Understanding Auditory Context Effects and their Implications

A THESIS

SUBMITTED TO THE FACULTY OF

UNIVERSITY OF MINNESOTA

BY

NINGYUAN WANG

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ADVISOR: PROF. DR. ANDREW J. OXENHAM

DECEMBER 2015

Acknowledgements

Foremost, it is my proud privilege to release the feelings of my gratitude to my advisor Prof. Andrew J. Oxenham for the endless support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. I would never have been able to complete my thesis without his guidance. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating.

I would like to thank the members in my Preliminary exam committee, Prof. Neal Viemeister, Prof. Peggy Nelson, Prof. Sheng He, as well as those in my Dissertation committee, Prof. Peggy Nelson, Prof. Magdalena Wojtczak and Prof. Daniel Kersten, for their insightful comments and encouragement, but also for the hard questions which incented me to widen my research from various perspectives.

I would like to thank Advanced Bionics LLC and Dr. Leonid Litvak for financially supporting my study and supplying the internship opportunities.

I would like to thank Heather Kreft, my lab mate, who helped me collect the experimental data from the cochlear-implant subjects.

My sincere thanks also go to Lei Feng, Emily Allen, and all my lab mates, for the stimulating discussions and all their cares and helps.

Last but not the least, I would like to thank my family: my parents, for giving birth to me at the first place and supporting me spiritually throughout my life.

Abstract

Our perception of sound at any point in time is dependent not only on the sound itself, but also on the acoustic environment of the recent past. These auditory context effects reflect the adaptation of the auditory system to the ambient conditions, and provide the potential for improving coding efficiency as well as providing the basis for some forms of perceptual invariance in the face of different talkers, different room environments, and different types of background noise. Despite their obvious importance for auditory perception, the mechanisms underlying auditory context effects remain unclear. The overall goal of this thesis was to investigate different auditory context effects in both normal-hearing listeners and cochlear-implant (CI) users, to shed light on the potential underlying mechanisms, to reveal their implications for auditory perception, and to investigate the effects of hearing loss on these context effects. In Chapters 2, 3 and 4, different context effects, known respectively as the loudness context effect (LCE), induced loudness reduction (ILR), and spectral motion contrast effect, are examined. Another context effect, known as auditory enhancement, is introduced in Chapter 5 with a vowel enhancement paradigm, and is further explored in Chapter 6 by treating it as process of frequency-selective gain control. Finally, a simplified neural model is proposed in Chapter 7 to explain the basis of auditory enhancement, while remaining consistent with the results from the studies of other context effects. The results reveal both similarities and differences between normal-hearing listeners and CI users in responses to auditory context effects, and suggest a role of peripheral processes played in auditory context effects and a potential opportunity to improve current CI speech processing strategies through a restoration of normal auditory context effects.

Table of Content

List of Figures

List of Tables

List of Abbreviations

AGC	Automatic gain control
ANOVA	Analyses of variance
CF	Characteristic frequency
CI	Cochlear implant
DR	Dynamic range
EE	Enhancement effect
ERB	Equivalent rectangular bandwidth
HL	Hearing level
IC	Inferior colliculus
ICS	Internal cochlear stimulator
ILR	Induced loudness reduction
ISI	Interstimulus interval
LCE	Loudness context effect
MAL	Maximum acceptable level
MCL	Most comfortable level
MOC	Medial olivocochlear
NH	Normal hearing
PSE	Point of subjective equality
SNR	Signal-to-noise ratio
SPL	Sound pressure level
TEN	Threshold equalizing noise
THS	Threshold
VC	Vocoder

CHAPTER I: GENERAL INTRODUCTION

In the natural environment, our perceptual systems receive and encode continuous stimulation from the world. The auditory processing in each moment in time is dependent on the past, meaning that our perception of incoming stimuli depends strongly on contextual information. One well-known example of context effects in vision is "color constancy," whereby the perceived color of an object remains relatively constant under various illumination conditions, despite the fact that the illumination conditions result in very different light spectra reaching the retina (Land, 1977; Bloj *et al.*, 1999). Another interesting example is termed the "negative afterimage": after staring at a picture for a while, the same pattern filled with opposing colors can be perceived if the observer focuses on a white sheet of paper right away (Shimojo *et al.*, 2001). Both examples may reflect a more general principle and function of any perceptual system, which is to adapt to the ambient conditions, thereby potentially improving coding efficiency and/or facilitating perceptual invariance.

Similarly, in audition, the perception of one sound, whether it is speech, music, or other natural sound, can be affected by the context in which it is presented. Some of these effects have been studied mainly with artificial laboratory stimuli, such as tones and noise. These including forward masking (e.g. Jesteadt *et al.*, 1982), loudness context effects (e.g. Marks, 1994; Arieh and Marks, 2003a), auditory enhancement effects (e.g. Viemeister, 1980; Viemeister and Bacon, 1982), and "overshoot" (e.g. Kimberley *et al.*, 1989b; McFadden, 1989), while others have been studied primarily in the context of speech

perception (e.g. Holt *et al.*, 2000; Holt, 2006a; b). Although several auditory context effects have been identified and explored, the underlying mechanisms remain for the most part unclear. Different effects may either reveal different properties of auditory system, or may reflect diverse reflections of the same property (mechanism).

Cochlear implants (CIs) provide a potential window into the role of peripheral processing in auditory context effects, because they bypass the cochlea and directly stimulate the auditory nerves. Although little is known about the physiological basis of context effects in hearing, one possibility is that some effects may be mediated by frequency-selective, time-varying changes in cochlear gain, produced by the medial olivocochlear (MOC) efferent system. If so, then certain context effects should be reduced or eliminated in CI users. Investigating context effects in CI users is therefore important for two reasons: First, it may provide us with insights into the underlying mechanisms of context effects in normal (acoustic) hearing, and second, it may provide us with important information on how to restore missing or abnormal context effects in CI users via signal processing.

The overall goal of this thesis is to investigate different auditory context effects, to shed light on the potential underlying mechanisms, to reveal their implications for auditory perception, and to investigate the effects of hearing loss on these context effects. The following sections outline the content of each chapter of the thesis. Of the six chapters that describe experimental or modeling studies, three have been published, two are currently under review, and one is in preparation for submission for publication.

In Chapter 2, a study of loudness context effects (e.g. Elmasian and Galambos, 1975; Marks, 1994) in both acoustic and electric hearing is introduced. This study is presented first in the series of studies of auditory context effects because of its relative simplicity and the inspiration it provides for some of the following studies. Chapter 3 aims to disentangle two potential factors of loudness context effects by using a comparison sound that is remote in frequency (or cochlear location) from the target and conditioning sounds (e.g. Arieh and Marks, 2003b; Oberfeld, 2007). This chapter also compares the results from normal-hearing listeners and CI users.

In Chapter 4, a more complicated context effect is explored involving spectral motion aftereffects. Previous studies of context effects of speech sounds have concentrated exclusively on the effects of static (long-term) spectral effects (e.g. Holt *et al.*, 2000; Holt, 2006a). To our knowledge, ours is the first to study dynamic aftereffects in a speech context. This work serves to highlight the complexity and multi-dimensionality of auditory context effects and how they may influence speech perception in everyday environments.

In Chapter 5, a classic auditory aftereffect, known as auditory enhancement (e.g. Viemeister, 1980; Viemeister and Bacon, 1982) is examined in both normal-hearing subjects and CI users with a vowel identification paradigm (Summerfield *et al.*, 1984a). This paradigm takes the basic psychophysical effect into a context where it can potentially affect speech perception.

In Chapter 6, a careful examination of the auditory enhancement effect in terms of its effective change in gain is carried out to provide new data with which to test more rigorous models of the effect. These data and others are considered in Chapter 7, which introduces a theoretical framework and preliminary model for explaining the auditory enhancement effect in a way that remains consistent with the other auditory context effects explored in this thesis.

The last chapter provides a summary and discussion based on the findings from the previous chapters, and point out possible future directions of research in auditory context effects.

CHAPTER 2: LOUDNESS CONTEXT EFFECTS IN NORMAL-HEARING LISTENERS AND COCHLEAR-IMPLANT USERS[1]

Abstract

Context effects in loudness have been observed in normal auditory perception, and may reflect a general gain control of the auditory system. However, little is known about such effects in cochlear-implant (CI) users. Discovering whether and how CI users experience loudness context effects should help us better understand the underlying mechanisms. In the present study, we examined the effects of a long-duration (1-s) intense precursor on the loudness relations between shorter-duration (200-ms) target and comparison stimuli. The precursor and target were separated by a silent gap of 50 ms, and the target and comparison were separated by a silent gap of 2 s. For normal-hearing listeners, the stimuli were narrowband noises; for CI users, all stimuli were delivered as pulse trains directly to the implant. Significant changes in loudness were observed in normal-hearing listeners, in line with earlier studies. The CI users also experienced some loudness changes but, in contrast to the results from normal-hearing listeners, the effect did not increase with increasing level difference between precursor and target. A "dual-process" hypothesis, used to explain earlier data from normal-hearing listeners, may provide an account of the present data by assuming that one of the two mechanisms, involving "induced loudness reduction," was absent or reduced in CI users.

2.1 Introduction

Our perception of a stimulus or event is dependent in large part on the context in which it is presented. Much has been learned about perceptual processing through the study of context effects and their neural correlates. In auditory perception, judgments of the loudness of a sound can be affected by the intensity relation between that target sound and sounds that precede it. Early studies showed that when an intense auditory stimulus precedes a weaker one, the loudness of the weaker stimulus can be judged to have increased by as much as 30 dB, whereas when the preceding signal, or precursor, is less intense than the following target signal, the loudness of the target decreases somewhat from its loudness in isolation (Galambos *et al.*, 1972; Elmasian and Galambos, 1975; Elmasian *et al.*, 1975). This phenomenon, known as loudness enhancement or decrement, respectively, was considered to reflect a general principle of intensity coding and gain control in the auditory system.

These early studies generally involved a three-tone paradigm, with a conditioning (or precursor) tone, followed by a target tone and then a comparison tone, which subjects adjusted in level to match the loudness of the target tone. All three tones were presented at the same frequency. Manipulations of the presentation ear revealed that loudness enhancement was strongest when all tones were presented to the same ear (binaural or monaural presentations). In a dichotic situation (with the precursor and target presented to opposite ears), less, but still significant, loudness enhancement was observed (Elmasian and Galambos, 1975). In contrast, loudness decrement effects seemed

relatively insensitive to ear of presentation (Elmasian *et al.*, 1980), suggesting that loudness enhancement may involve some monaural, possibly peripheral, processing components, whereas loudness decrement may primarily involve more central sites. A finding that raised fundamental questions concerning the peripheral nature of loudness enhancement was that enhancement (and decrement) could also be observed when the conditioning tone was presented *after* the target tone in time (Elmasian *et al.*, 1980).

Loudness enhancement and decrement are considered "assimilative" effects, in that the loudness of the target is drawn towards that of the conditioner (and presumably vice-versa). Other studies of loudness context effects have reported the opposite, namely that an intense precursor tone can *reduce* the loudness of a subsequent tone that is presented at a lower level. In contrast to loudness enhancement, this "loudness recalibration" (e.g. Marks, 1994; Mapes-Riordan and Yost, 1999) or "induced loudness reduction" (e.g. Scharf *et al.*, 2002), seems to be a relatively long-lasting effect. It is generally observed when the precursor and target are at the same frequency, but the comparison tone is presented at a frequency that is remote from that of the precursor and target. As with loudness enhancement, the effect can be relatively large, ranging from about 10 to 20 dB, depending on the measurement method and stimulus parameters used. Interestingly, maximum loudness recalibration is not obtained directly after the precursor, but instead builds up to reach a maximum at a delay of around 1 s, and is still observable at a delay of 3 s (Mapes-Riordan and Yost, 1999).

As proposed by Scharf *et al.* (2002), and supported by Arieh and Marks (2003b), the build-up and relatively long time constants associated with loudness recalibration suggest a possible reinterpretation of the earlier loudness enhancement studies, where all three tones were presented at the same frequency. In particular, it may be that the comparison tone is reduced in loudness by the precursor, rather than the target tone being increased in loudness. To investigate this issue, Oberfeld (2007) used a four-tone task, with the first three tones (precursor, target, and comparison) at the same frequency and fourth tone at a remote frequency. He asked listeners to compare the loudness of the original comparison (third) tone with that of the fourth tone. According to Oberfeld's results, it seems that both enhancement and adaptation contribute to loudness recalibration. Results from his study support an earlier hypothesis of Arieh and Marks (2003b), that loudness recalibration reflects a dual-process mechanism. On one hand, when an intense auditory signal (precursor) precedes a weaker one (target) by a short gap (less than 100 ms), the loudness of the following signal can be enhanced (Elmasian and Galambos, 1975; Marks, 1988); on the other hand, when the time interval between precursor and target (close in frequency) exceeds 200 ms, the target signal will be reduced, perhaps due to adaptation (Arieh and Marks, 2003b). These properties of loudness recalibration could be explained by the interaction between a fast onset and fast decay enhancement process and a fast onset but slower decay adaptation process (Oberfeld, 2007).

There are many potential sources of both enhancement and adaptation along the auditory pathways, and few attempts have been made to constrain the locus or nature of these sources. One of the potential sources of an adaptation-like process is the medial olivocochlear (MOC) efferent system, which acts to reduce both the gain and frequency selectivity of the basilar membrane response to sound, by affecting the action of the outer hair cells (Nieder *et al.*, 2003; Guinan, 2006; Jennings *et al.*, 2009). As such, an MOC-based effect could, in principle, help explain why loudness effects transfer only partially across the ears: MOC effects are activated bilaterally, but are strongest for ipsilateral activation (Guinan, 2006). Although the time constants associated with the MOC fast effect are not thought to extend to several seconds, the slow effect of MOC may at least contribute to loudness changes (Cooper and Guinan, 2003).

In this study, we investigated context effects on loudness using both normal-hearing listeners and cochlear-implant (CI) users with a three-tone paradigm similar to that used in early loudness enhancement studies. We use loudness context effect (LCE) as a relatively neutral term to avoid any assumption regarding whether the effect measured reflects an enhancement of the target or adaptation of the comparison (or both). The stimuli were presented as high-rate pulse trains to single electrodes of the CIs. In the normal-hearing listeners, narrowband noises were used (rather than tones) to better simulate the spread of excitation produced by single-electrode stimulation in CIs (e.g. Bingabr *et al.*, 2008). In addition, we varied the frequency (or electrode) of the precursor relative to that of the target and comparison tones. The rationale was that if two different

mechanisms are responsible for the time course of LCE, then the two mechanisms might have different frequency selectivity. The comparison of normal-hearing listeners and CI users allowed us to test the role of the MOC efferent system. Because MOC efferent activation affects cochlear gain, it requires an intact cochlea. Therefore, any portion of the effect due to MOC efferent effects should not be observed in CI users. Thus, if CI users show some LCE, we could conclude that LCE cannot be due solely to MOC activation (although it may still play some role). As a result, investigating LCE in CI users may provide us with important information about the potential underlying mechanisms. Some researchers have suggested that the cochlear gain changes induced by the MOC efferent system may be important for speech perception in noise (e.g. Guinan, 2010; Garinis *et al.*, 2011; Clark *et al.*, 2012; de Boer *et al.*, 2012; Mishra and Lutman, 2014). Therefore, any differences in the results between normal-hearing listeners and CI users may provide guidance for future CI signal processing systems to restore normal context effects for auditory and speech perception.

2.2 Experiment 1: loudness context effects in normal-hearing listeners

2.2.1 Methods

Subjects

Seven listeners (two males, five females) participated in this experiment and were compensated for their time. Their ages ranged from 18 to 63 years (mean age 26.1 years; only one subject older than 45). All listeners had normal hearing, as defined by audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz. All participants provided written informed consent, and all protocols were approved by the Institutional Review Board of the University of Minnesota.

Stimuli

A schematic diagram of the stimuli used in this experiment is shown in Fig. 2.1A. Each trial consisted of three sounds: a precursor, a target, and a comparison. The temporal properties of the stimuli remained constant for the entire experiment. The total duration of the precursor was 1 s, and the total durations of both the target and the comparison were 200 ms. The precursor and target were separated by a silent gap of 50 ms, which was sufficient to trigger both loudness enhancement and ILR effects according to Arieh and Marks (2003b), and the target and comparison were separated by a silent gap of 2 s. All the stimuli were gated on and off with 10-ms raised-cosine ramps. All the stimuli

were narrowband noises, created by filtering a Gaussian white noise with a second-order IIR peaking filter in the time domain, with slopes of either 24 or 96 dB/octave. The use of bandpass noise was intended to simulate the spread of current produced by CIs, and the different slopes were intended to simulate different degrees of current spread produced by monopolar and bipolar stimulation modes. The 24 dB/octave slopes were chosen to be within the range provided by Bingabr *et al.* (2008) to simulate monopolar stimulation (although shallower slopes have also been assumed; see Oxenham and Kreft (2014); the 96 dB/octave slopes were chosen to be in the range of the values provided by Bingabr *et al.* (2008) to simulate bipolar stimulation.

The level of the target was always 60 dB SPL. A precursor level of 70 dB SPL was tested in conjunction with filter slopes of both 24 and 96 dB/octave. The 10 dB level difference between the precursor and target was selected because it was deemed large enough to produce some effect, based on previous studies (Elmasian and Galambos, 1975; Elmasian *et al.*, 1980), but not so large as to make a comparison with CI users difficult, based on their more limited dynamic range (Hong *et al.*, 2003). The center frequency of the precursor within each block was selected from one of five values (455, 762, 1278, 2142, or 3590 Hz), approximately logarithmically spaced around the center frequency of the target and comparison, which was always 1278 Hz. The spacing between adjacent components corresponds to 3.5 to 4.5 equivalent rectangular bandwidths (ERBs) of the auditory filters (Glasberg and Moore, 1990). The level of precursor and target remained

constant within each block. The level of the comparison varied between trials within a specific range centered around the target level, from 57 to 63 dB SPL in 1-dB steps.
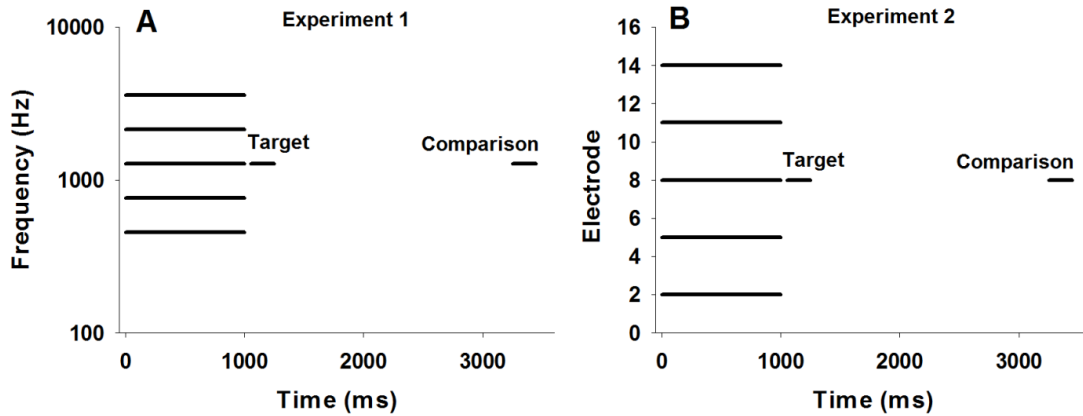
FIG. 2.1. Schematic diagrams of stimuli used in Experiments 1 and 2. Panel A shows the stimuli for Experiment 1, where the precursor was presented at one of five center frequencies of 455, 762, 1278, 2142, or 3590 Hz, selected from the standard Advanced Bionics 16-channel map, corresponding to the center frequencies of electrodes E2, E5, E8, E11, or E14, respectively, and the target and comparison stimuli had a center frequency of 1278 Hz, corresponding to electrode 8 of the standard CI map. Panel B shows the stimuli from Experiment 2, where a pulse train was delivered directly to those selected electrodes (E2, E5, E8, E11, or E14) of the CI via a clinical research platform.

Additional data were collected with an 85-dB SPL precursor and a 60-dB SPL target, with filter slopes of 96 dB/octave and only one precursor center frequency of 1278 Hz, corresponding to the center frequency of the target and comparison. The comparison level range was from 55 to 65 dB SPL, in 2-dB steps. A larger step size was used with the

13

higher precursor level, because a larger effect was expected, based on previous literature (Elmasian and Galambos, 1975).

The stimuli were generated digitally and played out diotically from a LynxStudio L22 24-bit soundcard at a sampling rate of 22.5 kHz via Sennheiser HD650 headphones to listeners seated in a double-walled sound-attenuating chamber.

Procedure

A training session was run prior to the actual experiment, involving the target and comparison sounds, but no precursor. Listeners were instructed to respond to the question, "Which sound is louder?" via virtual buttons on the computer display. As in the actual experiment, the target was always 60 dB SPL. The comparison was presented at one of six levels: 57, 58, 59, 61, 62, and 63 dB SPL. Each level was presented 10 times, resulting in 60 trials per training block. Feedback was provided throughout the training session. Listeners were required to reach 80% correct to proceed to the actual experiment. All of the participants achieved this level of performance within two blocks of training.

In the actual experiment, listeners were asked to ignore the precursor (if present), and to again judge which of the two short sounds (the target and the comparison) was louder. A reference condition with no precursor (similar to the training condition) was also included. Each precursor condition was repeated five times in random order within each of three sessions. The first session involved the 70-dB SPL precursor at one of five

center frequencies with the 24-dB/octave filter slopes; the second session involved the 70-dB SPL precursor at one of five center frequencies with the 96-dB/octave filter slopes; the third session involved the 85-dB SPL precursor at only a single center frequency with the 96-dB/octave filter slopes. In the first and second sessions, each block comprised one precursor frequency (or no precursor) with seven comparison levels, repeated 10 times in random order, making a total of 70 trials per block. Each session contained 30 blocks (five repetitions for each of the six precursor conditions, with trials in a new random order in each block), for a total of 50 repetitions of each stimulus per subject. In the final session, with the 85-dB precursor, six comparison levels were each repeated 10 times, for a total of 60 trials per block. A total of 10 blocks of trials were presented per subject in the last session (reference and on-frequency condition, five times for each condition), for a total of 50 repetitions of each stimulus per subject. No feedback was provided in the test sessions. The whole experiment lasted about 6 to 8 hours, divided into 2-h sessions.

2.2.2 Results and discussion

The mean results are presented in Fig. 2.2. In each panel, the proportion of trials in which the comparison was judged to be louder than the target is plotted as a function of the comparison level. The upper panels show the results with a 70-dB SPL precursor, with panel A and B showing data from the 24 and 96 dB/octave filter slopes, respectively. Different symbols represent the different precursor center frequencies, as shown in the legend. Fig. 2.2D shows the data using the precursor level of 85 dB SPL and filter slopes

of 96 dB/octave. Fig. 2.2C replots the on-frequency-precursor and no-precursor conditions from Fig. 2.2B for ease of comparison.

Consider first the conditions with no precursor (filled circles). In all three conditions, the point of subjective equality (PSE), i.e., the level at which the comparison was judged louder than the target 50% of time, was reached at a comparison level between 58 and 60 dB SPL. In other words, the two stimuli were judged equally loud when the target was 0-2 dB higher in level than the comparison. Perceptual biases of this kind have occurred in other loudness comparison studies, although the direction of the bias does not appear to be always consistent. For instance, in Elmasian *et al.* (1980), for baseline conditions, the 50-dB target alone was matched with a comparison tone level of around 52 dB, whereas the 70-dB target alone was matched with a comparison tone level nearer 66 dB SPL.

Consider next the effect of adding a precursor. In general, the addition of a precursor resulted in the target being judged louder (and/or the comparison being judged quieter), as shown by the fact that the filled circles (precursor absent) lie above the other symbols in all conditions. Moreover, the on-frequency precursor produced the largest effects, as shown by the fact that the open circles generally fall below all the other symbols. In general, the effect of the precursor diminished with increasing spectral distance between the precursor and target. This trend is particularly apparent in the case of the 24 dB/octave slopes, where the progression from no difference to a large difference in center frequency was more systematic; in the condition with 96 dB/octave slopes, the

16

on-frequency precursor produced the largest effect, but all other precursor conditions produced similarly small effects.

Finally, consider the effect of precursor level. As expected from previous studies (Elmasian and Galambos, 1975; Mapes-Riordan and Yost, 1999), the overall effect (difference between no precursor and on-frequency precursor) seems greater with the higher-level than with the lower-level precursor (compare Fig. 2.2C and 2.2D).

Probit analysis was used to fit each of the curves shown in Fig. 2.2 for each subject individually. The fitted curves from each subject and each condition were then used to calculate the comparison level at the PSE. A level higher than 60 dB SPL implies that the comparison required a higher level than the target to be judged equally loud.
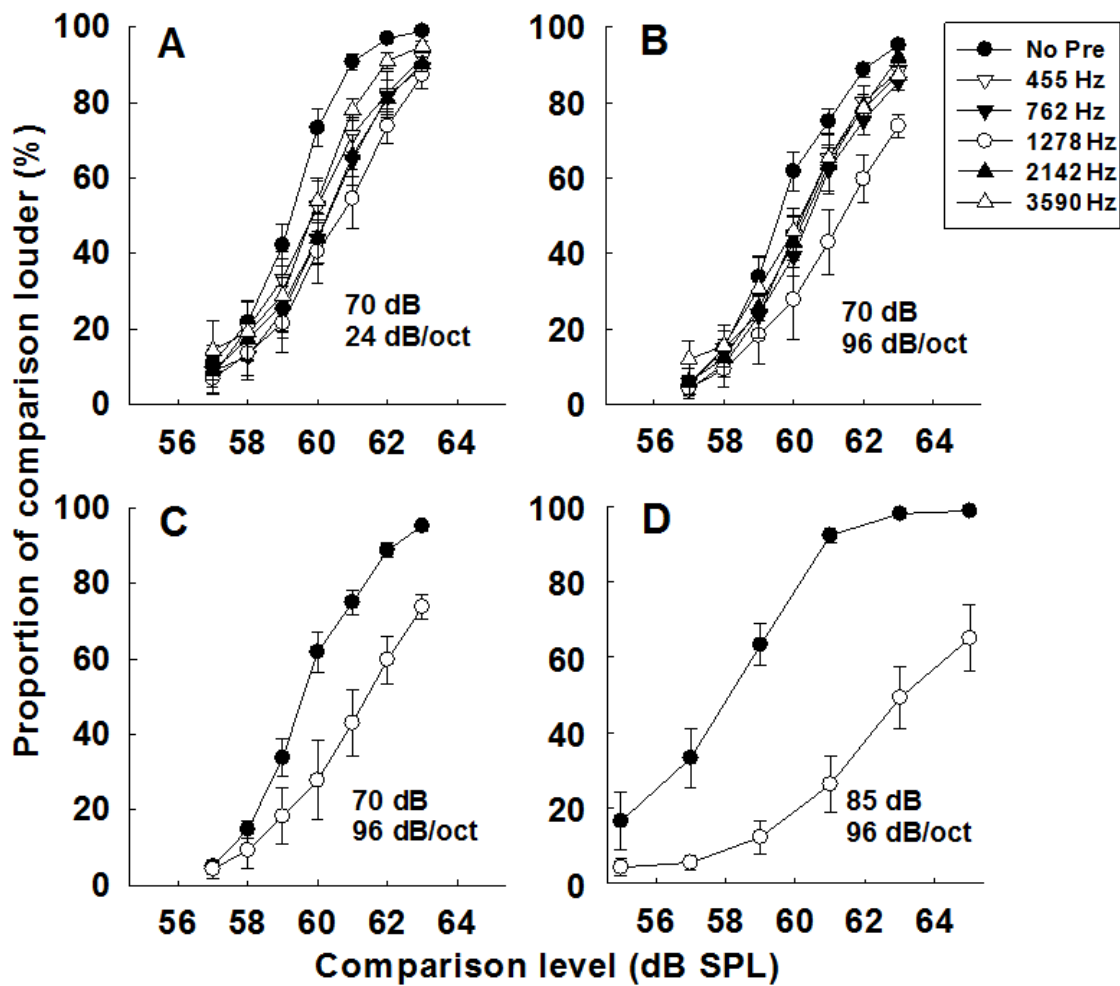
FIG. 2.2. Results from normal-hearing listeners. The proportion of trials (%) in which the comparison was judged louder than the target is plotted as a function of the comparison level (dB SPL). The target level was always 60 dB SPL. Panels A and B show results using a precursor level of 70 dB SPL with filter slopes of 24 and 96 dB/octave, respectively. Panel C replots the on-frequency and no precursor conditions from panel B for ease of comparison with Panel D, which shows data using a precursor level of 85 dB SPL and filter slopes of 96 dB/octave. Error bars represent 1 s.e. of the mean.

To confirm the statistical significance of the trends described above, within-subjects analyses of variance (ANOVA) were carried out with Huynh-Feldt corrections for lack of sphericity applied where appropriate, using the fitted PSEs as the dependent variable. In the first analysis considering just the conditions with the 70-dB precursor, the factors were filter slope (24 or 96 dB/oct) and precursor (6 levels – 5 frequencies or no precursor). Significant main effects were observed for both precursor [$F(5,30) = 8.86$; $p = 0.001$] and filter slope [$F(1,6) = 6.94$; $p = 0.039$]. There was also a significant interaction between filter slope and precursor type [$F(5,30) = 3.24$; $p = 0.019$]. A planned comparison found a significant difference between PSE in the no-precursor condition and the PSE in the on-frequency condition [$F(1,6) = 14.5$; $p = 0.009$]. In addition, when the no-precursor condition was removed, contrast analysis revealed a quadratic trend was revealed for precursor frequency [$F(1,6) = 23.4$; $p = 0.003$]. These two findings indicate that the precursor affected loudness judgments and that the effect appeared to be frequency selective, with the effect decreasing with increasing frequency distance between the precursor and the target frequency. Although the effect of filter slope and its interaction with precursor frequency reached significance, the effects appear small and not easily interpretable.
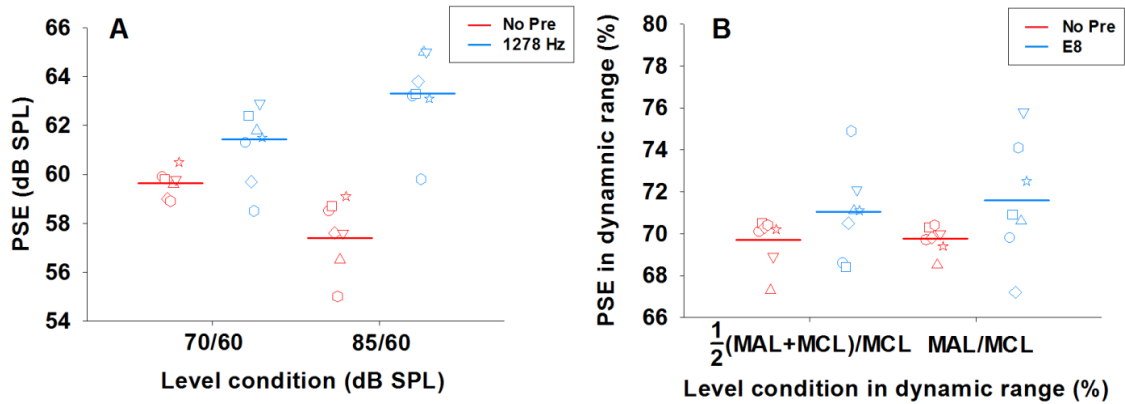
FIG. 2.3. Derived PSE for the individual subjects. Panel A shows PSEs for the normal-hearing subjects, and Panel B shows the results from CI users. Different symbols denote different subjects in the two panels, but there is no relationship between the symbols across the two panels. Symbols of CI users are indicated in Table 2.1. The levels of precursor and target (precursor/target) are shown on the x-axis. The results from no-precursor baseline conditions are shown as red unfilled symbols, and those from the on-frequency precursor condition are shown in blue unfilled symbols. The horizontal bars indicate the mean of each condition. In panel A, the difference in LCEs for the 70- and 85-dB precursor conditions was significant [t(6) = 5.08, p = 0.002]. However, in panel B, no significant effect of precursor level was found [t(6) = -0.207; p = 0.843].

To assess the effect of precursor level, the difference in PSE between the no-precursor condition and the on-frequency precursor condition was calculated from the data from session 2 (70 dB SPL precursor) and session 3 (85 dB SPL precursor). These differences, which represent the effect of the precursor on the loudness comparison, or LCE (in dB), were subjected to a paired-samples (within-subjects) t-test. As illustrated in

20

Fig. 2.3A, the difference in LCEs, which were 1.52 dB and 5.74 dB for the 70- and 85-dB precursor, respectively, was significant [$t(6) = 5.08$, $p = 0.002$].

One puzzling aspect of the data is that the larger LCE with the higher-level precursor is not just due to the higher PSE in the precursor condition, but seems to be also due to the lower PSE in the no-precursor condition. It is not clear why the no-precursor PSE was lower in the session that tested the higher-level precursor. It is conceivable that having blocks with the higher-level precursor interspersed with the no-precursor blocks led to an "over-compensation" of responses in the no-precursor blocks, in order for subjects to keep the overall number of "louder" and "quieter" responses more equal, when averaged over the session. However, the effect was relatively small, and further study would be needed to test this speculation.

In summary, significant LCE was observed in normal-hearing listeners. The effect exhibited frequency selectivity: it was greatest when the precursor was at the same frequency as the target and decreased with increasing spectral distance between the precursor and the target. The effect was also level-dependent, as it was greater for the 85-dB precursor than for the 70-dB precursor. Although the effect of filter slope reached statistical significance when all conditions were included, the overall amount of LCE and the effect of frequency separation between precursor and target were similar for both filter slopes tested.

2.3 Experiment 2: loudness context effect in cochlear-implant users

2.3.1 Methods

Subjects

Seven post-lingually deafened CI users participated in this study and were compensated
for their time. Information regarding the individual CI users is provided in Table 2.1. All
participants provided written informed consent, and all protocols were approved by the
Institutional Review Board of the University of Minnesota.

Table 2.1 CI patients information

Subject code	Gender	Age (Yrs)	CI use (Yrs)	Etiology	Duration HL prior to implant (Yrs)	THS* (μA)	MCL* (μA)	MAL* (μA)
D02(△)	F	63.9	12.1	Unknown	1	73	356	538
D10(○)	F	59.4	10.8	Unknown	8	68	509	630
D19(□)	F	54.1	9.4	Unknown	11	86	475	933
D24(◇)	M	63.3	5.9	Unknown progressive	27	91	413	505
D28(⬡)	F	64.6	10.6	Familial Progressive SNHL	27	186	766	1400
D33(▽)	M	74.4	1.0	Noise Exposure; Trauma	<1	55	637	900
D36(☆)	F	54.5	1.5	High Fever	Unknown	173	863	927

* The level was measured on target electrode (electrode 8)

Stimuli

The design was similar to that of Experiment 1. All the stimuli were delivered directly to the Internal Cochlear Stimulator (ICS) system based on a clinical research platform, BEDCS, provided by Advanced Bionics. All the center frequencies in Experiment 1 were converted to corresponding electrodes as shown in Fig. 2.1B (compare left and right y-axis labels). The durations of all the signals and gaps were exactly the same as those used in Experiment 1, with the exception that no onset and offset ramps were used. All stimuli were presented as trains of 32 µs/phase, cathodic-first biphasic pulses, presented in monopolar mode at a rate of 2000 pulses per second (pps).

Procedure

Before the experiment, for each selected electrode of each subject, three parameters were measured to calculate the level of stimuli. The three parameters were threshold (THS), most comfortable level (MCL) and maximum acceptable level (MAL). Stimuli were 200-ms pulse trains. The THS was measured using a three-interval, three-alternative forced-choice (3IFC/3AFC) procedure with a two-down, one-up adaptive tracking rule and correct-answer feedback. The THS estimates from two tracks were averaged to obtain a final THS value for each electrode and each subject. The MAL was measured using a one-up, one-down adaptive tracking procedure in which the sound was presented, followed by the question "Was it too loud?" A subject's "no" and "yes"

23

choices led to increases and decreases in signal level, respectively. The track terminated when the subject had responded that the intensity was too loud six times, and the average current level at the last six reversals was calculated. The MCL estimates from two tracks were averaged to obtain a final value of MCL for each electrode for each subject. The procedure to obtain MCL was the same as that for the MAL, except that the subject's question was "Was the sound medium loud/comfortable?"

A similar training session with the same criteria was set up as in Experiment 1. The target was always presented at 70% of the dynamic range (DR) of MCL in units of μA, and the comparison level was selected from 64%, 66%, 68%, 72% ,74% and 76% DR of MCL, based on pilot data. Within one training block, 10 repetitions for each level were presented in random order. No precursor was included. Feedback was provided and listeners were required to reach 80% correct to proceed to the actual experiment. All of the participants achieved this level of performance within two blocks of training.

The task of subjects was again to compare the loudness of the two brief sounds, the target and comparison. In the first session, a no-precursor reference condition and five precursor conditions (with precursors presented to electrode E2, E5, E8, E11, or E14) were tested. The target level was 70% DR of MCL, and the precursor (if present) was presented at MCL. In the second and third sessions, two conditions (no-precursor reference and E8) and two level relationships were investigated. The target level was fixed at MCL for both sessions. In the second session, the precursor level was the mean value of MCL and MAL in μA. In the third session the precursor was presented at MAL.

Each condition was repeated five times in random order in all three sessions, which resulted in 30 blocks in the first session and 10 blocks in each of the last two sessions. There were seven comparison levels symmetrically distributed around the target level in all sessions, from 64% to 76% DR of MCL in the first session, and from 94% to 106% in the second and third session, with a stepsize of 2% DR. Each comparison level was run 10 times in each block, for a total of 70 trials in each block and 50 repetitions of each stimulus per subject. No feedback was provided in the test sessions. The entire experiment lasted about 6-8 hours, divided into 2-hour sessions.

2.3.2 Results and discussion

Fig. 2.4 shows the mean results of Experiment 2. Panels A and B show the results with a MCL precursor and a target at 70% DR of MCL. Panels C and D show the data with the precursor at the level corresponding to the mean of MCL and MAL, and at MAL respectively. Different symbols represent the different precursor electrode numbers, as shown in the legend.

In general, some trends found in CI users were similar to those in normal-hearing subjects, with the presence of the precursor resulting in the target being judged louder than the comparison at equal levels. However, in contrast to the findings with normal-hearing listeners, the higher-level precursor did not result in a larger LCE.
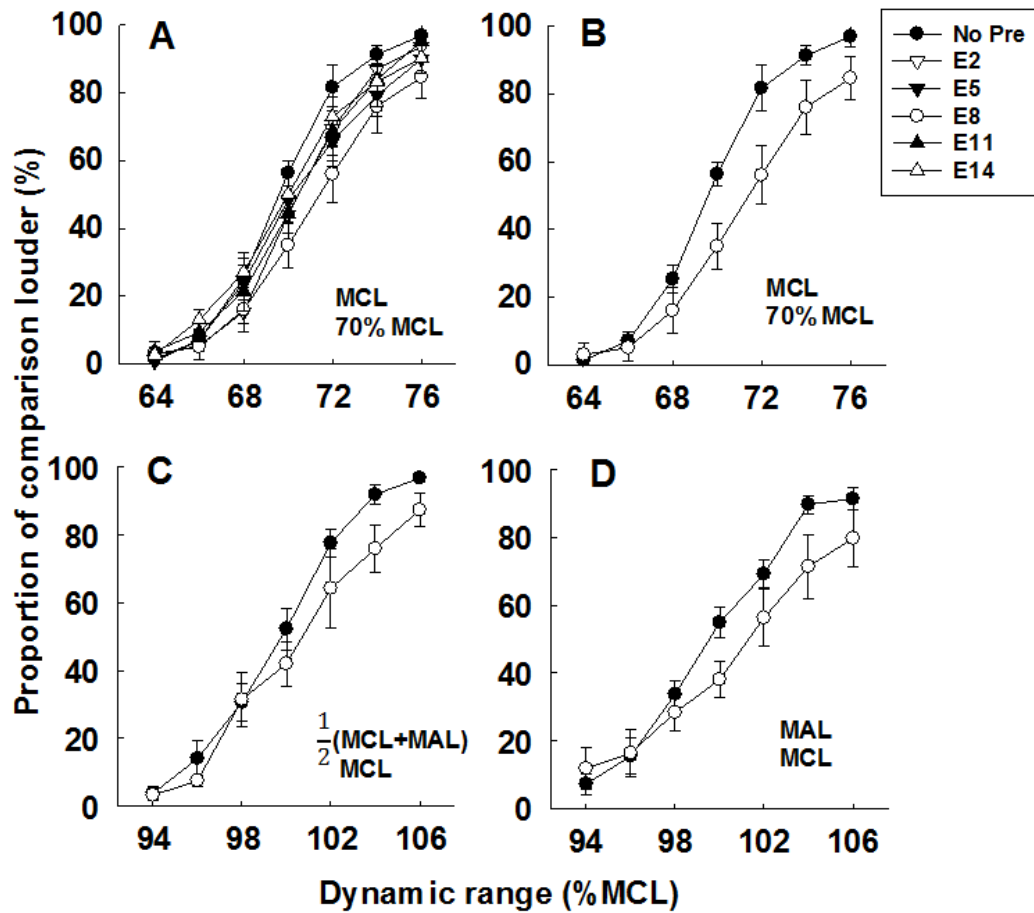
FIG. 2.4. Mean results from CI users. The proportion of trials (%) in which the comparison was judged louder than the target is shown as a function of the comparison level. Panels A and B show results from the first session, with the target level at 70% DR of MCL and precursor level at MCL. Panels C and D show results from the second and third sessions, respectively, with the target presented at MCL and the precursor presented at the midpoint between MCL and MAL (C) or at MAL (D). Error bars represent ±1 s.e. of the mean.

As in Experiment 1, a probit analysis was carried out on the data from the individual CI users in each condition, and the PSEs were derived from those fits. A one-way within-subjects ANOVA on the PSE data revealed a significant main effect for the precursor at MCL (Fig. 2.4A) [$F_{(5,30)}$ = 2.76; p = 0.043]. A planned analysis comparing the on-precursor condition with the no-precursor condition revealed a significant effect [$F_{(1,6)}$ = 11.7; p = 0.021]. Also, contrast analysis from an ANOVA with only the precursor conditions revealed a significant quadratic trend of electrode number [$F_{(1,6)}$ = 12.0; p = 0.013], reflecting the observation that the amount of LCE decreased with increasing electrode distance from the target. Considering the individual data, only one of the seven CI users showed a negative effect, with a lower PSE in the on-frequency precursor condition than in the no-precursor condition.

To assess the effect of precursor level, a paired-samples t-test was used to compare the LCE with the precursor at MAL and at the mean of MCL and MAL, as shown in Fig. 2.3B. No significant effect of precursor level was found [$t_{(6)}$ = -0.207; p = 0.843], reflecting the similar difference between the precursor and no-precursor conditions shown in Fig. 2.3B and seen also by comparing Figs. 2.4C and 2.4D. Considering the expansive or at least linear loudness growth function of CI users with current level in µA, the increment in the current level of the precursor from the MCL/MAL midpoint to MAL should have resulted in a considerable change in loudness (Hong and Rubinstein, 2006). Yet, this relatively large change in the presumed loudness of the precursor failed to produce a change in the size of LCE.

In summary, a significant LCE was observed in CI users. In line with earlier data using artificial vowel stimuli (Chapter 5), the data suggest that some auditory context effects can be observed in CI users. In addition, CI users also demonstrated some spectral (or cochlear spatial) selectivity, in that the effect was greatest when the precursor and target were presented to the same electrode. One apparent difference between the data from normal-hearing listeners and CI users is the apparent lack of a level effect in the CI users. However, the lack of a level effect, along with any conclusions about the size of LCE, is tempered by the fact that a direct comparison between normal-hearing listeners and CI users is made difficult by the different units (dB SPL vs. μA) and by the uncertain nature of the relationship between these variables and loudness. The final section attempts to provide a more quantitative comparison of the data from normal-hearing listeners and CI users by equating the results in terms of dynamic range.

2.4 Comparing loudness context effects in normal-hearing listeners and cochlear-implant users

To provide a more direct comparison between the LCE measured in normal-hearing listeners and CI users, we scaled the amount of LCE for both groups, relative to their respective dynamic ranges. For this calculation, the currents in μA of the CI users were converted to dB values (re: 1 μA). Then, the differences in dB (or ratio in μA) between

the PSEs with and without precursors were divided by the total dynamic range in dB of the individual CI users, or by an assumed dynamic range of 100 dB for the normal-hearing listeners. The resulting ratio was then treated as a percentage. For instance, a 10-dB difference in PSE for the normal-hearing group, given their 100-dB dynamic range, would be regarded as a 10% PSE. Fig. 2.5A shows the mean normalized PSE shift for the two groups, calculated in this manner.

For the normal-hearing listeners, the small level-difference condition refers to the condition in which the precursor was 70 dB SPL, and the large level-difference condition refers to the condition in which the precursor was 85 dB SPL (both with the 96 dB/octave filter slopes and a target level of 60 dB SPL). For the CI users, the small level-difference condition refers to the condition in which the precursor was presented at a level corresponding to the midpoint between MCL and MAL, and the large level-difference condition refers to the condition in which the precursor was presented at MAL (in both cases the target was presented at MCL).

A mixed-model ANOVA on the normalized PSEs with group (normal-hearing or CI) as a between-subjects factor and level difference (small or large) as a within-subjects factor revealed a significant effect of group [$F(1,12) = 18.1$; $p = 0.001$], a significant effect of level difference [$F(1,12) = 39.4$; $p < 0.001$], and a significant interaction [$F(1,12) = 29.5$; $p < 0.001$], emphasizing the observation that the normalized LCE seems generally smaller in the CI group, and that there is less effect (if any) of level difference in the CI group.
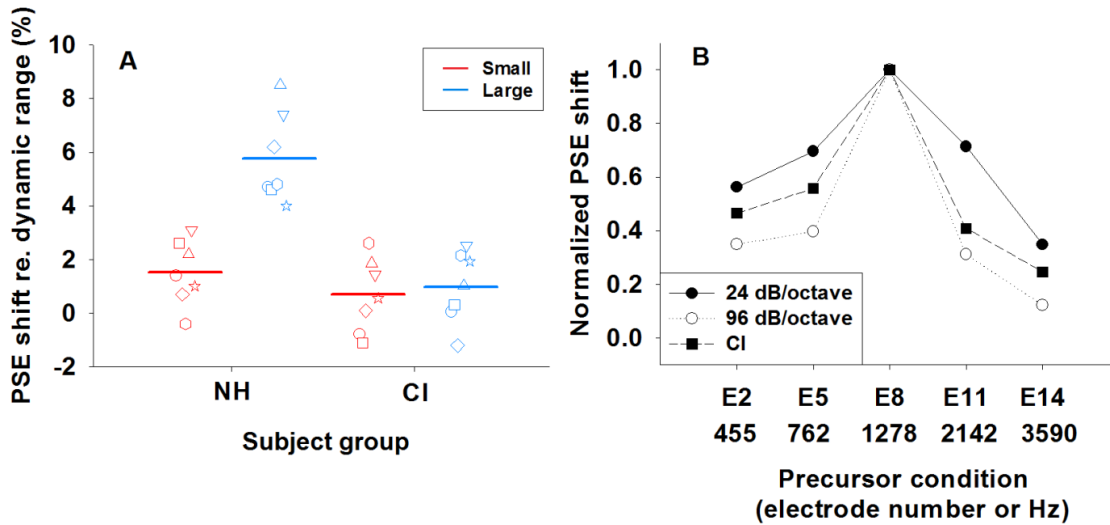
FIG. 2.5. Normalized PSE shifts due to the presence of the precursor in normal-hearing listeners and CI users. In Panel A, the effects of the level difference between the precursor and the target (red: small level difference; blue: large level difference) are compared in normal-hearing listeners and CI users, as a proportion of the overall dynamic range. Each symbol represents data from individual subjects. Symbols of CI users are indicated in Table 2.1. The horizontal bars indicate the means in each condition. A mixed-model ANOVA on the normalized PSEs with group (normal-hearing or CI) as a between-subjects factor and level difference (small or large) as a within-subjects factor revealed a significant effect of group $[F(1,12) = 18.1; p = 0.001]$, a significant effect of level difference $[F(1,12) = 39.4; p < 0.001]$, and a significant interaction $[F(1,12) = 29.5; p < 0.001]$. Panel B shows the mean frequency selectivity of LCE in normal-hearing listeners and CI users. The amount of LCE is normalized to the maximum amount of LCE for each group. Different symbols represent different conditions and subject groups, as shown in the legend.

30

Previous studies have discussed the frequency selectivity of potential underlying processes (Elmasian and Galambos, 1975; Marks, 1994; Oberfeld, 2007), and have concluded that maximal LCE occurs when all stimuli are presented at the same (or similar) frequencies. We observed similar results in both the normal-hearing listeners and CI users in the present experiments. To compare the frequency selectivity across the groups, we used the normalized LCE, as calculated above and plotted it in Fig. 2.5B, as a proportion of the maximum amount of LCE, observed in the on-frequency conditions. To obtain these values, we first obtained the PSEs (in dB, as described above) for each subject in each precursor condition. For each precursor condition, we then individually normalized the PSE as: $PSE_{ni} = (PSE_i - PSE_{ref})/(PSE_{on} - PSE_{ref})$, where PSE_i is the original PSE of condition i, and PSE_{on} and PSE_{ref} are PSE from on-frequency precursor condition and no-precursor reference condition, respectively. Finally, the averaged values for each subject group were calculated. For the two filter-slope conditions with the normal-hearing listeners, the outcomes are as expected: narrower excitation patterns result in greater frequency selectivity. Interestingly, the frequency selectivity observed in the CI group is quite similar to that found in the normal-hearing group, with frequency selectivity falling between the two curves from the normal-hearing listeners.

2.5 General discussion

This study measured how the loudness relationship between two brief (200-ms) sounds (the target and comparison), spaced 2 s apart, is affected by the presence of a longer (1-s) precursor, preceding the target and separated by a gap of 50 ms. Both normal-hearing listeners and CI users were tested with the precursor presented at various spectral (or electrode) positions relative to the target.

Our findings from Experiment 1, using normal-hearing listeners, are consistent with those of previous studies. A more intense precursor resulted in the target sound being judged louder than the comparison signal when they were presented at equal levels (e.g. Galambos *et al.*, 1972; Elmasian and Galambos, 1975), and an increase in precursor level resulted in an increased effect (Elmasian *et al.*, 1980; Zeng, 1994; Arieh and Marks, 2003a). Finally, the effect of the precursor depended on the frequency proximity of the precursor to the target and comparison, with the maximum effect occurring when the center frequencies of the precursor and target were the same.

In Experiment 2 using CI users, significant LCE and similar frequency-selectivity effects to those in normal-hearing listeners were found. The fact that LCE was observed at all in CI users suggests that at least part of LCE originates from a stage of processing higher than the cochlea. This observation implies that the MOC cannot be the only source of LCE. Thus, to the extent that LCE reflects auditory gain control, it must occur at least in part beyond the cochlea. However, one potentially important difference between the normal-hearing and CI results was that the LCE observed in CI users did not seem

dependent on precursor level, in contrast to the large level effects observed in normal-hearing listeners.

It is difficult to make quantitative comparisons between the results from normal-hearing listeners and those from CI users, because of the different units (dB SPL vs. μA), and the different (and uncertain) relationship between these units and the underlying neural responses and percepts. We provided one possible approach here, by normalizing the units in terms of overall dynamic range (on an individual basis for the CI users and with an assumption of 100 dB for the normal-hearing listeners). However, the conclusions based on these comparisons must be treated with caution. In addition, although the differences in level between the precursors and the targets in the CI users were substantial, it remains unknown whether the differences in the results were due to smaller CI effects, or because the differences in current levels between the precursor and the target were not sufficient to induce large effects. Future studies using wider ranges of level differences should help resolve this question.

Several theories have been proposed to explain aspects of LCE. Taking account of the fact that the loudness of a target is enhanced if the precursor is more intense than the target, and that its loudness is reduced if the precursor is less intense (Zwislock and Sokolich, 1974; Elmasian *et al.*, 1980), a "mergence hypothesis" was proposed, whereby the loudness of the target is derived from a weighted average of the intensities of both precursor and target (Elmasian *et al.*, 1980). The fact that the effect is only observed for precursor-target gaps of less than 400 ms provides an upper bound to the time window

associated with such mergence (Zwislock and Sokolich, 1974). This framework explained many aspects of LCE, except for the "mid-difference hump", that the maximal of effect size was gained only when the level difference was moderate (e.g. 20-30 dB), which was proposed by Oberfeld (2007). According to mergence theory, the effect size should increase monotonically with the level of precursor, in contrast to the data (Zeng, 1994; Plack, 1996b; Mapes-Riordan and Yost, 1999). To account for the mid-difference hump, Oberfeld (2008) proposed the "similarity model". The idea is that mergence in the auditory perceptual system will become more effective when two sounds are more similar perceptually. Therefore, if the precursor is presented at a level that is too different from that of the target, the mergence between the precursor and target would be weaker. With appropriately selected parameters, this model can quantitatively predict some of the LCE patterns observed in behavioral studies (Oberfeld, 2008).

As mentioned earlier, studies that use only a single frequency for the target and comparison sounds cannot distinguish between an enhancement of the target and a decrease in the loudness of the comparison (Scharf *et al.*, 2002). Studies using loudness comparisons across frequency have resulted in the proposal of a dual-process (Arieh and Marks, 2003b; Oberfeld, 2007). The first process is described as a fast-onset and fast-decay process, which is basically the "similarity model" discussed above. The second process is assumed to be a fast-onset, slow-decay process, which is responsible for the reduction of the comparison signal. This process has been termed "induced loudness reduction" by Scharf *et al.* (2002) and could last for seconds (Arieh and Marks, 2003b;

Arieh *et al.*, 2005). According to Arieh and Marks (2003b), this effect could monotonically increase to as much as 11 dB within about 1 s and then level off. In Elmasian and Galambos (1975), the amount of loudness enhancement was about 4 dB, with the precursor and target tones presented at 80 and 70 dB SPL respectively, which was comparable to what we measured here. Considering the short gap (100 ms) between the precursor and target in their study, two processes may have been partially cancelled out by each other, which presumably also occurred in the current study. In our experiment, the effect of the precursor reached a maximum of about 6 dB in normal-hearing listeners. The equivalent effect in CI users appeared to be smaller, when calculated in comparable units (based on overall dynamic range), and the amount of enhancement was less (or not) dependent on precursor level. This outcome, which suggests that at least one of the mechanisms underlying LCE may be different or absent in CI users, is intriguing. Further insights into the respective contributions of the two processes in CI users might be gained by applying the method proposed by Oberfeld (2007) to separate the two processes. However, any direct comparison between the results of normal-hearing listeners and CI users must be treated with caution, given the uncertainties surrounding the mapping of acoustic sound pressure level in normal-hearing listeners to electrical current in CI users. Further insights may be gained by tracking the time course of LCE in these two populations and by separating the effects of the precursor on the target and the comparison stimulus.

CHAPTER 3: INDUCED LOUDNESS REDUCTION IN ACOUSTIC AND ELECTRIC HEARING[1]

Abstract

The loudness of a tone can be reduced by preceding it with a more intense tone. This effect, known as induced loudness reduction (ILR), has been reported to last for several seconds. The underlying neural mechanisms are unknown. One possible contributor to the effect involves changes in cochlear gain via the medial olivocochlear complex (MOC) efferents. Since cochlear implants (CIs) bypass the cochlea, investigating whether and how CI users experience ILR should help provide a better understanding of the underlying mechanisms. In the present study, ILR was examined in both normal-hearing listeners and CI users by examining the effects of an intense precursor (50 or 500 ms) on the loudness of a target (50 ms) as judged by comparing it to a spectrally remote comparison sound (50 ms). The interstimulus interval (ISI) between the precursor and the target was varied between 10 and 1000 ms to estimate the time course of ILR. In general, the patterns of results from the CI users were similar to those found in the normal-hearing listeners. However, in the short-precursor short-ISI condition, an enhancement in the loudness of target was observed in CI subjects that was not present in the normal-hearing listeners, consistent with the effects of an additional attenuation present in the normal-hearing listeners but not in the CI users. The results suggest that the MOC may play a role but that it is not the only source of ILR.

3.1 Introduction

As with many other aspects of perception, the loudness of a sound depends not only on its physical properties, but also on the context in which the sound is presented. One loudness context effect that received early attention was termed "loudness enhancement." This effect was demonstrated by presenting three tones at the same frequency: a precursor, a target tone, and a comparison tone presented in sequence. Listeners were instructed to adjust the level of the comparison tone to match its loudness to that of the target. When the precursor was more intense than the target, the listeners often adjusted the level of the comparison tone to be higher than that of the target at equal loudness, leading to the conclusion that the precursor had enhanced the loudness of the target tone (Zwislocki and Sokolich, 1974; Elmasian and Galambos, 1975).

The effects of loudness recalibration (LR; e.g. Arieh and Marks, 2003b) or induced loudness reduction (ILR; e.g. Nieder *et al.*, 2003) are also measured by using a precursor at the same frequency as the target, but the comparison tone is at a frequency remote to that of the target and precursor. Using this paradigm, the opposite effect is usually reported: a precursor that is more intense than the subsequent target tone can reduce the target's loudness (Marks, 1994). Mapes-Riordan and Yost (1999) found that the effect was strongest when the precursor was 10-20 dB higher than the target and that the effect was smaller or non-existent when the level difference exceeded 40 dB or when the target was presented at or near its detection threshold.

Scharf *et al.* (2002) suggested that the results from earlier loudness enhancement studies should be reinterpreted in light of the ILR findings. They noted that when all three tones were presented at the same frequency (as in the traditional enhancement studies), it may be that the precursor reduced the loudness of the comparison tone rather than enhancing the loudness of the target tone. This reinterpretation was supported by Arieh and Marks (2003b), who found that ILR did not occur immediately after a precursor, but reached a maximum at a delay of around 1 s, and lasted for at least 3s (Arieh and Marks, 2003b), which was longer than the gaps between precursor and comparison in the previous loudness enhancement studies.

To further test this interpretation, Oberfeld (2007) measured loudness context effects with the traditional three-tone paradigm where all tones were presented at the same frequency, along with a novel four-tone paradigm, in which a fourth tone was presented at a different frequency to measure more directly the perceived loudness of the third (comparison) tone. Results from Oberfeld (2007) showed that not only was the target tone enhanced in loudness, but the loudness of comparison was reduced, in line with a "dual-process" model proposed by Arieh and Marks (2003b). In this model, a fast-onset and fast-decay "enhancement" process, accompanied by a fast-onset and slow-decay reduction process, contribute to the "loudness enhancement" of the target.

One way in which the rapid enhancement process could occur is through "assimilation", or "over-integration" of the loudness of the precursor with that of the target (Plack, 1996a). This is thought to occur only when the precursor and target are

38

perceptually similar (Oberfeld, 2008). Some evidence in favor of the assimilation hypothesis is that a decrement in the judged loudness of the target can occur when the precursor is lower in level than the target (Elmasian *et al.*, 1980). One potential mechanism of ILR involves the medial olivocochlear (MOC) efferent system, which can reduce cochlear gain by controlling the action of the outer hair cells (Stankovic and Guinan, 1999; Guinan, 2006). Although the time constants associated with the MOC fast effect are not thought to extend to several seconds, the slow effect of MOC may potentially contribute to loudness changes (Cooper and Guinan, 2003).

Cochlear implants (CI) may provide a way to examine the role of the MOC efferents in ILR. If the MOC system is the sole source of ILR, then ILR should not be observed in CI users. In Chapter 2, loudness context effect in CI users was investigated using the traditional three-stimulus technique, with all three stimuli presented to the same electrode. Both similarities and differences were found between the results from CI users and those from normal-hearing listeners. In particular, it was found that, in both normal-hearing and CI subjects, a more intense precursor resulted in the target sound being judged louder than the comparison signal when they were presented at equal levels, and frequency selectivity was observed in this effect. The effect size was found larger with stronger precursor in results from normal-hearing subjects, whereas it was not significantly affected by precursor level in CI users. However, because the target and the comparison tone were presented to the same electrode, it was not possible to separate potential loudness enhancement from ILR effects. In the present study, we measured ILR

in both normal-hearing listeners and CI users, with a moderately intense precursor and a fixed-level target, presented at the same frequency (or same electrode), and a varying-level comparison, presented at a spectrally remote frequency (or electrode) from the precursor and the target. Listeners were asked to compare the loudness of the comparison with that of the target. Pure tones were used as stimuli for the normal-hearing listeners (Experiment 1), whereas fixed-rate electrical pulse trains were presented directly to the CI users (Experiment 2).

3.2 Experiment

3.2.1 Methods

Subjects

Normal-hearing listeners. Ten listeners (3 males, 7 females) participated in this experiment and were compensated for their time. Their ages ranged from 19 to 63 years (mean age 25.3 years, only one subject older than 45). All listeners had audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz.

Table 3.1 CI user information. MCL denotes maximum comfortable loudness.

Subject code	Gender	Age (Yrs)	CI use (Yrs)	Etiology	Duration of hearing loss prior to implant (Yrs)	Absolute threshold (μA)	MCL (μA)
D02(★)	F	63.9	12.1	Unknown	1	73	356
D10(◆)	F	59.4	10.8	Unknown	8	68	509
D19(■)	F	54.1	9.4	Unknown	11	86	475
D24(▲)	M	63.3	5.9	Unknown progressive	27	91	413
D28(⬢)	F	64.6	10.6	Familial Progressive SNHL	27	186	766
D33(▼)	M	74.4	1.0	Noise Exposure; Trauma	<1	55	637
D36(✦)	F	54.5	1.5	High Fever	Unknown	173	863

Cochlear-implant users. Seven post-lingually deafened CI users participated in this study and were compensated for their time. Information regarding the individual CI users is provided in Table 3.1.

Stimuli

Normal-hearing listeners. Figure 1 provides a schematic diagram of the stimuli used in this experiment. In each trial, three pure tones, a precursor at 1278 Hz, a target also at 1278 Hz, and a comparison at 455 Hz, were presented in sequence. The frequencies of test tones were selected from the standard Advanced Bionics 16-channel map for CIs, corresponding to the center frequencies of the channels mapped to electrodes E8 and E2,

respectively. The total duration of precursor could be either 50 or 500 ms, and the duration of two following tones was always 50 ms. The ISI between the precursor and the target could be 50, 250, or 1000 ms, and the ISI from the target to the comparison was fixed at 1 s. All the stimuli used in Experiment 1 were gated on and off with 10-ms raised-cosine ramps. The levels of the precursor and the target were always 75 and 60 dB SPL, respectively. The level of comparison tone was modified according to subjects' responses in an adaptive procedure. The stimuli were generated digitally and played out diotically from a LynxStudio L22 24-bit soundcard at a sampling rate of 48 kHz via Sennheiser HD650 headphones to listeners seated in a double-walled sound-attenuating chamber.
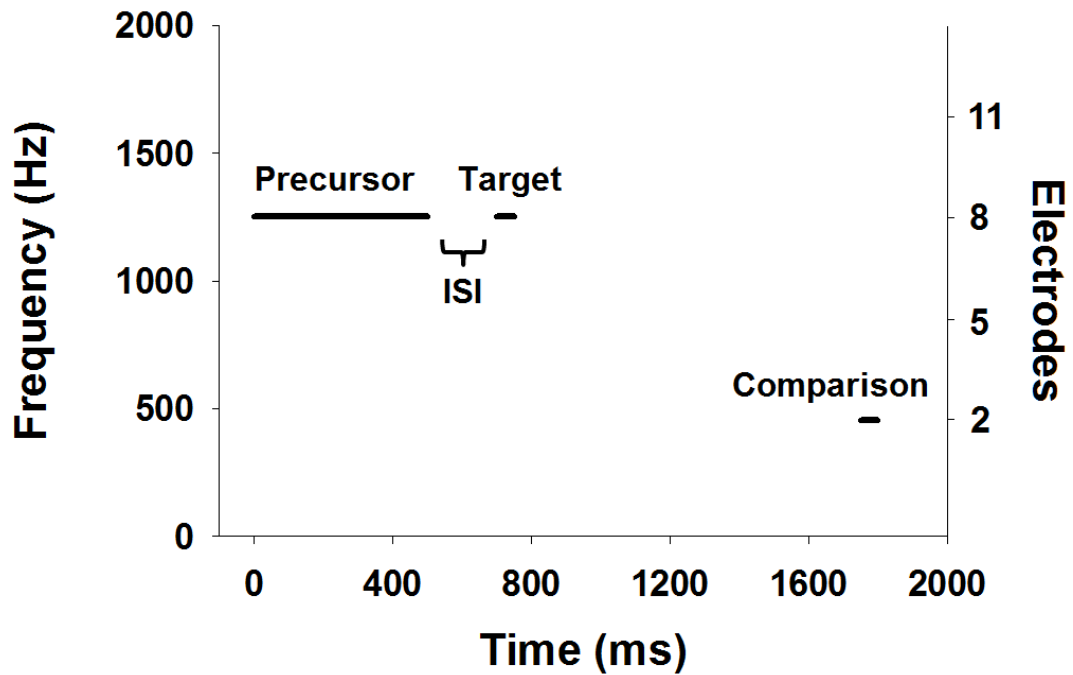
FIG. 3.1. Schematic diagram of the stimuli used in Experiment 1 and 2. In Experiment 1, the precursor was a pure tone at 1278 Hz, presented for either 50 or 500 ms. A 50-ms target tone at the same frequency followed it after a gap of 50, 250, or 1000 ms. After a further 1-s gap, a 50-ms comparison tone was presented at 455 Hz. In Experiment 2, the stimuli had the same overall durations, but were presented as pulse trains to different electrodes. The precursor and target were presented to a middle electrode (E8), whereas the comparison was presented to a more apical electrode (E2).

Cochlear-implant users. The stimuli were similar to those used for normal-hearing listeners (Fig. 3.1) All the stimuli were delivered directly to the Internal Cochlear Stimulator (ICS) system based on a clinical research platform, BEDCS, provided by Advanced Bionics. The durations of all the signals and gaps were the same as those used

with normal-hearing listeners, with the exception that no onset and offset ramps were used. The stimuli were all pulse trains, consisting of 32 µs/phase, cathodic-first biphasic pulses, presented in monopolar mode at a rate of 2000 pulses per second (pps). Electrode 8 was selected to present the precursor and target, and the comparison was presented from Electrode 2. These electrodes correspond to frequencies of 1278 Hz and 455 Hz, respectively, according to the standard Advanced Bionics 16-channel map. Presentation levels were determined for each subject individually by setting the target level at 70% of the dynamic range (DR), defined as the range from absolute threshold (THS) to most comfortable level (MCL) in µA, and by setting the precursor level at MCL. The level of the comparison stimulus was varied using the same adaptive procedure as for the normal-hearing listeners in units of percent in DR. Subjects were again asked to compare the loudness of the two brief sounds, the target and the comparison, with the same adaptive procedure as was used for the normal-hearing listeners, in all 28 blocks (7 conditions, 4 repetitions with different pairs of starting points).

Procedure

Normal-hearing listeners. Listeners were instructed to ignore the first stimulus, the precursor (if present), and to judge whether the comparison was louder or quieter than the target. A total of seven conditions (two precursor-duration conditions combined with three ISI conditions, as well as a baseline condition with no precursor) were tested. An interleaved tracking procedure (Jesteadt, 1980; Leek et al., 1991), consisting of a 2-down

1-up track and a 2-up 1-down track, was employed to estimate to the point of subjective equality (PSE) in loudness between the target and comparison tones. The two adaptive procedures track the 70.7% and 29.3% points on the psychometric function, so that the mean of the two tracks approximates the 50% point on the psychometric function, i.e., the point at which the target and comparison were judged to be equally loud. Each trial was selected at random with equal priori probability from one of the two tracks. Each condition was repeated four times in random order, with a different random order selected for each subject and each repetition. However, for each repetition, a different pair of starting points was selected for the tracking procedure (51/60, 54/63, 57/66 and 60/69 dB SPL), to avoid any potential response bias generated by the starting points (e.g. Marks, 1994). In each track, the initial step size was 5 dB. The step size was reduced to 3 dB after the first two reversals, and to 2 dB after the fourth reversal. A block of trials ended when four reversals at the final step size occurred in both tracks. If the stopping rule for one track was met before the other, the "completed" track would continue, but the levels were not incorporated into the threshold estimates. The final measured threshold (i.e., the level at which the comparison was judged louder than the target 50% of the time) was the mean of last four reversal points from both tracks.

Cochlear-implant users. The experimental procedures for the CI users were the same as those used for the normal-hearing listeners, with the following exceptions. First, the THS and MCL levels for each subject were determined for each subject individually using 200-ms pulse trains on each of the test electrodes (E2 and E8), as described in Chapter 2.

Second, the different pairs of starting point for each of the four repetitions of the adaptive

tracking procedure were 55/70, 60/75, 65/80 and 70/85% DR. Third, the initial step size

in the adaptive procedure was 5% DR, which was reduced to 3%DR after two reversals

and to 2% DR after four reversals.

3.2.2 Results

Normal-hearing listeners. A one-way within-subjects ANOVA revealed no significant

difference in comparison level when taking starting point as the main factor [$F(3,27) =$

1.13, p = 0.353], suggesting the starting points of the adaptive procedure did not affect

the actual matches of subjects, so the results were averaged across starting points for the

remainder of the analysis. The mean results are presented in the left panel of Fig. 3.2. The

black bar at the left represents the results from baseline condition with no precursor. The

mean level of the 455-Hz comparison tone was 58.3 dB SPL, which was not significantly

lower than the 60 dB SPL of the 1278-Hz target tone [paired-samples t-test: t(9) = -1.08,

p = 0.307]. The lack of a level difference is expected, given the relatively similar

expected loudness of a 455-Hz and 1278-Hz tones, based on the 60-phon curve from

current iso-loudness contours (ISO:226, 2003). The presence of the precursor generally

reduced the level of the comparison tone at the PSE, indicating a reduction in the

loudness of the target tone, as expected. The right panel of Fig. 3.2 displays the amount

of loudness reduction, calculated simply by subtracting the level of the comparison in the

presence of the precursor from its level in the absence of the precursor. The maximum

ILR of about 7 dB was found for ISIs of 1 s for both the 50-ms and the 500-ms precursor. At shorter ISIs, the longer precursor continued to produce ILR, whereas the amount of ILR produced by the shorter (50-ms) precursor decreased with decreasing ISI, reaching an average of less than 1 dB at the shortest ISI of 50 ms.
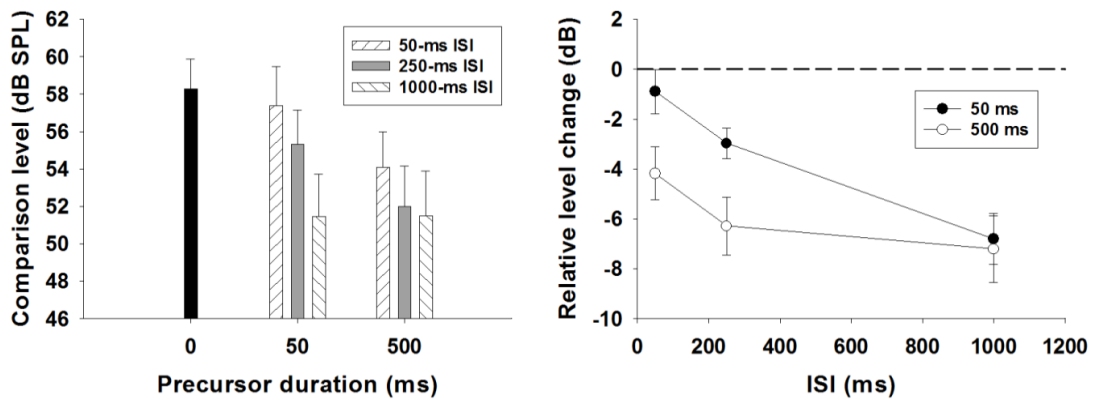
FIG. 3.2. Mean results from normal-hearing listeners. In the left panel, from left to right, the matched levels of comparison tone in no-precursor (baseline), short-precursor (50 ms) and long- precursor (500 ms) conditions are displayed. In the right panel, the relative level changes in each condition comparing to the baseline are shown. The filled and open circles represent results from conditions with 50-ms or 500-ms precursor, respectively. The dash line indicates the baseline, which is 0 dB. Error bars represent 1 s.e. of the mean.

The increase in ILR with increasing ISI out to 1 s is consistent with the results from previous studies (Arieh and Marks, 2003b; Nieder *et al.*, 2003). The maximum

effect of 7 dB is somewhat less than that reported in earlier studies (around 10 dB), although this may be due to the relatively small level difference we used between the 60-dB target and the 75-dB precursor. Most previous studies have used precursor levels of 80 dB SPL, with level differences between the precursor and target of 20 dB or more. The amount of ILR has generally been found to reach a maximum with a level difference of around 20-30 dB between the precursor and target (Mapes-Riordan and Yost, 1999; Oberfeld, 2007). A two-way within-subjects ANOVA was conducted, with the amount of ILR as the dependent variable and precursor duration and ISI as the two factors. A significant main effect was found for precursor duration [$F(1,9) = 9.59$, $p = 0.013$], in line with the observation that the longer precursor induced a larger effect overall. The main effect of ISI was also significant [$F(2,18) = 16.38$, $p < 0.001$], as was the interaction between precursor duration and ISI [$F(2,18) = 4.74$, $p = 0.022$], reflecting the observation that the effect of precursor duration was greatest at the smallest ISI and became much smaller at the longest ISI.

Cochlear-implant users. The individual matched levels of the comparison tone were converted into dB re. 1 µA and were then averaged to get the mean results. As in Experiment 1, a one-way within-subjects ANOVA revealed no significant difference in the response level with different starting points of the adaptive tracking procedure [$F(3,18) = 1.46$, $p = 0.26$], so the results were averaged across the different starting levels.

The mean results from CI users are shown in Fig. 3.3. In the left panel, the baseline condition with no precursor is shown with the black bar. It is not informative to compare the current levels of the comparison and target in the baseline condition, as they were presented to different electrodes, and so likely have different loudness-level relationships. In general, the effects of the precursor seem greater for the 50-ms precursor than for the 500-ms precursor. There was a trend for decreasing thresholds with increasing ISI for the 50-ms precursor, but the trend was less apparent for the 500-ms precursor. The effect of the precursor was again calculated by subtracting the comparison level at threshold in the no-precursor condition from the comparison level at threshold in the with-precursor conditions (right panel of Fig. 3.3). In one case (50-ms precursor and 50-ms ISI), the comparison level was higher in the presence of the precursor, suggesting some form of enhancement. In all other cases, the mean levels with the precursor were the same as, or lower than, the levels without the precursor, suggesting either no effect or a reduction in loudness.
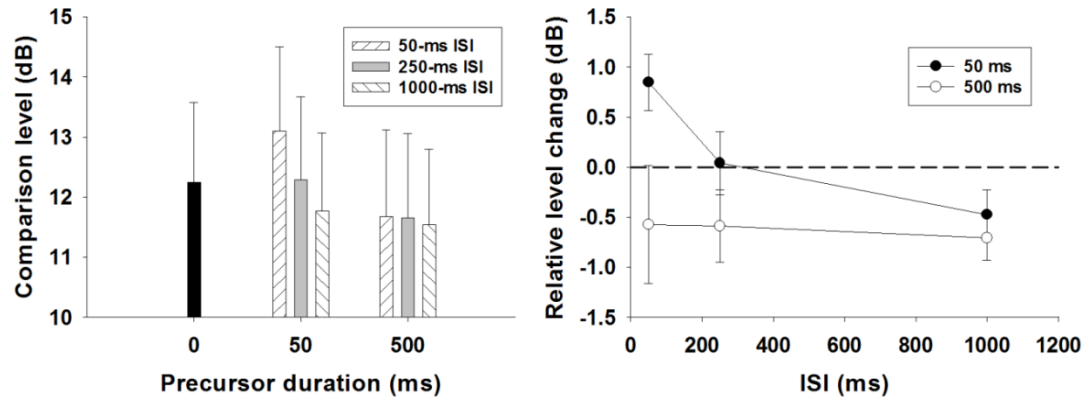
FIG. 3.3. Mean results from CI users. In the left panel, the individual matched levels of comparison tone were converted into dB re. 1 μA, and averaged to the mean matched levels. From left to right, the matched levels of comparison tone in no-precursor (baseline), short-precursor (50 ms) and long- precursor (500 ms) conditions are displayed. In the right panel, the relative level changes in each condition comparing to the baseline are plotted. The filled and open circles represent results from conditions with 50-ms and 500-ms precursor, respectively. The dash line indicates the baseline, which is 0 dB. Error bars represent 1 s.e. of the mean.

A two-way within-subjects ANOVA was performed with the amount of enhancement or reduction (in dB) as the dependent variable and precursor duration and ISI as the two factors. A significant main effect was obtained for precursor duration [$F(1,6) = 12.3$, $p = 0.013$]. The main effect of ISI failed to reach significance [$F(2,12) = 3.522$, $p = 0.063$], but the interaction was significant [$F(2,12) = 4.71$, $p = 0.031$], reflecting the greater effect of ISI for the 50-ms precursor than for the 500-ms precursor.

3.3 Comparison of results from normal-hearing and cochlear-implant subjects

When considered in isolation, the patterns of results from the CI group with each of the two precursors look reasonably similar to those found in the normal-hearing subjects: with the 50-ms precursor, the matched comparison level decreased with increasing ISI, and with the 500-ms precursor the effect of ISI was reduced. However, when comparing the absolute effects of the presence of the precursor, some differences between the data from the CI users and the normal-hearing subjects emerge. As shown in the right panel of Fig. 3.3, the mean difference between the precursor and no-precursor threshold levels is positive for the 50-ms precursor and 50-ms ISI, implying enhancement, rather than loudness reduction. In contrast, as shown in the right panel of Fig. 3.2, the normal-hearing listeners showed no enhancement in any of the conditions tested.

A more direct or quantitative comparison of the data from the normal-hearing and CI groups is hampered by the differences in overall dynamic range, and by the uncertainty regarding the appropriate units in which to compare the data. One way to provide such a comparison is to convert the amount of change in matching stimulus into a proportion of the overall dynamic range (Chapter 2). We calculated these normalized values by considering the dynamic range of the CI users to be difference between MCL and THS (in dB) for each subject individually, and then to convert any changes in level into a proportion of the dynamic range. For instance, if the overall dynamic range was 10 dB, then a change in the comparison level of 1 dB was considered a 10% change. For the normal-hearing listeners, the dynamic range was assumed to be 100 dB. Using these

conversions, the individual normalized effects of the precursors are shown in Fig. 3.4, with the normal-hearing listeners on the left and the CI users on the right.
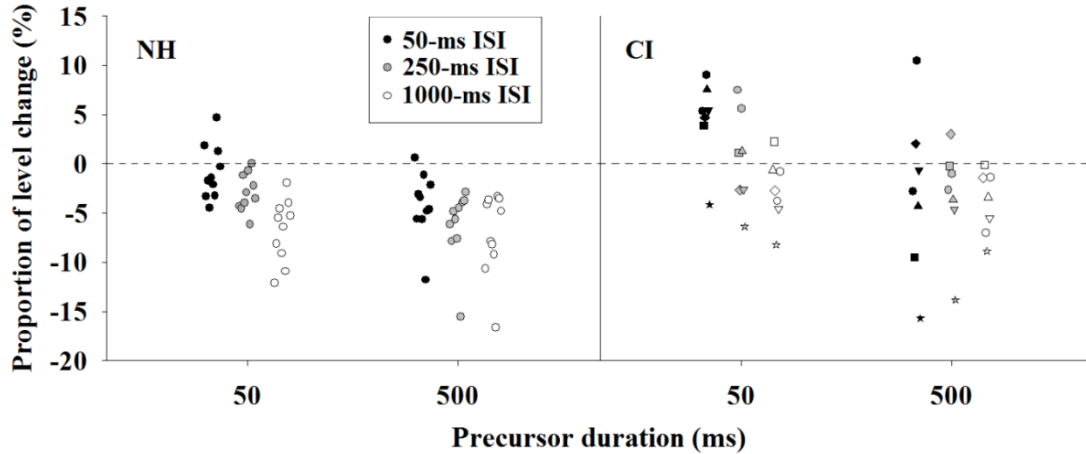
FIG. 3.4. Individual normalized proportion of level change in normal-hearing listeners and CI users. In the left panel, black, gray and unfilled circles indicate results from 50, 250 and 1000-ms ISI conditions, respectively. Each circle represents the normalized effect of precursor of an individual subject. The right panel shows individual results of CI users. The corresponding symbols of CI users are indicated in Table I. The dashed line indicates the baseline (0 %).

In the left panel of Fig. 3.4, individual results from normal-hearing subjects are displayed. In the 50-ms ISI and 50-ms precursor condition, only three out of ten subject showed some enhancement effect by the precursor, and the average proportion of level change is slightly below the baseline as shown in Fig. 3.2. In other conditions, almost no enhancement was observed, indicating ILR dominated the effect in those conditions. In

the right panel of Fig. 3.4, which shows the individual results from CI users, some enhancement effects were observed. Specifically, in the 50-ms ISI and 50-ms precursor condition, six out of seven subjects showed enhancement other than ILR, resulting in a significant enhancement effect, as confirmed by a one-sample t-test between the normalized level change and 0% [t(6) = 2.85, p = 0.029]. In the 50-ms ISI and 500-ms precursor condition, a large variation in the result was observed, indicating some individually different interactions between the long precursor and short target in CI users when ISI was short.

A mixed-model ANOVA on the normalized proportion of level change with group (normal-hearing or CI) as a between-subjects factor, and ISI and precursor duration as two within-subjects factors revealed a significant effect of group [F(1,15) = 4.69; p = 0.047], a significant effect of ISI [F(2,30) = 14.6; p < 0.001], and a significant effect of precursor duration [F(1,15) = 23.7; p < 0.001]. Significant interaction between ISI and precursor duration was also found [F(2,30) = 9.24; p = 0.001], whereas no significant interaction between subject group with each of them was observed. The main effect of group, and lack of significant interactions with group, support the observation that the effect of the precursors in the CI group was vertically shifted up relative to the effect found in normal-hearing listeners. This could be ascribed to a loss in the CI users of a fast-acting decrease in gain produced by the precursor.

3.4 Discussion

In the present study, we investigated ILR in both normal-hearing listeners and CI users. When considering the two precursor conditions in isolation, the patterns of results from the CI users look reasonably similar to those found in the normal-hearing listeners: with the 50-ms precursor, the matched comparison level decreased with increasing ISI, and with the 500-ms precursor the effect of ISI was reduced. However, when comparing the results with and without a precursor, some differences between the results from the CI users and the normal-hearing subjects emerge. For the CI users, in the 50-ms precursor and 50-ms ISI condition, the mean difference between the precursor and no-precursor threshold levels was positive, implying an enhancement effect on the loudness of the target, rather than loudness reduction. In contrast, the normal-hearing listeners showed no enhancement in any of the conditions tested.

The ability to distinguish between an enhancement of the target's loudness and a reduction of the comparison stimulus's loudness is only possible through the use of a different frequency (or electrode) for the comparison stimulus. The assumption is that any effects of the precursor will be frequency selective, and so will not extend to the frequency of the comparison stimulus. This assumption is generally well supported by the work of Marks and colleagues, who have shown that the effects of ILR, or loudness recalibration, are highly frequency selective and are reduced or absent once the two frequencies differ by more than about 15%, or a "critical band" (e.g. Marks, 1994). It is known that CI users generally exhibit much poorer frequency selectivity than normal-

hearing listeners (e.g. Zeng, 2004). An earlier study of spectral enhancement of vowels (Chapter 5) found that CI users showed generally less enhancement than normal-hearing listeners, but that the difference was reduced once poorer spectral resolution was simulated in normal-hearing listeners using vocoder techniques. It is therefore possible that the differences observed in the present study between normal-hearing listeners and CI users may be due to the CI users' poorer spectral resolution. This explanation seems unlikely to account for the whole effect, however, given the earlier results in Chapter 2. They found that loudness context effects in CI users decreased with increasing electrode difference between the precursor and target, and were generally negligible when the precursor was presented to electrode 8 and the target was presented to electrode 2. Thus, the reduction in spectral resolution in CI users is unlikely to account fully for the differences in ILR observed here.

Another difference between normal-hearing listeners and CI users, is the presence of the MOC efferent system in the normal auditory system (e.g. Liberman, 1988; Stankovic and Guinan, 1999). The CI results suggest less reduction in initial gain, relative to the normal-hearing listeners, leading to an enhancement effect for the short precursor with the shortest ISI. It is possible that the gain reduction "missing" from the CI data may reflect the absence of the MOC-induced gain reduction, which in this case led to some initial loudness enhancement for the CI users. The fact that ILR was observed at all in CI users is consistent with previous findings suggesting that loudness context effects

55

(Chapter 2), as well as auditory enhancement effects (Goupell and Mostardi, 2012; Chapter 5) cannot be mediated solely by the MOC efferent system.

CHAPTER 4: SPECTRAL MOTION CONTRAST AS A SPEECH CONTEXT EFFECT[1]

Abstract

Spectral contrast effects may help "normalize" the incoming sound and produce perceptual constancy in the face of the variable acoustics produced by different rooms, talkers, and backgrounds. Recent studies have concentrated on the after-effects produced by the long-term average power spectrum. The present study examined contrast effects based on spectral motion, analogous to visual-motion after-effects. In Experiment 1, the existence of spectral-motion after-effects with word-length inducers was established by demonstrating that the identification of the direction of a target spectral glide was influenced by the spectral motion of a preceding inducer glide. In Experiment 2, the target glide was replaced with a synthetic sine-wave speech sound, including a formant transition. The speech category boundary was shifted by the presence and direction of the inducer glide. Finally, in Experiment 3, stimuli based on synthetic sine-wave speech sounds were used as both context and target stimuli to show that the spectral-motion after-effects could occur even with inducers with relatively short speech-like durations and small frequency excursions. The results suggest that spectral motion may play a complementary role to long-term average power spectrum in inducing speech context effects.

4.1 Introduction

Perceptual systems encode stimuli in a way that is highly dependent on contextual information. Speech is no exception to this general rule, and our perception of individual speech sounds can depend strongly on the context in which they are presented. In a pioneering study, Ladefoged and Broadbent (1957) tested 60 subjects in a word identification task. They observed that altering the first two formants within a context sentence ("Please say what this word is") dramatically changed subjects' identification of the following tests words. For example, a test word was perceived as "bit" by 53 subjects out of 60 when the unfiltered sentence was presented as the context, whereas the same word was perceived as "bet" by 54 of the subjects after the first formant (F1) of the preceding sentence was lowered somewhat. In a later example, Mann (1980) found that ambiguous syllables along a /ga/-/da/ continuum were generally perceived as /ga/ when preceded by the syllable /al/ and were perceived as /da/ when preceded by the syllable /ar/.

Since these early studies, it has been debated whether such context effects are specific to speech, or whether they reflect more general auditory processes. Soon after Mann's study, Fowler (1981) suggested that this "compensation for coarticulation" must reflect speech processes, since subjects' strategy for perceiving vowels was tightly coupled to their strategy for producing them. However, other researchers have since argued that such context effects may reflect more general auditory processes (Diehl *et al.*, 2004). For instance, Lotto and Kluender (1998) observed a smaller but significant effect

58

even when using sine-wave tones or glides corresponding to F3 of /al/ and /ar/ as the precursor, demonstrating that it was not necessary for the precursor to be perceived as speech for context effects to occur. In addition, Lotto *et al.* (1997) found similar context effects in a behavioral study of Japanese quails, suggesting that knowledge of speech was also not necessary. Both these and other studies (e.g., Holt, 2006a), have suggested that the average power spectrum of the preceding sound plays a dominant role in determining context effects, and that the effects are contrastive. Summerfield *et al.* (1984a) found that listeners were able to identify a flat-spectrum harmonic tone complex as a vowel, if it followed a sound with a similar spectrum, but with components at frequencies corresponding to the first three formants of the vowel omitted. In Chapter 5, similar effects were also observed with cochlear-implant users. Such contrastive effects are common in other sensory modalities (Gibson, 1933), and may reflect the tendency of perceptual systems to normalize or "whiten" the incoming stimuli to improve coding efficiency (e.g., Barlow, 1961; Dean *et al.*, 2008).

Aside from average power spectrum, other stimulus properties may also induce after-effects that may be relevant to speech perception. For instance, both speech and non-speech contexts affect the perception of the fundamental-frequency (F0) contour of lexical tones in a contrastive way: following a context with a higher mean F0, the target syllable is more likely to be identified as a lexical tone starting from a lower frequency and vice versa (Huang and Holt, 2012).

In addition to spectral contrast effects, temporal contrast effects also occur in speech perception (e.g., Diehl and Walsh, 1989; Wade and Holt, 2005). For instance, Wade and Holt (2005) measured the influence of the presentation rate of a preceding sequence of pure tones on perception of stimuli generated from a continuum between /ba/ and /wa/, as defined by the duration of formant transitions. They observed that a rapid presentation rate of the preceding pure tones resulted in more /wa/ responses, corresponding to the perception of a longer formant transition, while a slower presentation rate resulted in more /ba/ responses, corresponding to the perception of a shorter formant transition. Thus, contrastive after-effects have been shown in speech in both the spectral and temporal domains.

Dynamic spectral changes may also play a role in inducing context effects. In a demonstration with some similarities to the visual motion after-effect (Gibson, 1933), often referred to as the "waterfall effect," Shu *et al.* (1993) found that preceding glides in the center frequency of narrowband noise induced the perception of spectral motion in the opposite direction, such that a downward sweep, repeated over 2-3 minutes, caused listeners to hear a stationary noise band as increasing in frequency, and vice versa. Beyond that initial report on the spectral motion after-effect, little is known concerning the underlying mechanisms, or its relevance to everyday auditory perception. One earlier study (Holt *et al.*, 2000) reported that preceding contexts that included formant transitions had a larger effect on synthesized vowel identification than conditions with only a steady-

state spectral context, suggesting that spectral motion may also play a role in speech context effects.

The present study investigates spectral-motion after-effects and their influence on the perception of non-speech and synthesized-speech sounds. The first experiment confirms the presence of spectral-motion after-effects with stimulus durations closer to those approximating speech sounds. The second experiment reports after-effects of spectral motion on perceptual judgments of speech sounds. Finally, the third experiment examines possible trade-offs between average spectrum and spectral motion, using precursors that were designed to more closely resemble speech sounds.

4.2 Experiment 1: auditory spectral-motion after-effects with word-length inducers

4.2.1 Methods

Subjects

Eight (2 males, 6 females) native speakers of American English participated in this experiment and were compensated for their time. Their ages ranged from 18 to 28 years (mean age 23.6 years). They had normal hearing, as defined by audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz.

Stimuli

Each trial consisted of a single 500-ms precursor tone, followed by a single 50-ms target tone. The precursor and target were separated by a 50-ms silent gap. All the stimuli were gated on and off with 20-ms raised-cosine ramps. As illustrated in Fig. 4.1, the precursor was centered in the high (2200 Hz), middle (2000 Hz), or low (1800 Hz) frequency region, and was a rising or falling linear frequency glide, or remained at the same frequency. The combination of three frequency regions and three temporal patterns resulted in a total of nine precursor conditions. The nominal beginning and end frequencies of the precursors are listed in Table 4.1. The nominal beginning frequency of target stimulus was selected from the range between 1920 Hz and 2080 Hz in steps of 20 Hz, and the nominal end frequency was always 2000 Hz. The overall frequency content of both precursor and target was roved together by ±10% across trials, so that the frequency relationship between the precursor and the target remained constant. The rove was designed to discourage listeners from using potential cues based on absolute frequency.
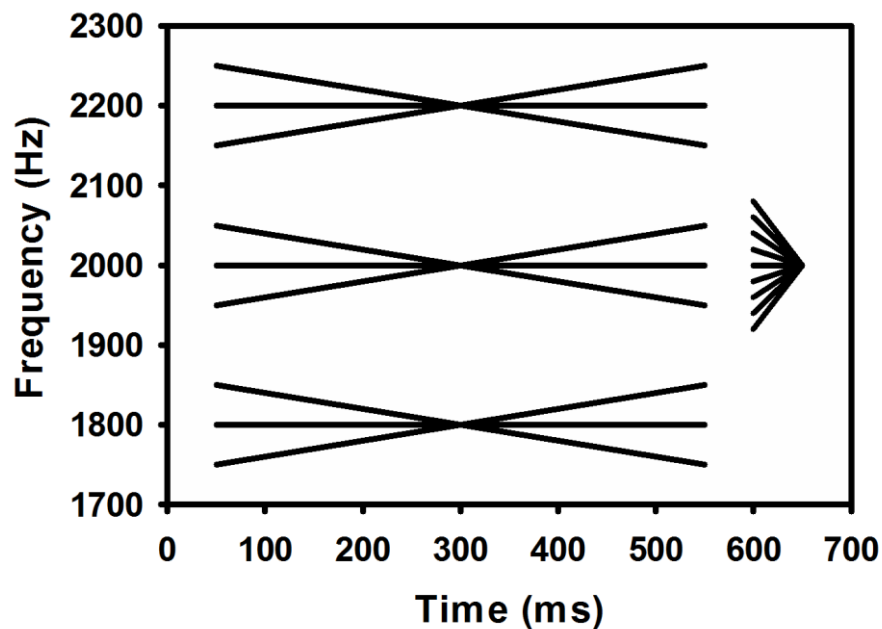
FIG. 4.1. Schematic diagram of the stimuli used in Experiment 1. The precursor, or inducer, was a rising, falling, or steady 500-ms glide that was centered at one of three frequencies. The test stimulus, or target, was a 50-ms tone, selected from one of the rising, falling, or steady lines shown at the right of the figure.

The stimuli were generated digitally and played out diotically from a LynxStudio L22 24-bit soundcard at a sampling rate of 22.5 kHz via Sennheiser HD650 headphones to subjects seated in a double-walled sound-attenuating chamber. The equivalent diffuse-field presentation level for all the sounds was 65 dB SPL.

Table 4.1 Onset and offset frequencies of each precursor condition

Conditions	No Pre	High-Rising	High-Flat	High-Falling	Middle-Rising	Middle-Flat	Middle-Falling	Low-Rising	Low-Flat	Low-Falling
Onset (Hz)	N/A	2150	2200	2250	1950	2000	2050	1750	1800	1850
Offset (Hz)	N/A	2250	2200	2150	2050	2000	1950	1850	1800	1750

Procedure

Subjects were asked to judge whether the target tone was "rising" or "falling" and to respond via virtual buttons on the computer display. Prior to the actual experiment, all subjects underwent a training session, during which they were presented with just the target and no precursor. Eight target conditions were tested, including all the target conditions tested in the actual experiment, with the exception of the "flat" target. Each of the conditions was presented 10 times within a block of trials. Feedback was provided during training. In order to progress to the actual experiment, subjects had to achieve at least 80% correct responses on average within 3 blocks in discriminating rising from falling glides. Two of the initial 10 subjects failed to reach this criterion, so only the remaining 8 were tested further. In the actual experiment, all 9 target conditions were tested 10 times each within each block in random order, for a total block length of 90 trials with a single precursor condition. The 10 precursor conditions (9 precursors and 1 no-precursor reference condition) were presented in separate blocks and were repeated 5 times, each in random order, for a total of 50 blocks. Thus, each of the 90 conditions (9 target by 10 precursor conditions) was repeated 50 times, and the proportion of "rising" and "falling" responses was calculated for each subject and condition from these 50

64

responses. No feedback was provided in the test sessions. All subjects provided informed written consent prior to participating, and the experimental protocols were approved by the Institutional Review Board of the University of Minnesota.

4.2.2 Results

The mean results are shown in Fig. 4.2. The left, middle, and right panels show the results using the precursor in the low, middle, and high spectral region, respectively. For comparison, the results from the condition with no precursor are shown as circles in all three panels. Considering first the condition with the precursor in the middle spectral region (Fig. 4.2, middle panel), it seems that on average the rising precursor led to more "falling" responses, and the falling precursor led to more "rising" responses, relative to the "flat" precursor condition. In other words, the results from the precursor in the middle region are consistent with predictions based on a contrastive spectral-motion after-effect. Similar differences between the falling and rising precursor can be observed in the lower and higher spectral regions (Fig. 4.2, left and right panels, respectively), although the relationship between those responses and the responses to the flat or no precursor are not so clear cut.
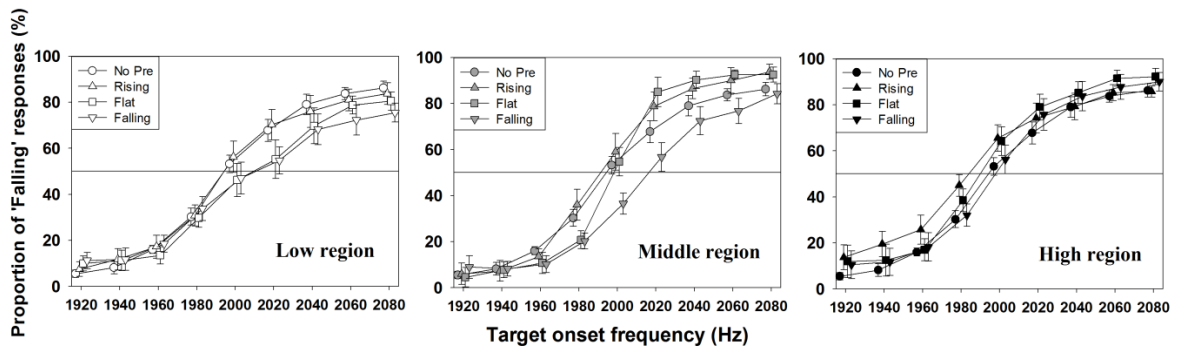
FIG. 4.2. Psychometric functions showing the average proportion of 'Falling' responses in percent as a function of the onset frequency of the target glide. The left, center, and right panels show results from the precursors in the low, medium, and high spectral regions, respectively. Upward- and downward-pointing triangles denote conditions with rising and falling precursors, respectively. Squares denote conditions with the constant-frequency (flat) precursors. The same data from the condition with no precursor (circles) are shown in each panel for ease of comparison. Error bars represent 1 s.e. of the mean across subjects. The horizontal lines mark the category response boundary of 50% of 'Falling' responses. Symbols in the different conditions are offset slightly in the horizontal direction for clarity.

To quantify the effects of the precursor, we used probit analysis to fit each of the curves shown in Fig. 4.2 for each subject individually. Then we calculated the point at which each curve crossed the 50% point (i.e., the point at which a "falling" response was as likely as a "rising" response), which is termed the "category response boundary." The mean category response boundaries, averaged across subjects, are shown in Fig. 4.3. A boundary value of 2000 Hz implies that a flat target was perceived veridically; higher boundary values imply that flat targets were more likely to be reported as rising, whereas

66

lower boundary values imply that flat targets were more likely to be reported as falling. The category response boundaries were subjected to a two-way within-subjects ANOVA, with precursor glide direction (up, down, or flat) and spectral region (low, medium, or high) as the two factors. Significant main effects were observed for both glide direction [$F(2,14) = 5.6$; $p = 0.016$] and frequency region [$F(2,14) = 12.6$; $p = 0.001$], and for their interaction [$F(4,28) = 4.05$; $p = 0.01$]. The main effect of glide direction reflects the trend visible in Fig. 4.3 that the rising precursor tended to lead to lower boundary values than the falling precursor. Post-hoc contrast analysis showed that the response boundary in the rising condition was significantly different from that in the falling condition ($p = 0.049$). However, no significant difference was observed between the response boundary in the flat condition and that in either the rising or falling condition. The main effect of spectral region reflects the trend for decreasing boundary value from low to high precursor spectral region. The interaction presumably reflects the impression that the effect of spectral motion seems greater in the middle spectral region than in the low or high region.
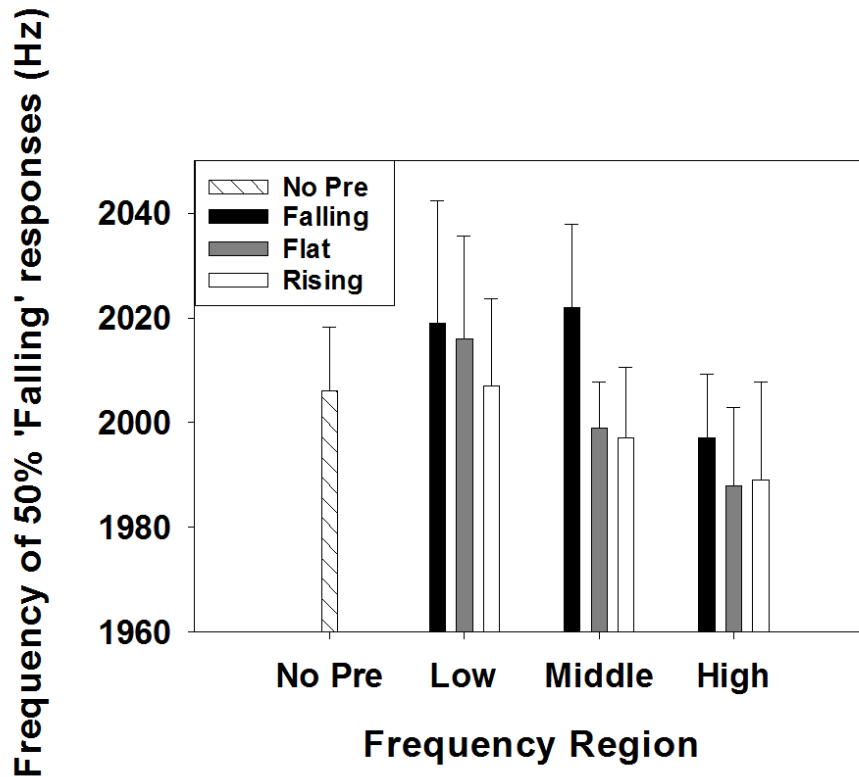
FIG. 4.3. Mean category response boundary frequencies for each condition. The different bar shadings represent the different precursor motion conditions, as shown in the legend. The results from the three spectral regions are shown in separate groups, as listed along the horizontal axis. Error bars represent 1 s.e. of the mean.

4.2.3 Discussion

The results from this experiment, showing a rising precursor leading to more "falling" responses, and vice versa, is consistent with the original report of a contrastive spectral-motion after-effect (Shu *et al.*, 1993), and extends the original finding by showing that a

relatively short, word-length, precursor of 500 ms is sufficient to produce a measurable effect. Relatively short spectral motion on this time scale could come from pitch glides in speech, particularly in tone languages, where it has already been shown that F0 contrast effects can be measured (Huang and Holt, 2009).

The effect of spectral region produced an interesting trend, which might be described as "continuity": if the precursor was in the high spectral region, then the target was more likely to be reported as "falling," i.e., moving from the region of the precursor to the center, whereas if the precursor was in the low spectral region, the target was more likely to be reported as "rising." This is the opposite of what would be expected based on spectral contrast, where a high precursor would be expected to lower the perceived beginning of the precursor. One potential reason for why our results are not consistent with expectations based on spectral contrast was that the target consisted of just a short glide, whereas earlier studies have used speech-like sounds that began with a short glide, simulating a formant transition, and ended with a longer steady-state portion. The lack of a steady-state portion at the end of the glide may have reduced the extent to which spectral contrast differentially affected the beginning and end of the target sound.

We have assumed that the differences produced by the rising and falling precursors, particularly in the middle spectral region, are due to their spectral-motion properties. It is clear that the *average* spectrum of the precursor in the middle region cannot explain the effects, as the average frequency of the rising, falling and flat precursors are the same. Nevertheless, it is possible that the results reflect primarily the

end frequency of the precursor, rather than spectral motion *per se*. This interpretation is rendered less likely by the fact that the end frequency does not provide a good predictor of all the results. Progressing from the low spectral region to the high, there is a 100-Hz difference between the end frequency of the falling and rising precursor within each spectral region, and between the rising precursor of one spectral region and the falling precursor of the next (going from left to right in Fig. 4.3, ignoring the flat precursor conditions). Therefore, if the end frequency of each precursor predicted the results, the category response boundary should monotonically (and perhaps linearly) decrease with increasing end frequency. Although this pattern holds within each of the three spectral regions, it does not hold across spectral regions; for instance, going from low-rising to middle-falling leads to an increase in category response boundary, rather than the expected decrease predicted by the end frequency of the precursor. However, the results are somewhat variable, leaving potential room for doubt. In the next experiment we used sine-wave speech targets where the perceived glide direction of a synthetic formant changed the identity of the speech sound. Based on earlier studies, we expected long-term spectral contrast effects to predict the opposite pattern of results from spectral-motion after-effects, thereby making it easier to distinguish between the two.

4.3 Experiment 2: spectral-motion after-effects with synthetic sine-wave speech targets

4.3.1 Methods

Subjects

Eight (3 males, 5 females) native speakers of American English participated in this experiment and were compensated for their time. Their ages ranged from 18 to 61 years (mean age at 29.3 years). They had normal hearing, as defined by audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz. Three of them had also participated in Experiment 1.

Stimuli

A synthetic syllable identification task (/ba/-/da/), similar to that of Holt and Lotto (2002), was used, as shown in Fig. 4.4. Target syllables of 250 ms duration were synthesized with sine waves representing the first three formants. The frequency of the tone representing F1 began at 450 Hz and was swept linearly to 700 Hz over the first 50 ms, where it remained for the final 200 ms. The tone representing F3 remained steady at 2600 Hz. The onset frequency of F2 varied from 800 Hz to 1600 Hz in steps of 100 Hz. During its first 50 ms, the frequency of the F2 tone was swept linearly to 1200 Hz, where it remained for the final 200 ms.

Two precursor durations were tested. The first was 500 ms, which was shown to produce a spectral-motion after-effect in Experiment 1. The second was 100 ms, which is of a more relevant duration for formant transitions in speech. The precursor frequency always began at 1200 Hz, and was swept linearly over its entire duration to an end frequency that varied parametrically between 800 and 1600 Hz in steps of 200 Hz. The gap between the precursor and target was always 50 ms.
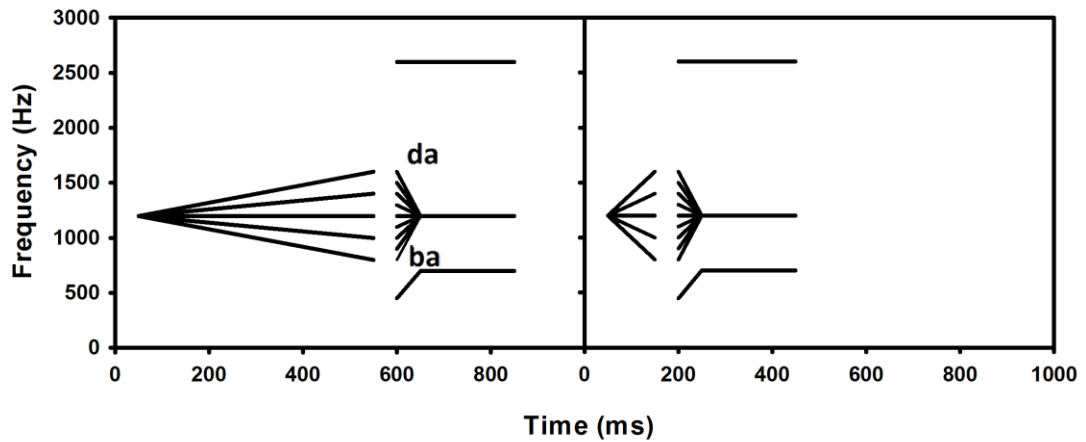
FIG. 4.4. Schematic diagram of the stimuli used in Experiment 2. A 500-ms (left) or 100-ms (right) precursor was followed by a synthetic target syllable that listeners were asked to categorize as either /da/ or /ba/.

Procedure

Initially, subjects took part in a training session in which no precursor was presented. Only the two endpoints of the /ba/-/da/ continuum were presented, with beginning F2 frequencies of 800 and 1600 Hz for /ba/ and /da/, respectively. Subjects were required to achieve at least 80% correct identification in the training phase in order to proceed to the test phase. Within each of the training blocks, there were 20 repetitions for each of the two target conditions, resulting in 40 trials per block. All eight subjects passed the training phase. In the test phase, there were 5 repetitions of each of the 9 targets (4 falling, 4 rising, 1 flat) per block (45 trials). Each precursor was tested in a separate block and each of these blocks was presented 8 times for a total of 80 blocks, and 3600 total trials per subject (40 per subject and condition). All conditions were presented in random order, selected independently for each subject. Feedback was provided for the training blocks, but not during the test blocks.
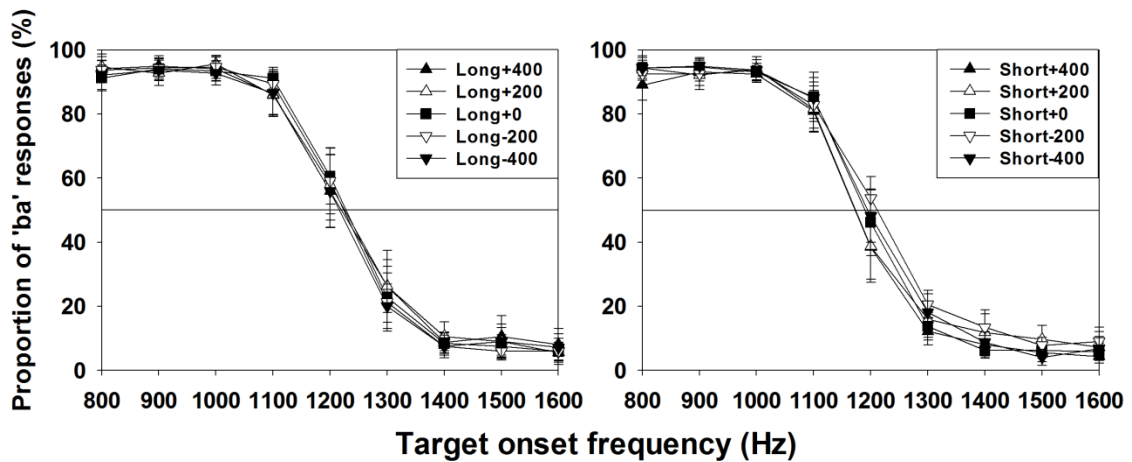
FIG. 4.5. Proportion of trials identified as /ba/ as a function of the F2 onset frequency. Left and right panels show results from long and short precursor conditions, respectively. Error bars represent 1 s.e. of the mean across subjects. The horizontal lines mark 50% of /ba/ responses. Numbers in the legend represent the frequency difference between the beginning and end of the precursor glide, with negative numbers indicating a falling glide and positive numbers indicating a rising glide.

4.3.2 Results

The results of Experiment 2 are presented in Figs. 5 and 6, using the same format as those of Experiment 1. Fig. 4.5 shows the average identification curves in terms of proportion of /ba/ responses as a function of the F2 onset frequency with the preceding glides for both long (left panel) and short (right panel) precursor conditions. As with the results from Experiment 1, a probit analysis was performed using the psychometric functions from the individual listeners to derive a 50% category response boundary for each listener and condition. A one-way within-subjects ANOVA was conducted, with the

frequency at the category response boundary as the dependent variable, and precursor glide slopes (difference between start and end points of 400, 200, 0, -200, -400 Hz) as the factor for both long and short precursor conditions separately (Fig. 4.6). No significant main effect of precursor was observed for 500-ms precursor conditions [$F_{(4,28)} = 0.388$; $p = 0.815$]. However, a significant main effect was found for the 100-ms precursor conditions [$F_{(4,28)} = 2.85$; $p = 0.042$]. In pairwise comparison contrast tests, the Short+400 condition differed significantly from the Short-200 ($p = 0.013$) and Short-400 ($p = 0.04$) conditions. A further contrast analysis revealed a significant linear trend [$F_{(1,7)} = 6.03$; $p = 0.044$], confirming that there was a systematic trend for increasing boundary value with decreasing slope value of the precursor.
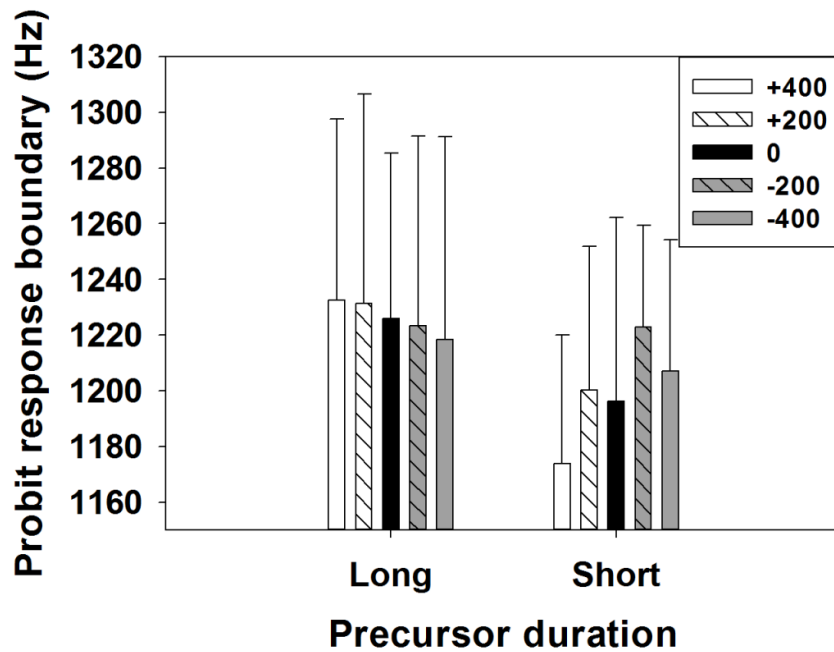
FIG. 4.6. Mean 50% boundary response frequencies from Experiment 2. Numbers in the legend represent the frequency difference between the onset and offset frequency of the precursor, with negative numbers indicating a falling glide and positive numbers indicating a rising glide. Error bars represent 1 s.e. of the mean across subjects.

4.3.3 Discussion

The main finding from Experiment 2 is the existence of a spectral-motion after-effect using a synthesized sine-wave speech sound as a target. The effect found with the shorter precursor is the same direction as the spectral-motion after-effect found in Experiment 1: a falling precursor glide led to a greater proportion of responses corresponding to the

rising target, which in this case corresponds to the syllable /ba/. Thus, as in Experiment 1, the results are consistent with a contrastive after-effect of spectral motion.

It is not clear why the shorter, but not the longer, precursor resulted in a measurable after-effect. The frequency excursion of the longer, 500-ms precursor, relative to that of the target, was similar to what was used in Experiment 1, also with a 500-ms precursor. However, there are also multiple differences between the two experiments. First, the nature of the task was different, with glide direction identification in Experiment 1, compared with consonant identification in Experiment 2. Second, there were large differences in the stimuli, including a much wider range of frequency excursion for both the precursor and the target (+/- 4% of the end frequency in Experiment 1, compared to +/- 33% of the end frequency in Experiment 2), a lack of overall frequency roving in Experiment 2, and the addition of F1 and F3 in Experiment 2. Third, the long-term (average) spectrum of the three precursors was the same in Experiment 1, whereas in Experiment 2 the rising precursor was higher, and the falling precursor was lower, in average spectrum than the steady (flat) precursor. The difference in average spectrum might have counteracted part of the effects of spectral motion, through a spectral contrast effect, whereby a higher average spectrum would be expected to lead to more /ba/ responses. Thus, the effects of long-term spectrum may have reduced the (opposite) effect produced by spectral motion.

Overall, the results suggest that spectral-motion after-effects can affect speech category boundaries. The outcome cannot be easily explained in terms of long- or short-

term average spectrum of the precursor, and instead seems to reflect genuine spectral-motion after-effects. It is possible that similar phenomena could arise with speech sounds, not just artificial glides, as context or precursors. In the final experiment, materials based on synthetic sine-wave speech sounds were used as both context and target stimuli, to investigate whether relatively small formant transitions could themselves produce spectral-motion after-effects in speech, beyond long-term spectral contrasts.

4.4 Experiment 3: spectral-motion after-effects with synthesized speech context and target

4.4.1 Methods

Subjects

Eight (2 males, 6 females) native speakers of American English participated in this experiment and were compensated for their time. Their ages ranged from 18 to 61 years (mean age at 28.1 years). They had normal hearing, as defined by audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz. Five of them had also participated in Experiment 2.

Stimuli

The same nine target stimuli were used as in Experiment 2. A total of 7 different precursors were used. All precursors consisted of three 100-ms tones, resembling formant frequencies. The lowest and highest tones remained constant at 870 and 2300 Hz, respectively. The middle tone began at one of three frequencies, as shown in Table 4.2 and illustrated in Fig. 4.7. The final 30 ms of the middle precursor tone was either constant, or was a linear rising or falling sweep. Two of the precursors were designed to resemble the speech sounds /i/ and /u/, as indicated in Table 4.2. The others were variations of these speech sounds that were designed to test the relative importance of the average spectrum and the sweep direction. The precursor and the target were separated by a 50-ms silent gap.
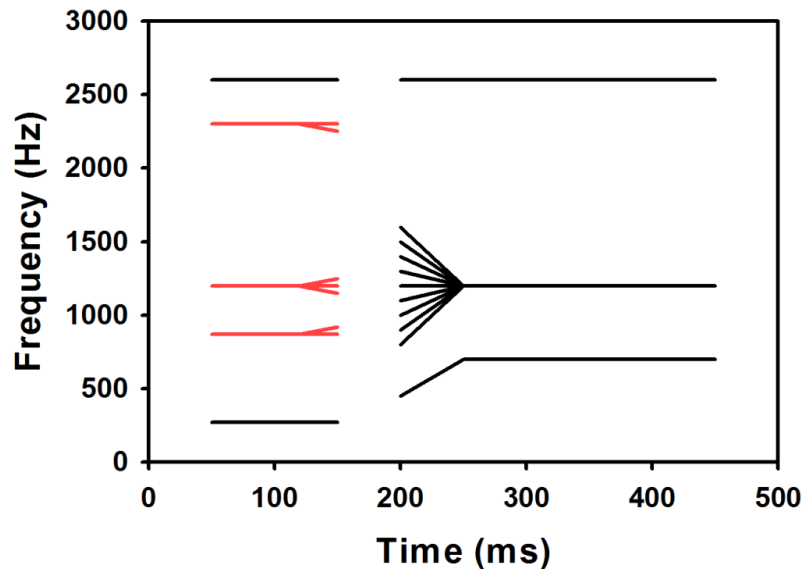
FIG. 4.7. Schematic diagram of stimuli used in Experiment 3.

79

Procedure

The subjects first completed the same training session that was used in Experiment 2. They were again required to achieve at least 80% correct in the /ba/-/da/ identification task before progressing to the test phase. Again, all eight subjects passed the training phase. In the test phase, each block tested a single precursor and each of the 9 targets was presented 5 times in random order. Each of the 7 precursor conditions was tested in 8 blocks for a total of 56 blocks, and 40 (8x5) repetitions of each condition per subject. As in the previous experiments, feedback was provided only in the training phase, and not in the test phase.

Table 4.2 Stimulus conditions from Experiment 3. The upper and lower rows show the onset and offset frequencies of F2 for each precursor condition separately.

Conditions	High-Flat	High-Falling (/i/)	Middle-Rising	Middle-Flat	Middle-Falling	Low-Rising (/u/)	Low-Falling
Onset (Hz)	2300	2300	1200	1200	1200	870	870
Offset (Hz)	2300	2250	1250	1200	1150	920	870

4.4.2 Results

The results from Experiment 3 are shown in Fig. 4.8. Again, probit analysis was used to derive the 50% point of the psychometric functions for each subject in each condition. These category response boundaries, averaged across subjects, are shown in Fig. 4.9 and were used as the dependent variable in two separate ANOVAs. Consider first the

conditions where the precursor F2 was in the highest or lowest spectral region (Fig. 4.8 left panel).
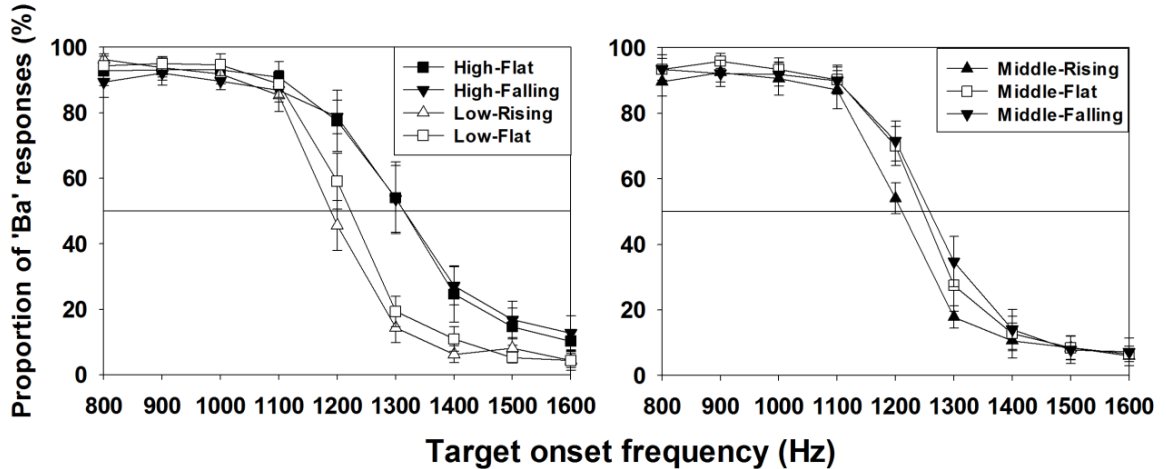
FIG. 4.8. Proportion of trials identified as /ba/, as a function of the target F2 onset frequency. The left panel shows responses in conditions where the precursor F2 was in the high or low spectral region. The right panel shows responses in conditions where the precursor F2 was in the middle spectral region. Error bars represent 1 s.e. across subjects. The horizontal lines mark 50% of /ba/ responses.

A two-way repeated-measures ANOVA with main factors of spectral location (high or low) and spectral motion (moving or steady) showed a significant main effect of spectral location [$F(1,7) = 9.11$, $p = 0.02$] but no effect of spectral motion [$F(1,7) = 1.53$, $p = 0.26$], and no interaction [$F(1,7) = 0.036$, $p = 0.86$]. Thus, in cases where the spectral motion in the precursor was remote from the target, no effect of spectral motion was

found. Instead, an effect of spectral contrast was observed, with the high-frequency precursor resulting in the target formant being perceived as beginning from a lower frequency, and the low-frequency precursor resulting in the target formant being perceived as beginning from a higher frequency.
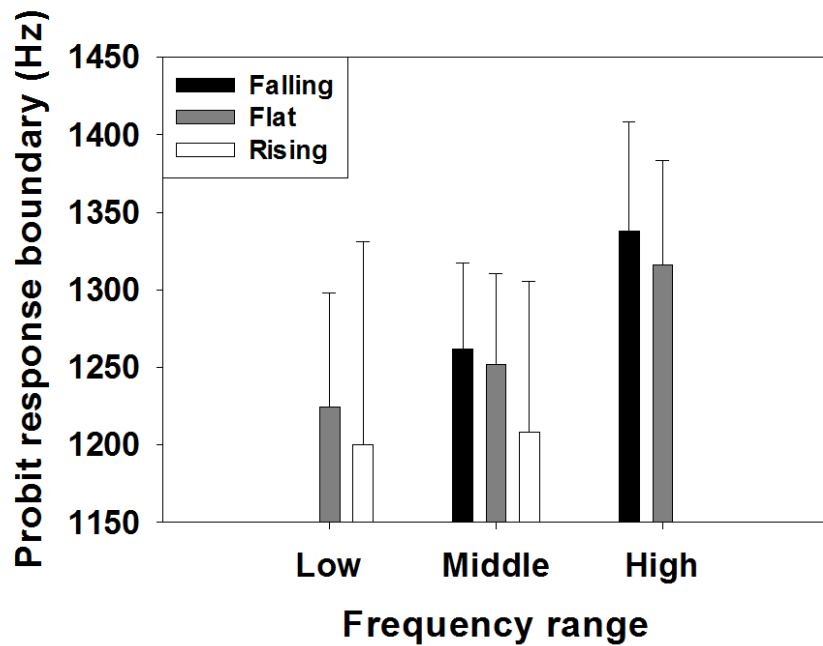
FIG. 4.9. Mean category response boundaries, at which /ba/ and /da/ responses are equally likely. The x-axis indicates the frequency region of the precursor F2. Different shaded bars represent the different spectral motion of the patterns of formant transition. Error bars represent 1 s.e. of the mean across subjects.

Consider next the three conditions with the precursor F2 in the central spectral region (Fig. 4.8 right panel). A one-way repeated-measures ANOVA found a significant

main effect of spectral motion [$F(2,14) = 6.9$, $p = 0.008$]. Pairwise comparisons revealed that the precursor with the rising formant transition produced a significantly different boundary than the precursor with the falling formant transition ($p = 0.025$). Similarly, a contrast analysis revealed a significant linear trend, confirming that the boundary value increased with decreasing slope value [$F(1,7) = 13.19$, $p = 0.008$]. Thus, for the precursor at the target frequency, a small but contrastive spectral-motion after-effect was observed.

4.4.3 Discussion

When the precursor was in the same spectral region as the target, a small but significant spectral-motion after-effect was observed, even though the precursor motion was as short as that of the target itself, and the frequency excursions of the precursor were actually smaller than that of the smallest frequency excursion of the target. The lack of an effect of spectral motion with the low- and high-region precursors suggests that, as in Experiment 1, the strongest effects of spectral motion are observed when the precursor and target fall in the same spectral region. Again, as in the previous two experiments, the results cannot be easily explained in terms of the average long- or short-term spectrum of the precursor; in fact, as in Experiment 2, any averaging of the precursor would lead to predictions in the opposite direction of those observed in the results, with the higher precursor predicted to produce more rising responses. Thus, as in the previous experiments, the outcome is more easily explained in terms of sensitivity to spectral motion *per se*.

83

Comparing the results from the low- and high-region precursors, the effect is similar to that observed in earlier studies (Holt and Lotto, 2002; Holt, 2006a) in that the higher precursor led to the report of more /ba/ responses, corresponding to a lower perceived starting frequency of the target glide, and the lower precursor led to the report of more /da/ responses, corresponding to a higher perceived starting frequency of the target glide. Note that this outcome, although in line with earlier studies, is not consistent with the results from Experiment 1, where a "continuity" effect was observed. As mentioned in the discussion of Experiment 1, one explanation for this apparent discrepancy relates to the nature of the target stimulus: in earlier studies (and in the current experiment), the target remained at a constant frequency after the initial glide, whereas in Experiment 1 the target consisted only of a brief glide. It may be necessary to have a longer target (or stable end frequency) for the precursor to have a differential effect on the beginning and end of the target. Another difference between the experiments is that the target glide in Experiment 3 led to the perception of one of two speech categories, /ba/ and /da/, for which subjects have established long-term category representations. Further experiments will be required to test the possibility that the nature of the target can affect the direction of the context effects.

The duration and size of the precursor's spectral motion were chosen in this experiment to be representative of the motion found in speech. Therefore, the fact that spectral-motion after-effects were found suggests that they may also play some role in more natural situations involving speech perception.

4.5 General discussion

The three experiments presented here all provide evidence that spectral motion on frequency and time scales that are relevant to speech can produce contrastive after-effects. The after-effects are relatively small in absolute terms, but can be induced with surprisingly small frequency excursions and short precursor durations. The after-effects also appear to be spectrally local, in that precursor glides that are remote in frequency from the target do not produce significant after-effects. The finding that the effects are spectrally local can be compared to the conclusions drawn from an earlier study of temporal contrast effects. As mentioned earlier, Wade and Holt (2005) found that a sequence of pure tones presented at a rapid rate resulted in more /wa/ responses, whereas a sequence presented at a slower rate resulted in more /ba/ responses to the target. The results were therefore consistent with a temporal contrast effect, in which a faster precursor rate led listeners to judge the following transitions as slower. When comparing their effect with a null result reported in earlier study by Summerfield (1981), Wade and Holt (2005) suggested that one important difference might have been the lack of spectral and temporal continuity between the relevant precursor dimensions and the target.

It is important to establish whether the spectral-motion after-effect is in fact mediated by spectral motion, rather than the long- or short-term spectrum of the precursor. In this respect the results from all three experiments converge to suggest that it is the spectral motion per se, rather than the spectrum that determines the effect, as outlined below.

In the first experiment, the average spectra of all the precursors were the same. If one assumes that only the final part of the precursor contributes to the aftereffect, then the direction of the after-effect was the same as that predicted by just the spectrum, as illustrated by the effect of the precursors in the higher and lower spectral regions. However, as discussed in Experiment 1, considering just the end points of the precursor frequency cannot explain why the rising precursor in the low spectral region produced a lower category response boundary than the falling precursor in the middle spectral region. Instead, the overall pattern of results are more consistent with an explanation based on spectral motion within the local spectral region of the target.

In the second experiment, all the precursors began at the same frequency, so that the precursor with the upward spectral motion also had a higher long-term (and short-term) spectrum than the precursor with the downward spectral motion. In this case, spectral-motion contrast predicts that an upward precursor glide should lead to more perceived downward target glides and a downward precursor glide should lead to more perceived upward target glides, consistent with the obtained data, whereas an explanation based on simple spectral contrast predicts the opposite. Therefore, in this case it is clear that an explanation based on spectral-motion contrast provides a better account of the data.

In the third experiment, spectral-motion contrast and simple spectral contrast make opposite predictions. Presenting the precursor in different spectral regions resulted in outcomes consistent with spectral contrast, whereas the precursor in the spectral region

of the target produced results consistent with the predictions of spectral-motion contrast. Thus, a parsimonious explanation of all three experiments is that the spectral motion of the precursor can induce after-effects beyond those predicted by the long-term (or short-term) spectrum.

Having established the existence of a spectral-motion after-effect that may be relevant for speech perception, a next step is to determine the underlying mechanisms. Just as spectral-contrast context effects could be explained in terms of neural adaptation or forward suppression of frequency-selective cortical and/or sub-cortical neurons, spectral-motion after-effects could be explained in terms of adaptation or forward suppression of neurons that are tuned to the direction of spectral motion. Such neurons have been identified in other mammals (Weinberger and Mckenna, 1988; McKenna *et al.*, 1989; Brosch and Schreiner, 2000). In addition, there are other psychophysical results involving tone detection and discrimination experiments that have led researchers to propose the presence of "pitch-shift detectors" (Demany and Ramos, 2005; Demany *et al.*, 2009), which could also be invoked to explain the results of the present experiment. Further studies could explore in more detail the parametric effects of precursor duration and rate of frequency change to better define the nature of these hypothetical frequency glide detectors.

4.6 Summary

This study explored the potential role of spectral motion in inducing context effects in non-speech and synthesized-speech stimuli. Experiment 1 confirmed the existence of a contrastive spectral-motion after-effect in judging the motion of a target tone glide, and extended previous findings by showing that significant after-effects could be produced using a relatively short (500-ms) inducer.

Experiment 2 found that the glide direction of a shorter inducer, of only 100 ms, could influence phonemic judgments along a /ba/-/da/ continuum in a way that was also consistent with spectral-motion contrast, although the longer 500-ms precursor had no significant influence on the phonemic judgments.

In Experiment 3 a precursor was constructed with three tones to resemble the formant structure of two vowels, /i/ and /u/, along with more artificial variants. Consistent with previous studies, the long-term spectrum of the precursor affected judgments of artificial stimuli constructed along the /ba/-/da/ continuum. In addition, when the spectral motion of the precursor was in the same spectral region as the formant transition of the target, a contrastive spectral-motion after-effect was observed.

Overall, the results demonstrate that spectral motion can induce changes in the responses to both non-speech and speech-like stimuli, and suggest that spectral-motion after-effects may play a role in more natural situations involving speech perception.

CHAPTER 5: VOWEL ENHANCEMENT EFFECTS IN COCHLEAR-IMPLANT USERS[1]

Abstract

Auditory enhancement of certain frequencies can occur through prior stimulation of surrounding frequency regions. The underlying neural mechanisms are unknown, but may involve stimulus-driven changes in cochlear gain via the medial olivocochlear complex (MOC) efferents. Cochlear implants (CIs) bypass the cochlea, to stimulate the auditory nerve directly. If the MOC plays a critical role in enhancement then CI users should not exhibit this effect. Results using vowel stimuli, with and without preceding sounds designed to enhance formants, provided evidence of auditory enhancement in both normal-hearing listeners and CI users, suggesting that vowel enhancement is not mediated solely by cochlear effects.

5.1 Introduction

A classical psychophysical finding, analogous to the negative afterimage, has been termed the "auditory afterimage" (Wilson, 1970) or, more commonly, the auditory "enhancement effect" (EE) (Viemeister, 1980; Viemeister and Bacon, 1982; Byrne *et al.*, 2011a). An example of EE is the finding that the perception of a single target frequency component is enhanced from within a complex tone if it is preceded by a precursor that consists of the same complex tone but with the target component removed (Schouten, 1940).

A few studies have sought direct neural correlates of the enhancement effect. Palmer *et al.* (1995a) studied responses in the auditory nerve of guinea pig to stimuli that were likely to produce an enhancement effect. Consistent with more recent studies of adaptation in the auditory nerve (Wen *et al.*, 2009) and inferior colliculus (IC) (Dean *et al.*, 2005), they found that the response to the non-target tones was reduced by the precursor, so that the target tone was enhanced relative to the other tones in the complex. However, they found that the precursor produced no absolute increase in the response to the target, and therefore concluded that the mechanisms responsible for any increase in effective gain were probably located more centrally than the auditory nerve. More recently, Nelson and Young (2010) explored the EE at the level of inferior colliculus (IC) in two awake and passively listening female marmoset monkeys. Both significant EE and suppression were observed in single neurons. Moreover, effects of many other stimulus parameters on enhancement and suppression were qualitatively similar to those observed

in psychophysical measurements. The difference between the results of the Palmer *et al.* (1995a) and Nelson and Young (2010) studies may reflect the transformations that occur between the auditory nerve and IC, but may also reflect other differences, such as the use of anesthetized animals in the Palmer et al. study. For instance, if EE is mediated primarily by the medial olivocochlear (MOC) efferent system (Kawase *et al.*, 1993; Guinan, 2006; Jennings *et al.*, 2009; Zhou *et al.*, 2010), then anesthesia may result in a lack of MOC efferent activation, and hence a lack of observed enhancement in the auditory nerve of anesthetized animals.

In this study, we tested the hypothesis that mechanisms involving the cochlea, such as the MOC efferents, play a critical role in auditory enhancement by studying listeners with cochlear implants (CIs). Because the cochlea is bypassed by the CI, any effects elicited by MOC efferent activity should be absent in CI users. Some support for this hypothesis is found in earlier work that suggests that listeners with cochlear hearing loss show reduced or absent EE (Kimberley *et al.*, 1989a; Thibodeau, 1991). On the other hand, the "Zwicker tone" (Zwicker, 1964), an auditory afterimage which may be related to EE, have been postulated to be more central in origin (Wiegrebe *et al.*, 1996). In a recent study, Goupell and Mostardi (2012) reported that CI users were able to detect a target pulsing electrode from a background of continuously stimulated electrodes, and related that to the EE. However, because their paradigm had no control or reference condition, it is not clear whether and to what extent the target was enhanced, relative to the unpulsed condition. In the present study, we adapted a paradigm described by

Summerfield *et al.* (1984b). They found that listeners were able to identify a flat-spectrum harmonic tone complex as a vowel, if it followed a precursor with a similar spectrum, but with components at frequencies corresponding to the first three formants of that vowel omitted. The present study tested CI users' ability to identify similar artificial vowel stimuli with and without a precursor. Two control groups were also included: Normal-hearing (NH) listeners, and the same NH listeners presented with the vowels passed through a noise-excited envelope vocoder (VC), in order to simulate certain aspects of CI processing, in particular poorer spectral resolution.

5.2 Experiment

5.2.1 Method

Subjects

Twelve normal-hearing (NH) listeners and 8 post-lingually deafened CI users participated in this study, all of whom were native speakers of American English. The ages of the NH listeners ranged from 19 to 29 yr (mean age: 22.3 yr). Information regarding the CI users is shown in Table 5.1.

Table 5.1. Cochlear-implant subject information

Subject	Gender	Age (Yrs)	CI use (Yrs)	Etiology	Duration HL prior to implant (Yrs)	Electrode Type	Speech Strategy
N32	M	44.2	14.4	Maternal rubella	<1		SPEAK
D02	F	61.7	9.9	Unknown	1	HFII	HiRes-P 120
D05	F	81.2	9.6	Unknown	3	HFP	HiRes-S
D08	F	59.9	9.1	Otosclerosis	13	HF	HiRes-S 120
D10	F	57.5	8.9	Unknown	8	HFP	HiRes-S 120
D19	F	51.9	7.2	Unknown	11	HF	HiRes-S 120
D26	F	52.2	2.6	Unknown	11	HF1j	HiRes-P 120
D27	F	59.8	2.3	Otosclerosis	13	HF1j	HiRes-S 120

Stimuli

The paradigm and stimuli used in the current study were similar to those used by Summerfield *et al.* (1984b). All the subjects were tested in a double-walled sound-attenuating booth. Synthetic vowel identification was studied in CI users listening to the vowels acoustically through their speech processor from a loudspeaker (JBL 2500) located 1 meter in front of them, and in NH subjects listening diotically via headphones (Sennheiser HD580) either to the unprocessed stimuli (the same as were presented to the CI users) or to the stimuli after they were passed through an 8-channel noise-excited envelope vocoder (VC). In the VC condition, the input stimuli were first applied to 8 band-pass filters with contiguous passbands and cutoff frequencies at 250, 494, 697, 983, 1387, 1958, 2762, 3898, and 6800 Hz. The envelopes of the outputs were then full-wave

rectified and low-pass filtered (with a 400 Hz cutoff frequency). All the filters were 4th order Butterworth filters. Each envelope was used to modulate a broadband noise (50-8000 Hz), which was subsequently filtered by 2nd order bandpass Butterworth filters with the same center frequencies as the analysis filters. And the bandwidths were computed using the Greenwood (1990) function, as implemented by Bingabr *et al.* (2008) to simulate spread of excitation similar to that typically found in CI users with monopolar stimulation.

Table 5.2. **Formant frequencies for vowel stimuli**

Word	Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
Heed	/i/	342	2322	3000	3657
Hod	/ʌ/	768	1333	2522	3687
Who'd	/u/	378	997	2343	3357
Head	/ɛ/	580	1799	2605	3677
Heard	/3/	474	1379	1710	3334

The same five vowels were used as in Qin and Oxenham (2005). The vowels were generated by producing a "baseline" harmonic tone complex with a fundamental frequency (F0) of 100 Hz. The first 50 harmonics were generated with random starting phases, and were lowpass filtered to produce a spectral envelope with a constant slope of -6 dB/oct. To produce the formants, we increased the amplitudes of the harmonics in the baseline complex around the formant frequencies. The amplitudes of two adjacent

harmonics were increased for F1 and F2, and the amplitudes of four adjacent harmonics were increased for F3 and F4. The amount by which the amplitudes were increased was defined as the formant level. The formant frequencies are shown in Table 5.2. Each vowel had a total duration of 200 ms, including 10-ms raised-cosine onset and offset ramps. The vowels were presented either in isolation or 20 ms after a 1-s long precursor that consisted of the baseline harmonic complex, with the components corresponding to the formant frequencies of the target vowel attenuated by 24 dB. The vowels and precursors were presented at an overall level of 65 dB SPL (see Fig. 6.1A). Formant levels were set so as to achieve performance above chance but below ceiling, as determined in initial pilot experiments. For the NH group, formant levels were set at 2, 4, and 6 dB without the precursor, and at 0, 2, and 4 dB with the precursor. In the CI group, the formant levels were 2, 6, and 10 dB, and in the VC group the formant levels were 2, 4, and 6 dB, both with and without the precursor.
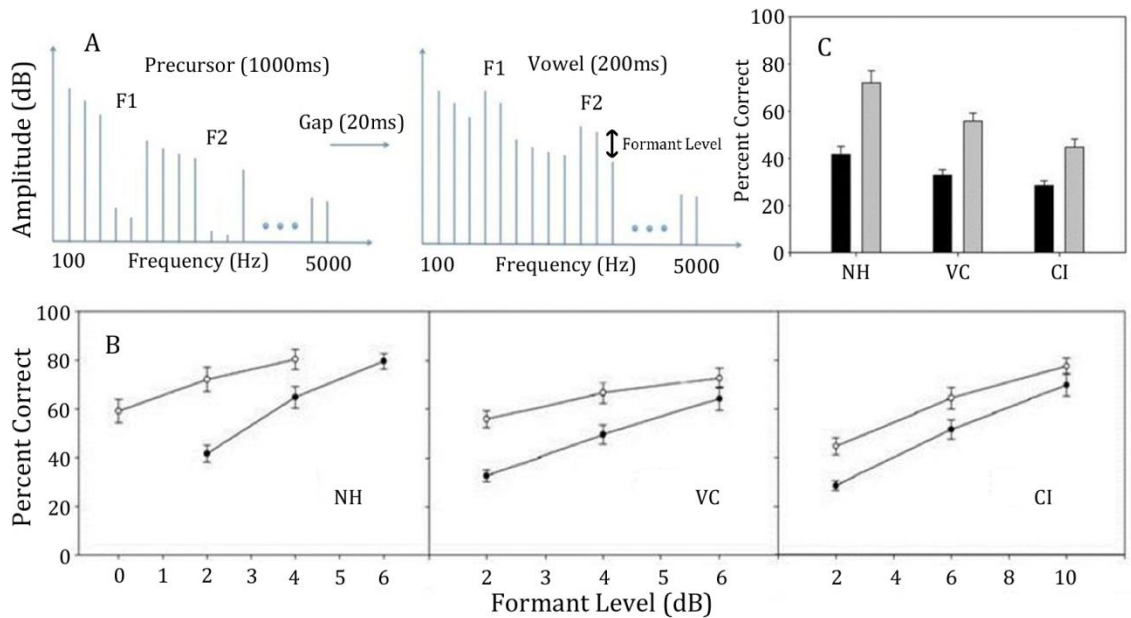
FIG. 5.1. Experimental stimuli and results. Panel A (upper left): Schematic diagram of precursor and vowel stimuli. Panel B (bottom): Mean vowel identification scores across subjects for NH, VC, and CI groups, plotted as a function of formant increment level (note the different abscissa for the CI group). Filled symbols represent performance for vowel only; open symbols represent performance with precursor and vowel. Panel C (upper right): Mean data from all three subject groups at the 2-dB formant contrast level. Dark bars represent conditions without a precursor; light bars represent conditions with a precursor. Error bars represent 1 s.e. of the mean.

5.2.2 Procedure

Subjects were asked to identify the synthetic vowels and were provided with 5 virtual "buttons" on a computer screen, each marked with one of the 5 target vowels. After each presentation, subjects were asked to press the button corresponding to the vowel they had

96

just heard. All subjects received initial training with correct-answer feedback to ensure that they could recognize the vowels at high formant levels without a precursor. For the NH condition, the formant level during training was fixed at 12 dB, while for the VC condition and CI users, it was fixed at 20 dB. The training involved only the vowel-alone conditions, i.e. the precursor was never presented during training. A training session consisted of twenty vowel presentations, so that each vowel was presented four times in random order. During the initial training period, subjects only progressed to the actual experiment if they achieved at least 90% (NH subjects) or 80% (VC and CI subjects) correct. Particularly for the CI and VC groups, subjects often needed several training blocks before they reached the criterion level of performance.

In the actual experiment, no feedback was provided. Each block of trials contained 20 randomly ordered vowels (four presentations of each of the five vowel), as in the training block. Each vowel was either presented in isolation (as in the training) or together with a precursor. Every subject completed 5 blocks of trials for each test condition, so that each stimulus was presented a total of 20 times. A training block (as described above) was inserted after every four experimental blocks to help subjects maintain the correct mapping from the synthetic materials to the vowels categories. Six NH subjects were tested in the NH condition (unprocessed synthetic vowels) first, then the VC condition. The other six NH subjects ran the VC condition first, followed by the NH condition.

5.2.3 Results

Enhancement of the vowels was defined as the improvement in recognition scores in conditions with the precursor, compared to scores without the precursor. Figure 1B shows the percent-correct scores for each group, averaged across subjects and vowels, as a function of the formant levels. Performance with the precursor was higher overall than without. For the NH group, performance with the flat-spectrum vowels (0-dB contrast), at 55.2% correct, was significantly higher than chance (20%) [t (11) = 7.107, p < 0.001], consistent with the results of Summerfield *et al.* (1984). Overall, the effect of the precursor was found to be significant for all three groups [NH: F(1, 11) = 100.1, p < 0.001; VC: F(1, 11) = 67.8, p < 0.001; CI: F(1, 7) = 18.4, p = 0.004], reflecting the fact that scores were generally better in the presence of a precursor. The improvement was found not only at the group level, but for almost all individuals as well; seven of eight CI users exhibited a positive effect of the precursor.

Figure 1C shows the effect of the precursor at the single formant contrast level (2 dB) that was measured in all groups, and indicates a somewhat reduced EE in the CI and VC groups, relative to the NH group. Part of this difference may be due to the different levels of baseline performance. To help account for the effect of baseline performance, performance of the three groups was compared at contrast levels that produced more similar levels of performance (2, 4, and 6 dB for the NH, VC, and CI groups, respectively). The percent-correct scores without the precursor were 41.6%, 49.4% and 51.6%, respectively, while the improvement caused by the precursor was 30.5%, 17.2%

and 12.8%, for the NH, VC, and CI groups, respectively. The fact that a difference in benefit between the NH group and the two other groups remained, suggests that the amount of enhancement was indeed somewhat greater for the NH than for the CI and VC groups.

5.3 Discussion

A significant EE was observed in CI users. This result suggests that EE is not mediated solely or primarily by gain changes in the cochlear via the MOC efferent system, and instead suggests a more central locus. In this respect our human behavioral data are consistent with the conclusions of the physiological studies of Palmer *et al.* (1995a) and Nelson and Young (2010). The fact that the EE was generally smaller in CI users than in NH users may reflect the limited spectral resolution available to CI users. This hypothesis is supported by the fact that the VC condition, which uses vocoding to simulate the loss of spectral resolution in NH listeners, produced very similar results to those found in the CI group.

In the present study, we did not include a test condition with precursor alone to eliminate the effect of possible useable cues provided by precursor. However, results from Summerfield *et al.* (1984b) suggested that listeners were unable to use the "negative" spectra provided by the precursors to identify the vowels. They used both

precursor-target and target-postcursor combinations as stimuli, and found that percent-correct scores were 25% in target-postcursor condition, which was only slightly higher than the 20% chance level. Note also that in our experiment feedback was never provided in conditions that included the precursors, making it less likely that subjects were able to learn the vowel categories based on the precursors alone.

One complicating factor may be the "automatic gain control" (AGC) circuit in CIs, which could potentially change the gain and spectral distribution across frequency, and which may have been affected by the precursor. However, all our CI users had broadband AGC circuits, rather than multi-channel AGCs, meaning that any changes in gain due to the AGC are the same at all frequencies (Zeng, 2004). Beyond the AGC, the mapping of sound input level to electrical output level (determining threshold and maximum comfortable level) is instantaneous, and so should be not be affected by the precursor.

In summary, CI users show significant EE in a vowel identification task, suggesting that EE with these vowel stimuli is not mediated solely by changes in cochlear gain via the MOC system. The smaller amount of EE overall, relative to the NH group, may reflect poorer spectral resolution provided by the CI. There may therefore be a potential opportunity to improve current CI speech processing strategy by enhancing the EE using channel-dependent forms of AGC that simulate the time constants associated with EE as determined in earlier psychophysical studies (Viemeister, 1980).

CHAPTER 6: EFFECTS OF AUDITORY ENHANCEMENT ON THE LOUDNESS OF MASKER AND TARGET COMPONENTS[1]

Abstract

Auditory enhancement refers to the observation that the salience of one spectral region (the "signal") of a broadband sound can be enhanced and can "pop out" from the remainder of the sound (the "masker") if it is preceded by the broadband sound without the signal. The present study investigated auditory enhancement as an effective change in loudness, to determine whether it reflects a change in the loudness of the signal, the masker, or both. In the first experiment, the 500-ms precursor, an inharmonic complex with logarithmically spaced components, was followed after a 50-ms gap by the 100-ms signal or masker alone, the loudness of which was compared with the same signal or masker presented 2 s later. In the second experiment, the loudness of the signal embedded in the masker was assessed with and without a precursor using the same method, as was the loudness of the entire signal-plus-masker complex. The results suggest that the precursor does not affect the loudness of the signal or the masker alone, but enhances the loudness of the signal in the presence of the masker, while leaving the loudness of the surrounding masker unaffected. The results are consistent with the "adaptation of inhibition" hypothesis [Viemeister and Bacon (1982). J. Acoust. Soc. Am. 71, 1502-1507].

---

[1]This chapter has been submitted for publication to *Hearing Research*.

6.1 Introduction

The perceptual salience of a spectral region can be enhanced if it is preceded by its spectral complement. This auditory enhancement effect has been investigated using psychoacoustic masking techniques (e.g. Viemeister, 1980; Thibodeau, 1991; Wright *et al.*, 1993; Viemeister *et al.*, 2013), as well as with vowel-identification paradigms (Summerfield and Assmann, 1987; Chapter 5). The auditory enhancement effect is probably related to the spectral contrast effects that have also been reported using both speech (Holt and Lotto, 2002; Holt, 2006b) and non-speech (Holt, 2006a; Stilp *et al.*, 2010) stimuli. These phenomena demonstrate how the auditory system adapts to long-term spectral properties, and how any changes relative to the long-term spectrum of the preceding sounds are enhanced. More generally, enhancement and contrast effects can be interpreted in terms of a normalization process, which may help establish auditory perceptual invariance in the face of different talkers, changing acoustic environments, and varying background noises.

One possible neural implementation of auditory enhancement involves adaptation. Preceding sounds lead to adaptation of neurons responding to those spectral regions that are most stimulated. Thus, when spectral energy appears in new spectral regions, the neurons responding to the new energy are not in an adapted state and so respond more strongly than the neurons that responded to the preceding sound. In terms of auditory enhancement experiments, this implies that a precursor with the same spectral properties as the masker will lead to a reduced neural response to the masker but not the signal

(Viemeister, 1980; Summerfield *et al.*, 1987; McFadden and Wright, 1990). Note that this "pure adaptation" account implies that the precursor does not produce an *absolute* enhancement of the signal, relative to its response in the absence of the precursor, but rather an enhancement *relative* to the response to the masker. Relative enhancement of this nature has been observed in neural responses at the level of the auditory nerve (Palmer *et al.*, 1995b). However, it is difficult to explain all the available psychophysical results just with pure adaptation. For instance, Viemeister and Bacon (1982) found that the amount of forward masking produced by the signal component increased when a precursor (which itself produced little or no forward masking) was added, suggesting an *absolute* enhancement of the signal component. To account for this phenomenon, Viemeister and Bacon (1982) proposed an "adaptation of suppression" or "adaptation of inhibition" hypothesis, whereby the inhibition usually produced by adjacent components adapts over time, so that when the signal is introduced, it is not inhibited as much as it would have been if all the components began at the same time. This hypothesis was further supported by a recent study (Byrne *et al.*, 2011b). Neural responses consistent with this hypothesis have been identified at the level of the inferior colliculus (IC) (Nelson and Young, 2010), and predictions of a model based on adaptation of inhibition have been tested directly with psychophysical data (Shen and Richards, 2012).

Although both adaptation and adaptation of inhibition may combine to produce the overall auditory enhancement effect, their relative contributions remain unknown. To gain more insight into the mechanisms underlying auditory enhancement, the present

103

study investigates enhancement in terms of the changes in loudness produced by preceding stimuli. The effects of precursors on loudness have been studied over many decades (e.g. Elmasian and Galambos, 1975; Elmasian *et al.*, 1980; Scharf *et al.*, 2002; Arieh and Marks, 2003b; Oberfeld, 2007; Chapter 2). These effects have been termed "loudness enhancement," "loudness decrement," and "loudness recalibration," but have not often been related to the literature on auditory enhancement effects discussed above.

One popular method to measure the effects of a precursor on the loudness of a tone has been to present a sequence of three tones at the same frequency: a precursor, followed by the signal, followed some time later by the comparison tone. The subject's task is to judge the loudness of the signal relative to the comparison tone (e.g. Elmasian and Galambos, 1975; Elmasian *et al.*, 1980). In general, a loud tone preceding a quieter tone can lead to substantial increases in the perceived loudness of the quieter tone, relative to the comparison tone, termed "loudness enhancement" (Elmasian and Galambos, 1975). Experiments using a comparison tone at a different frequency from the precursor and signal have suggested that the precursor may enhance tones close in time to a louder precursor, but may also reduce the loudness of tones that follow more than about 100 ms after the precursor (Scharf *et al.*, 2002; Oberfeld, 2007). The reduction in loudness has been termed "loudness recalibration." These effects appear to be greatest when the precursor is about 20 dB higher in level than the signal (e.g. Elmasian and Galambos, 1975; Oberfeld, 2007). In all cases, precursor tones presented at the *same* level as the signal tone seem to have very little effect on the loudness of the signal.

Because studies of loudness context effects have found little effect of a precursor on a signal if they are presented at the same level, it may be tempting to conclude that loudness context effects have little or no relation to auditory enhancement effects, where large effects are observed when the precursor and masker (and sometimes target) are all presented at the same level. However, there is one important difference between the two paradigms as studied so far: studies of loudness context effects have used pure tones in isolation, whereas auditory enhancement studies have used broadband stimuli. To the extent that auditory enhancement relies on lateral inhibition or suppression, such effects would not be observed in the studies of loudness that have only used pure tones.

The present study investigates auditory enhancement using a paradigm similar to those used in previous studies of loudness context effects, with the important distinction that complex (broadband) stimuli are used. The use of broadband sounds allows an assessment of the potential effects and interactions of suppression or inhibition, and allows us to test some basic properties of the loudness of the stimuli used in auditory enhancement studies. Four basic hypotheses are proposed: 1) The precursor enhances the loudness of the signal in isolation; 2) the precursor reduces the loudness of the masker in isolation; 3) the precursor enhances the loudness of the signal in the presence of the masker; and 4) the precursor reduces the loudness of the masker in the presence of the signal. Experiment 1 measures the effect of the precursor on the signal tone, when it is presented in isolation (i.e., without the flanking masker tones), and the effect of the precursor on the flanking masker tones, when the signal is not present. These two

conditions address hypotheses 1 and 2. Experiment 2 measures the effect of the precursor on the relative loudness of the signal compared with the flanking masker tones, and on the overall loudness of the signal-plus-masker complex, thereby addressing hypotheses 3 and 4. Our results rule out hypotheses 1, 2, and 4, and provide constraints concerning hypothesis 3.

6.2 Experiment 1: effects of a precursor on the loudness of the signal and masker in isolation

6.2.1 Method

Subjects

Experiment 1 was divided into two parts. In experiment 1A, eleven subjects (6 males, 5 females) with normal hearing participated. Their ages ranged from 18 to 22 years (mean age 20.8 years). In experiment 1B, six subjects (2 males, 4 females) with normal hearing participated. Their ages ranged from 20 to 65 years (mean age 29.7 years). Normal hearing was defined as audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz. All subjects were compensated for their time and all provided written informed consent. All protocols were approved by the Institutional Review Board of the University of Minnesota.

Stimuli

Schematic diagrams of the stimuli used in this experiment are shown in Fig. 6.1. Three sounds were presented on each trial: a precursor, a target, and a comparison. Their durations were 500 ms, 100 ms, and 100 ms, respectively, including 10-ms raised-cosine onset and offset ramps. The precursor and target were separated by a silent gap of 50 ms. This gap is well within the range known to elicit strong enhancement effects (e.g. Carcagno et al., 2012), and the target and comparison were separated by a relatively long silent gap of 2 s, to reduce any potential interactions between the precursor and the comparison. The precursor was an inharmonic complex tone, similar to the one in Byrne et al. (2011b), consisting of pure tones with nominal frequencies (before roving) evenly spaced on a logarithmic scale between 250 and 8000 Hz. In experiment 1A, the precursors had spacing between components of 0.1, 0.3, and 0.5 octaves. Three spacings were selected to investigate the effects of spectral gap between components on loudness comparison. The central (median) on-frequency component (corresponding to the signal frequency) could be present or absent in the precursor. In experiment 1B only the 0.3-octave component spacing was tested, as this was reported by Viemeister et al. (2013) to produce the largest auditory enhancement effects. In experiment 1A, both the target and comparison stimuli were pure tones at the frequency of the central component in the precursor; in experiment 1B, the target and comparison were complex tones with the same spectral content as the precursor. The precursor and the target were presented at 50

107

dB SPL per component, and the level of the comparison varied between trials. The overall frequency content of all three sounds was roved together across trials by ±0.5 octaves around the nominal frequencies, so that the frequency relationships between all tones within a trial remained constant. The rove was designed to discourage listeners from using potential cues other than the loudness of the stimuli on each trial (such as a long-term representation of the target), and to avoid any long-term (across-trial) effects of enhancement.
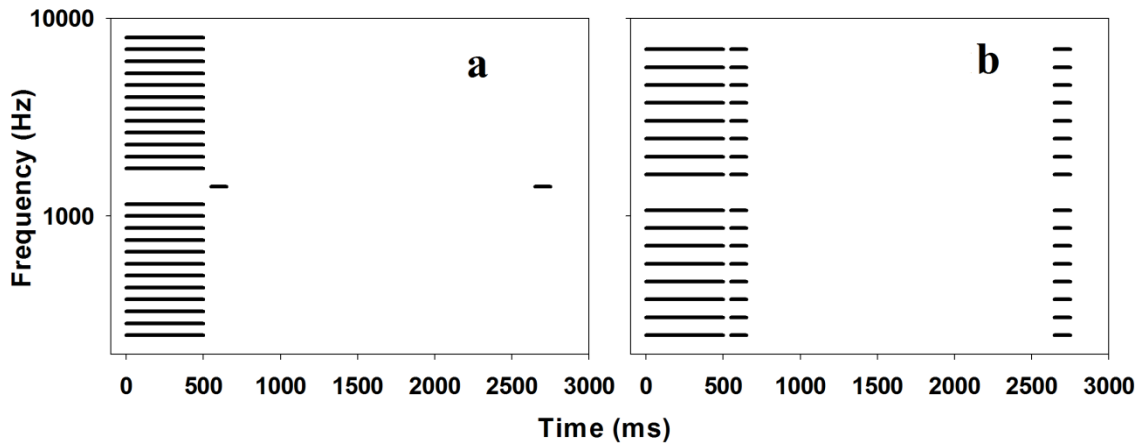
FIG. 6.1. Schematic diagrams of stimuli used in experiment 1. a. The stimuli used in experiment 1A (0.1-octave spacing, without the on-frequency component condition). b. The stimuli used in experiment 1B (0.3-octave spacing). The first stimulus (0-500 ms) is the precursor, the second stimulus (550 to 650 ms) is the target, and the third stimulus (2650-2750 ms) is the comparison.

The stimuli were generated digitally and presented diotically from a LynxStudio L22 24-bit soundcard at a sampling rate of 48 kHz via Sennheiser HD650 headphones to subjects seated in a double-walled sound-attenuating chamber.

Procedure

In both experiment 1A and 1B, a reference condition was tested first. In the reference condition, no precursor was presented. After the reference condition was completed, all other conditions were presented in random order, with each condition being repeated four times per subject. An adaptive two-interval forced-choice task was employed in all conditions. Subjects were instructed to judge which sound was louder, the target or the comparison. The target and precursor levels were fixed. The level of comparison was adaptively varied based on the listener's response, using an interleaved tracking procedure (Jesteadt, 1980; Leek et al., 1991), consisting of a 2-down 1-up track and a 2-up 1-down track. The two procedures track the points at which the comparison is judged louder than the target 70.7% and 29.3% of the time, respectively. The initial step size for each track was 5 dB. After two reversals in one track, the step size for that track was reduced to 2 dB and the track continued for another four reversals. The threshold for each track was defined as the mean comparison level of the last four reversal points. If the stopping rule for one track was met before the other, the "completed" track would continue, but the levels were not incorporated into the threshold estimates. The threshold level of the comparison for each subject was estimated by taking the mean of the

thresholds from the two interleaved tracks, which corresponds approximately to the point of subjective equality (PSE), i.e., the level at which the comparison was judged louder than the target 50% of the time.

In experiment 1A, the on-frequency component in the precursor was either present or absent. In combination with three frequency spacings, this led to a total of 6 conditions, plus the reference condition. In experiment 1B, only one condition was tested along with the reference condition.

6.2.2 Results and discussion

The mean PSE levels for the comparison tone in the different conditions from experiment 1A are shown in Fig. 6.2A. Overall, the PSE levels were somewhat lower than the level of the target tone (50 dB SPL), even in the reference condition, where no precursor was present. However, there was no clear effect of the precursor on the loudness of the target tone, with or without a spectral gap in the precursor. Effects of the precursor and component spacing were investigated in a two-way within-subjects (repeated-measures) ANOVA, excluding the reference condition, with the comparison PSE levels as the dependent variable and component spacing (0.1, 0.3, and 0.5 octaves) and precursor spectral gap (yes or no) as factors. The main effect of component spacing was found to be significant [$F(2,20) = 11.0$, $p = 0.001$, partial $\eta^2 = 0.524$], but neither the main effect of precursor spectral gap [$F(1,10) = 0.388$, $p = 0.547$, partial $\eta^2 = 0.037$] nor the interaction

term [$F(2,20) = 1.59$; $p = 0.228$, partial $\eta^2 = 0.137$] was significant. Pairwise comparisons for component spacing revealed that results from the 0.1-octave spacing condition were significantly different from those of the 0.3-octave ($p = 0.001$) and 0.5-octave ($p = 0.004$) conditions, but these differences were generally less than 2 dB. Paired t-tests showed no significant differences between the reference condition and any of the other six conditions ($p > 0.05$ in all cases). These results indicate that the presence of precursor has little or no effect on the loudness of target tone when presented in isolation.

The results from experiment 1B are shown in Fig. 6.2B. Here again, the loudness of the target (in this case, the tones that act as the masker in a traditional auditory enhancement experiment) does not appear to be affected by the presence of the precursor. A paired-sample t-test confirmed that there was no significant difference between the comparison PSE with or without the precursor [$t(5) = 0.127$; $p = 0.904$].

In both experiment 1A and 1B, the comparison was between 1 and 3 dB lower in level on average than the target at the PSE, even in the absence of a precursor. The reason for this systematic bias is unclear. However, similar biases have been reported in the past. For instance, Elmasian *et al.* (1980) found that a 50-dB SPL target tone alone was matched with a comparison tone near 52 dB SPL, whereas a 70-dB SPL target tone was matched with a comparison tone near 66 dB SPL. They ascribed the differences to a central tendency. Because we only used a single target level (50 dB SPL), the bias observed in our study does not seem likely to be due to a central tendency. Nevertheless, systematic biases in loudness judgments were also observed by Botte (1992), albeit in the

opposite direction (where the comparison tone level was higher than the target when their loudnesses were judged most equal).
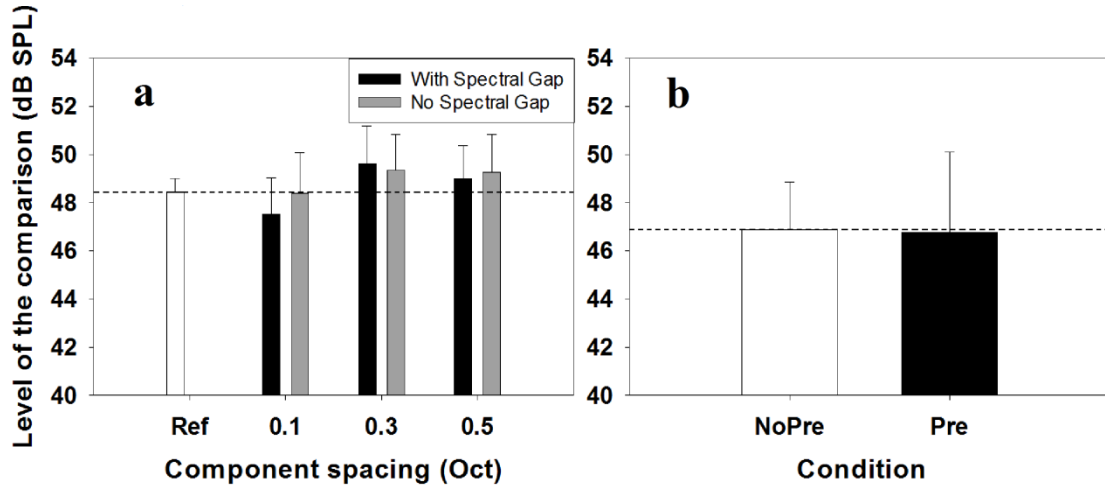
FIG. 6.2. Average level of the comparison stimulus, representing the PSE between the target and comparison stimuli. a. The results from experiment 1A, where the single center tone was the target. The unfilled bar represents the PSE for the reference condition with no precursor; black bars represent results from conditions with a spectral gap in the precursor; and grey bars represent conditions with no spectral gap in the precursor. b. Results from experiment 1B. Again, the white bar represents the PSE without a precursor, and the black bar represents the PSE with a precursor (including a spectral gap at the center frequency). Error bars represent 1 s.e. of the mean. The dotted line indicates the PSE for the reference condition.

In a typical auditory enhancement experiment, the loudness of the signal tone is enhanced, relative to the loudness of the flanking masker tones. The results from

experiment 1 show that neither the loudness of the signal in isolation (experiment 1A) nor the loudness of the flanking masking tones alone (experiment 1B) is affected by the presence of a precursor when presented at the same level as the signal and masker. This outcome suggests that the effect of the precursor in typical auditory enhancement experiments is to alter the interactions between the signal tone and the flanking masker tones. This possibility is explored further in experiment 2.

6.3 Experiment 2: effects of a precursor on the loudness of the masker-signal complex

6.3.1 Methods

Subjects

The same six subjects (2 males and 4 females, age from 20 to 65 with a mean of 29.7 years) were tested in experiment 2 as in experiment 1B.
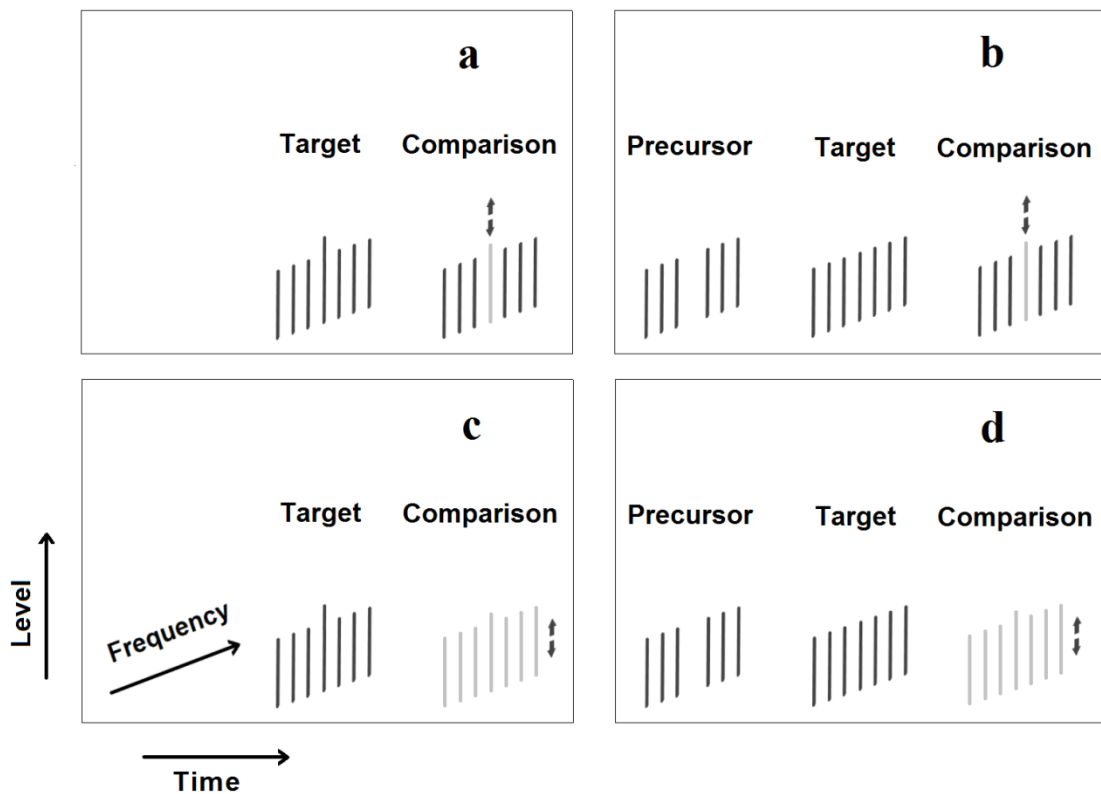
FIG. 6.3. Schematic diagrams of stimuli used in experiment 2A and 2B. a. The stimuli used in the reference condition of experiment 2A. b. The stimuli used in the test condition of experiment 2A. c. The stimuli used in the reference condition of experiment 2B. d. The stimuli used in the test condition of experiment 2B. Black lines represent fixed-level components. Gray lines indicate components in the comparison tone that are varied in level (in isolation or together coherently) by the adaptive tracking procedure.

Stimuli and Procedure

Schematic diagrams of the stimuli used in experiment 2 are shown in Fig. 6.3, with the stimuli for experiment 2A and 2B shown in panels a and b, respectively. As in

experiment 1, subjects first completed a reference condition, in which the target and comparison stimuli were presented without a precursor. In the training phase of experiment 2A, the signal component in the target sound was increased by 10 dB (i.e. 60 dB SPL), and subjects were asked to compare the loudness of the signal component in the target with the loudness of the signal component in the comparison, and judge which was louder. All the flanking masker components in the target and the comparison were presented at 50 dB SPL. A level difference between the signal and masker tones of 10 dB was selected after some pilot testing indicated that 10 dB was sufficient to allow the signal to be heard out from the complex, while still being within the plausible realm of enhancement.

In the test phase of experiment 2A, a precursor was added prior to the target. The signal component in the target sound was presented at the same 50 dB SPL as all of the other target (and precursor) components. The subjects' task was again to compare the loudness of the signal components in the target and comparison stimuli. The same interleaved adaptive procedure that was used in Experiment 1 was employed here to determine the PSE. The average amount of enhancement (i.e., the level increase in the signal tone in the comparison stimulus relative to its level in the target stimulus) was calculated from four repetitions of each condition.

In experiment 2B, for the reference condition, the flanking masker tones in the target stimulus were fixed at 50 dB SPL per component, and the signal component in the target was adjusted to match the amount of enhancement measured for each subject

individually in experiment 2A (5.8 dB on average). On average, this led to a signal component level of about 56 dB SPL in the target stimulus. Subjects were then asked to compare the overall loudness of the target stimulus (signal and masker tones) with the overall loudness of the comparison stimulus. The level relationship between the signal and masker components within the comparison stimulus was the same as for the target stimulus (adjusted for each subject individual, but 5.8 dB higher for the signal component on average). All the components in the comparison were then adjusted simultaneously to compare the overall loudness of two stimuli. In the main condition, the task was still to compare the overall levels, but in the presence of the precursor, so the level of the signal component in the target sound remained at 50 dB SPL. However, the comparison stimulus maintained the same level relationship between the signal and masker components as in the reference condition. Both reference and main conditions were repeated four times each in experiment 2A and 2B.

6.3.2 Results and discussion

Average PSEs for all conditions were calculated in the same way as in experiment 1. As shown in Fig. 6.4A, the signal component level at PSE in the comparison stimulus was 56.2 dB SPL in the reference condition of experiment 2A, which was lower than the one in the target (60 dB SPL). This apparent mismatch in the absence of a precursor is in the same direction as that observed in experiment 1. When the precursor was present, the signal component in the target was presented at the same level as the masker components

116

(i.e. 50 dB SPL), but the level of the signal in the comparison was adjusted to a mean level of 55.8 dB SPL (s.d. 2.91 dB), revealing a significant enhancement in the loudness of signal component in the target sound [paired-sample t-test; $t(5) = 5.24$; $p = 0.003$]. The PSE level for the 60-dB signal component without the precursor not significantly different from that of the 50-dB signal component with the precursor [paired-sample t-test; $t(5) = 0.265$; $p = 0.802$], suggesting effective enhancement by the precursor of around 10 dB.

In experiment 2B, subjects were asked to compare the overall loudness of the target (signal and masker components) with that of the comparison stimulus. The average level of the masker components within the comparison at PSE with and without a precursor is shown in Fig. 6.4B. A paired-sample t-test revealed no significant difference between the masker component levels at PSE between the no-precursor and precursor conditions [$t(5) = 1$; $p = 0.36$], suggesting that the precursor had no effect on the overall loudness of the signal-plus-masker complex. Again, however, we observed that the matched level of the comparison was a few dB below the actual level of the target stimulus. This bias was relatively consistent across conditions and subjects and was observed in both experiments 1 and 2.
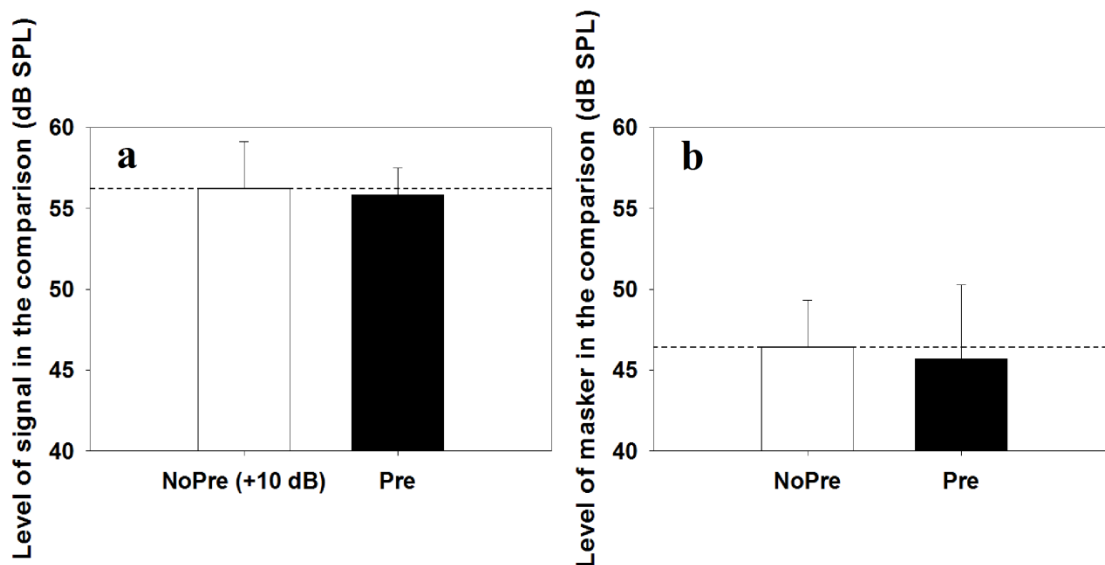
FIG. 6.4. Average level of the comparison stimulus, representing PSE between the target and comparison stimuli. a. The results from experiment 2A, the average PSE of the signal component in the comparison sound in the no-precursor reference condition (with the signal component in the target raised by 10 dB) and in precursor test condition (where all the components in the target, including the signal component, were set to 50 dB SPL). b. The average threshold levels of the masker components in the comparison stimuli in both reference (no precursor) and test (precursor) conditions of experiment 2B. Error bars represent 1 s.e. of the mean across subjects. The dotted line indicates the PSE for the reference condition.

In experiment 2A, we observed an enhancement of the signal component in the target sound, which is consistent with previous studies (Viemeister and Bacon, 1982; Byrne *et al.*, 2011b). After our analysis of experiment 1, showing that the precursor affected neither the masker nor the signal in isolation, two potential explanations of auditory enhancement remained. The first explanation was that the precursor reduces the

loudness of the masker in the presence of the signal. The second explanation was that the precursor enhances the loudness of the signal in the presence of the masker. The results from experiment 2 contradict the first explanation and support the second explanation: experiment 2A showed that the loudness of the signal component was enhanced, relative to the surrounding masker components, and experiment 2B showed that the loudness of the masker components was not reduced by the presence of the precursor. The idea that the signal component is partially masked by the masker, and then released from this partial masking effect by the influence of a precursor is in line with the original idea of "adaptation of inhibition" (Viemeister and Bacon, 1982; Carcagno *et al.*, 2013b).

6.4 General discussion

In this study, we investigated auditory enhancement from the perspective of loudness changes to the signal and flanking masker components in the target sound under the influence of a precursor. The four hypotheses to be tested were: 1) The precursor enhances the loudness of the signal in isolation; 2) the precursor reduces the loudness of the masker; 3) the precursor enhances the loudness of the signal in the presence of the masker; and 4) the precursor reduces the loudness of the masker in the presence of the signal.

In experiment 1, we examined the influence of the precursor on the loudness of signal component in isolation, and on the loudness of the masker components in isolation. It was found that the precursor produced no significant change in the loudness of either. The situation might have been different had the level of the precursor been substantially lower or higher than that of the target. In such cases, effects known as "loudness enhancement" or "loudness decrement" have been shown to influence loudness judgments (Galambos *et al.*, 1972; Elmasian *et al.*, 1975). However, in the present study, the levels of the precursor and target components were always the same, making it less likely that such assimilative or contrastive effects would be observed. In any case, the lack of an effect of the precursor on either the signal or masker components alone rules out hypotheses 1 and 2.

In experiment 2A, the comparison stimulus was judged most similar to the target stimulus (with all components at equal amplitude) when the comparison's signal component was 5.8 dB higher than that of the flanking masker components. This amount of "enhancement" is very similar to that obtained in previous studies using a variety of similar tasks. In Byrne *et al.* (2011b), the estimated amount of enhancement was generally in the range of 4 to 6 dB, whether it was estimated in a masker enhancement paradigm or in a binaural centering paradigm, for stimulus levels of 50 and 70 dB SPL. The "true" amount of enhancement may be somewhat greater than that, based on the systematic bias of 2-3 dB found in our reference conditions. In fact, the PSE found with the precursor was not significantly different from that found with a 10-dB increment in

the signal component, suggesting that the enhancement could be as great as 10 dB. The relative increase in the loudness of the signal is consistent with both hypotheses 3 and 4. In experiment 2B, we added the estimated amount of enhancement to the signal component in the comparison sound, to make the comparison stimulus sound as similar as possible to the target in terms of spectral quality or timbre, and we asked listeners to directly compare the overall loudness of the target and comparison stimuli. According to hypothesis 3, the loudness of the masker components should be unaffected by the precursor, and so listeners would be expected to match the target stimulus with a comparison stimulus that had the same level per masker component. In contrast, hypothesis 4 predicts that listeners should match the target stimulus with a comparison that has the same signal level as the target (and hence lower masker component levels). The results showed that the precursor had no significant effect on the loudness of the overall target stimulus, ruling out hypothesis 4 and supporting hypothesis 3.

What neural mechanisms could produce the observed pattern of outcomes? The proposal of "adaptation of inhibition," originally suggested by Viemeister and Bacon (1982), predicts that a given component within a complex sound is normally inhibited by spectrally surrounding components, but that this inhibition adapts over time. Thus, when a component that is missing in the precursor is then introduced into the target stimulus, the inhibition from all the surrounding components has adapted, leading to an enhancement in the representation of the introduced component. It is not obvious what predictions such a scheme would make regarding the loudness of the masker components

themselves. In fact, no computational models of the auditory enhancement effect have been tested for their ability to predict the effects of enhancement on the perceived loudness of the stimuli. Our study provides critical new data against which future models can be tested.

CHAPTER 7: AUDITORY ENHANCEMENT AS A RELEASE FROM PARTIAL MASKING: DATA AND MODELING[1]

Abstract

The auditory enhancement effect is an example of an auditory process that assigns perceptual salience to new events in an ongoing acoustic environment. Despite its clear importance in auditory perception, the mechanisms underlying auditory enhancement are not fully understood. In this study, auditory enhancement was investigated using a loudness comparison paradigm. The same notched noise was used as the precursor and the background (or masker) and the target was a pure tone spectrally centered in the notch. The target and masker were gated simultaneously and were 100-ms long. The duration of the precursor varied between 100 and 1000 ms, and the inter-stimulus interval (ISI) between the end of the precursor and the target varied between 10 and 500 ms. Subjects judged the loudness of the target relative to a comparison tone at the same frequency, presented after a silent gap of 2 s. Results showed that the loudness of the target was reduced by the masker, relative to its loudness alone, and that the precursor reduced the effect of partial masking, maximally enhancing the target to near its loudness without the masker present. A model that combines "adaptation of lateral inhibition" and short-term neural plasticity with an excitation-pattern-based predictions of partial loudness was able to reproduce the main aspects of these and earlier data sets.

7.1 Introduction

Selective auditory attention is critical for the survival of both prey and predators. Auditory attention can be directed either via "top-down" goal-driven processing, or via "bottom-up" stimulus-driven processing, whereby a sound attracts attention by way of its salience, relative to the background sounds. A sound can be made salient by its physical characteristics, such as greater intensity or a different spectral distribution relative to the background (Kayser *et al.*, 2005). In addition, the auditory system may enhance the salience of novel sounds when they are introduced in an ongoing acoustic environment. The perception of new sounds as separate and salient auditory objects is part of what was described by Bregman (1990) as the "old-plus-new heuristic" and has also been studied in the context of auditory change detection (e.g. Cervantes Constantino *et al.*, 2012).

A classic psychoacoustic effect, known as auditory enhancement (Viemeister, 1980; Viemeister and Bacon, 1982; Summerfield *et al.*, 1984a), can be thought of in terms of enhanced salience of a target sound, making it stand out from an ongoing background sound. The effect is illustrated in Fig. 7.1A, where the red dashed line represents the target tone, the detection threshold of which is reduced by the presence of the precursor (Viemeister and Bacon, 1982; Summerfield *et al.*, 1984a; Summerfield *et al.*, 1987; Shen and Richards, 2012; Carcagno *et al.*, 2013a; Viemeister *et al.*, 2013). Some proposed explanations of auditory enhancement are related to short-term neural adaptation (Viemeister, 1980; Summerfield *et al.*, 1987; McFadden and Wright, 1990). The hypothesis is that the firing rate of auditory nerve fibers stimulated by components in

the precursor declines over time because of short-term adaptation, while nerve fibers with characteristic frequencies corresponding to the target frequency remain unadapted until the target component is introduced. As a result, the neural response to the target component will be greater than that to the others. The relative enhancement of the target may explain the initial results found using simultaneous masking, but does not explain the finding of increased forward masking produced by the target tone (Viemeister and Bacon, 1982). To account for the apparent increment in the absolute response to the target, the "adaptation of inhibition" hypothesis was proposed (Viemeister and Bacon, 1982). According to this hypothesis, the lateral inhibition produced by adjacent components adapts over time, and when the target is introduced, it is not inhibited as much as it would have been if all the components had started at the same time. This hypothesis was in line with results from most of the behavioral studies (Byrne *et al.*, 2011a; Shen and Richards, 2012; Carcagno *et al.*, 2013a; Viemeister *et al.*, 2013). In addition, some physiological support for this proposition was provided by Nelson and Young (2010), who reported that some neurons in the inferior colliculus (IC) of the guinea pig with characteristic frequencies (CFs) close to the target frequency responded more strongly to the target when it was preceded by the precursor.

A qualitative model derived from the "adaptation of inhibition" theory can explain both relative and absolute enhancement in the loudness of the target tone, but only if the adaptation of lateral inhibition is stronger than neural adaptation (Nelson and Young, 2010). This model was implemented and evaluated in Shen and Richards (2012),

125

in which the effects of interstimulus interval (ISI) and the level of precursor were investigated using a simultaneous masking paradigm. Their results provided relatively accurate predictions from the model on the temporal dynamics of the auditory enhancement. However, the model predicted enhancement in all level conditions, which was inconsistent with their psychophysical data, where enhancement existed primarily when the precursor and masker were presented at similar levels. A similar trend was also found by Viemeister *et al.* (2013). This property of auditory enhancement is in contrast to that of other context effects, such as "induced loudness reduction" (Nieder *et al.*, 2003; Arieh *et al.*, 2005) and "loudness enhancement" (Elmasian and Galambos, 1975), whereby the loudness of a target is affected by a spectrally similar precursor. These effects are observed only if the level of the precursor is higher than the target level by around 10-20 dB; if the precursor and the target are presented at the same level, little or no effect is observed on the loudness of the target (Elmasian and Galambos, 1975; Oberfeld, 2007), implying that in auditory enhancement the loudness of the masker should not be affected by the precursor if they are presented at the same level. In Chapter 6, the lack of an effect with equal-level precursors has been confirmed and been extended to broadband sounds. Specifically, when a complex tone (precursor) was followed by a pure tone (target), the loudness of the pure tone was not changed by the presence of the complex (Fig. 7.1D). Both the results from Chapter 6, where a complex tone and a pure tone were employed as the precursor and the target, and results from loudness context studies (Elmasian and Galambos, 1975; Oberfeld, 2007), in which both the precursor and the target were pure tones, contradict the predictions from the model of Nelson and

Young (2010): according to their model, the loudness of the target should be reduced by within-channel neural adaptation when there is no cross-channel inhibition.

The aims of this study were two-fold. The first aim was to better define the amount of enhancement produced by a spectrally notched precursor, in terms of its effect on the loudness of a target tone. A spectrally notched noise was used as both the precursor and the masker to provide a clear qualitative difference between the noise masker and the tone target. Enhancement was measured as a function of both precursor duration and precursor-masker interval. The second aim was to develop a model, in terms of partial masking, that could account not only for the amount of enhancement observed in this and previous studies, but that was also consistent with the findings from other loudness context effects (such as induced loudness reduction), showing that the precursor does not affect the loudness of either the masker or the target in isolation.

7.2 Experiment

7.2.1 Methods

Subjects

Ten (3 males, 7 females) listeners participated in this experiment. Their ages ranged from 18 to 63 years (mean age 27.7 years; only one subject was older than 40).
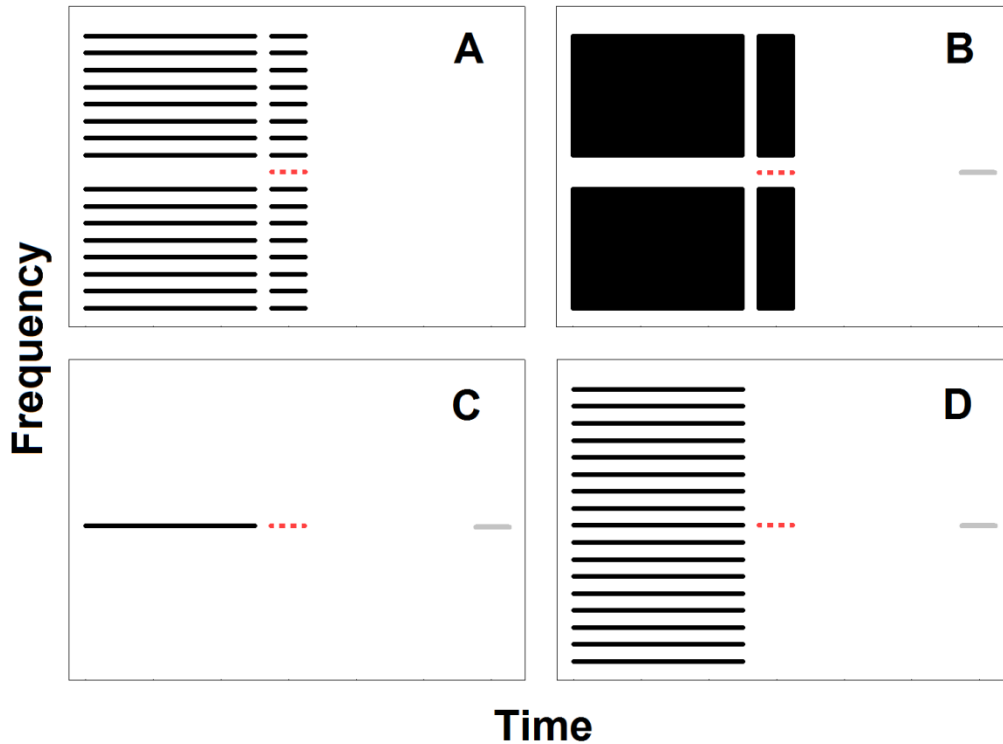
Stimuli

A schematic diagram of the stimuli used in the loudness comparison experiment is shown in Fig. 7.1B. Each trial consisted of three sounds: a precursor, a target tone simultaneously presented with a masker, and a comparison tone. Both the precursor and the masker were threshold equalizing noise (TEN) (Moore et al., 2000) with a notch of 0.6 octaves, based on the fact that it was expected to produce maximum auditory enhancement, according to an earlier study (Viemeister et al., 2013). The nominal passband of the noise ranged from 250 Hz to 8000 Hz, and the spectral notch was geometrically centered around 1414 Hz. The target and comparison tones were presented at the geometric center of the notch (1414 Hz). The overall frequency content of all three sounds was roved together across trials over a 1-octave range (+/- 0.5 octaves) with uniform distribution. The rove was designed to discourage listeners from using potential cues other than the loudness of the tone in each trial, and to avoid any long-term (across-trial) effects of enhancement. Because the rove was implemented across trials (rather than across different sounds within a trial) the frequency relationships between all stimuli within a trial remained constant. The total duration of the precursor was 100, 250 or 1000 ms, and the duration of all the other stimuli was 100 ms. The ISI between the precursor and the target was 10, 100 or 500 ms, and the silent gap between the target and the comparison was 2 s. The level of the target was 50 dB SPL. The level of the noise was 50 dB SPL within the equivalent rectangular bandwidth (ERB) centered around 1 kHz, and the level of the comparison tone was adjusted in an adaptive procedure according to

subjects' responses. All sounds were gated on and off with 10-ms raised-cosine ramps. The stimuli were generated digitally and played out diotically from a LynxStudio L22 24-bit soundcard at a sampling rate of 48 kHz via Sennheiser HD650 headphones to subjects seated in a double-walled sound-attenuating chamber.

Procedure

Training session. A training session was set up prior to the loudness comparison experiment, to make sure listeners could hear out the target tone from the flanking noise. The level of masker (no precursor) was decreased to 40 dB SPL per ERB and the notchwidth was broadened to 0.8 octaves. In the training phase, there were four blocks of trials tested. The level of the comparison was adaptively varied according to the listener's response, using an interleaved tracking procedure consisting of a 2-down 1-up track and a 2-up 1-down track (Jesteadt, 1980; Leek et al., 1991). The two procedures track the points at which the comparison is judged louder than the target 70.7% and 29.3% of the time, respectively. The step size for each track was initially 5 dB, and was reduced to 2 dB after two reversals. Each track was terminated after four reversals at the 2-dB step size, and the threshold for each track was defined as the mean comparison level at the last four reversal points. If one track was completed before the other, the "completed" track would continue, but the levels were not taken into account for the threshold estimates. The threshold level of the comparison for each subject was estimated by taking the mean threshold of the two interleaved tracks, which corresponds approximately to the point of

subjective equality (PSE), i.e., the level at which the comparison was judged louder than the target 50% of the time.

FIG. 7.1. Schematic diagrams of auditory enhancement and loudness context effect. Panel A shows the stimuli used in many previous experiments, where both the precursor (first sound) and the masker (second sound) are complex tones. Both precursor and masker are presented in solid black and the target is in dashed red. Panel B shows the stimuli used in the present experiment, in which the complex tones are replaced by notched noise, and a comparison tone is added to compare with the loudness of the target tone. In the bottom panels, the illustrations of the context effects without lateral inhibition of the target are shown. In panel C, the precursor consists of a pure tone followed by the target. In panel D, the precursor consists of the target and flanking tones together (no spectral gap). No change in the loudness of the target is produced by the precursors in either C or D.

We expected the loudness of the target to be somewhat below 50 dB SPL, based on the partial masking effect of the notched noise. Therefore, we interpreted matches exceeding 50 dB SPL to reflect an inability of the subjects to judge the loudness of the target, independent of the noise masker. Subjects could pass the training and move to the real test session only when the average comparison level was lower than 50 dB SPL in all four blocks, implying that they were able to differentiate the target from the masker. If a subject could not pass training within three training phases, he/she was excluded from the main experiment. Two of the original 12 subjects did not pass the training.

Loudness comparison session. In the loudness comparison session, the procedure for each block of trials was the same as in the training. There were nine precursor conditions (three precursor durations by three ISIs), along with two baseline conditions (masker plus target condition and a target only condition), resulting in eleven conditions altogether. Each condition was repeated four times and all conditions were tested in random order for each subject before any were repeated.

7.2.2 Results

The mean results are shown in Fig. 7.2. The results show that in the absence of the precursor (MT), the loudness of the target was reduced, presumably due to partial masking by the masker (Scharf, 1964). The level of the comparison tone in MT condition was significantly lower than that in the target-along (T) condition [paired t-test; $t(9) = -$

3.49; p = 0.007]. When the precursor was added, the level of the comparison progressively recovered to the level measured in T condition from the level measured in MT condition as shown in Fig. 7.2A, to the point where, with the longest precursor and shortest ISI, the comparison level was not significantly different from its level in the absence of a masker [paired t-test; $t(9) = -0.522$; $p = 0.614$].

The amount of enhancement was calculated by subtracting the comparison level in the absence of a precursor (MT) from the comparison level in the presence of the different precursor configurations. A two-way within-subjects analysis of variance (ANOVA) on the amount of enhancement revealed significant main effects of both the precursor duration [$F(2,18) = 3.62$; $p = 0.048$] and ISI [$F(2,18) = 22.9$; $p < 0.001$], but no interaction [$F(4,36) = 0.709$; $p = 0.591$], confirming that both precursor duration and ISI affect the effect size of auditory enhancement.
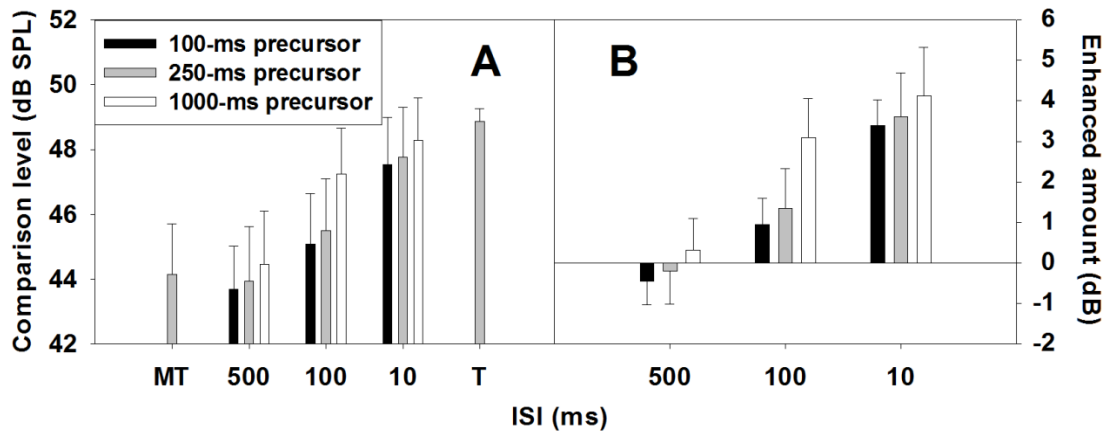
FIG. 7.2. Average level of the comparison stimulus and the amount of enhancement. Panel a shows average level of the comparison tone measured in all conditions (500-ms ISI: black; 100-ms ISI: grey; 10-ms ISI: unfilled), including the target along (T) and the target-plus-masker (MT) condition. Error bars indicate 1 s.e. of the mean.

7.2.3 Discussion

In this experiment, auditory enhancement was measured as a change in the loudness of the target tone in the presence of a spectrally notched noise. The target's loudness was reduced by the presence of the notched-noise masker, as expected based on the effects of partial masking (e.g. Scharf, 1964). The loudness of the target in the presence of the masker was enhanced by the precursor, with the amount of enhancement increasing with increasing precursor duration and with decreasing ISI between the precursor and the target. With maximal enhancement, using a long precursor and short ISI, the loudness of the target was statistically indistinguishable from the loudness of the target presented in

isolation. At the longest ISI of 500 ms, the enhancement approached zero, even with the 1000-ms precursor.

The results from this study, taken together with those from Chapter 6, provide important constraints for any model of auditory enhancement. First, the model should predict no effect on the loudness of the target in isolation, whether it is preceded by a copy of itself (Nieder *et al.*, 2003; Arieh *et al.*, 2005), by a spectrally notched precursor (Chapter 6), or by a spectrally broadband precursor (with no spectral notch (Chapter 6)). This lack of an effect of a same-level precursor on the loudness of the target in isolation suggests that adaptation of responses plays a very small role in auditory enhancement. Second, the enhancement of the target may be implemented as a release from the (partial) masking produced by the simultaneous masker. If so, the enhancement should never result in a target loudness that exceeds the loudness of the target presented alone. These constraints are not met by the model of Nelson and Young (2010), as implemented by Shen and Richards (2012). In the following section, a new modeling approach is considered, which treats auditory enhancement as a release from partial masking.

7.3 A model of auditory enhancement

7.3.1 A theoretical framework based on single neuron response

Short-term neural facilitation, or paired pulse facilitation, refers to the fact that the postsynaptic potential evoked by an impulse is increased when that impulse closely follows a prior impulse (e.g. Zucker, 1989; Regehr, 2012). It is a kind of short-term neural plasticity, the time constant of which is around a hundred milliseconds and approaches that of auditory enhancement, as shown in our data above. In addition, it has been shown that during short-term neural plasticity, the total synaptic charge transfer, which refers to the product of the firing rate and the postsynaptic potential amplitude, is not strongly affected by firing rate, because changes in firing rate are offset by changes in postsynaptic potential arising from neural plasticity (Abbott *et al.*, 1997).

In previous physiological studies, the firing rate of neurons has been measured to evaluate the correspondence between neural activity and loudness, but it has been found that simple sum of spike counts in the auditory nerve does not correspond well with loudness (Relkin and Doucet, 1997b). Findings on short-term neural plasticity may provide a way to understand the unchanged loudness of a target following a precursor when no lateral inhibition is involved (e.g. Fig. 7.1C and 7.1D): even though the firing rate is decreased during the presentation of the second tone, the postsynaptic potential generated by each spike has been enhanced by prior stimulation. Since postsynaptic potential is graded, unlike the action potential, the total amount of postsynaptic potential produced in unit time can hold as a constant (Abbott *et al.*, 1997), leading to an increase

in the sensitivity of a neuron to subtle changes in the firing patterns of its afferents. The probability of a neuron firing is assumed to depend on the total synaptic current it receives from all its neighboring neurons (e.g. Torres and Kappen, 2013), so a relatively constant number of neurons can be stimulated in the next layer. Thereby, in the two-tone condition (Fig. 7.1C), the excitation level of a single neuron stimulated by the precursor and the masker are equivalent, as shown in Eq. (7.1), where e_p and e_m indicate the average electrical potential generated by each spike during the precursor and the masker, respectively. f_p and f_m stand for the mean corresponding firing rates of the neuron during the stimulation of the precursor and the masker. E represents the amount of excitation or the amount of input to the next neuron produced by this single unit. According to neural adaptation and facilitation, f_p is larger than f_m, whereas e_p is smaller than e_m.

$$E = e_p f_p = e_m f_m \quad (e_p < e_m, f_p > f_m) \tag{7.1}$$

In the complex tone condition, the firing rate of a neuron is reduced by lateral inhibition from excited neighboring neurons (e.g. Shapley, 1971). The excitation of the neuron stimulated by a single component in the precursor and the masker, E_p and E_m, are presented in Eq. (7.2) and (7.3), where Δf_p and Δf_m are firing rates reduced by lateral inhibition in precursor and masker, respectively.

$$E_p = e_p(f_p - \Delta f_p) \tag{7.2}$$

$$E_m = e_m(f_m - \Delta f_m) \tag{7.3}$$

$$E_p = E_m \tag{7.4}$$

$$e_p \Delta f_p = e_m \Delta f_m \quad (e_p < e_m, \Delta f_p > \Delta f_m) \tag{7.5}$$

As illustrated in **Chapter 6**, the loudness of masker alone is not affected by the presence of precursor, which supports Eq. (7.4). Eq. (7.5) can be derived from Eq. (7.1) through (7.4). The value of Δf_p must be larger than Δf_m, implying that lateral inhibition adapts over time, in line with the idea of "adaptation of lateral inhibition".

The neural response to the target tone is shown in Eq. (7.6), and is derived in a similar way to that of a single component in the precursor and the masker. Note that when the precursor is presented, the amount of lateral inhibition applied to the neurons responding to the target is the same as that applied to the neurons stimulated by the masker components, while their average firing rate and postsynaptic potential are equal to those of the neurons responding to the precursor components, because no neural plasticity is applied to the freshly introduced target. The absolute enhancement in neural response to the target tone compared with that to precursor components is calculated in Eq. (7.7), which is positive only if there is adaptation of lateral inhibition by the precursor ($\Delta f_p >$

Δf_m). When no precursor is presented, the neural response to the target (E_t) in the masking complex should be the same as E_p and E_m. The relative enhancement in neural response to the target is derived from (7.8), which equals to the result in (7.7) according to the relationships in (7.1) and (7.5). If a pure tone which is at the same level and frequency as the target tone is introduced in the precursor, neural-plasticity will be applied to the target neuron, and E_t will approach to E_m as the result.

$$E_t = \begin{cases} e_p(f_p - \Delta f_m) & with\ precursor \\ e_p(f_p - \Delta f_p) & without\ precursor \end{cases} \qquad (7.6)$$

$$E_t - E_p = \begin{cases} e_p(\Delta f_p - \Delta f_m) > 0 & w/p \\ 0 & w/o\ p \end{cases} \qquad (7.7)$$

$$E_t - E_m = \begin{cases} \Delta f_m(e_m - e_p) = e_p(\Delta f_p - \Delta f_m) > 0 & w/p \\ 0 & w/o\ p \end{cases} \qquad (7.8)$$

This framework can in principle account for the major trends found in the psychoacoustic data. Generally, longer precursors and shorter ISIs generate a larger effect. A longer precursor indicates stronger adaptation and shorter ISIs mean less recovery from adaptation, both of which imply a larger difference between Δf_p and Δf_m (adaptation of lateral inhibition), referring to a stronger effect due to the constancy of e_p. Viemeister *et al.* (2013) investigated effects of the width of the spectral gap in the precursor and masker.

138

A moderate gap of around 0.6 octave was reported to generate the largest effect. If the spectral gap is too narrow, the neurons responding to the target will be adapted by the precursor, resulting in E_t approaching E_m. When the spectral gap is too wide, little lateral inhibition is applied to the target ($\Delta f_m \to 0$ and $\Delta f_p \to 0$ in Eq. (7.6)), which means E_t approaches to E no matter the precursor exists or not and the absolute enhancement declines to 0.

7.3.2 An extended excitation-pattern model of loudness of auditory enhancement

Chen *et al.* (2011) have described an excitation-pattern model of loudness for steady sounds, which predicts the partial loudness of a target tone in noise with good accuracy. The model takes the spectrum of the stimulus and calculates the corresponding auditory excitation patterns, based on auditory filters and a compressive function that calculates the specific loudness (i.e., the loudness contribution from each frequency channel). The overall loudness is then calculated from the total amount of auditory excitation. In partial masking, only the excitation of the filters in which the excitation evoked by the target is stronger than the inhibition generated by the masking noise is counted, as shown in Eq. (7.9). In this model, $E_t(i)$ and $E_m(i)$ indicate the total amount of excitation stimulated by the target and the masker in the ith auditory filter, respectively. K refers to the signal-to-noise ratio (SNR) at the output of the auditory filter required for the signal to reach threshold at high noise levels (Patterson and Moore, 1986), the product of which and

$E_m(i)$ is assumed to be the amount of lateral inhibition from the masker. C is a constant, representing the mapping from neural excitation to loudness.

$$L_t = C \sum_{i:\, E_t(i) > K \cdot E_m(i)} (E_t(i) - K(i) \cdot E_m(i)) \tag{7.9}$$

$$L_t = C \sum_{i:\, E_t(i) > K \cdot E_m(i)} (E_t(i) - A \cdot K(i) \cdot E_m(i)) \tag{7.10}$$

Here we describe how the model of Chen *et al.* (2011) can be extended to predict the loudness of the target tone in auditory enhancement as described in (7.10). According to our theoretical framework, only lateral inhibition is adapted by the precursor, so an adaptation factor (A) is introduced into the original model. The value of A is determined by both precursor duration and ISI, which is determined to be consistent with the data collected in the psychoacoustic experiment described above. A value of $A = 1$ implies no adaptation of inhibition and $A = 0$ implies that lateral inhibition is fully adapted. The duration of precursor (d) determines A_0 [Eq. (7.11)], the initial value of A at the end of precursor, with a longer precursor producing a smaller A_0 (more adaptation). When there is no precursor ($d = 0$), A_0 is 1, implying no adaptation triggered ($A_0 \leq A \leq 1$). The ISI leads to the recovery of lateral inhibition from adaptation, which is represented by t here. The time constant of recovery is τ, which is also controlled by the duration of precursor as in Eq. (7.12). Finally, A is calculated by t, τ and A_0 as shown in Eq. (7.13). A short ISI

implies less recovery and more adaptation, so a smaller t produces a smaller A. When t is

0, A is equivalent to A_0. δ (0.03), γ (1000), α (-7.5e-006) and β (0.003) are four free

parameters, which were adjusted to provide a good model fit to the data from our

experiment.

$$A_0 = e^{-d \cdot \delta} \tag{7.11}$$

$$\tau = (d - \gamma) \cdot \alpha + \beta \tag{7.12}$$

$$A = 1 - e^{(-t \cdot \tau + \ln(1 - A_0))} \tag{7.13}$$

The comparisons between the prediction of the extended model and the real data

from the psychophysical experiments are shown in Fig. 7.3. With the proper selection of

free parameters, the extended model gives a fairly good prediction on results of auditory

enhancement effect. The comparison level of the target in the no-precursor baseline

condition is predicted as in Fig. 7.3A (dashed line). The difference between the colored

curves and the dashed line serves as the predicted amount of absolute enhancement in

different precursor conditions. When ISI is prolonged to 1000 ms, the adaptation of

lateral inhibition has been fully recovered, resulting in the comparison level measured in

the MT (rightmost black unfilled circle) condition of the experiment approaches to that

predicted in the baseline condition of the model. As ISI declines to 0 ms, the lateral

inhibition has been strongly adapted, causing the predicted comparison level to move towards that measured in T condition (leftmost black unfilled circle).
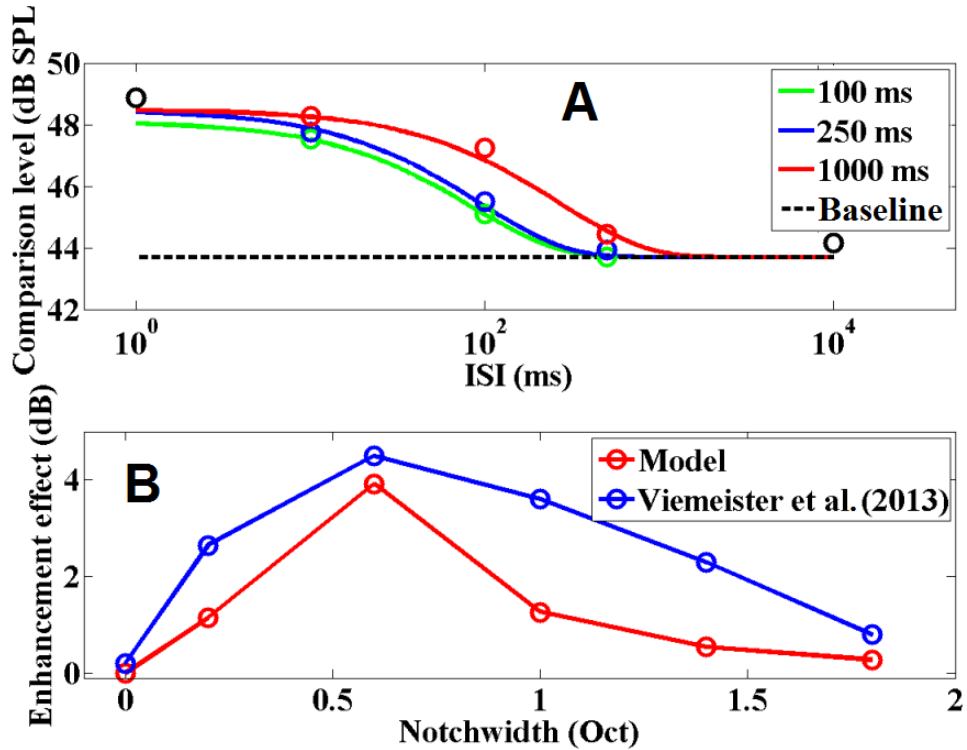
FIG. 7.3. Results predicted by the model. In panel A, comparison levels of the target predicted by the model (curve) as a function of ISI are compared with that measured in the psychoacoustic experiment in this study (unfilled circle). The leftmost and rightmost black circles represent the results from T and MT conditions, respectively. The dashed line indicates the predicted level of the target tone in the no-precursor baseline condition. Different color represents different precursor condition (100 ms: green; 250 ms: blue; 1000 ms: red). In panel B, the frequency selectivity of auditory enhancement is predicted by the model (red), compared with the results in Viemeister *et al.* (2013) measured with simultaneous masking paradigm (blue).

In addition, this model is capable of predicting the frequency selectivity of auditory enhancement. In Fig. 7.3B, the auditory enhancement predicted by this model is plotted as a function of the notch width of the precursor and the masker, and compared with the results from Viemeister *et al.* (2013). The maximum predicted effect is obtained at the notch width of 0.6 octave. Note that the absolute amount of auditory enhancement predicted by the model is somewhat less than shown in the data. This may be due to the fact that the Viemeister *et al.* (2013) data show simultaneously masked thresholds, whereas the model was optimized for a paradigm involving loudness judgments. A further goal remains to better to reconcile both paradigms in enhancement.

7.4 Summary

Results from a psychoacoustic experiment demonstrated that auditory enhancement can act as a release from full or partial masking of the target tone. The effect size increased with the duration of precursor and decreased with ISI, in line with predictions based on an adaptation of lateral inhibition. A theoretical neural framework was proposed, which considered the effects from both firing rate and electric potential produced by each spike during short-term neural facilitation. This framework can explain the patterns of results found in traditional auditory enhancement experiments, while remaining consistent with

other data from studies of loudness context effects, which suggest little or no effect of a precursor on a target when they are matched in level and spectrum and/or when the target is presented in isolation. Finally, an extended excitation-pattern model of loudness was developed, which captured the main trends in data from psychophysical experiments. Future work should implement a more general computational version of this model, based on known properties of the auditory nerve and brainstem to predict results not only from normal-hearing listeners, but also from listeners with hearing loss and cochlear implants.

CHAPTER 8: CONCLUSIONS AND FUTURE DIRECTIONS

8.1 Conclusions

8.1.1 Auditory context effects in normal-hearing listeners

In the present study, auditory context effects were examined in normal-hearing listeners and CI users. Some underlying mechanisms were investigated and their potential implications were discussed.

From the results collected in all experiments on auditory context effects introduced above, it has been shown that the duration, level, and frequency of the precursor, as well as its temporal proximity to the target, can affect how listeners perceive a target sound. In Chapter 2, by examining loudness context effects (LCEs) in normal-hearing listeners, it was found that a more intense precursor resulted in the target sound being judged louder than the comparison signal when they were presented at equal levels and that an increase in precursor level produced an increased effect, in line with previous studies (Elmasian *et al.*, 1980; Arieh and Marks, 2003a). The effects of the precursor also depended on the frequency proximity of the precursor to the target and the comparison, with the maximum effect occurring when the center frequencies of the precursor and target were the same. In Chapter 3, it was shown that longer inter-stimulus intervals resulted in larger effects in normal-hearing listeners. Therefore, the effects of precursor on target in LCEs depend upon the relationships between them in the dimensions of level, frequency and time.

Early studies on spectral contrast effects had revealed that the average power spectrum of the preceding sound played a dominant role in determining context effects, and that the effects were contrastive. It had also been found that the spectral statistics of precursor can affect the recognition of target syllable in a contrastive way (e.g. Holt, 2006a; b). Aside from this spectral contrast, dynamic spectral changes may also play a role in inducing context effects in speech perception. Results from Chapter 4 demonstrated that spectral motion was able to induce changes in responses to both non-speech and speech-like stimuli, and suggested that spectral-motion aftereffects may play a role in more natural situations involving speech perception.

The auditory system seems to enhance the salience of novel sounds when they are introduced into an ongoing acoustic environment, even they are presented at the same overall intensity as ongoing sounds. Auditory enhancement can be thought of as a reflection of this principle. In Chapter 5, the ability to recognize synthetic vowels of listeners was improved by the presence of the precursor. Chapter 6 investigated auditory enhancement as an effective change in loudness of the signal, the masker, or both. It was found that the precursor produced a significant change in the loudness of neither of them in isolation, but instead enhanced the loudness of the target in the presence of the precursor, without affecting the loudness of the masker. Overall the pattern of results supported the hypothesis of "adaptation of lateral inhibition". In Chapter 7, auditory enhancement was measured with a loudness comparison paradigm, in which notched noise was employed as the precursor and the flanking masker. The results revealed that

the perceptual enhancement in the loudness of the target provided a release from the partial masking effect, and that a longer precursor or longer ISI generated larger effect. We proposed a simplified neural model, trying to connect well-known neural properties (e.g. adaptation, inhibition and plasticity) with the mechanisms underlying auditory enhancement in the auditory system.

8.1.2 Comparisons between normal-hearing listeners and CI users in auditory context effects

In general, the pattern of results from CI users was similar to that of normal-hearing listeners. In Chapter 2, using CI users, significant LCEs were found, along with frequency selectivity that was comparable to that found in normal-hearing listeners. In Chapter 3, similar effects of ISI on ILR were observed in both groups of subjects. In Chapter 5, CI subjects showed significant vowel enhancement, as was found in vocoder simulations as well as normal-hearing conditions. These similarities imply that the MOC efferent system cannot be the sole explanation of auditory enhancement effects.

Nevertheless, some differences between the results from normal-hearing listeners and CI users were also observed. In Chapter 2, the results from normal-hearing listeners showed that the effects of the precursor on the target increased with increasing precursor level – an effect that was absent in the results from the CI users. Also, in Chapter 3, enhancement in the loudness of the target was observed in the short-ISI short-precursor

condition in CI users, whereas no enhancement was found in any conditions in normal-hearing listeners. A comparison of the results from CI users and normal-hearing listeners supported the "dual-process" hypothesis, whereby LCE can be explained by the interaction between a fast onset and fast decay enhancement process and a fast onset but slower decay adaptation process (Arieh and Marks, 2003b; Oberfeld, 2007). Our results suggest that at least part of the adaptation process (perhaps related to MOC efferent effects) may be reduced or absent in CI users. In Chapter 5, vowel enhancement effects were smaller in CI users than that measured in normal-hearing listeners. However much of that difference was eliminated by presenting NH listeners with vocoded stimuli, designed to simulate certain aspects of CI processing. The results suggest that the reduction in auditory enhancement in CI users may be caused by poorer spectral resolution, rather than any changes in the mechanisms underlying the enhancement itself.

In summary, poor spectral resolution and a lack of certain peripheral processes (e.g. MOC efferents) may result in the differences between CI users and normal-hearing listeners in auditory context effects, although MOC efferent system cannot be the sole explanation for auditory enhancement effects.

8.2 Future directions

8.2.1 Categorization of auditory context effects

The primary aim of this thesis was to gather empirical data in both normal-hearing listeners and CI users to provide important new information regarding auditory context effects, including loudness effects, such as loudness context effect and induced loudness reduction, and contrastive effects, such as spectral contrast, spectral-motion contrast effects and auditory enhancement.

According to Chapters 2 to 7, there are three factors determining the effect of a precursor on the perception of the target in auditory context effects: the temporal, spectral, and level relationships between the precursor and the target. Most fundamental auditory context effects can be divided into two categories. The first category is the within-channel context effects, in which the target and the precursor have to be presented within the same "critical band;" these include loudness context effects, such as loudness enhancement and induced loudness reduction, and forward masking. These effects decline as the spectral distance between the precursor and the target increases and can disappear completely at large spectral distances (e.g. Moore and Glasberg, 1983; Marks, 1994; Oxenham, 2001), implying the potential involvement of auditory neural adaptation following initial peripheral (cochlear) filtering. For loudness context effects, a "mergence" process (or loudness enhancement) is suspected to happen when the ISI is shorter than about 200 ms (Arieh and Marks, 2003b); this effect was also observed in CI users (see Chapter 3). Similarly, an "integration" hypothesis describes forward masking as a

peripheral nonlinearity followed by linear temporal integration at higher levels in the auditory system (Oxenham, 2001). Time constants of loudness enhancement and forward masking are generally shorter than 100-200 ms, whereas induced loudness reduction dominates the loudness context effect after an ISI of 200 ms, indicating that multiple processes at different stages of auditory system contribute to within-channel context effects. According to psychoacoustic studies, only if there is a difference in level between the precursor and the target can loudness effects be observed (Elmasian and Galambos, 1975; Marks, 1994; Mapes-Riordan and Yost, 1999). In most cases, the maximum effect occurs when the precursor is about 20-30 dB higher in level than the target, and the effect almost disappears when the level difference exceeds 40 dB (Mapes-Riordan and Yost, 1999).

The second category involves across-channel context effects, in which the presence of the precursor with a spectral notched around the target frequency may enhance the gain of the frequency band corresponding to the target, producing an enhancement or spectral shift (spectral contrast) in the perception of the target. Auditory enhancement and overshoot are considered as two examples of the target-enhancement condition, the difference of which is in overshoot no spectral gap is required in the precursor and the masker, whereas a spectral gap is necessary to observe the auditory enhancement. Considering the similar time constants of auditory enhancement and overshoot (generally shorter than 300-400 ms) (e.g. McFadden, 1989; Shen and Richards, 2012) and the similar stimuli used, it may be that they share at least some of the same

150

underlying mechanisms. In overshoot, when a precursor is presented, lateral inhibition from masking noise is adapted. However, unlike the target in auditory enhancement, neural plasticity also occurs in the neurons evoked by the target tone in overshoot, therefore the theoretical framework proposed in Chapter 7 needs to be extended to explain the underlying mechanism. In auditory enhancement, the effect of precursor level has been investigated (Shen and Richards, 2012; Viemeister *et al.*, 2013), revealing that the maximum effect is obtained when the precursor and the masker are presented at the same level. When the level of the precursor is not consistent with that of the target, within-channel processes either at a peripheral level or at higher stages of the auditory system may affect the perception of the target, if the precursor is not spectrally remote from the target. Spectral contrast effects have generally been treated as being different from auditory enhancement and overshoot. However, if contrast effects are considered as coming from a frequency-selective gain change induced by the precursor (Holt *et al.*, 2000; Holt, 2006a), then it may be possible to reconcile these multiple effects with the same underlying mechanisms, suggesting in turn that it may be possible to devise a unified model to account for the effects.

8.2.2 Models of auditory context effects

In Chapter 7, an extended excitation-pattern model of loudness was proposed to capture the data from the psychoacoustic studies of auditory enhancement. The model predicted the effects of precursor duration, ISI and spectral notch on the comparison level of the

target tone, but was not able to take the level relationship between the precursor and the masker or the target into account.

One way to overcome this defect is to develop a computational model of within-channel context effects, in which the effects of neural adaptation, temporal integration and perceptual mergence can be introduced. The time constants of different processes may be derived from both physiological and psychophysical studies. In general, the time constants of temporal integration and perceptual mergence are shorter than 100-200 ms, as mentioned above (Oxenham, 2001; Arieh and Marks, 2003b). Level relationships between the precursor and the target also affect the loudness of the target, which has been well documented (e.g. Elmasian and Galambos, 1975; Marks, 1994; Mapes-Riordan and Yost, 1999; Arieh and Marks, 2003b; Chapter 2). Once developed, this model of within-channel context effects could be integrated with the model of auditory enhancement, proposed in Chapter 7, so that the effects of spectral distance, temporal distance and loudness relationship between the precursor and the target can be represented quantitatively in multiple context effects.

Overshoot is another context effect that has been extensively studied, but remains elusive in terms of the potential underlying mechanisms (e.g. Bacon, 1990; Bacon and Smith, 1991; Strickland, 2001). The model framework outlined above could also be tested against the extensive data sets available with overshoot.

The theoretical framework supporting the model of auditory enhancement discussed in Chapter 7 also raises the question of the relationship between neural

excitation and loudness perception. In previous physiological studies on auditory enhancement effect, firing rate of neurons was measured to evaluate the perceptual level of the input stimulus (Nelson and Young, 2010). However, in auditory context effects, those evoked neurons may adjust their firing rates by the mechanism of neural plasticity (e.g. Zucker, 1989; Regehr, 2012), even when the level and the perceived loudness of the input stimulus stay relatively constant. The hypothesis given in Chapter 7 is described as when a neuron is excited by a stimulus, the product of its firing rate and the postsynaptic potential generated by each spike reflects the perceptual level or the loudness of the stimulus. This neural excitation-loudness hypothesis requires further evaluation through physiological studies. However, it will be difficult to assess loudness in animals during simultaneous physiological experiments. In human subjects it may be possible using non-invasive methods, such as EEG, or potentially more invasive studies using ECoG, to relate neural responses to loudness. As with the animal studies (e.g. Relkin and Pelli, 1987), such studies so far have not produced strong correlations between summed neural activity and perceived loudness (Relkin and Doucet, 1997a).

8.2.3 Restoring auditory context effects in cochlear-implant users

In this study, a critical concern is the comparison between the performance of normal-hearing listeners and that of CI users in auditory context effects. According to the discussion in categorization of context effects, contrastive cross-channel context effects seem to be more critical in producing perceptual constancy and improving speech

perception. Since reduced auditory enhancement was found in Chapter 5, there may be a potential opportunity to improve current CI speech processing strategy by restoring the cross-channel context effects.

A general property of auditory enhancement is that a novel sound is enhanced in loudness and salience within an ongoing background. One potential way to improve the gain of a novel sound is to manually enhance the gain of the channel that is freshly activated within a time period. If a channel is freshly activated comparing to other channels within a time period, it may imply the presence of a novel sound, so the gain of this channel could be enhanced to make it salient from others. This manipulation would require a fast attack and slow release process according to experimental results in Chapter 7. The time constant of release could be measured via a modified simultaneous masking paradigm with a short ISI. In terms of the fitting procedure, the amount of enhancement introduced by this scheme could be individually tailored, depending on how much enhancement a subject shows before compensation.

The effect of compensation in auditory enhancement (cross-channel context effects) in CI users could then be assessed by running psychoacoustic and speech experiments involving auditory context effects, to determine whether the signal processing results in CI outcomes that more closely resemble the data found for normal-hearing listeners. Finally, it would then be possible to evaluate whether such a scheme results in improved speech perception in CI users, particularly in noisy and variable acoustic backgrounds.

REFERENCES

Abbott, L. F., Varela, J. A., Sen, K., and Nelson, S. B. (**1997**). "Synaptic depression and cortical gain control," Science **275**, 220-224.

Arieh, Y., Kelly, K., and Marks, L. E. (**2005**). "Tracking the time to recovery after induced loudness reduction," J. Acoust. Soc. Am. **117**, 3381-3384.

Arieh, Y., and Marks, L. E. (**2003a**). "Recalibrating the auditory system: a speed-accuracy analysis of intensity perception," J. Exp. Psychol. Hum. Percept. Perform. **29**, 523-536.

Arieh, Y., and Marks, L. E. (**2003b**). "Time course of loudness recalibration: implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550-1556.

Bacon, S. P. (**1990**). "Effect of masker level on overshoot," J. Acoust. Soc. Am. **88**, 698-702.

Bacon, S. P., and Smith, M. A. (**1991**). "Spectral, intensive, and temporal factors influencing overshoot," Q. J. Exp. Psych. **43A**, 373-400.

Barlow, H. B. (**1961**). "Possible principles underlying the transformations of sensory message," Sensory Communication, W. Rosenblith, ed. (MIT press, Cambridge, MA), 217-234.

Bingabr, M., Espinoza-Varas, B., and Loizou, P. C. (**2008**). "Simulating the effect of spread of excitation in cochlear implants," Hear. Res. **241**, 73-79.

Bloj, M. G., Kersten, D., and Hurlbert, A. C. (**1999**). "Perception of three-dimensional shape influences colour perception through mutual illumination," Nature **402**, 877-879.

Botte, M.-C., Baruch, C., and Mönikheim, S (**1992**). "Memory for loudness: The role of loudness contour," Adv. Biosci., edited by Y. Cazals, L. Demany and K. Horner (Pergamon Press, Oxford) **7**, 305-311.

Bregman, A. S. (**1990**). Auditory Scene Analysis: The Perceptual Organisation of Sound (MIT Press, Cambridge, MA).

Brosch, M., and Schreiner, C. E. (**2000**). "Sequence sensitivity of neurons in cat primary auditory cortex," Cereb. Cortex. **10**, 1155-1167.

Byrne, A. J., Stellmack, M. A., and Viemeister, N. F. (**2011a**). "The enhancement effect: Evidence for adaptation of inhibition using a binaural centering task," J. Acoust. Soc. Am. **129**, 2088-2094.

Byrne, A. J., Stellmack, M. A., and Viemeister, N. F. (**2011b**). "The enhancement effect: evidence for adaptation of inhibition using a binaural centering task," J. Acoust. Soc. Am. **129**, 2088-2094.

Carcagno, S., Semal, C., and Demany, L. (**2012**). "Auditory enhancement of increments in spectral amplitude stems from more than one source," J. Assoc. Res. Otolaryngol. **13**, 693-702.

Carcagno, S., Semal, C., and Demany, L. (**2013a**). "Enhancement of increments in spectral amplitude: further evidence for a mechanism based on central adaptation," Adv. Exp. Med. Biol. **787**, 175-182.

Carcagno, S., Semal, C., and Demany, L. (**2013b**). "No Need for Templates in the Auditory Enhancement Effect," PLOS ONE **8**, e67874.

Cervantes Constantino, F., Pinggera, L., Paranamana, S., Kashino, M., and Chait, M. (**2012**). "Detection of appearing and disappearing objects in complex acoustic scenes," PLOS ONE **7**, e46167.

Chen, Z., Hu, G., Glasberg, B. R., and Moore, B. C. (**2011**). "A new method of calculating auditory excitation patterns and loudness for steady sounds," Hear. Res. **282**, 204-215.

Clark, N. R., Brown, G. J., Jurgens, T., and Meddis, R. (**2012**). "A frequency-selective feedback model of auditory efferent suppression and its implications for the recognition of speech in noise," J. Acoust. Soc. Am. **132**, 1535-1541.

Cooper, N. P., and Guinan, J. J., Jr. (**2003**). "Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity," J. Physiol. **548**, 307-312.

de Boer, J., Thornton, A. R., and Krumbholz, K. (**2012**). "What is the role of the medial olivocochlear system in speech-in-noise processing?," J. Neurophysiol. **107**, 1301-1312.

Dean, I., Harper, N. S., and McAlpine, D. (**2005**). "Neural population coding of sound level adapts to stimulus statistics," Nat. Neurosci. **8**, 1684-1689.

Dean, I., Robinson, B. L., Harper, N. S., and McAlpine, D. (**2008**). "Rapid neural adaptation to sound level statistics," J. Neurosci. **28**, 6430-6438.

Demany, L., Pressnitzer, D., and Semal, C. (**2009**). "Tuning properties of the auditory frequency-shift detectors," J. Acoust. Soc. Am. **126**, 1342-1348.

Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: perceiving shifts in nonperceived pitches," J. Acoust. Soc. Am. **117**, 833-841.

Diehl, R. L., Lotto, A. J., and Holt, L. L. (**2004**). "Speech perception," Ann. Rev. Psychol. **55**, 149-179.

Diehl, R. L., and Walsh, M. A. (**1989**). "An auditory basis for the stimulus-length effect in the perception of stops and glides," J. Acoust. Soc. Am. **85**, 2154-2164.

Elmasian, R., and Galambos, R. (**1975**). "Loudness enhancement: monaural, binaural, and dichotic," J. Acoust. Soc. Am. **58**, 229-234.

Elmasian, R., Galambos, R., and Bernheim, A., Jr. (**1980**). "Loudness enhancement and decrement in four paradigms," J. Acoust. Soc. Am. **67**, 601-607.

Elmasian, R., Morgan, R., and Galambos, R. (**1975**). "Time Course of Loudness Enhancement and Intensity Discrimination," J. Acoust. Soc. Am. **58**, S35-S35.

Fowler, C. A. (**1981**). "Production and perception of coarticulation among stressed and unstressed vowels," J. Speech Hear. Res. **24**, 127-139.

Galambos, R., Bauer, J., Picton, T., Squires, K., and Squires, N. (**1972**). "Loudness enhancement following contralateral stimulation," J. Acoust. Soc. Am. **52**, 4.

Garinis, A. C., Glattke, T., and Cone, B. K. (**2011**). "The MOC reflex during active listening to speech," J. Speech Lang. Hear. Res. **54**, 1464-1476.

Gibson, J. J. (**1933**). "Adaptation, after-effect and contrast in the perception of curved lines," J. Exp. Psychol. **16**, 1-31.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103-138.

Goupell, M. J., and Mostardi, M. J. (**2012**). "Evidence of the enhancement effect in electrical stimulation via electrode matching," J. Acoust. Soc. Am. **131**, 1007-1010.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species - 29 years later," J. Acoust. Soc. Am. **87**, 2592-2605.

Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589-607.

Guinan, J. J., Jr. (**2010**). "Cochlear efferent innervation and function," Curr. Opin. Otolaryngol. Head Neck Surg. **18**, 447-453.

Holt, L. L. (**2006a**). "The mean matters: effects of statistically defined nonspeech spectral distributions on speech categorization," J. Acoust. Soc. Am. **120**, 2801-2817.

Holt, L. L. (**2006b**). "Speech categorization in context: joint effects of nonspeech and speech precursors," J. Acoust. Soc. Am. **119**, 4016-4026.

Holt, L. L., and Lotto, A. J. (**2002**). "Behavioral examinations of the level of auditory processing of speech context effects," Hear. Res. **167**, 156-169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710-722.

Hong, R. S., and Rubinstein, J. T. (**2006**). "Conditioning pulse trains in cochlear implants: effects on loudness growth," Otol. Neurotol. **27**, 50-56.

Hong, R. S., Rubinstein, J. T., Wehner, D., and Horn, D. (**2003**). "Dynamic range enhancement for cochlear implants," Otol. Neurotol. **24**, 590-595.

Huang, J., and Holt, L. L. (**2009**). "General perceptual contributions to lexical tone normalization," J. Acoust. Soc. Am. **125**, 3983-3994.

Huang, J., and Holt, L. L. (**2012**). "Listening for the norm: adaptive coding in speech categorization," Front. Psychol. **3**, 10.

ISO:226 (**2003**). "Normal equal-loudness-level contours," Acoustics Geneva, Switzerland.

Jennings, S. G., Strickland, E. A., and Heinz, M. G. (**2009**). "Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking," J. Acoust. Soc. Am. **125**, 2172-2181.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophy. **28**, 85-88.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950-962.

Kawase, T., Delgutte, B., and Liberman, M. C. (**1993**). "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones," J. Neurophysiol. **70**, 2533-2549.

Kayser, C., Petkov, C. I., Lippert, M., and Logothetis, N. K. (**2005**). "Mechanisms for allocating auditory attention: an auditory saliency map," Curr. Biol. **15**, 1943-1947.

Kimberley, B. P., Nelson, D. A., and Bacon, S. P. (**1989a**). "Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1660-1665.

Kimberley, B. P., Nelson, D. A., and Bacon, S. P. (**1989b**). "Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1660-1665.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98-104.

Land, E. H. (**1977**). "The retinex theory of color vision," Sci. Am. **237**, 108-128.

Leek, M. R., Hanna, T. E., and Marshall, L. (**1991**). "An interleaved tracking procedure to monitor unstable psychometric functions," J. Acoust. Soc. Am. **90**, 1385-1397.

Liberman, M. C. (**1988**). "Response properties of cochlear efferent neurons: monaural vs. binaural stimulation and the effects of noise," J. Neurophysiol. **60**, 1779-1798.

Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: effect of preceding liquid on stop consonant identification," Percept. Psychophy. **60**, 602-619.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica)," J. Acoust. Soc. Am. **102**, 1134-1140.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophy. **28**, 407-412.

Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506-3511.

Marks, L. E. (**1988**). "Magnitude estimation and sensory matching," Percept. Psychophy. **43**, 511-525.

Marks, L. E. (**1994**). ""Recalibrating" the auditory system: the perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382-396.

McFadden, D. (**1989**). "Spectral differences in the ability of temporal gaps to reset the mechanisms underlying overshoot," J. Acoust. Soc. Am. **85**, 254-261.

McFadden, D., and Wright, B. A. (**1990**). "Temporal decline of masking and comodulation detection differences," J. Acoust. Soc. Am. **88**, 711-724.

McKenna, T. M., Weinberger, N. M., and Diamond, D. M. (**1989**). "Responses of single auditory cortical neurons to tone sequences," Brain Res. **481**, 142-153.

Mishra, S. K., and Lutman, M. E. (**2014**). "Top-down influences of the medial olivocochlear efferent system in speech perception in noise," PLOS ONE **9**, e85756.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise," J. Acoust. Soc. Am. **73**, 1249-1259.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205-224.

Nelson, P. C., and Young, E. D. (**2010**). "Neural correlates of context-dependent perceptual enhancement in the inferior colliculus," J. Neurosci. **30**, 6577-6587.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test- and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846-2855.

Oberfeld, D. (**2007**). "Loudness changes induced by a proximal sound: loudness enhancement, loudness recalibration, or both?," J. Acoust. Soc. Am. **121**, 2137-2148.

Oberfeld, D. (**2008**). "The mid-difference hump in forward-masked intensity discrimination," J. Acoust. Soc. Am. **123**, 1571-1581.

Oxenham, A. J. (**2001**). "Forward masking: Adaptation or integration?," J. Acoust. Soc. Am. **109**, 732-741.

Oxenham, A. J., and Kreft, H. A. (**2014**). "Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing," Trends. Hear. **18**.

Palmer, A. R., Summerfield, Q., and Fantini, D. A. (**1995a**). "Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement," J. Acoust. Soc. Am. **97**, 1786-1799.

Palmer, A. R., Summerfield, Q., and Fantini, D. A. (**1995b**). "Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement," J. Acoust. Soc. Am. **97**, 1786-1799.

Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in Frequency Selectivity in Hearing, edited by B. C. J. Moore (Academic, London), pp. 123-177.

Plack, C. J. (**1996a**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024-1030.

Plack, C. J. (**1996b**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024-1030.

Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451-460.

Regehr, W. G. (**2012**). "Short-term presynaptic plasticity," Cold Spring Harb. Perspect. Biol. **4**, a005702.

Relkin, E. M., and Doucet, J. R. (**1997a**). "Is loudness simply proportional to the auditory nerve spike count?," J. Acoust. Soc. Am. **101**, 2735-2741.

Relkin, E. M., and Doucet, J. R. (**1997b**). "Is loudness simply proportional to the auditory nerve spike count?," J. Acoust. Soc. Am. **101**, 2735-2740.

Relkin, E. M., and Pelli, D. G. (**1987**). "Probe tone thresholds in the auditory nerve measured by two-interval forced-choice procedures," J. Acoust. Soc. Am. **82**, 1679-1691.

Scharf, B. (**1964**). "Partial masking," Acustica **14**, 16-23.

Scharf, B., Buus, S., and Nieder, B. (**2002**). "Loudness enhancement: Induced loudness reduction in disguise? (L)," J. Acoust. Soc. Am. **112**, 807-810.

Schouten, J. F. (**1940**). "The residue and the mechanism of hearing," Proc. Kon. Akad. Wetenschap. **43**, 991-999.

Shapley, R. (**1971**). "Effects of lateral inhibition on fluctuations of the impulse rate," J Gen. Physiol. **57**, 557-575.

Shen, Y., and Richards, V. M. (**2012**). "Investigating the auditory enhancement phenomenon using behavioral temporal masking patterns," J. Acoust. Soc. Am. **132**, 3363-3374.

Shimojo, S., Kamitani, Y., and Nishida, S. (**2001**). "Afterimage of perceptually filled-in surface," Science **293**, 1677-1680.

Shu, Z. J., Swindale, N. V., and Cynader, M. S. (**1993**). "Spectral motion produces an auditory after-effect," Nature **364**, 721-723.

Stankovic, K. M., and Guinan, J. J., Jr. (**1999**). "Medial efferent effects on auditory-nerve responses to tail-frequency tones. I. Rate reduction," J. Acoust. Soc. Am. **106**, 857-869.

Stilp, C. E., Alexander, J. M., Kiefte, M., and Kluender, K. R. (**2010**). "Auditory color constancy: calibration to reliable spectral properties across nonspeech context and targets," Atten. Percept. Psychophy. **72**, 470-480.

Strickland, E. A. (**2001**). "The relationship between frequency selectivity and overshoot," J. Acoust. Soc. Am. **109**, 2062-2073.

Summerfield, A. Q., and Assmann, P. (**1987**). "Auditory enhancement in speech perception," Psychophys. Speech Percept. pp. 140-150.

Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074-1095.

Summerfield, Q., Haggard, M., Foster, J., and Gray, S. (**1984a**). "Perceiving vowels from uniform spectra: phonetic exploration of an auditory aftereffect," Percept. Psychophy. **35**, 203-213.

Summerfield, Q., Haggard, M. P., Foster, J., and Gray, S. (**1984b**). "Perceiving vowels from uniform spectra: Phonetic exploration of an auditory after-effect," Percept. Psychophys. **35**, 203-213.

Summerfield, Q., Sidwell, A., and Nelson, T. (**1987**). "Auditory enhancement of changes in spectral amplitude," J. Acoust. Soc. Am. **81**, 700-708.

Thibodeau, L. M. (**1991**). "Performance of hearing-impaired persons on auditory enhancement tasks," J. Acoust. Soc. Am. **89**, 2843-2850.

Torres, J. J., and Kappen, H. J. (**2013**). "Emerging phenomena in neural networks with dynamic synapses and their computational implications," Front. Comput. Neurosci. **7**, 30.

Viemeister, N. F. (**1980**). "Adaptation of masking," in Psychophysical, physiological and behavioural studies in hearing, edited by G. van den Brink, and F. A. Bilsen, pp. 190-198.

Viemeister, N. F., and Bacon, S. P. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**, 1502-1507.

Viemeister, N. F., Byrne, A. J., and Stellmack, M. A. (**2013**). "Spectral and level effects in auditory signal enhancement," Adv. Exp. Med. Biol. **787**, 167-174.

Wade, T., and Holt, L. L. (**2005**). "Perceptual effects of preceding nonspeech rate on temporal properties of speech categories," Percept. Psychophy. **67**, 939-950.

Weinberger, N. M., and Mckenna, T. M. (**1988**). "Sensitivity of Single Neurons in Auditory-Cortex to Contour - toward a Neurophysiology of Music Perception," Mus. Percept. **5**, 355-389.

Wen, B., Wang, G. I., Dean, I., and Delgutte, B. (**2009**). "Dynamic range adaptation to sound level statistics in the auditory nerve," J. Neurosci. **29**, 13797-13808.

Wiegrebe, L., Kossl, M., and Schmidt, S. (**1996**). "Auditory enhancement at the absolute threshold of hearing and its relationship to the Zwicker tone," Hear. Res. **100**, 171-180.

Wilson, J. P. (**1970**). "An auditory after-image," in Frequency Analysis and Periodicity Detection in Hearing, edited by R. P. a. G. F. Smoorenburg (Sijthoff, Leiden, The Netherlands), pp. 303-318.

Wright, B. A., McFadden, D., and Champlin, C. A. (**1993**). "Adaptation of suppression as an explanation of enhancement effects," J. Acoust. Soc. Am. **94**, 72-82.

Zeng, F. G. (**1994**). "Loudness growth in forward masking: relation to intensity discrimination," J. Acoust. Soc. Am. **96**, 2127-2132.

Zeng, F. G. (**2004**). "Trends in cochlear implants," Trends Amplif. **8**, 1-34.

Zhou, X., Henin, S., Thompson, S. E., Long, G. R., and Parra, L. C. (**2010**). "Sensitization to masked tones following notched-noise correlates with estimates

of cochlear function using distortion product otoacoustic emissions," J. Acoust. Soc. Am. **127**, 970-976.

Zucker, R. S. (**1989**). "Short-term synaptic plasticity," Annu. Rev. Neurosci. **12**, 13-31.

Zwicker, E. (**1964**). "'Negative afterimage' in hearing," J. Acoust. Soc. Am. **36**, 2413-2415.

Zwislock, J. J., and Sokolich, W. G. (**1974**). "Loudness Enhancement of a Tone Burst by a Preceding Tone Burst," Percept. Psychophy. **16**, 87-90.

Zwislocki, J. J., and Sokolich, W. G. (**1974**). "On loudness enhancement of a tone burst by a preceding tone burst," Percept. Psychophys. **16**, 87-90.