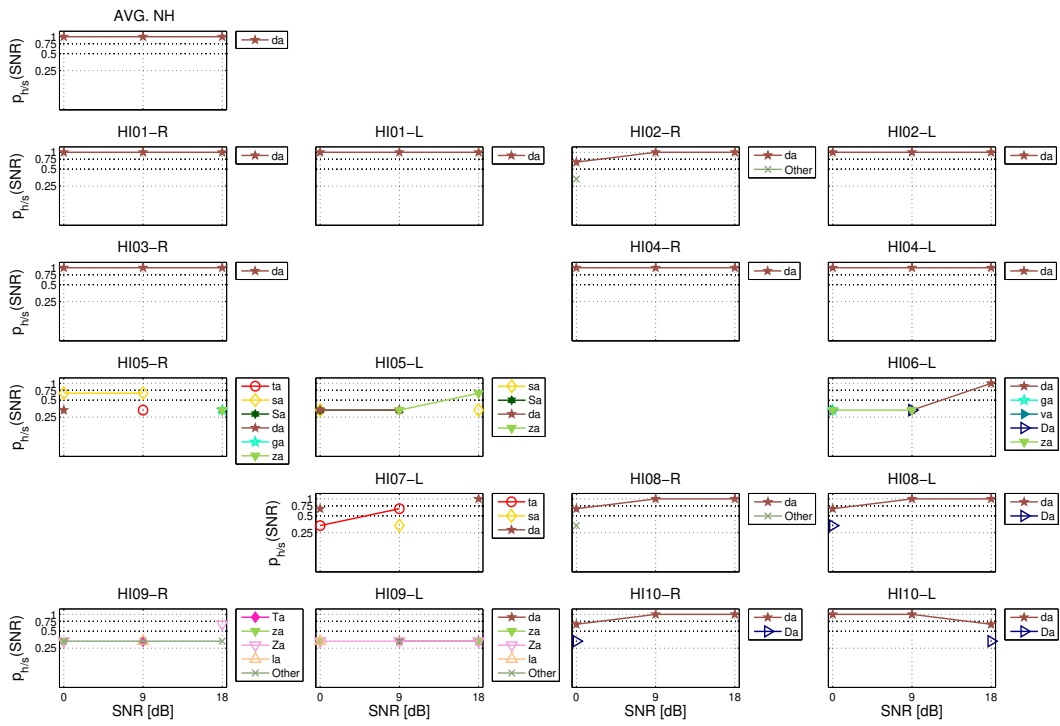


Figure B.45: f105 da confusions patterns when primary cues are unmodified.

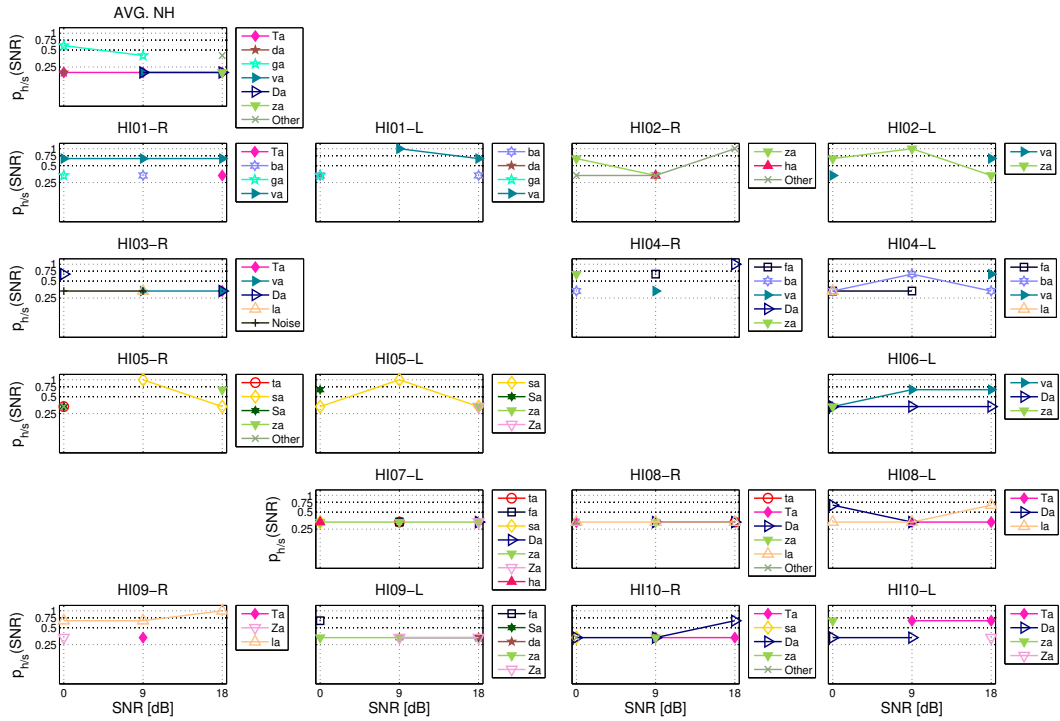


Figure B.46: f105 da confusions patterns when primary cues are removed.

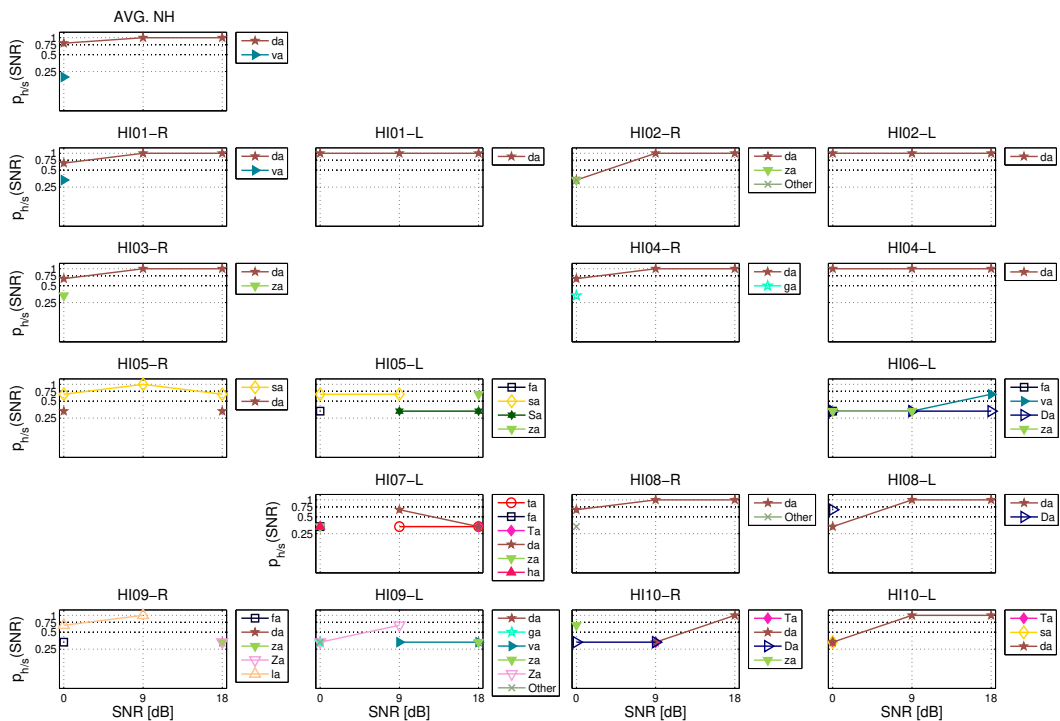


Figure B.47: f105 da confusions patterns when primary cues are attenuated by -6 [dB].

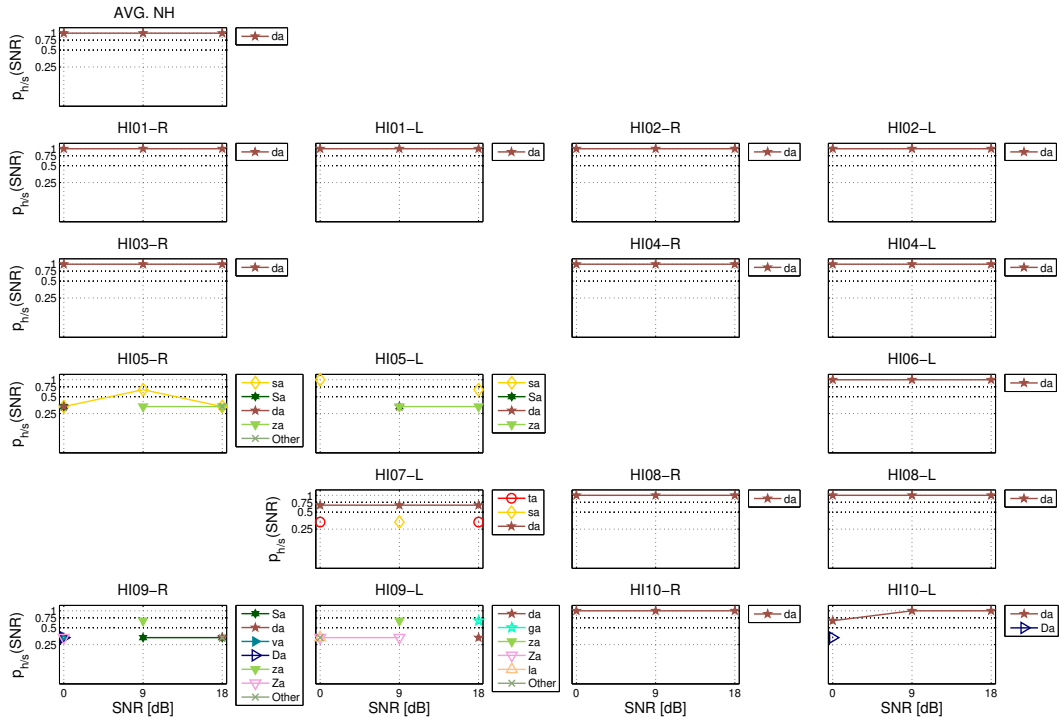


Figure B.48: f105 da confusions patterns when primary cues are amplified by 6 [dB].

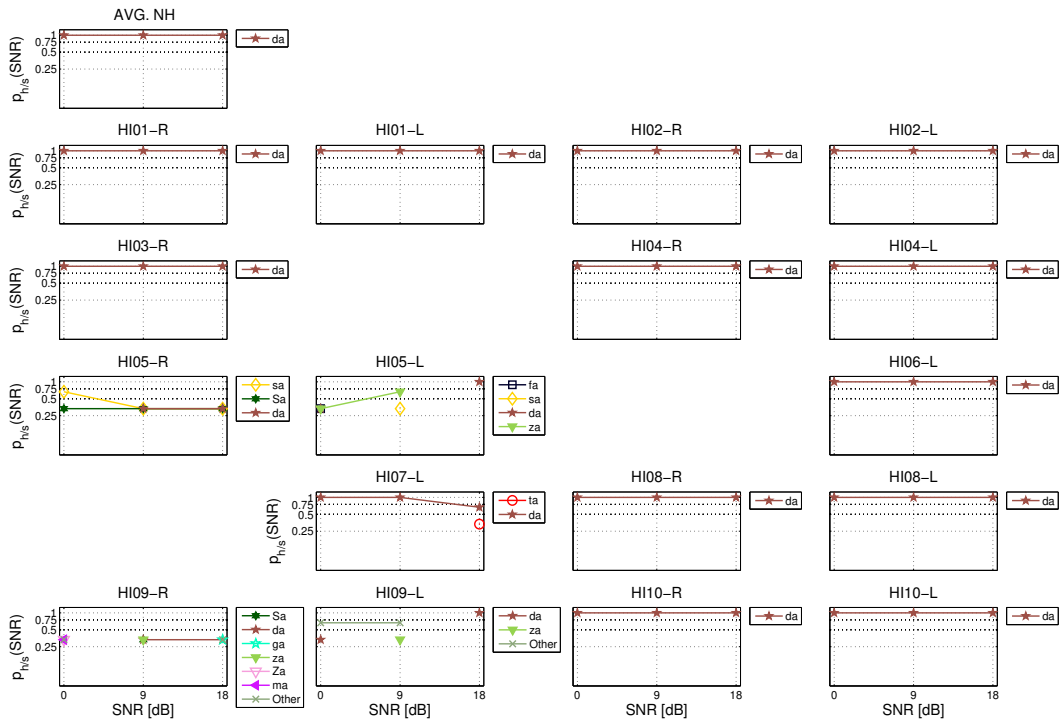


Figure B.49: f105 da confusions patterns when primary cues are amplified by 12 [dB].

f119da

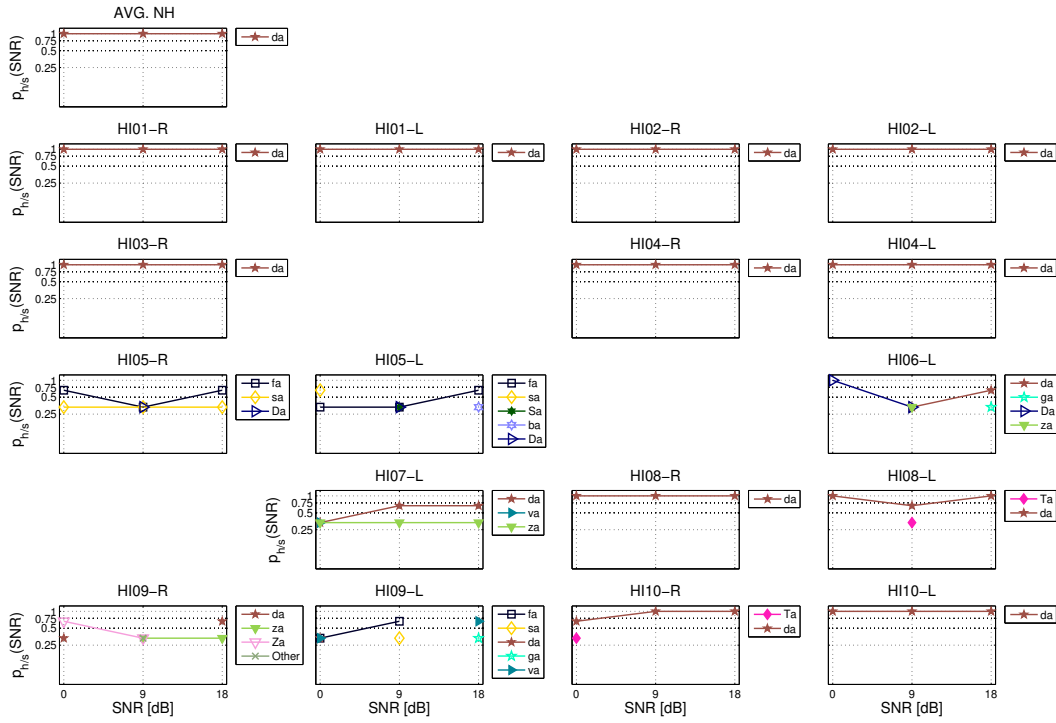


Figure B.50: f119 da confusions patterns when primary cues are unmodified.

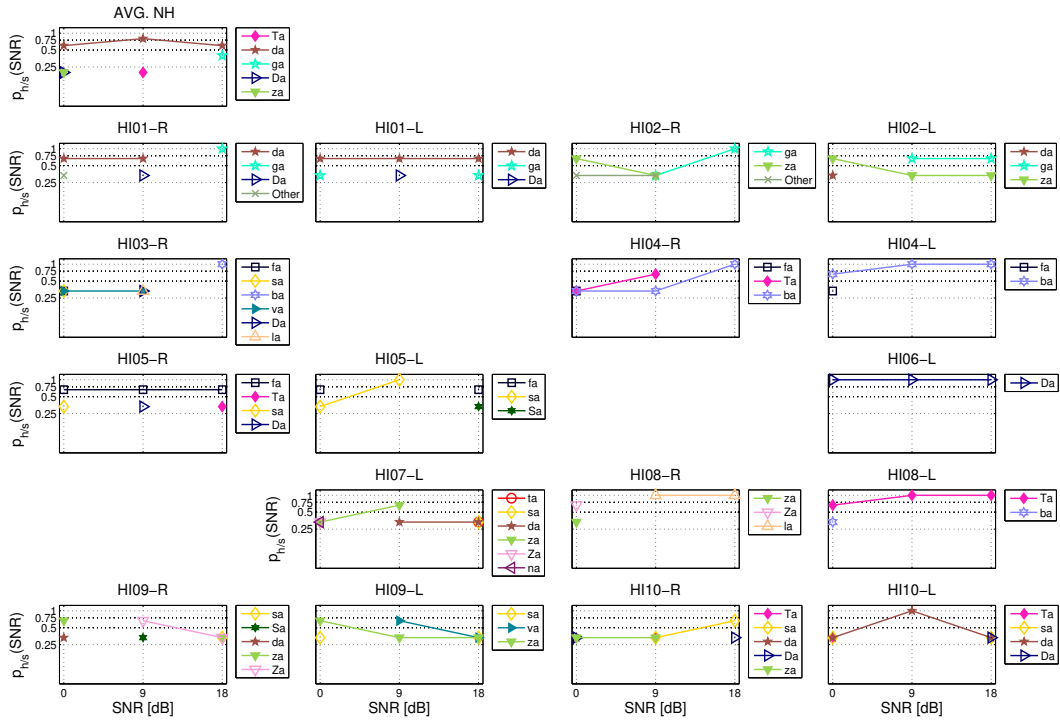


Figure B.51: f119 da confusions patterns when primary cues are removed.

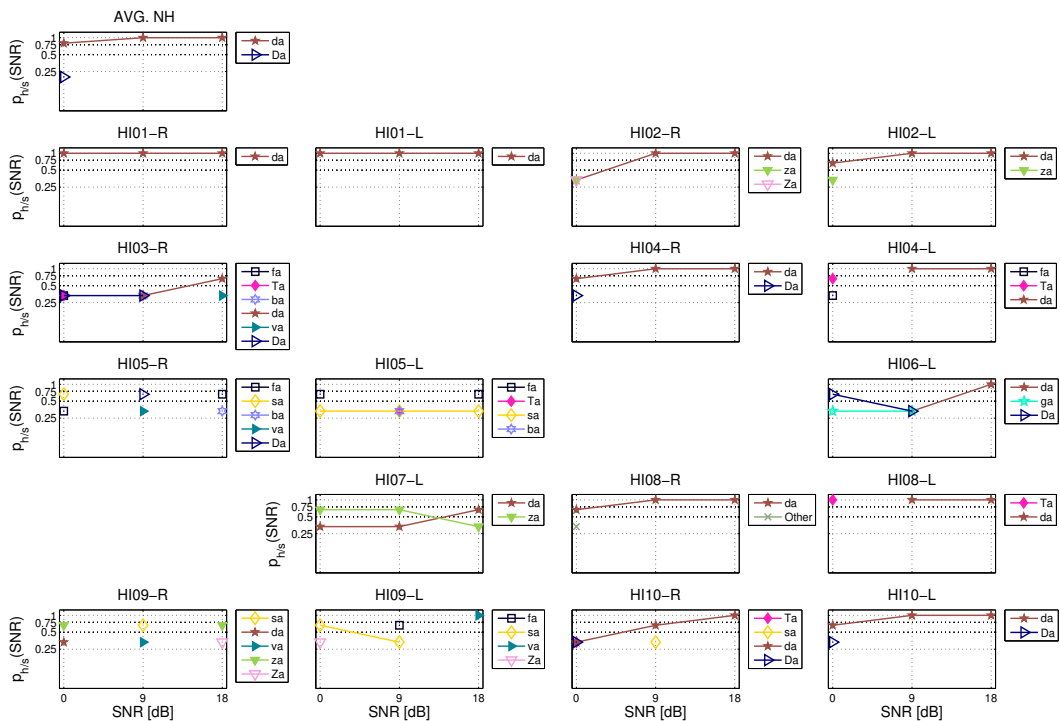


Figure B.52: f119 da confusions patterns when primary cues are attenuated by -6 [dB].

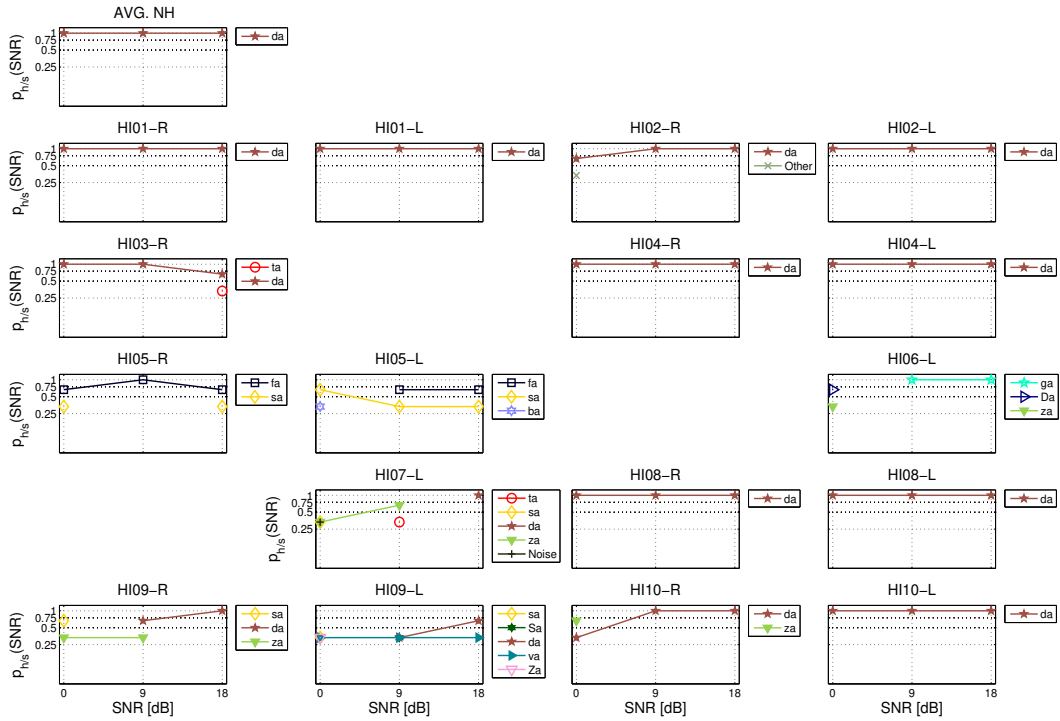


Figure B.53: f119 da confusions patterns when primary cues are amplified by 6 [dB].

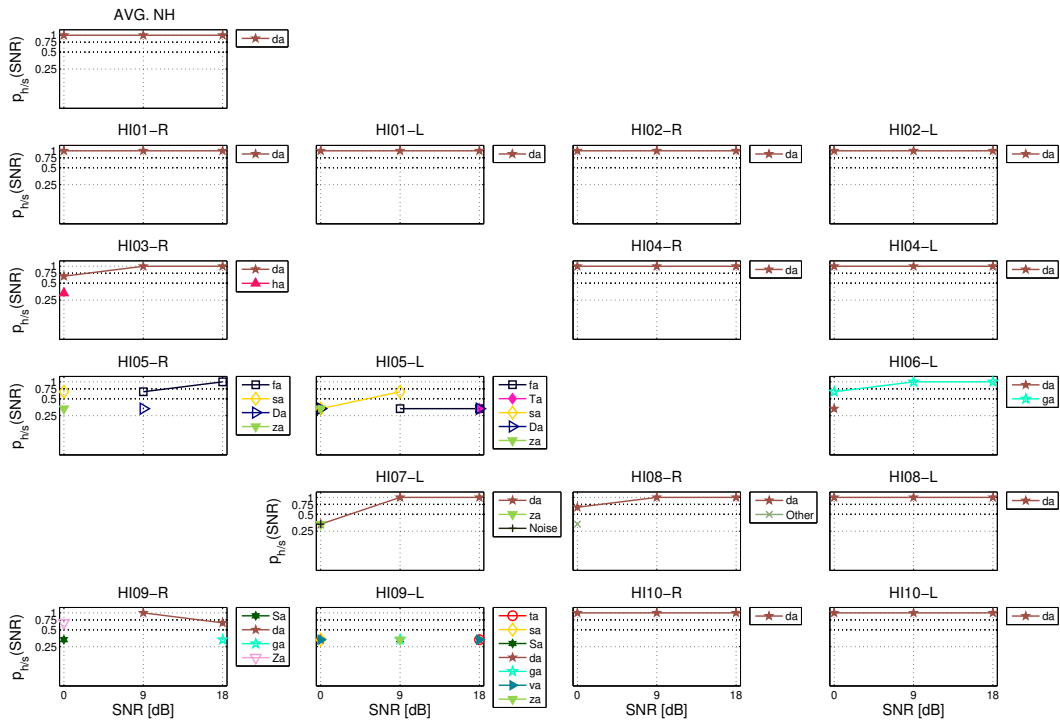


Figure B.54: f119 da confusions patterns when primary cues are amplified by 12 [dB].